

Original Article

Time series analysis of rubella incidence in Chongqing, China using SARIMA and BPNN mathematical modelsQi Chen¹, Han Zhao², Hongfang Qiu¹, Qiyin Wang¹, Dewei Zeng³, Mengliang Ye¹¹ School of Public Health, Research Center for Medicine and Social Development and Innovation Center for Social Risk Governance in Health, Chongqing Medical University, Chongqing, China² Chongqing Municipal Center for Disease Control and Prevention, Chongqing, China³ Nan'an district center for disease control and prevention, Chongqing, China**Abstract**

Introduction: Chongqing is among the areas with the highest rubella incidence rates in China. This study aimed to analyze the temporal distribution characteristics of rubella and establish a forecasting model in Chongqing, which could provide a tool for decision-making in the early warning system for the health sector.

Methodology: The rubella monthly incidence data from 2004 to 2019 were obtained from the Chongqing Center of Disease and Control. The incidence from 2004 to June 2019 was fitted using the seasonal autoregressive integrated moving average (SARIMA) model and the back-propagation neural network (BPNN) model, and the data from July to December 2019 was used for validation.

Results: A total of 30,083 rubella cases were reported in this study, with a significantly higher average annual incidence before the nationwide introduction of rubella-containing vaccine (RCV). The peak of rubella notification was from April to June annually. Both SARIMA and BPNN models were capable of predicting the expected incidence of rubella. However, the linear SARIMA model fits and predicts better than the nonlinear BPNN model.

Conclusions: Based on the results, rubella incidence in Chongqing has an obvious seasonal trend, and SARIMA (2,1,1) × (1,1,1) 12 model can predict the incidence of rubella well. The SARIMA model is a feasible tool for producing reliable rubella forecasts in Chongqing.

Key words: incidence; rubella; forecasting; SARIMA; BPNN.

J Infect Dev Ctries 2022; 16(8):1343-1350. doi:10.3855/jidc.16475

(Received 17 February 2022 – Accepted 15 June 2022)

Copyright © 2022 Chen *et al.* 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.

Introduction

Rubella is an eruptive, highly contagious, and mild viral infection transmitted through direct or droplet contact from nasopharyngeal secretions [1,2]. Humans are the only known host [3]. It typically begins with low-grade fever and lymphadenopathy, followed by characteristic brief appearance of a generalized erythematous, maculopapular rash [4]. Maternal rubella infection, especially during the first trimester of gestation, can cause miscarriage, stillbirths, or congenital rubella syndrome (CRS), a constellation of congenital disabilities that often includes the classic triad of cataracts, sensorineural deafness, and congenital heart defects [1,4,5]. Rubella occurs worldwide [6-8], and China is among the countries that have experienced frequent and widespread rubella outbreaks [9].

China, the most populous country, has reported 88% of rubella cases in the Western Pacific Region since 2004, when it began reporting rubella cases [10]. The rubella-containing vaccine (RCV) was licensed and

made available in China in 1993 and was introduced nationwide by the government's Expanded Program on Immunization (EPI) in 2008 [11]. The overall incidence of rubella in China has declined in recent years, with a historically low level in 2017, during which the incidence was 1.16 per million total population, the accumulation of susceptible individuals led to a rebound in the epidemic, and almost 30,000 cases were reported in 2019 [9], according to the data from the National Notifiable Disease Reporting System (NNDRS). In fact, the number of rubella cases worldwide had more than doubled in 2019, compared to the year before [12]. Rubella remains a public health concern worldwide [3]. Thus, understanding the epidemiological patterns and establishing the accurate morbidity prediction model of rubella are critical for risk analysis and resource allocation in the health sector.

Many mathematical models have been developed to successfully forecast infectious diseases [13-16]. The autoregressive integrated moving average (ARIMA)

model, which is the most popular linear modeling technique for forecasting time series [17], integrates changing trends, periodic changes, and random disturbances in time series and can reveal the quantitative relationship between the research object and other objects with the development and transformation of time [18]. The artificial neural network (ANN) model, on the other hand, has usually been preferred for nonlinear time series [19]. The backpropagation neural network (BPNN) model is one of the most widely used ANNs; it can approximate any complex nonlinear mapping and is especially suitable for solving complicated internal mechanisms [20].

Chongqing is one of the areas with the highest rubella incidence rates in China. A total of 187 rubella public health emergencies were reported nationwide in 2019, including 47 in Chongqing [21]. Although there are some epidemiological rubella studies, no previous research has been conducted to predict the incidence of rubella in Chongqing. Consequently, we developed a seasonal autoregressive integrated moving average (SARIMA) model, an extension of the ARIMA model with seasonal effect, and a BPNN model to forecast the incidence of rubella. The results of this study may help predict the epidemic trends of rubella and provide reference information for rubella public health intervention in Chongqing.

Methodology

Data Collection

In this study, the incidence data from 2004 to 2019 were primarily gained from the Chongqing Center of Disease and Control (CDC), while population data were collected from the Chongqing Statistics Bureau. All rubella cases were initially diagnosed by clinical symptoms and confirmed by laboratory examination.

SARIMA model construction

The SARIMA model integrates the seasonal effects, long-term trend effects, cyclical variations, and disturbances of stochastic perturbations of the series. The general structure of the SARIMA model could be expressed as SARIMA (p, d, q) × (P, D, Q)_s, in which p, d, and q are non-negative integers that indicate orders of non-seasonal autoregressive (AR) terms, non-seasonal differencing and non-seasonal moving average (MA), respectively; P, D, and Q are also non-negative integers that indicate orders of seasonal AR terms, seasonal differencing and seasonal MA terms, respectively; and S suggests the length of seasonal period [18]. The values of the above six parameters will be determined based on the sequence autocorrelation

function (ACF) and the partial autocorrelation function (PACF). The complete structure of the model fitted to the sequence of observations is as follows [18]:

$$(1 - B)^d (1 - B^S)^D x_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)} \varepsilon_t$$

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\Theta_S(B) = 1 - \theta_1 B^S - \dots - \theta_Q B^{QS}$$

$$\Phi_S(B) = 1 - \phi_1 B^S - \dots - \phi_P B^{PS}$$

In the above equation, *B* represents the backward shift operator, ε_t denotes the residual at time *t*, the mean of ε_t is zero and the variance is constant, x_t is the observed value at time $t (t = 1, 2, \dots, k)$.

Akaike information criterion (AIC) and Schwarz Bayesian criterion (BIC) were employed for model optimization. The tested model with the lowest AIC and BIC values was regarded as the relatively optimal one.

BPNN model construction

The BPNN is a typical multilayer feed-forward neural network consisting of an input layer, at least one hidden layer, and an output layer [22]. The basic idea of the BP algorithm is that the learning process consists of two processes: forward propagation of the signal and backward propagation of the error. In the forward propagation process, the input samples are passed in from the input layer, and after nonlinear processing at each hidden layer, they are passed to the output layer. The actual output value is compared with the expected value, if the given output requirement is not satisfied, and the error is propagated back. Error back-propagation is to back-propagate the output error in some form to the hidden and input layers, gradually reducing the error by adjusting the weight of each neuron in the hidden layer until the error between the actual and the expected output meets the accuracy requirement or reaches the maximum learning count [23].

A single hidden layer feed-forward network is the most widely used for time series modeling and forecasting [22]. The relationship between the output (x_t) and the inputs ($x_{t-1}, x_{t-2}, \dots, x_{t-n}$) is based on the following mathematical representation:

Table 1. Comparison of rubella incidence before and after the introduction of RCVs into EPI.

Time	Total number of cases	Average annual incidence (1/100000)	Person χ^2 -value	p value
2004-2008	15181	10.76	6062.98	< 0.001
2009-2019	14902	4.51		

$$x_t = f_2\left(\sum_{j=1}^M w_j f_1\left(\sum_{k=1}^n v_{jk} x_{t-k} + \theta_j\right) + w_0\right) + \varepsilon_t$$

Where $v_{jk} (j = 1, 2, \dots, M, k = 1, 2, \dots, n)$ is the weight of the k^{th} neuron in the input layer to the j^{th} neuron in the hidden layer; $w_j (j = 1, 2, \dots, M)$ is the weight of the j^{th} neuron in the hidden layer corresponding to the output layer neuron; $\theta_j (j = 1, 2, \dots, M)$ is the threshold of the j^{th} neuron in the hidden layer; w_0 is the threshold of the output layer neuron; n is the number of input nodes; M is the number of hidden nodes. Both $f_1(\cdot)$ and $f_2(\cdot)$ are activation functions.

Performance statistic index

The mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute scaled error (MASE) were used to determine the fitting and forecasting effect [24]. The calculation formulae are as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_t| \quad RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2}$$

$$MASE = \text{mean}\left(\left|q_j\right|\right)$$

$$q_j = \frac{\hat{x}_j - x_j}{\frac{1}{T-m} \sum_{t=m+1}^T |x_t - x_{t-m}|}$$

Where x_t and \hat{x}_t denote the actual incidence value and the estimated incidence value, respectively, at time t , m is the seasonal period. Generally, lower MAE, RMSE and MASE values suggest better fitting and predicting performance, and vice versa.

Statistical analysis

At first, we used Excel 2016 software to count the number of reported rubella incidences/fatalities in Chongqing for each year from 2004 to 2019, organize the data into two periods: before (2004-2008) and after (2009-2019) the introduction of RCVs into China’s nationwide EPI system, and conduct a descriptive analysis. Next, we analyzed the time series analysis of the rubella incidence sequence. We used the R4.0.5 software, to develop a SARIMA model and a BPNN model and predict the incidence of rubella. In this study, the incidence of rubella from 2004 to June 2019 was used as a training dataset to fit the model, predict the incidence of rubella from July to December in 2019, and verify the predicted effect. Statistical significance was determined as $p < 0.05$.

Ethics approval and consent to participate

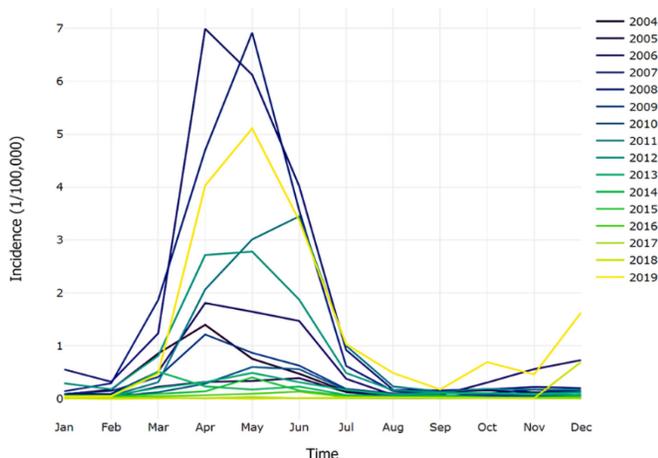
This study protocol was approved by the Ethics Committee of Chongqing Medical University. Since all data collected in this study are was in accordance to the Law of the People’s Republic of China on the Prevention and Treatment of Infectious Diseases, and we did not include any patients’ personal information, written informed consent is not needed for this study.

Results

Descriptive analyses

During this study (2004-2019), a total of 30083 rubella cases were reported in Chongqing, with an average annual incidence of 6.47 per 100000, and no deaths. Table 1 shows the average annual incidence rate after EPI introduction decreased from 10.76 per 100000 before implementation to 4.51 per 100000, with

Figure 1. Seasonal distribution of rubella in Chongqing.



statistically significant differences ($\chi^2=6062.98$, $p<0.001$).

Figure 1 presents the temporal distribution of rubella incidence in Chongqing from 2004 to 2019, mainly showing a unimodal feature, with a peak incidence from April to June and a slight increase again in December, with obvious seasonal characteristics.

The sequence diagram also shows that the incidence of rubella from 2004 to 2019 was unstable (Figure 2). Between 2004 and 2008, regional rubella incidence ranged from 2.08 per 100000 in 2005 to a peak of 21.09 per 100,000 in 2007. Since 2013, the incidence of rubella has shown a downward trend, and the 2017 incidence was at a historic low level, during which only 52 rubella cases were reported (incidence rate of 0.17

per 100000). However, his rate increased again in 2019, with an incidence rate of 17.52 per 100000.

SARIMA (p, d, q) × (P, D, Q)_S model

Figure 3 shows the decomposition diagram of the rubella incidence sequence, which exhibits a fluctuating downward trend and prominent seasonal characteristics. Given the periodicity and seasonality of rubella, a first difference ($d=1$) and a seasonal difference ($D=1$) with a period of 12 were performed to eliminate non-stationarity. The sequence diagram after the difference appeared to be stationary (Figure 4), and the ADF test results remained significant ($p < 0.05$).

Figure 5 shows that ACF and PACF of the stationary sequence were both trailing. Based on the

Figure 2. Reported monthly incidence of rubella from January 2004 to December 2019.

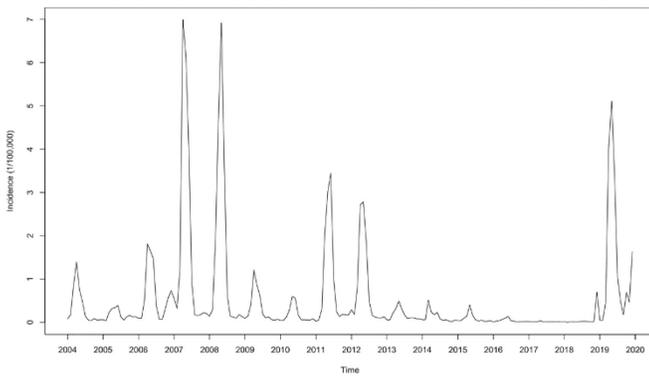


Figure 3. Monthly data cases of rubella incidence from 2004 to 2019 with multiplicative decomposition of rubella incidence time-series data.

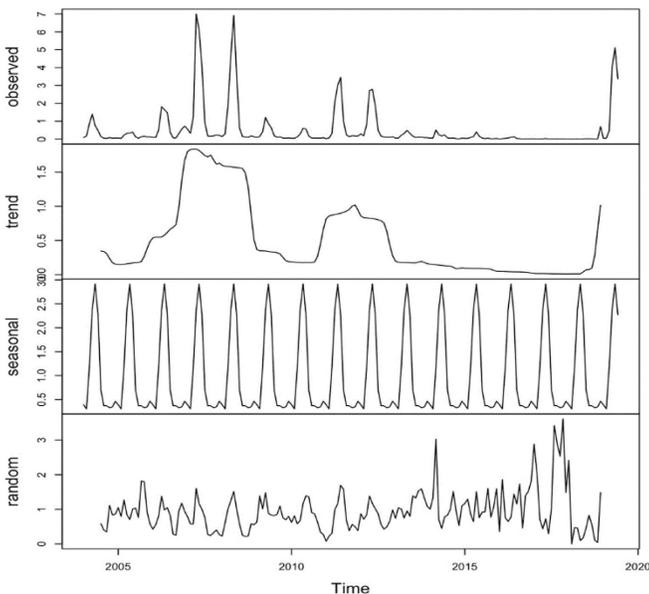


Figure 4. Sequence diagram after a 1-step difference and seasonal difference with a period of 12.

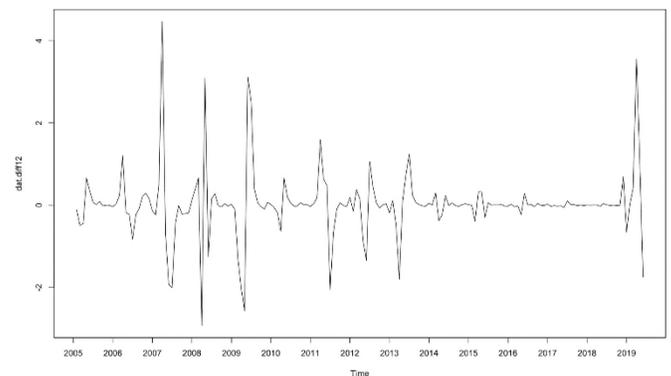
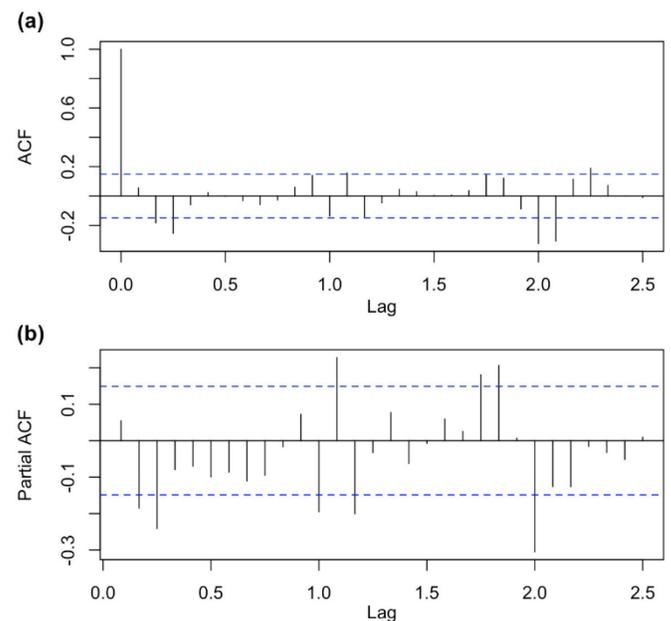


Figure 5. Autocorrelation function (ACF) and partial ACF charts of monthly rubella incidence numbers. (a) ACF chart; (b) Partial ACF chart.



ACF, we determined the possible values of q ($q = 1, 2$ or 3) and Q ($Q = 0$ or 1) of SARIMA (p, d, q) \times (P, D, Q)s, and based on PACF, we determined the possible values of p ($p = 1, 2$ or 3) and P ($P = 0$ or 1). All primary models were used to simulate and model the monthly rubella incidence. Several models passed the test and the model parameter test. The AIC and BIC values of the six candidate models are listed in Table 2, and we finally confirmed SARIMA (2, 1, 1) \times (1, 1, 1)₁₂ as the optimal model, which had resulted in the minimum AIC (358.85) and BIC (377.77) values.

Table 3 shows the parameter estimate results of the SARIMA model, and all the parameter estimates were significant ($p < 0.001$). The Ljung-Box test also confirmed that the residuals of the model were a white noise sequence ($p > 0.05$). The model equation is given as

$$(1 - B)(1 - B^{12})x_t = \frac{1 - 0.8877B}{1 - 0.9477B + 0.4019B^2} \times \frac{1 - B^{12}}{1 - 0.2396B^{12}} \varepsilon_t$$

BPNN model

After normalizing rubella incidence and converting all values to intervals [0, 1], we used the data of the past 12 months as the input data and the data of the thirteenth month as the output data. For example, the observed value from January to December 2004 was selected to predict that in January 2005, the practical value from February 2004 to January 2005 was chosen to indicate

Table 2. AIC and BIC values for candidate SARIMA models.

Candidate Models	AIC	BIC
SARIMA (1, 1, 2) \times (1, 1, 1) ₁₂	364.71	383.63
SARIMA (1, 1, 2) \times (0, 1, 1) ₁₂	373.38	389.15
SARIMA (2, 1, 0) \times (1, 1, 1) ₁₂	379.05	394.81
SARIMA (2, 1, 1) \times (0, 1, 1) ₁₂	364.57	380.33
SARIMA (2, 1, 1) \times (1, 1, 1) ₁₂	358.85	377.77
SARIMA (3, 1, 0) \times (1, 1, 1) ₁₂	367.42	386.34

that in February 2005, and so on. Thus, the number of nodes in the input and output layers could be determined, with $n = 12, m = 1$. The number of neurons in the hidden layer could be calculated by the empirical formula: $M = \sqrt{m + n} + \alpha$, ranging from 5 to 14, where α is the regulation constant with values between 1 and 10 [23]. We set the target error of the training of BPNN as 0.001, the training steps as 2000, the transfer function of the hidden layer as “tansig”, the transfer function of the output layer as “purelin”. When the difference between the target estimate and the actual value is less than 0.001, error back-propagation is not necessary; otherwise, back-propagation is required to further adjust the weights. During the entire modeling, when the difference between the two residual squares is less than 0.001, it can be considered converged, and the iteration can be skipped to end the algorithm [23]. Predictions are made based on the fitted weight matrix and threshold values for the validation set.

We constructed ten different BPNN models in terms of the number of neurons in the hidden layer.

Table 3. Estimation of parameters of the SARIMA (2,1,1) \times (1,1,1)₁₂ model.

Variable	Coefficient	Standard Error	t value	p value
AR (1)	0.9477	0.0868	10.9192	< 0.001
AR (2)	-0.4019	0.0788	-5.0998	< 0.001
MA (1)	-0.8877	0.0680	-13.0497	< 0.001
Seasonal AR (1)	0.2396	0.0872	2.7472	< 0.001
Seasonal MA (1)	-1.0000	0.0755	-13.2513	< 0.001

Table 4. Reported and forecasted incidence of rubella from July to December in 2019.

Month	Actual incidence (1/100000)	Forecasted incidence by SARIMA model (1/100000)	Forecasted incidence by BPNN model (1/100000)
Jul 2019	1.0306	1.3600	1.1479
Aug 2019	0.4865	0.4621	0.3611
Sep 2019	0.1760	0.3052	0.3979
Oct 2019	0.6914	0.4691	0.4647
Nov 2019	0.4609	0.6905	0.5141
Dec 2019	1.6292	1.0498	0.2668

Table 5. Comparison of the fitting and prediction performance of the two model.

Evaluation index	Fitting performance		Forecasting performance	
	SARIMA	BPNN	SARIMA	BPNN
MAE	0.3015	0.3477	0.3065	0.3511
RMSE	0.5850	0.6458	0.2524	0.5758
MASE	0.6317	0.7009	0.4099	0.5703

Comparing the RMSE value of the testing set in each model, we finally chose the 12-8-1 BPNN model, which had the minimum RMSE value of 0.5758.

Comparative analysis

The SARIMA (2, 1, 1) × (1, 1, 1)₁₂ model and the 12-8-1 BPNN model were employed for forecasting the incidence of rubella from July to December 2019. Table 4 shows the value of the prediction, showing that the predicted values obtained by the SARIMA model were closer to the actual values than those obtained by BPNN. We compared the observed rubella notification rate with the fitted and predicted ones in a point-to-point manner to further compare the validity of the two models (Table 5). The SARIMA model was superior to the BPNN model for either fitting or forecasting performance.

Figure 6 shows the actual incidence and fitted incidence of the two models; the SARIMA model had a fit value closer to the actual value than the BPNN model. Both Figure 6 and Table 4 show that the tendency and epidemics from predicted incidence are close to the real value of incidence and epidemic circumstance of rubella. These two prediction methods can be considered to predict the incidence of rubella in Chongqing. In terms of the fitting effect of the model and the accuracy of the prediction results, the SARIMA model would be much more suitable.

Discussion

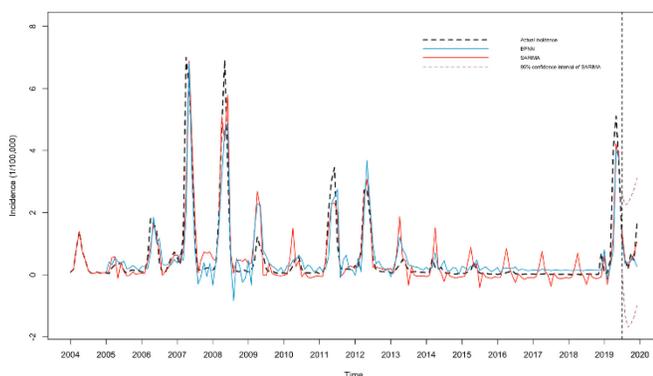
In this study, we found that the incidence of rubella in Chongqing from 2004 to 2019 was not stable, but based on the annual trend of the national rubella incidence [9,25], with a peak incidence every 5-8 years [26]. In 2007 and 2008, the incidence of rubella in Chongqing was as high as 21.09 per 100,000 and 18.76 per 100000, respectively. A further peak of 10.81 per 100000 in incidence occurred in 2011. While RCV was

introduced nationwide in 2008, its full implementation required four years due to sporadic vaccine supply constraints. Nationwide implementation was achieved in 2012, with routine immunization coverage exceeding 95% [25]. Since 2013, the incidence of rubella has exhibited a downward trend, and the 2017 incidence was at a historically low level. But it broke out again in 2019, with an incidence rate of 17.52 per 100,000, which may be caused by the accumulation of susceptible populations and the spread of imported strains [10,27]. Despite fluctuations, the introduction of rubella into China's EPI system in 2008 has indeed contributed to the control and elimination of rubella and the prevention of CRS in Chongqing, with a significant difference in the average annual incidence rate before and after the nationwide introduction of RCVs. Large-scale routine vaccination with RCVs should be continued to close known immunity gaps and immunize health workers and populations at risk for rubella transmission [28]. The health department should continuously carry out virological surveillance work, timely detection of epidemic viruses, and genotype identification to scientifically prevent and control the spread of imported virus strains and further reduce the incidence of rubella.

We analyzed the rubella incidence rates and observed a fluctuating downward trend and seasonal characteristics in this area, with a single peak from April to June, and troughs from August to February, similar to the studies of other areas [9,25]. This may be related to the climate and human behaviors [29]. As temperatures rise, interpersonal activity and contact tend to increase during the spring and summer months. Thus, the virus is more likely to spread throughout the population, leading to widespread epidemics. Therefore, sanitary inspection should be done in the peak period to control the source of infection, cut off the transmission route, and establish a complete emergency plan.

Time series analysis extracts valuable information from historical data, identifies recursive mechanisms, and expresses the current observations as a function of their historical observations to make predictions about future trends. Rubella is a globally important public health issue. Accurate prediction of rubella is helpful for policy-makers to develop effective intervention plans and efficient allocation of public health resources. We demonstrated that both SARIMA and BPNN models can be used to predict the monthly incidence of rubella. The linear SARIMA model fits and predicts better than the nonlinear BPNN model, indicating that the rubella incidence data in Chongqing has good linear

Figure 6. Model fitting, verification and forecasting of rubella incidence in Chongqing from January 2004 to December 2019.



characteristics [19]. Since most epidemiological data are seasonal and cyclical [15-17,30], the SARIMA model considers seasonal effects and is suitable for analyzing sequences with significant seasonality and periodicity [31], which appears more appropriate for the prediction of rubella incidence. In addition, the establishment of SARIMA model is simple. To the best of our knowledge, this is the first study that assesses both linear and nonlinear models to predict rubella report rates in Chongqing. The results indicate that model SARIMA $(2, 1, 1) \times (1, 1, 1)_{12}$ is the relatively more optimal prediction model in this study; the fitted values were in good agreement with the observed. The SARIMA model in this study can reasonably contribute to rubella surveillance in Chongqing and has a good short-term predictive effect. It provides a basis for predicting future rubella epidemics and for health authorities to strengthen public health measures to prevent and control the disease. It is worth noting that SARIMA is a short-term forecasting model with poor long-term forecasting ability, and the model needs to be updated or replaced by continuously incorporating new monitoring data during subsequent use. Therefore, model updating and the selection of new models remain essential elements of future research work.

Several limitations also exist in the present study. Firstly, both models were based on the time-series data for preliminary modeling prediction, without considering other factors affecting the rubella incidence, such as meteorological [32] and socio-geographical [33]. Incorporating more potential influencing factors may allow for more accurate mathematical modeling. Secondly, neither model can capture both linear and nonlinear patterns of the data equally well. Thus, in both cases, one of these parts is not considered, and can lead to deceptive results. Therefore, various hybrid approaches have been suggested [19]. In future research, more models with good time-series prediction results can be tested. A deep optimization combination of different models can make the advantages complement each other and achieve further improvement in the prediction accuracy rate.

Conclusions

The results of this study suggest that it is feasible to apply both SARIMA and BPNN models to predict the incidence of rubella in Chongqing; however SARIMA $(2, 1, 1) \times (1, 1, 1)_{12}$ showed better performance. The short-term prediction is effective, and it is helpful for the understanding of rubella epidemiology and resources allocation in Chongqing. Meanwhile, timely

and effective countermeasures can be taken for possible epidemic peaks.

Acknowledgements

The authors express their gratitude to the Chongqing Municipal Center for Disease Control and Prevention for the disease data as well as the help from teachers of Chongqing Medical University.

Funding

This work was financially supported by the Application and Research of Public Health Emergency Management and On-site Emergency Management System (2020MSXM018).

Authors' Contributions

Q.C. and H.Z. contributed equally to this study. Q.C. conceived and designed the study. H.Z. collected data, Q.C. and H.Q. performed the statistical analyses and wrote the main manuscript. H.Q., Q.W., D.Z., and M.Y. helped to draft the manuscript. All authors read and approved the final manuscript.

References

1. Yazigi A, de Pecoulas AE, Vauloup-Fellous C, Grangeot-Keros L, Ayoubi JM, Picone O (2017) Fetal and neonatal abnormalities due to congenital rubella syndrome: a review of literature. *J Matern Fetal Neonatal Med* 30: 274-278.
2. Chang C, Ma H, Liang W, Hu P, Mo X, An Z, Zheng H (2017) Rubella outbreak and outbreak management in a school setting, China, 2014. *Hum Vaccin Immunother* 13: 772-775.
3. Lambert N, Strebel P, Orenstein W, Icenogle J, Poland GA (2015) Rubella. *Lancet* 385: 2297-2307.
4. Leung AKC, Hon KL, Leong KF (2019) Rubella (German measles) revisited. *Hong Kong Med J* 25: 134-141.
5. Oishi K, Satoh H, Tanaka-Taya K (2020) Re-emerging rubella epidemic and public health measures in Japan. *Yakugaku zasshi* 140: 901-904.
6. Patel MK, Gibson R, Cohen A, Dumolard L, Gacic-Dobo M (2018). Global landscape of measles and rubella surveillance. *Vaccine*. 36: 7385-7392.
7. Orenstein WA, Cairns L, Hinman A, Nkwane B, Olivé JM, Reingold AL (2018) Measles and rubella global strategic plan 2012-2020 midterm review report: background and summary. *Vaccine*. 36: A35-A42.
8. Bozick BA, Worby CJ, Metcalf CJE (2019) Phylogeography of rubella virus in Asia: vaccination and demography shape synchronous outbreaks. *Epidemics*. 28: 100346.
9. Ma Y, Liu K, Hu W, Song S, Zhang S, Shao Z (2021) Epidemiological characteristics, seasonal dynamic patterns, and associations with meteorological factors of rubella in Shaanxi Province, China, 2005-2018. *Am J Trop Med Hyg* 104: 166-174.
10. Knapp JK, Mariano KM, Pastore R, Grabovac V, Takashima Y, Alexander JP Jr, Reef SE, Hagan JE (2020) Progress toward rubella elimination - Western Pacific Region, 2000-2019. *MMWR Morb Mortal Wkly Rep* 69: 744-750.
11. Yu W, Lee LA, Liu Y, Scherpbier RW, Wen N, Zhang G, Zhu X, Ning G, Wang F, Li Y, Hao L, Zhang X, Wang H (2018)

- Vaccine-preventable disease control in the People's Republic of China: 1949–2016. *Vaccine* 36: 8131-8137.
12. Ujiie M (2019) Rubella Resurgence in Japan 2018-2019. *J Travel Med.* 26: taz047.
 13. Zhang YQ, Li XX, Li WB, Jiang JG, Zhang GL, Zhuang Y, Xu JY, Shi J, Sun DY (2020) Analysis and predication of tuberculosis registration rates in Henan Province, China: an exponential smoothing model study. *Infect Dis Poverty* 9: 123.
 14. Saxena A (2021) Grey forecasting models based on internal optimization for novel corona virus (COVID-19). *Appl Soft Comput* 111: 107735.
 15. Zhang X, Zhang T, Young AA, Li X (2014) Applications and comparisons of four time series models in epidemiological surveillance data. *PLoS One.* 9: e88075.
 16. Wang Y, Xu C, Wang Z, Yuan J (2019) Seasonality and trend prediction of scarlet fever incidence in mainland China from 2004 to 2018 using a hybrid SARIMA-NARX model. *Peer J* 7: e6165.
 17. Wang KW, Deng C, Li JP, Zhang YY, Li XY, Wu MC (2017) Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiol Infect* 145: 1118-1129.
 18. Yi DH, Wang Y (2019) *Applied time series analysis*, 5th ed. Beijing: China Renmin University Press. [Book in Chinese].
 19. Yolcu U, Egrioglu E, Aladag CH (2013) A new linear and nonlinear artificial neural network model for time series forecasting. *Decision Support Systems.* 54: 1340-1347.
 20. Gu KY, Zhou Y, Sun H, Zhao LM, Liu SK (2020) Prediction of air quality in Shenzhen based on neural network algorithm. *Neural Comput Appl* 32: 1879–1892.
 21. Li YQ, Ma C, Hao LX, Wang FZ, Wen N, Yang H, Tang L, An ZJ, Yin ZD (2021) Epidemiology of rubella in China, 2018-2019. *Chinese Journal of Vaccines and Immunization* 27: 62-66. [Article in Chinese].
 22. Wang J, Wang J (2017) Forecasting stochastic neural network based on financial empirical mode decomposition. *Neural Netw.* 90: 8-20.
 23. You HL (2016) *R language prediction practice*. Beijing, China: Publishing House of Electronics Industry pp 325-326. [Book in Chinese].
 24. Hyndman, RJ, Koehler, AB (2006) Another look at measures of forecast accuracy. *International Journal of Forecasting.* 22: 679-688.
 25. Su Q, Ma C, Wen N, Fan C, Yang H, Wang H, Yin Z, Feng Z, Hao L, Yang W (2018) Epidemiological profile and progress toward rubella elimination in China. 10 years after nationwide introduction of rubella vaccine. *Vaccine* 36: 2079-2085.
 26. Cao BN (2019) Analysis of the epidemiological and spatial distribution characteristic of rubella in Jilin province during 2004-2018. Thesis. Jilin University. [Article in Chinese]
 27. Liu Y, Cui AL, Guo JY, Deng LL, Fan LX, Wei H, Liu L, Cong YL, Tang XM, Lei Y, Zhou SJ, Peng XF, Zhou J, Feng Y, Zhang Y, Mao NY, Xu WB, Zhu Z (2021) Epidemiology and genetic characteristics of rubella virus circulating in China in 2020. *Chinese Journal of Virology* 37: 638-647. [Article in Chinese].
 28. WHO Regional Office for the Western Pacific ((2019 Eighth annual meeting of the Regional Verification Commission for measles and rubella elimination in the western Pacific, Hanoi, Vietnam, 16-20 September 2019. Available: <https://apps.who.int/iris/handle/10665/330708>. Accessed 29 July 2021.
 29. Wesolowski A, Metcalf CJ, Eagle N, Kombich J, Grenfell BT, Bjørnstad ON, Lessler J, Tatem AJ, Buckee CO (2015) Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc Natl Acad Sci U S A.* 112: 11114-11119.
 30. Zhai M, Li W, Tie P, Wang X, Xie T, Ren H, Zhang Z, Song W, Quan D, Li M, Chen L, Qiu L (2021) Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: a time series predictive analysis. *BMC Infect Dis* 21: 280.
 31. Arun Kumar KE, Kalaga DV, Sai Kumar CM, Chilkoor G, Kawaji M, Brenza TM (2021) Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Appl Soft Comput* 103: 107161.
 32. Zhang X, Ma SL, Liu ZD, He J (2019) Correlation analysis of rubella incidence and meteorological variables based on Chinese medicine theory of Yunqi. *Chin J Integr Med* 25: 911-916.
 33. Grant GB, Desai S, Dumolard L, Kretsinger K, Reef SE (2019) Progress toward rubella and congenital rubella syndrome control and elimination - worldwide, 2000-2018. *MMWR Morb Mortal Wkly Rep* 68: 855-859.

Corresponding author

Mengliang Ye, PhD
Chongqing Medical University 1# Yixueyuan Road,
Yu Zhong District 400016 China.
Tel: +86-153-1093-9053
Email: yemengliang@cqmu.edu.cn

Conflict of interests: No conflict of interests is declared.