CHINESE MEDICAL SCIENCES JOURNAL

ORIGINAL ARTICLE

Time Series Models for Short Term Prediction of the Incidence of Japanese Encephalitis in Xianyang City, P R China[△]

Rongqiang Zhang^{1,2}, Fengying Li³, Junli Liu³, Meining Liu³, Wenrui Luo³,
Ting Ma³, Bo Ma³, and Zhigang Zhang^{1*}

¹School of Public Health, Shaanxi University of Chinese Medicine, Xianyang, Shaanxi 712046, China

²Institute of Endemic Diseases, School of Public Health, Health Science Center,
Xi'an Jiaotong University, Key Laboratory of Trace Elements and Endemic
Diseases of National Health and Family Planning
Commission, Xi'an 710061, China

³Department of Immunology, Center for Disease Control and
Prevention of Xianyang City, Xianyang, Shaanxi 712046, China

Key words: Japanese encephalitis; time series models; incidence; prediction

Objective To construct a model of Seasonal Autoregressive Integrated Moving Average (SARIMA) for forecasting the epidemic of Japanese encephalitis (JE) in Xianyang, Shaanxi, China, and provide valuable reference information for JE control and prevention.

Methods Theoretically epidemiologic study was employed in the research process. Monthly incidence data on JE for the period from Jan 2005 to Sep 2014 were obtained from a passive surveillance system at the Center for Diseases Prevention and Control in Xianyang, Shaanxi province. An optimal SARIMA model was developed for JE incidence from 2005 to 2013 with the Box and Jenkins approach. This SARIMA model could predict JE incidence for the year 2014 and 2015.

Results SARIMA (1, 1, 1) (2, 1, 1)₁₂ was considered to be the best model with the lowest Bayesian information criterion, Akaike information criterion, Mean Absolute Error values, the highest R^2 , and a lower Mean Absolute Percent Error. SARIMA (1, 1, 1) (2, 1, 1)₁₂ was stationary and accurate for predicting JE incidence in Xianyang. The predicted incidence, around 0.3/100 000 from June to August in 2014 with low errors, was higher compared with the actual incidence. Therefore, SARIMA (1, 1, 1) (2, 1, 1)₁₂ appeared to be

reliable and accurate and could be applied to incidence prediction.

Conclusions The proposed prediction model could provide clues to early identification of the JE incidence that is increased abnormally ($\geq 0.4/100\ 000$). According to the predicted results in 2014, the JE incidence in Xianyang will decline slightly and reach its peak from June to August.

Chin Med Sci J 2017; 32(3):152-160. DOI: 10.24920/J1001-9294.2017.036

JAPANESE encephalitis (JE), a member of the genus flavivirus, is a mosquito-borne disease caused directly by the JE virus.¹⁻⁴ JE is an acute infectious disease with a high mortality rate of around 25%.⁵ The mortality rate of children is generally much higher. Estimated 67 900 cases develop each year in Asia; about half of the cases occur in China.³ If the virus reaches the central nervous system, it is likely to cause lifelong neurological defects, such as deafness, hemiparesis and emotional problems.

Since JE severely threatens people's health, it is highly crucial to identify a JE outbreak early in order to better plan for control and intervention. The seasonality of JE incidence makes it possible to use the existing data more efficiently and effectively. There have been many retrospective studies using surveillance data to predict the number of cases or incidence in the near future, taking tuberculosis and Hand, Foot, and Mouth Disease for example.⁶⁻⁸ Until now, no studies have been conducted in Xianyang, China to forecast the incidence of JE and help understand the possibility of an epidemic or provide valuable clues for prevention.

There are many models available that can be used to forecast infectious diseases. These include general regression models, Markov chain models, Seasonal Autoregressive Integrated Moving Average (SARIMA), Grey models, etc.⁹⁻¹⁰ Due to the strong seasonality of JE, SARIMA model is the most appropriate model. SARIMA model is a statistical method and prediction approach that is particularly useful if there is time dependence in each observation. It is assumed that each observation in the time series correlates with previous ones, which makes it possible to model a temporal structure and obtain a more reliable prediction, particularly for seasonal infections. In recent years, SARIMA models have been successfully employed to predict the incidence or death trends of malaria¹¹ and pneumonia¹² or daily patient numbers.¹³⁻¹⁴

The principal objective of our study was to construct a SARIMA model by which the reported JE incidence in Xianyang, Shaanxi, China could be predicted. The findings of this study will be useful for forecasting JE epidemics and providing valuable reference information for JE control and prevention in Xianyang City and even throughout China.

MATERIALS AND METHODS

Study area

Xianyang City is located in the hinterland of Guanzhong Plain in Shaanxi province, China, adjacent to Gansu province, with a population of about 5.4 million. Situated at 107°38' to 109°10' east longitude and 34°11' to 35°32' north latitude, Xianyang enjoys a semi-humid tropical climate with good sunshine and rainfall. The geographical features of Xianyang include mountains, hills, plains and gullies. In mountainous regions, some villages can only be reached by foot.

JE diagnosis

JE is a disease that is legally mandated for reporting in China. All reported JE cases in this study were confirmed according to the WHO's criteria.⁵ Specifically, laboratory diagnosis of JE virus infection should be performed by using a JE virus-specific IgM-capture ELISA. JE virus-specific IgM can be measured in serum by 7 days after the onset. A ≥ 4 -fold rise of JE virus-specific neutralizing antibodies in the serum of acute phase, compared to the convalescent phase, may be used to confirm recent infection.

To ensure diagnostic reliability, confirmatory serological testing was performed on all patients with suspected JE by detecting specific measles IgM antibodies with a commercially available ELISA kit (Yanhui Biotechnology Company, China), according to the manufacturer's instruction. Patients with a ≥ 4 -fold rise of JE virus-specific neutralizing IgM of the acute phase, compared to the convalescent phase, may be diagnosed as JE.

Data collection and management

Monthly incidence data on JE for the period from Jan 2005 to Sep 2014 were obtained from a passive surveillance system at the Center for Diseases Prevention and Control in Xianyang, Shaanxi province. The well-established Disease Prevention and Control Information System could ensure the integrity, accuracy and reliability of the data used in this study. The population information was obtained from the Statistical Yearbook of Shaanxi Province from 2005 to 2014.

Stability test of the incidence data

To explore information on variance stability and check for seasonal effects, the plot of monthly incidence was examined and an Augmented Dickey-Fuller (ADF) test was used.¹⁵

SARIMA model development

An optimal SARIMA model was developed for JE incidence from 2005 to 2013, with the Box and Jenkins approach established in 1974.¹⁶ This SARIMA model can predict JE incidence for the year 2014 and 2015.

SARIMA was modeled by the following four stages:¹⁷⁻¹⁸

Firstly, SARIMA requires a stationary time series, so that they could be tested for variance stability. If the variance is too large and unstable, an appropriate transformation, like logarithmic, square root or difference equations, can be used to stabilize the variance and the mean according to the profile of the original time series. In the modeling process, the seasonality and the trend are both highlighted by plotting the original time series.

Secondly, the SARIMA (p, d, q) (P, D, Q)_s model was fitted, where p is the order of auto regression (AR), d is the order of integration, q is the order of moving average (MA), P is the order of seasonal AR (SAR), D is the order of seasonal integration, Q is the order of seasonal MA (SMA), and s is the length of seasonal period. Then, the method with the maximum likelihood was employed to estimate the parameters of the SARIMA model. The equation of SARIMA (p, d, q) (P, D, Q)_s model can be expressed as:

$$\nabla^d \nabla_s^D = \frac{\Theta(B)\Theta_s(B)}{\Phi(B)\Phi_s(B)} \varepsilon_t$$

In this equation, D is the order of seasonal integration; d is the order of difference to extract all the information of the trend; ε_t is the estimated residual at time t with zero mean and constant variance; s is the length of the seasonal period.

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p;$$

$$\Phi_s(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps};$$

$$\Theta(B) = 1 + \Theta_1 B + \Theta_2 B^2 + \dots + \Theta_q B^q;$$

$$\Theta_s(B) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_q B^{qs}.$$

Thirdly, the following four tools were used to choose the best model: (1) A plot of JE incidence, which could provide information on the need for non-seasonal and seasonal differencing. (2) The autocorrelation (ACF) and partial autocorrelation (PACF) functions, which contain the temporal dependence structure information for the time series. (3) The residual ACF and p -values were used for the Portmanteau test to diagnose and identify the model.

(4) Some other measures or indexes, based on information theory, like Bayesian information criterion (BIC), Akaike information criterion (AIC), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and Theil Inequality Coefficient (TIC) were used to achieve an optimal model choice by balancing an adequate prediction and a minimum, or maximum of number of parameters. Meanwhile, residuals were analyzed by the Ljung-Box test to validate the final model, and residuals must be equivalent to white noise. Finally, the best fit model was identified.

Fourthly, the best fit model was used for forecasting the incidence or the number of cases of the next period.

Model construction in this study

This study was based on time series data of JE incidence covering the period from Jan 2005 to Sep 2014. Before the model was developed, a line plot was drawn to understand the trend of the incidence. From this line plot, we could find that the incidence had a strong seasonal trend and certain periodicity. Consequently, we considered that a SARIMA was needed to be established. Based on the four steps of the SARIMA fitting, possible SARIMA models with seasonal difference, were fitted to data series from Jan 2005 to Dec 2013. In order to compare the adequacy and performance of the possible SARIMA models constructed, residual analysis (Q-test) was conducted to determine whether white noise exists in the residuals or not. When the residuals are all centering on zero randomly in the ACF and PACF, the residuals are statistically independent (white noise) and the appropriateness of the model can be assured. Meanwhile, the AIC, BIC, and MAPE were calculated. When we could not judge which model was clearer between the PACF and ACF, forecasting models with the smallest AIC and BIC values could be selected as the final one. Graphs for impulse response and accumulated response were plotted to detect the stability of the model. MAE and MAPE represent the relative scale of error due to forecasting, which means the smaller the error, the more accurate the forecast. Then the forecasting accuracy of this chosen model was also verified using the data from January 2014 to September 2014, and 95% CI of the fitted values from 2014 to 2015. All statistical analyses were conducted with Eviews Software (IHS EViews, Irvine, CA, USA).¹⁹

RESULTS

The basic status of reported incidence of JE in Xianyang from Jan 2005 to Sep 2014

The monthly JE incidence of Xianyang covering the period from Jan 2005 to Sep 2014 demonstrated periodic fluctuations and there was a strong seasonal fluctuation that

peaked from June to August each year. On the other hand, the incidence data presented a long-term trend of decline.

Stability of the incidence data

The plots of monthly incidence showed a strong long-term trend. Test results of ACF, PACF and ADF showed that the original time series were not stationary ($ADF = -1.5603 > 1\%$ level critical value -3.493747 , $P < 0.01$). To achieve a stationary time series, monthly incidences were seasonally differentiated by replacing each observed incidence with the first order difference between itself and the observation of previous 12 months (s in the model was 12), and then ADF test was used again. After applying the seasonal difference, we achieve a stationary time series ($ADF = -12.0271 < 1\%$ level critical value -3.493747 , $P > 0.01$). From Figure 1 we can see that the decline and period trend of the incidence were relieved partially after the seasonal first order difference.

Establishment and identification of SARIMA ($p, 1, q$)($P, 1, Q$)₁₂ model

The data set from Jan 2005 to Dec 2013 was used to model fitting. For determining the main parameters ($p, d, q, P, D, and Q$) of SARIMA, ACF and PACF of monthly JE incidence after the seasonal first order difference was drawn (Fig. 1).

From ACF in Figure 1, we can see that the partial autocorrelation coefficient of the incidence sequence was rapidly tending towards zero after $k=2$, so we can consider it as $p=2$ or $p=1$. Similarly, the autocorrelation coefficient from PACF was rapidly changing to zero after $k=1$; from

this we can regard it as $q=1$. The parameters of P and Q were rarely more than 2, which meant that $P=0, 1$ or 2 and $Q=0, 1$ or 2 . Accordingly, we might obtain 18 models (showed in Table 1). The key indexes which could be used to judge the power of the models are seen in Table 1. From Table 1, SARIMA (1, 1, 1) (2, 1, 1)₁₂ was considered to be the best model (Fig. 2).

Figure 1. AC and PAC function of monthly incidence of Japanese Encephalitis in Xianyang, after the seasonal first order difference.

In figure 1, 1-24 means number of lags.

AC: autocorrelation coefficient; PAC: partial autocorrelation coefficient; Q-stat: value of Q-test; Prob: P value.

Table 1. Key measures to judge the powers of the models

Models	R^2	Adjusted R^2	BIC	AIC	MAE	MAPE	White noise
SARIMA(1, 1, 1)(0, 1, 0) ₁₂	0.3593	0.3523	-1.5550	-1.6091	0.0448	42.4177	No
SARIMA(1, 1, 1)(0, 1, 1) ₁₂	0.5253	0.5149	-1.8066	-1.8878	0.0349	20.2465	Yes
SARIMA(1, 1, 1)(0, 1, 2) ₁₂	0.7281	0.7281	-2.4714	-2.4983	0.0298	25.5302	No
SARIMA(1, 1, 1)(1, 1, 0) ₁₂	0.3921	0.3767	-1.7403	-1.8284	0.0388	48.3811	Yes
SARIMA(1, 1, 1)(1, 1, 1) ₁₂	0.5767	0.5604	-2.0486	-2.1660	0.0355	43.4823	Yes
SARIMA(1, 1, 1)(1, 1, 2) ₁₂	0.5229	0.5110	-3.1737	-2.0836	0.0336	34.8344	No
SARIMA(1, 1, 1)(2, 1, 0) ₁₂	0.6557	0.6507	-3.0530	-3.1167	0.0205	18.4136	Yes
SARIMA(1, 1, 1)(2, 1, 1) ₁₂	0.7675	0.7569	-3.3085	-3.4370	0.0181	21.1313	Yes
SARIMA(1, 1, 1)(2, 1, 2) ₁₂	0.7142	0.7012	-3.1024	-3.2309	0.0186	17.0829	Yes
SARIMA(2, 1, 1)(0, 1, 0) ₁₂	0.2399	0.2315	-1.3727	-1.4271	0.0670	45.8096	No
SARIMA(2, 1, 1)(0, 1, 1) ₁₂	0.4602	0.4482	-1.6661	-1.7478	0.0490	31.168	No
SARIMA(2, 1, 1)(0, 1, 2) ₁₂	0.8080	0.8038	-2.7001	2.7818	0.0327	24.6900	No
SARIMA(2, 1, 1)(1, 1, 0) ₁₂	0.3187	0.3012	-1.6125	-1.7012	0.0543	48.0692	No
SARIMA(2, 1, 1)(1, 1, 1) ₁₂	0.5012	0.4817	-1.8700	-1.9982	0.0505	34.5438	No
SARIMA(2, 1, 1)(1, 1, 2) ₁₂	0.8623	0.8569	-3.1569	-3.2752	0.0258	22.5722	No
SARIMA(2, 1, 1)(2, 1, 0) ₁₂	0.6558	0.6454	-2.9607	-3.0578	0.0211	48.9294	Yes
SARIMA(2, 1, 1)(2, 1, 1) ₁₂	0.7342	0.7220	-3.1580	-3.2874	0.0248	21.8292	No
SARIMA(2, 1, 1)(2, 1, 2) ₁₂	0.6763	0.6668	-3.3054	-3.1503	0.0278	16.7768	Yes

BIC: Bayesian information criterion; AIC: Akaike information criterion; MAE: Mean Absolute Error; MAPE: Mean Absolute Percent Error.

Figure 2. AC and PAC function of SARIMA (1, 1, 1) (2, 1, 1)₁₂.

In figure 2, 1-24 means number of lags.

Analysis and examination of SARIMA (1, 1, 1) (2, 1, 1)₁₂

In the SARIMA time series analysis, the optimal model generated from the incidence data set was SARIMA (1, 1, 1) (2, 1, 1)₁₂. The estimates of model parameters and their testing results are presented in Table 2. All the parameters are statistically significant except for the constant. The model equation was:

$$\nabla \nabla_{12} \text{incidence} = \frac{(1 + 0.2217B)(1 + 0.9290B^{12})}{(1 + 0.3323B)(1 - 0.9975B^{24})} \varepsilon_t$$

Table 2. Parameters of SARIMA (1, 1, 1) (2, 1, 1)₁₂

Variables	Coefficient	Std. error	t-Statistic	P value
Non-Seasonal AR(1)	0.3323	0.1271	2.6154	0.0110
Lags MA(1)	-0.2217	0.0473	-4.6867	0.0000
Seasonal SAR(2)	-0.9975	0.0001	-906103.6	0.0000
Lags SMA(1)	-0.9290	0.0418	-22.2026	0.0000

AR(1): an autoregressive model with 1 degree of differencing; MA(1): moving-average model with the first order; SAR(2): an seasonal autoregressive model with 2 degree of differencing; SMA(1): an seasonal moving-average model with the first order. R-square=0.7675; Adjusted R-square=0.7569; Akaike info criterion=-3.4370; Root Mean Squared Error=0.040984; Mean Absolute Error=0.01808; Mean Absolute Percent Error=21.1313; Theil Inequality Coefficient=0.2272; Bias Proportion=0.00001.

Assessment and diagnosis of SARIMA (1, 1, 1) (2, 1, 1)₁₂

The SARIMA model required that the ACF and PACF of the residuals were considered white noise to believe that the information of the original time series was extracted completely. From Figure 2, we can see that the ACF and PACF of the residuals of SARIMA (1, 1, 1) (2, 1, 1)₁₂ were white

noise and met this requirement. SARIMA (1, 1, 1)(2, 1, 1)₁₂ had 38 inverted roots, including 25 from autoregressive model and 13 from moving-average model, and all the roots were in the unit. What's more, Figure 3 suggested that the trend of actual incidence was in agreement with the theoretical incidence. Graphs for impulse response and accumulated response (Fig. 4) could show that the volatility of the impulse response was not obvious. The fitted monthly incidence of JE from Jan 2014 to Nov 2015 is shown in Figure 5, from which we could find that the fitted incidences were very consistent with the actual incidences except for two values with errors. Meanwhile, the fitting curve

Figure 3. Correlation of the actual incidence and theoretical incidence.

The abscissa represents the number of lags, and the ordinate represents the autocorrelation coefficient and partial autocorrelation coefficient.

Figure 4. Graphs for impulse response and accumulated response.

The abscissa represents the number of lags, and the ordinate represents the impulse response and accumulated response. The blue line means the average response, the red line above means the average response+2SE, the red line below means the average response-2 SE.

was basically in the range of 95%CI. The above results indicated that SARIMA (1, 1, 1) (2, 1, 1)₁₂ was stationary and inverted. Therefore, SARIMA (1, 1, 1) (2, 1, 1)₁₂ could be considered as an effective model to predict the JE incidence of Xianyang, Shaanxi Province, China.

Prediction of incidence by SARIMA (1, 1, 1) (2, 1, 1)₁₂

Figure 5. Observed incidence and fitted incidence of Japanese Encephalitis in Xianyang from Jan 2005 to Nov 2015.

The ordinate represents incidence of Japanese Encephalitis.

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

Table 3. Comparison of observed incidences and forecasting incidences from SARIMA (1, 1, 1) (2, 1, 1)₁₂

Months	Observed incidence (1/10 ⁵)	Forecast incidence (1/10 ⁵)	Upper limit of 95%CI	Low limit of 95%CI
Jan-14	0.0000	-0.0062	0.1413	-0.1538
Feb-14	0.0000	-0.0072	0.1430	-0.1575
Mar-14	0.0000	-0.0073	0.1431	-0.1577
Apr-14	0.0000	-0.0072	0.1432	-0.1576
May-14	0.0000	-0.0071	0.1434	-0.1575
Jun-14	0.0000	-0.0069	0.1435	-0.1574
Jul-14	0.0000	0.0220	0.1725	-0.1284
Aug-14	0.0203	0.1505	0.3010	0.0001
Sep-14	0.0203	0.0451	0.1955	-0.1054
Oct-14	-	-0.0064	0.1440	-0.1568
Nov-14	-	-0.0063	0.1442	-0.1567
Dec-14	-	-0.0061	0.1443	-0.1566
Jan-15	-	-0.0101	0.1705	-0.1907
Feb-15	-	-0.0106	0.1711	-0.1924
Mar-15	-	-0.0106	0.1713	-0.1924
Apr-15	-	-0.0103	0.1715	-0.1922
May-15	-	-0.0101	0.1717	-0.192
Jun-15	-	-0.0099	0.1719	-0.1917
Jul-15	-	0.0010	0.1828	-0.1809
Aug-15	-	0.1358	0.3176	-0.046
Sep-15	-	0.0427	0.2245	-0.1391
Oct-15	-	-0.0090	0.1728	-0.1908
Nov-15	-	-0.0088	0.1731	-0.1906
Dec-15	-	-0.0085	0.1733	-0.1903

"-" means no data.

Root Mean Squared Error=0.0906; Mean Absolute Error=0.0435; Mean Absolute Percent Error=21.3892; Theil Inequality Coefficient=0.9218; Bias Proportion=0.1056.

We attempted to predict the monthly JE incidences in 2014 using the SARIMA model, and compared them with the actual numbers (Table 3). MAE and MAPE were all low, and all of the forecasting incidences were in the range of 95% CI. We also found that the monthly JE incidences in Xianyang also showed a seasonal pattern; the incidence was highest from June to August and very low in other months in 2014 and 2015.

DISCUSSION

Early recognition of health related problems, especially infectious diseases, is highly important for their control and prevention. Thanks to the efforts of Box and Jenkins, the SARIMA models have become one of the most popular statistical approaches for predicting the occurrence of infectious diseases.¹⁷ The variations in annual incidences indicated that there was a seasonal variation in JE incidence in China. The monthly incidence data demonstrated a peak from June to August, and roughly 0 in other months, which provided the possibility to establish a model to forecast the incidence in the near future. Hence, we developed an optimal SARIMA model with the incidence data from January 2005 to December 2013 to predict monthly incidences of JE, in order to improve the surveillance system of JE in Xianyang City. Computer results showed that the multivariate SARIMA (1, 1, 1) (2, 1, 1)₁₂ had the best stability, with a comparison of the models' forecasting accuracy, which proved it to be the most accurate, with the lowest AIC, BIC, MAE values, the highest R^2 , and a lower MAPE.

The results in this study indicated that incidence of JE showed a decline in the long term, and a obvious seasonal change, which is highly similar to studies from other countries, such as India, Korea, and Australia.^{2-3, 18-21} One reason for this decline might be that the vaccines are available to improve immunization of the population and increase awareness to help control and prevent infectious diseases. On the other hand, the temperature and rainfall are believed to be the two major climatic factors affecting the replication and propagation of JE. Heffelfinger *et al*²² also reported that the genotype of JE was associated with climate, for example, GIII is significantly associated with temperate climates. The life cycle and population density of mosquitoes and reproduction rate of JE in mosquitoes could be affected by climate.⁴⁻⁵ Some researchers found that increased temperature and rainfall can reduce the maturation time of mosquito larvae, and rapidly increase the population density of mosquitoes.

Although our SARIMA model was based only on reported incidence of JE, which meant the real JE incidence of Xianyang could not be readily obtained, the web-based and case-based mandatory JE reporting system had been fully operational since 2002 and covered 100% of counties of Xianyang City. Therefore, the reported JE incidences should closely mirror the actual figures. What's more, levels of the incidence of JE in our results were very similar to other cities, such as Xi'an. Based upon the results of our study, we could believe that there would be no obvious

decrease in the incidence of JE in Xianyang in the near future. The results also indicated that the reported annual JE incidence in Xianyang City would decrease slightly.

When it comes to JE prevention, continuing to improve the vaccination coverage; especially for children from 5 to 15 years old, may play an important role in decreasing incidence numbers in our city, and even in our whole country, in the long run. Secondly, strengthening hygienic awareness of the whole society and eliminating mosquitoes will help control and prevent JE. However, possible biomechanics for the seasonal variations in JE incidence need to be further studied.

There are several limitations in present study. First, climate-related factors were not incorporated in the model construction because the data for these factors are still unavailable. The second limitation is that the models established upon the data from 2005 to 2013 were only verified against the data of 2014, which could introduce some errors. The third limitation is that we used a single SARIMA model. Although we chose the most accurate one, there is still the probability that a better hybrid model exists. The data collected and used in this paper is only from Xianyang, which was limited by the authority of Disease Prevention and Control Information System, and we can't obtain more data from a larger region. This is another limitation of this study. However, this study demonstrates the feasibility of the ARIMA model in the process of JE prevention, and the methods and conclusions of the present study can still provide a valuable reference for JE prevention in other areas.

In summary, as the result of our comparison of the possible models that could be used to predict the incidence numbers of JE in Xianyang, PR China, the model fitted in the present study incorporated explanatory variables (seasonal) and was the most appropriate for forecasting JE incidence in Xianyang. We have proved that SARIMA (1, 1, 1) (2, 1, 1)₁₂ appears to be reliable and accurate, and could be applied for incidence prediction. The model in our study showed that the season should be considered as an important factor affecting the occurrence of JE. The proposed prediction model can provide clues to early identification of the incidence that increases abnormally ($\geq 0.4/100\ 000$). According to the predicted results from 2014 to 2015, the JE incidence in Xianyang would decline slightly and reach its peak below 0.2/100 000 from June to August. However, our results still need to be re-examined with series data of a longer time.

Conflict of Interest Statement

The authors have no conflict of interest to disclose.

Acknowledgements

The authors wish to thank the staff from the CDCs from 13 counties of Xi'an Yang, Shaanxi province, China, for their contribution to Japanese encephalitis cases reporting. The authors would specifically like to thank Yongmin Xiong from Xi'an Jiaotong University, Yan Yang from CDC of Qindu District, Yajuan Jing from CDC of Weicheng District and Yaya Feng from CDC of Wugong County for their research support and help.

REFERENCES

1. Diagana M, Preux PM, Dumas M. Japanese encephalitis revisited. *J Neurol Sci* 2007; 262(1-2):165-70. doi: 10.1016/j.jns.2007.06.041.
2. Ghimire S, Dhakal S. Japanese encephalitis: challenges and intervention opportunities in Nepal. *Vet World* 2015; 8(1):61-5. doi: 10.14202/vetworld.2015.61-65.
3. Misra UK, Kalita J. Overview: Japanese encephalitis. *Prog Neurobiol* 2010; 91(2):108-20. doi: 10.1016/j.pneurobio.2010.01.008.
4. Jasik KP, Okła H, Słodki J, Rozwadowska B, Słodki A, Ruppik W. Congenital tick borne diseases: is this an alternative route of transmission of tick-borne pathogens in mammals? *Vector Borne Zoonotic Dis* 2015; 15(11):637-44. doi: 10.1089/vbz.2015.1815.
5. Ai J, Xie Z, Liu G, Chen Z, Yang Y, Li Y, et al. Etiology and prognosis of acute viral encephalitis and meningitis in Chinese children: a multicentre prospective study. *BMC Infect Dis* 2017; 17(1):494. doi: 10.1186/s12879-017-2572-9.
6. Feng H, Duan G, Zhang R, Zhang W. Time series analysis of hand-foot-mouth disease hospitalization in Zhengzhou: establishment of forecasting models using climate variables as predictors. *PLoS One* 2014; 9(1):e87916. doi: 10.1371/journal.pone.0087916.eCollection2014.
7. Cappelle J, Duong V, Pring L, Kong L, Yakovlev M, Prasetyo DB, et al. Intensive circulation of Japanese Encephalitis virus in peri-urban Sentinel Pigs near Phnom Penh, Cambodia. *PLoS Negl Trop Dis* 2016; 10(12):e0005149. doi: 10.1371/journal.pntd.0005149.
8. Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Serv Res* 2017; 17(1):469. doi: 10.1186/s12913-017-2407-9.
9. Bozkurt ÖÖ, Biricik G, Tayşi ZC. Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. *PLoS One* 2017; 12(4):e0175915. doi: 10.1371/journal.pone.0175915.
10. Sharafi M, Ghaem H, Tabatabaee HR, Faramarzi H. Forecasting the number of zoonotic cutaneous leishmaniasis cases in south of Fars province, Iran using seasonal ARIMA time series method. *Asian Pac J Trop Med* 2017; 10(1):79-86. doi: 10.1016/j.apjtm.2016.12.007.
11. Yousefzadeh-Chabok S, Ranjbar-Taklimie F, Malekpouri R, Razzaghi A. A time series model for assessing the trend and forecasting the road traffic accident mortality. *Arch Trauma Res* 2016; 5(3):e36570. doi: 10.5812/atr.36570.
12. Azeez A, Obaromi D, Odeyemi A, Ndege J, Muntabayi R. Seasonality and trend forecasting of tuberculosis prevalence data in Eastern Cape, South Africa, using a hybrid model. *Int J Environ Res Public Health* 2016; 13(8). pii: E757. doi: 10.3390/ijerph13080757.
13. Zhang G, Huang S, Duan Q, Shu W, Hou Y, Zhu S, et al. Application of a hybrid model for predicting the incidence of tuberculosis in Hubei, China. *PLoS One* 2013; 8(11):e80969. doi: 10.1371/journal.pone.0080969.
14. Kam HJ, Sung JO, Park RW. Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthc Inform Res* 2010; 16(3):158-65. doi: 10.4258/hir.2010.16.3.158.
15. Guo Chen, Amy K. Glasmeier, Min Zhang, et al. Urbanization and income inequality in post-reform China: a causal analysis based on time series data. *PLoS One* 2016; 11(7):e0158826. doi: 10.1371/journal.pone.0158826.
16. Box GE, Jenkins GM. Time series analysis: forecasting and control. editors. Oakland: Holden-Day; 1976. p. 300-33.
17. Ren X, Fu S, Dai P, Wang H, Li Y, Li X, et al. Pigsties near dwellings as a potential risk factor for the prevalence of Japanese encephalitis virus in adult in Shanxi, China. *Infect Dis Poverty* 2017; 6(1):100. doi: 10.1186/s40249-017-0312-4.
18. Yu L, Zhou L, Tan L, Jiang H, Wang Y, Wei S, et al. Application of a new hybrid model with seasonal auto-regressive integrated moving average (ARIMA) and nonlinear auto-regressive neural network (NARNN) in forecasting incidence cases of HFMD in Shenzhen, China. *PLoS One* 2014; 9(6):e98241. doi: 10.1371/journal.pone.0098241.
19. IHS EViews EViews 7. [cited 2016 Nov 1]. Available from: <http://www.eviews.com/download/download.html#eviews7> (English).
20. Wang X, Li SH, Zhu L, Nian QG, Yuan S, Gao Q, et al. Near-atomic structure of Japanese encephalitis virus reveals critical determinants of virulence and stability. *Nat Commun* 2017; 8(1):14. doi: 10.1038/s41467-017-00024-6.
21. Parmar KS, Bhardwaj R. Statistical, time series, and frac-

- tal analysis of full stretch of river Yamuna (India) for water quality management. *Environ Sci Pollut Res Int* 2015; 22(1):397-414. doi: 10.1007/s11356-014-3346-1.
22. Heffelfinger JD, Li X, Batmunkh N, Grabovac V, Diorditsa S, Liyanage JB, et al. Japanese Encephalitis Surveillance and Immunization-Asia and Western Pacific Regions, 2016. *MMWR Morb Mortal Wkly Rep* 2017; 66(22):579-83. doi: 10.15585/mmwr.mm6622a3.