

The research of SARIMA model for prediction of hepatitis B in mainland China

Daren Zhao, MM^a, Huiwu Zhang, MM^{a,*} , Qing Cao, MM^b, Zhiyi Wang, MM^c, Ruihua Zhang, MM^d

Abstract

Hepatitis B virus infection is a major global public health concern. This study explored the epidemic characteristics and tendency of hepatitis B in 31 provinces of mainland China, constructed a SARIMA model for prediction, and provided corresponding preventive measures.

Monthly hepatitis B case data from mainland China from 2013 to 2020 were obtained from the website of the National Health Commission of the People's Republic of China. Monthly data from 2013 to 2020 were used to build the SARIMA model and data from 2021 were used to test the model.

Between 2013 and 2020, 9,177,313 hepatitis B cases were reported in mainland China. SARIMA(1,0,0)(0,1,1)₁₂ was the optimal model and its residual was white noise. It was used to predict the number of hepatitis B cases from January to December 2021, and the predicted values for 2021 were within the 95% confidence interval.

This study suggests that the SARIMA model simulated well based on epidemiological trends of hepatitis B in mainland China. The SARIMA model is a feasible tool for monitoring hepatitis B virus infections in mainland China.

Abbreviations: AdaBoost = AdaBoost with decision tree, ARIMA = Autoregressive Integrated Moving Average Model, GM (1, 1) = Grey Model First Order One Variable, HBV = Hepatitis B virus, PMTCT = prevention of mother-to-child transmission, XGBoost = extreme gradient boosting decision tree.

Keywords: epidemic, hepatitis B, periodicity, prediction, SARIMA model, seasonality

1. Introduction

Hepatitis B, caused by the hepatitis B virus (HBV) infection, is transmitted through contact with percutaneous or per mucosal exposure to infected blood or body fluids.^[1,2] Although there are effective prophylactic vaccines and treatments for hepatitis B, they still face a high level of morbidity and mortality due to hepatitis B and a tremendous economic burden.^[3] It has been reported that 3.5% of the global population is infected with chronic hepatitis B virus.^[4] In 2013, there were 1.45 million deaths due to viral hepatitis, which was the seventh highest cause of mortality worldwide.^[5,6] Between 2015 and 2019, the number of confirmed cases of hepatitis B increased from 257 million to 296 million, with 1.5 million new infections each year.^[7,8]

Hepatitis B remains a global epidemic, particularly in Southeast Asia, sub-Saharan Africa, and China.^[9] In China, hepatitis B is classified as a Category B infectious disease. Over the past few decades, a series of comprehensive preventive measures in China, including HBV immunization programs, integrated prevention of mother-to-child transmission, and effective finance policies,^[10–12] have been implemented effectively and have achieved remarkable progress. However, it has been reported that China has a higher intermediate prevalence (5%–7.99%)^[13]; therefore, there are still many challenges in the prevention and control of hepatitis B.

Scientific prediction and analysis play a crucial role in the prevention and control of hepatitis B, allocation of health resources properly or methodically, provision of effective measures, and the promotion of public health.^[3] Meanwhile,

This study was funded by National Health Commission of the People's Republic of China (Grant No. YLZLXZ-2021-004).

The funders played no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

The authors have no conflicts of interests to disclose.

The datasets generated or analyzed during this study are included in this published article. The corresponding author may provide specified analyses or fully de-identified parts of the dataset upon reasonable request.

^a Department of Medical Administration, Sichuan Provincial Orthopedics Hospital, Chengdu, Sichuan, China., ^b Department of Medical Administration, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, Sichuan, China., ^c Department of Medical Administration, Sichuan Cancer Hospital & Institute, Chengdu, Sichuan, China., ^d School of Management, Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China.

* Correspondence: Huiwu Zhang, Department of Medical Administration, Sichuan Provincial Orthopedics Hospital, No. 132, West Section, First Ring Road, Chengdu, Sichuan 610041, China. (e-mail: cdzhanghuiwu@163.com).

Copyright © 2022 the Author(s). Published by Wolters Kluwer Health, Inc.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Zhao D, Zhang H, Cao Q, Wang Z, Zhang R. The research of SARIMA model for prediction of hepatitis B in mainland China. *Medicine* 2022;101:23(e29317).

Received: 20 September 2021 / Received in final form: 31 March 2022 / Accepted: 29 April 2022

<http://dx.doi.org/10.1097/MD.00000000000029317>

a reasonable prediction of hepatitis B can give full play to early warning links in advance to provide corresponding preventive measures.^[14] Therefore, it may remind us the prediction and analysis of hepatitis B are crucial for prevention and control.

A time series is a statistical method used to predict future trends based on historical data and time variables.^[15,13] Currently, various prediction approaches have been adopted to handle the temporal characteristics of infectious diseases,^[14] such as the exponential smoothing model,^[16] grey model first-order one-variable (GM(1,1) model),^[17] linear regression,^[18] and autoregressive integrated moving average model (ARIMA).^[19] The ARIMA model is widely used to predict infectious diseases. SARIMA is an extended version of the ARIMA model, which requires periodicity, seasonality, and randomness.

In this study, we collected monthly hepatitis B case data from 31 provinces in mainland China from 2013 to 2020, and applied the SARIMA model to construct a prediction model. The SARIMA(1,0,0)(0,1,1)₁₂ model was fitted to predict the hepatitis B epidemic trends in 2021.

2. Materials and methods

2.1. Data source

Monthly HBV case data for 31 provinces in mainland China from 2013 to 2020 were provided by the National Health Commission of the People’s Republic of China (<http://www.nhc.gov.cn/>). Hepatitis B is classified as a category B infectious disease since the “Regulations of the People’s Republic of China on the Administration of Acute Infectious Diseases” promulgated by the Ministry of Health of the People’s Republic of China in 1978. The diagnostic criteria for hepatitis B (WS 299–2008 Diagnostic Criteria for Hepatitis B) were formulated by the Ministry of Health of the People’s Republic of China (<http://www.nhc.gov.cn/wjw/s9491/200907/41983.shtml>). In China, once the diagnosis of hepatitis B is confirmed, physicians must report to the Internet-based surveillance system within 24 hours in local hygiene departments.

From 2013 to 2020 were used as basic data and materials for the study (Table 1). The research database was divided into training and validation datasets. Monthly data from 2013 to 2020 were used as the training set to build the SARIMA model, and monthly data from 2021 were used as the validation set to test the model.

2.2. Ethics statement

In this study, monthly hepatitis B case data for 31 provinces in mainland China from 2013 to 2020 were obtained from publicly

accessible sources (<http://www.nhc.gov.cn/>). The data were relatively uninvolved in detailed patient personal information. Therefore, this study did not require a formal ethical assessment.

2.3. SARIMA model

The SARIMA model, which is an extended version of the ARIMA model, is generally expressed as SARIMA (p, d, q) (P, D, Q).^[20] However, a distinctive difference is that the SARIMA model includes seasonal characteristics.^[21] In the expression, p, d, and q represent the order of auto-regression, the degree of trend difference, and the order of moving average, respectively; P, D, and Q represent the seasonal auto regression lag, degree of seasonal difference, and seasonal moving average, respectively; and s represents the length of the seasonal period,^[22] which was evaluated as 12 in our study.

The modeling process of the SARIMA model consists of 3 parts: series stability, model identification and estimation, and model diagnosis.^[23] Firstly, 2 methods can be used to solve this problem in the series stability process. One is the original sequence plot, which was observed by viewing the fitted curve to determine whether it was constant. The other is the augmented Dickey-Fuller test, which confirms the series stability. The ADF test was performed using the EViews 10.0 software. Second, model identification and estimation were determined by using autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. On this basis, the model parameters of the p, q, P, Q, and candidate SARIMA models were determined.^[24] Thirdly, model diagnosis was evaluated using a white noise test. The Ljung-Box Q test was used to determine whether the residuals were independently or normally distributed. A t-test and a P value were used to determine whether the model parameters were statistically significant. Finally, using the lowest Bayesian information criterion of Schwarz (BIC) values and their residual white noise, we identified the optimal model to predict future epidemic trends of hepatitis B.

2.4. Assessment of prediction performance

In this section, the prediction performance is evaluated using a mathematical formula that describes the goodness-of-fit between the predicted and observed values.^[23]

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (X_t - \hat{X}_t)^2}{n}} \tag{1}$$

Table 1
The number of Hepatitis B cases from 2013 to 2020 in mainland China.

Year	Month												Total
	January	February	March	April	May	June	July	August	September	October	November	December	
2013	102,367	78,884	107,535	97,225	96,978	86,335	97,401	97,430	87,915	86,161	88,479	87,609	1,114,319
2014	90,210	83,068	99,292	94,768	91,936	88,201	95,648	94,075	87,827	85,996	85,125	88,397	1,084,543
2015	96,649	72,869	104,427	94,350	91,194	89,224	93,586	89,228	89,806	86,393	87,284	90,103	1,085,113
2016	89,699	82,204	105,745	93,190	95,079	90,166	91,219	97,670	87,390	85,480	91,478	91,371	1,100,691
2017	86,657	99,417	110,717	98,123	101,783	100,155	98,501	103,977	96,856	89,105	97,694	97,560	1,180,545
2018	109,021	86,886	120,659	106,398	108,831	98,491	103,809	105,068	95,312	94,217	100,849	96,336	1,225,877
2019	107,754	90,985	113,941	110,266	106,431	97,362	112,454	106,985	97,815	98,774	102,174	102,151	1,247,092
2020	91,026	51,506	88,150	101,262	97,651	99,319	106,135	102,304	105,377	95,633	100,561	100,209	1,139,133

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{X_t - \hat{X}_t}{X_t} \right|}{n} \times 100 \tag{2}$$

$$MAE = \frac{\sum_{t=1}^n |X_t - \hat{X}_t|}{n} \tag{3}$$

where, \hat{X}_t is the predicted value, X_t is the observed value, and n is the sequence sample size.

2.5. Statistical analysis

EViews 10.0 and SPSS software (version 23.0) were used to construct the SARIMA model. The significance level was set at $P < .05$.

3. Results

3.1. Description of hepatitis B cases from 2013 to 2020 in mainland China

A total of 9,177,313 hepatitis B cases were reported in mainland China between 2013 and 2020, with an average annual growth rate of 0.05% (Table 1 and Fig. 1). Figure 2 shows that the lowest point of the monthly number of hepatitis B cases each year from 2013 to 2020 was February, whereas the highest point was March. It was dominated by seasonal and periodic fluctuations.

3.2. Construction of SARIMA model

As shown in Fig. 3, the original sequence seems to be non-stationary, suggesting that trend differenced and seasonally differenced or augmented Dickey-Fuller tests should be per-

formed. After the natural logarithm transformation, first-order differenced, and seasonally differenced with one period of 12 (Fig. 4), the time series was stationary. Thus, parameters $d, D,$ and s were 0, 1, and 12, respectively. Furthermore, the ADF test results ($t = -4.734, P < .001$) indicated that after the natural logarithm transformation and differenced, the time series was stationary.

Figures 5 and 6 shows the results of the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. Based on the analysis of the ACF plot, after a first-order differenced, the autocorrelation coefficient values exhibited slow decay. Meanwhile, the ACF was maximal in order 1, and it was significantly higher than that in order 2 or 3; therefore, we identified the parameter q as 0 or 1 and p as 1. Similarly, the PACF plot suggests that the parameters of Q as 1 and P are 0 or 1 because the values of the autocorrelation coefficients trail with a slow decay, and its maximum is on order 1 among orders 2 and 3. As described above, the parameters of the $p, q, P, Q,$ and candidate SARIMA models are determined (Table 2).

Further analysis by applying the white noise test revealed that the candidate SARIMA models passed the Ljung-Box Q test ($P > .05$). Based on the results of the estimates and standard errors of the candidate SARIMA model parameters (Table 3), as well as for the lowest Bayesian information criterion of Schwarz (BIC) values, we found that SARIMA(1,0,0)(0,1,1)12 was the optimal model, which passed the t -test ($P < .001$), and its residual was white noise (Fig. 7). Moreover, the predicted and observed values from 2013 to 2020 in mainland China were well simulated by the SARIMA(1,0,0)(0,1,1)12 model (Fig. 8).

Therefore, SARIMA(1,0,0)(0,1,1)12 was used to predict the number of HBV cases in mainland China between January and December 2021. The prediction results are presented in Table 4. Moreover, the predicted values were within the 95% confidence interval.

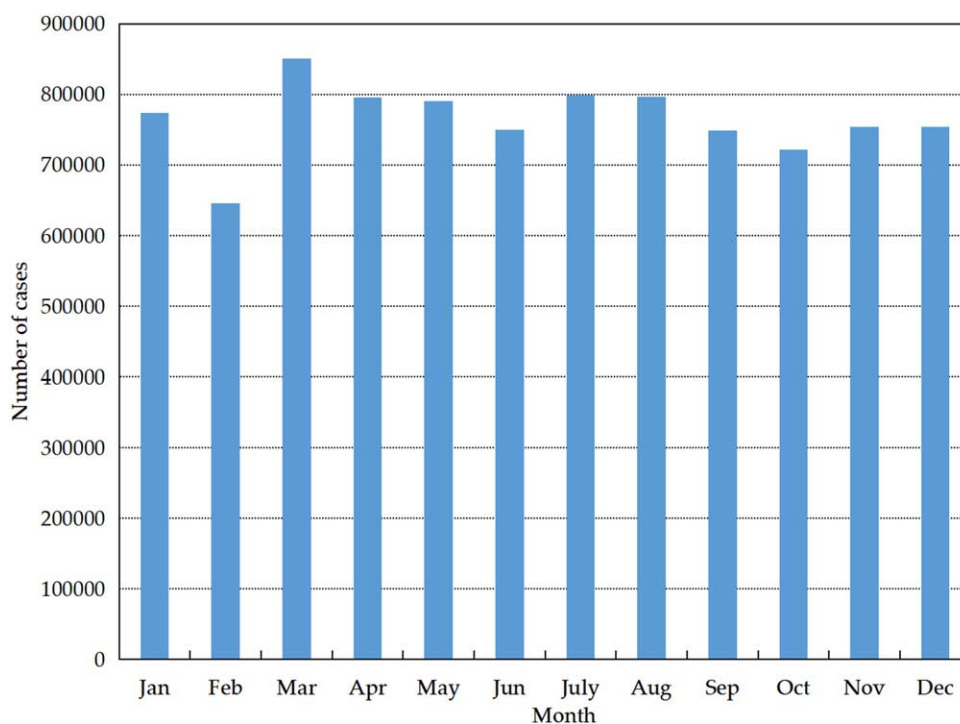


Figure 1. The monthly number of Hepatitis B cases from 2013 to 2020 in mainland China.

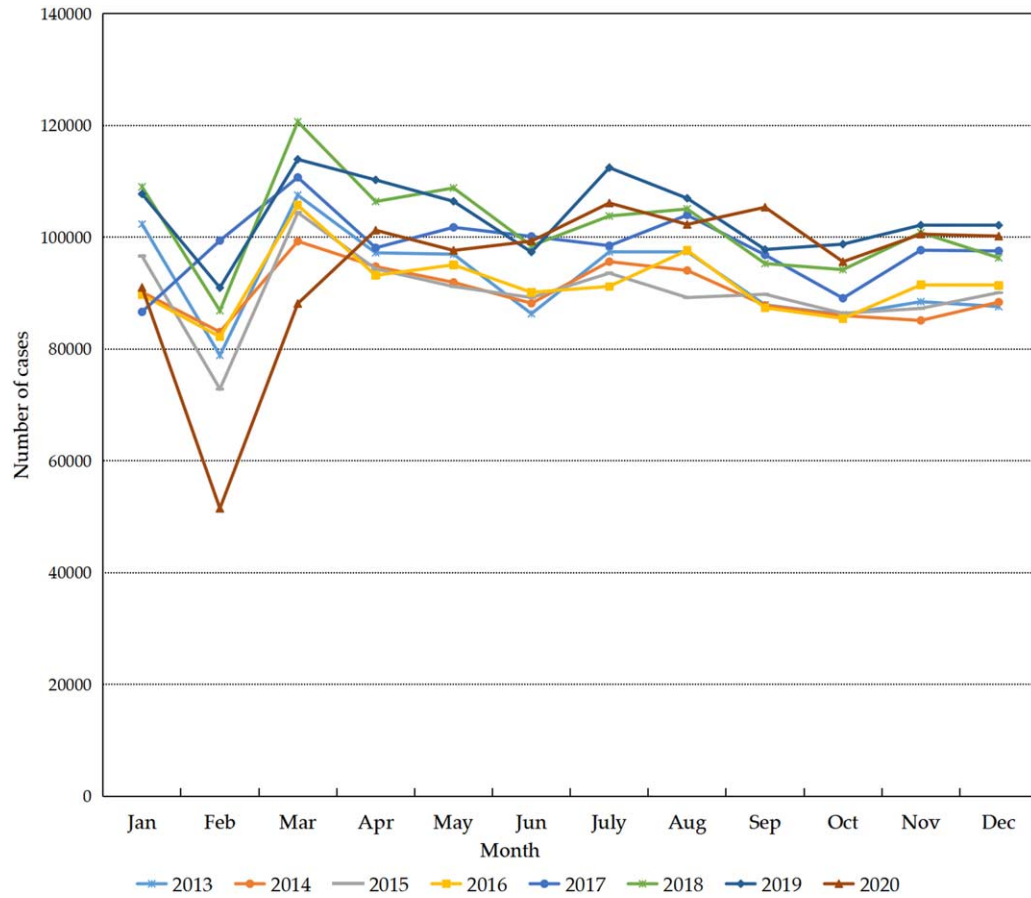
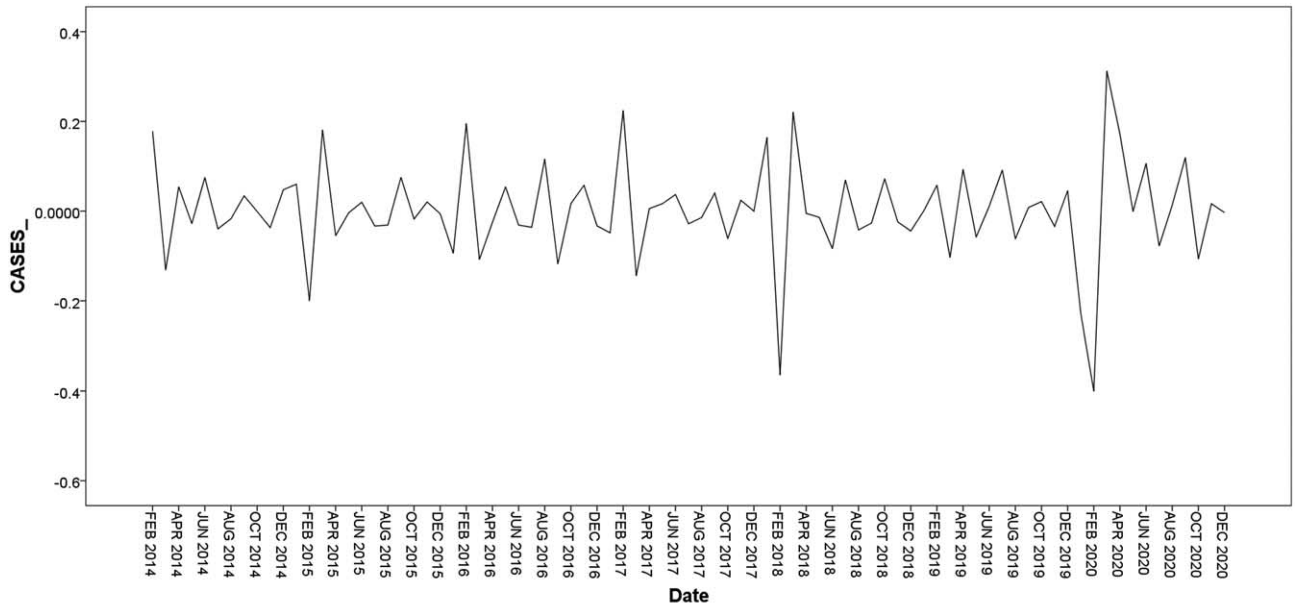


Figure 2. The monthly number of Hepatitis B cases each year from 2013 to 2020 in mainland China.



Figure 3. Plot of the original sequence.



Transforms: natural logarithm, difference(1), seasonal difference(1, period 12)

Figure 4. Plot of after the natural logarithm transformation and a first-order differenced and seasonal differenced with one-period of 12.

4. Discussion

Accurate prediction and estimation of infectious diseases is crucial for public health authorities to clearly understand their epidemic characteristics and take preventive and control measures in advance.^[25,26] Given the background of big data, the use of historical data to forecast future HBV incidence has recently become an important issue in public health research.

Therefore, it is important to apply predictive models to explore the regularity of HBV infections. It plays a significant role in the prevention and control of HBV infections.

Our findings suggest that the monthly number of hepatitis B cases from 2013 to 2020 has witnessed uniform growth and seasonal and periodic fluctuations. The reason for the uniform growth of hepatitis B cases is that China has a large population

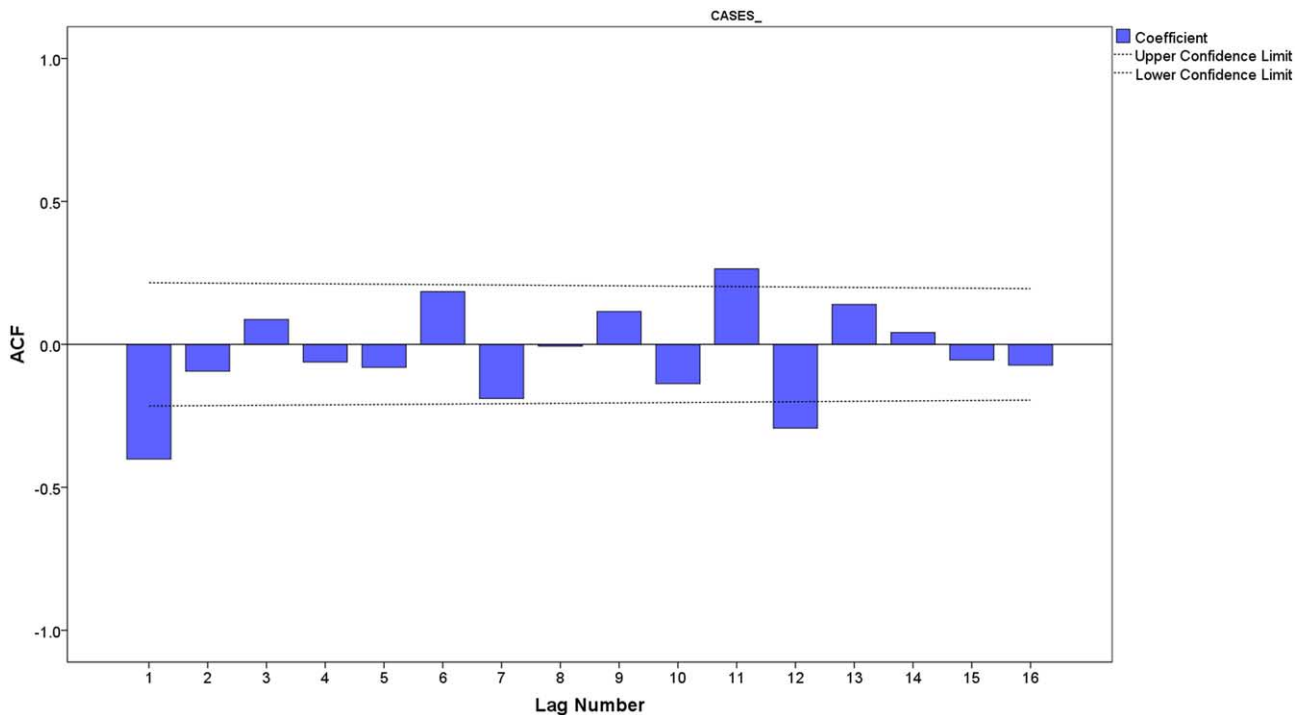


Figure 5. Plot of ACF after differenced sequence.

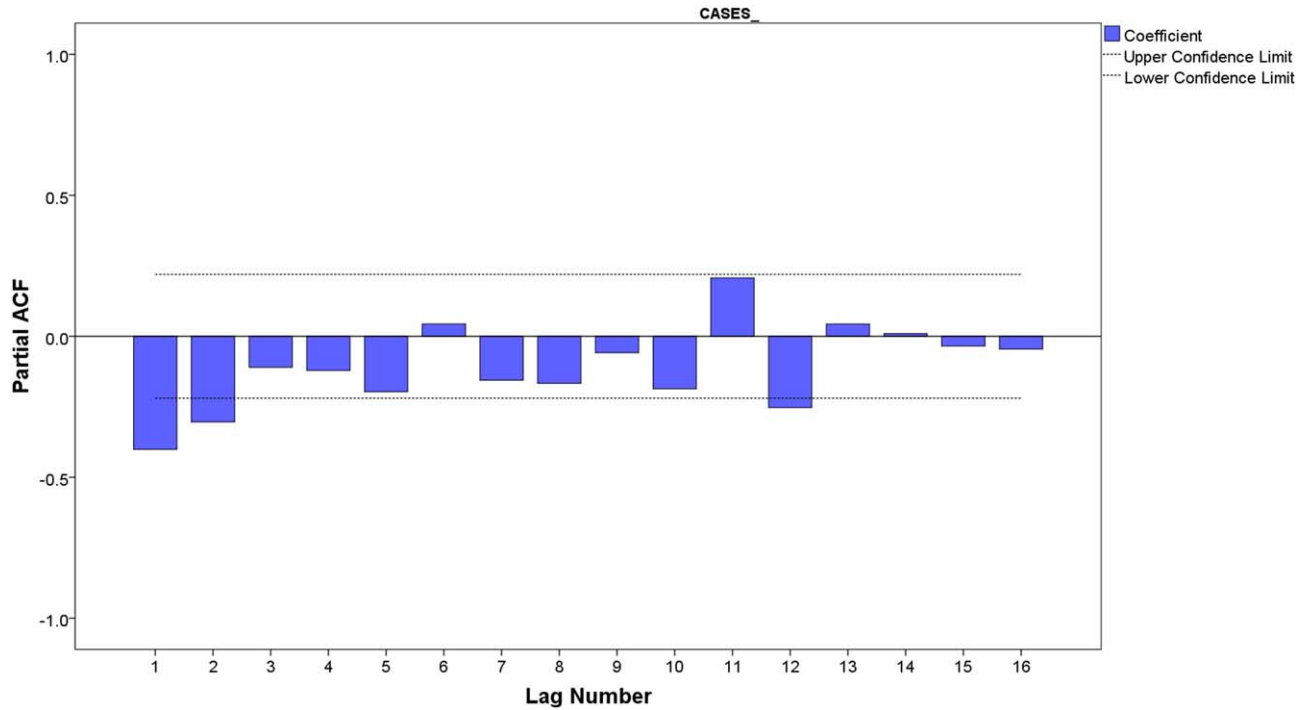


Figure 6. Plot of PACF after differenced sequence.

Table 2

Parameter estimation of candidate SARIMA models.

Model	R-squared	RMSE	MAPE	MAE	Normalized BIC	Ljung-Box Q(18)		
						Statistics	DF	P value
SARIMA(1,0,0)(0,1,1)12	0.441	7236.625	5.493	4980.087	17.879	21.036	16	.177
SARIMA(1,0,0)(1,1,1)12	0.489	7009.027	5.290	4728.593	17.921	16.843	15	.328
SARIMA(1,0,1)(0,1,1)12	0.496	6957.034	5.127	4545.904	17.906	13.915	15	.532
SARIMA(1,0,1)(1,1,1)12	0.496	7005.569	5.095	4513.149	17.973	12.559	14	.561

Table 3

Estimates and standard error of candidate SARIMA models parameter.

Model	Parameters	Estimate	SE	t	P value
SARIMA(1,0,0)(0,1,1)12	AR Lag 1	0.504	0.098	5.145	.000
	Seasonal Difference	1			
	MA, Seasonal Lag 1	0.611	0.147	4.150	.000
SARIMA(1,0,0)(1,1,1)12	Constant	1213.634	600.071	2.022	.046
	AR Lag 1	0.449	0.101	4.434	.000
	AR, Seasonal Lag 1	0.186	0.284	0.654	.515
SARIMA(1,0,1)(0,1,1)12	Seasonal Difference	1			
	MA, Seasonal Lag 1	0.984	5.270	0.187	.852
	Constant	1382.365	606.969	2.277	.025
SARIMA(1,0,1)(1,1,1)12	AR Lag 1	0.649	0.183	3.542	.001
	MA Lag 1	0.237	0.234	1.011	.315
	Seasonal Difference	1			
SARIMA(1,0,1)(1,1,1)12	MA, Seasonal Lag 1	0.932	0.847	1.101	.274
	Constant	1097.881	737.941	1.488	.141
	AR Lag 1	0.663	0.197	3.374	.001
SARIMA(1,0,1)(1,1,1)12	MA Lag 1	0.287	0.249	1.152	.253
	AR, Seasonal Lag 1	0.272	0.340	0.800	.426
	Seasonal Difference	1			
SARIMA(1,0,1)(1,1,1)12	MA, Seasonal Lag 1	0.998	42.748	0.023	.981

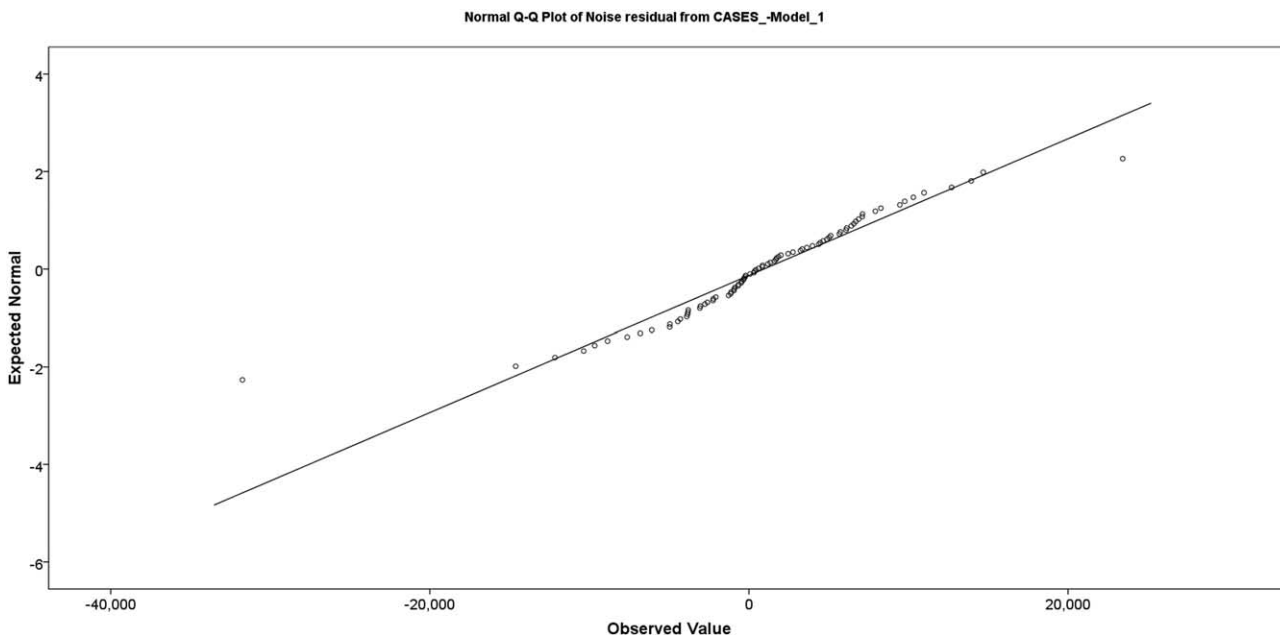


Figure 7. Normal Q-Q plot of noise residual from SARIMA(1,0,0)(0,1,1)12.

and an increasing frequency of large floating populations in urban and rural areas.^[27,28] Moreover, the reasons for the seasonal and periodic fluctuation tendencies in hepatitis B cases are as follows. First, the most important festival in China is the Spring Festival, which is celebrated in January or February each year, and people always return to their hometown earlier to attend a family reunion. Consequently, large-scale migrant workers continue to float between the urban and rural areas. During this time, people did not want to be in the hospital because they believed that it was unfortunate to go to the

hospital, according to traditional Chinese concepts. Therefore, the lowest monthly number of hepatitis B cases each year from 2013 to 2020 was recorded in February. Second, it is challenging to prevent and control HBV infection because migrant workers return to their original work after the Spring Festival. In addition, migrant workers who would spend time in entertainment places frequently enjoy nightlife delights.^[29] Therefore, the highest point in each year from 2013 to 2020 was March.

It has been proven that statistical mathematical models can help governments and at all levels in hygiene departments

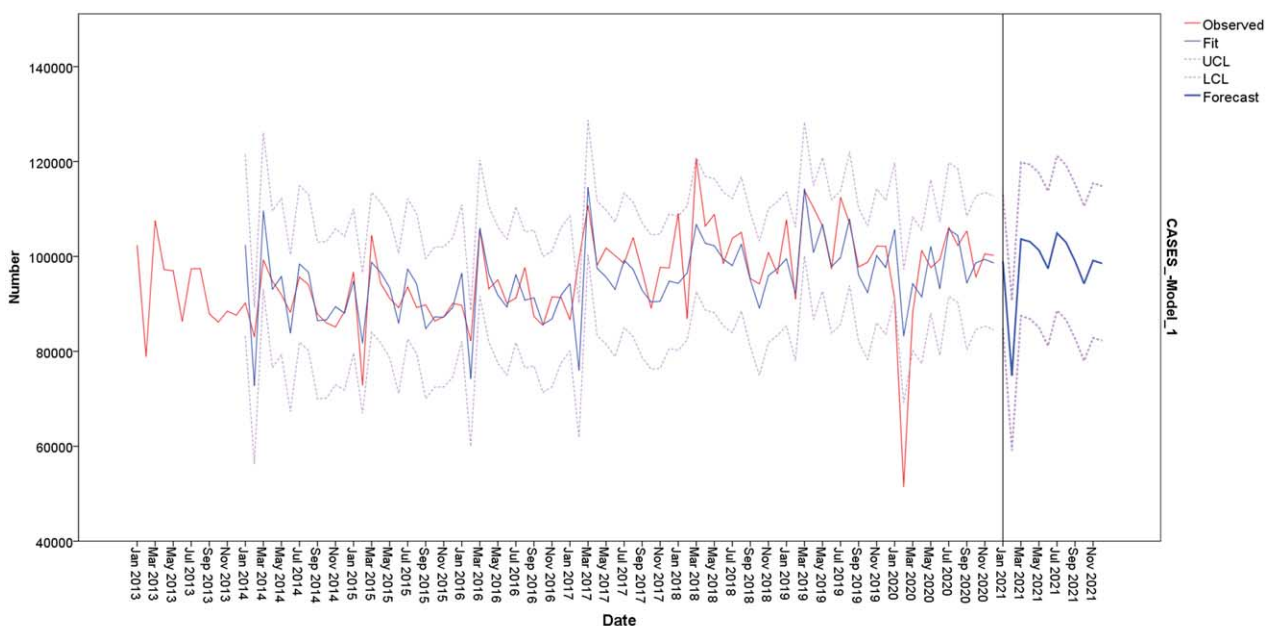


Figure 8. Comparison of the predicted value and observed value results from SARIMA(1,0,0)(0,1,1)12 model.

Table 4
Prediction result of SARIMA(1,0,0)(0,1,1)12 model.

Date	Predicted value	95% Lower confidence limit	95% Upper confidence limit
Jan-21	98968	84888	113049
Feb-21	74889	59123	90655
Mar-21	103648	87482	119814
Apr-21	103111	86844	119377
May-21	101303	85012	117595
Jun-21	97444	81146	113742
Jul-21	104920	88620	121220
Aug-21	102957	86657	119257
Sep-21	99006	82705	115306
Oct-21	94277	77977	110578
Nov-21	99150	82850	115450
Dec-21	98523	82223	114823

recognize epidemic behavior earlier^[30] and make better decisions for the prevention and control of infectious diseases.^[31] Therefore, time series analysis was developed to fit an appropriate prediction model that can provide a reference for the surveillance of infectious diseases. It should be noted that according to the data features of the study, choosing the right model is a prerequisite for exploring credible results. Based on a literature review and our previous research, we found that the ARIMA,^[3] GM (1, 1) model,^[3] Decision tree,^[32] Random forest,^[32] AdaBoost with decision tree (AdaBoost),^[32] extreme gradient boosting decision tree (XGBoost),^[32] Elman network,^[33] and generalized regression neural network (GRNN)^[34] can be used to predict hepatitis B. Among these, the GM (1,1) model is suitable for short-term prediction with small sample sizes.^[17] However, Decision tree, Random forest, AdaBoost, XGBoost, and GRNN prediction techniques, described as machine-learning-based forecasting models, specialize in dealing with big data and nonlinear problems but require a large number of samples.^[32] For example, Yin et al^[32] used Decision tree, Random forest, AdaBoost, XGBoost, and GRNN prediction techniques to fit and predict patients co-infected with hepatitis B virus (HBV) and human immunodeficiency virus (HIV) in Jiangsu, China, which provided 12469 samples sizes in the study.

Compared to GM (1, 1) and machine-learning-based forecasting models, the SARIMA model has numerous advantages for infectious disease prediction. It is much easier to construct a model that requires at least 36 samples^[3] rather than a large number of samples owing to its simple theoretical principle. However, the SARIMA model uses historical data and effectively captures the essence of the dependence between current and past observations, while considering the dynamic nature of infectious disease sequences.^[3] As a result, the SARIMA model is widely used to predict infectious diseases such as hemorrhagic fever,^[35] seasonal influenza,^[36] mumps,^[37] visceral leishmaniasis,^[38] hand-foot-mouth disease,^[39] brucellosis,^[40] and scarlet fever.^[41]

In our study, the prediction technique we selected was in accordance with data characteristics and sample sizes. The hepatitis B data displayed features of seasonality and periodicity, and the sample size was 96, which was in full accordance with the requirements of data characteristics and sample sizes for SARIMA. As a result, SARIMA was applied to the prediction of hepatitis B in mainland China from 2013 to 2020. The

prediction results showed well-simulated epidemiological trends for hepatitis B in mainland China. Therefore, this research suggests that the SARIMA model has a good short-term predictive performance, which is in accordance with previous studies.^[35–41] This study also found that the predicted values for 2021 were within the 95% confidence interval, which theoretically indicates that it is impossible to cause an outbreak of hepatitis B during this period. However, the outbreak of the hepatitis B epidemic has been affected by many social,^[42] sociocultural,^[43] and economic factors.^[44] Consequently, when taking preventive measures in the future, the potential factors that can cause HBV infection cannot be ignored.

To the best of our knowledge, this is the first detailed study to explore the construction of a SARIMA model to predict HBV cases in mainland China. However, our study had several limitations. First, only a linear relationship between the incidence of hepatitis B and its associated factors should be considered, leaving social, cultural, economic, and other factors out of account. Second, the SARIMA model has disadvantages: it is feasible only for short-term predictions and cannot handle nonlinear problems.^[33] To obtain accurate and long-term prediction results, a large amount of data must be continually collected and updated. We then apply the combined forecasting model for prediction, such as the SARIMA-NARNNX hybrid model,^[45] SARIMA-NAR hybrid model,^[46] and SARIMA-SDGM hybrid prediction model.^[47] Third, almost all prediction techniques have defects; therefore, the prediction results only take references.^[33] We should adopt a correct attitude towards the prediction results and use them to guide practical work, rather than regarding the prediction results as an exclusive policy-making reference frame.

5. Conclusions

This study showed that hepatitis B cases from 2013 to 2020 exhibited uniform growth and a seasonal and periodic fluctuation tendency. This warns us that we should take predictive and preventive measures to prevent outbreaks of this infectious disease, which peaks in March each year. SARIMA model simulated well by the reality of epidemiological trends of Hepatitis B in mainland China. It is helpful for governments and hygiene departments at all levels to provide appropriate measures and reasonably allocate health resources.

Acknowledgments

We thank the Project of the Hospital Management Institute, National Health Commission of the People's Republic of China (Grant No. YLZLXZ-2021-004) for funding this study.

Author contributions

Conceptualization: Daren Zhao, Huiwu Zhang, Ruihua Zhang

Formal analysis: Daren Zhao

Resources: Daren Zhao

Writing – original draft: Daren Zhao, Qing Cao, Zhiyi Wang

Writing – review & editing: Huiwu Zhang

References

- [1] Trépo C, Chan HL, Lok A. Hepatitis B virus infection. *Lancet* 2014;384:2053–63.

- [2] Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet* 2015;386:1546–55.
- [3] Wang YW, Shen ZZ, Jiang Y. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. *PLoS One* 2018;13:e0201987. DOI: 10.1371/journal.pone.0201987.
- [4] Yuen MF, Chen DS, Dusheiko G, et al. Hepatitis B virus infection. *Nat Rev Dis Primers* 2018;4:18036. DOI: 10.1038/nrdp.2018.36.
- [5] Stanaway JD, Flaxman AD, Naghavi M, et al. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. *Lancet* 2016;388:1081–8.
- [6] World Health Organization. Regional Office for Africa. (2017). End hepatitis by 2030: prevention, care and treatment of viral hepatitis in the African Region: framework for action, 2016-2020. World Health Organization. Regional Office for Africa. <https://apps.who.int/iris/handle/10665/259637>
- [7] World health statistics 2017: monitoring health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.
- [8] Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021 Accountability for the global health sector strategies 2016-2021: actions for impact. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.
- [9] Aspinal EJ, Hawkins G, Fraser A, Hutchinson SJ, Goldberg D. Hepatitis B prevention, diagnosis, treatment and care: a review. *Occup Med (Lond)* 2011;61:531–40.
- [10] Chen S, Mao W, Guo L, Zhang J, Tang S. Combating hepatitis B and C by 2030: achievements, gaps, and options for actions in China. *BMJ Glob Health*, 2020; 5:e002306. DOI: 10.1136/bmjgh-2020-002306.
- [11] Liu J, Liang W, Jing W, Liu M. Countdown to 2030: eliminating hepatitis B disease, China. *Bull World Health Organ* 2019;97:230–8.
- [12] Tao J, Zhang W, Yue H, et al. Prevalence of hepatitis B virus infection in Shenzhen, China, 2015-2018. *Sci Rep* 2019;9:13948. doi: 10.1038/s41598-019-50173-5.
- [13] Qiu H, Zhao H, Xiang H, et al. Forecasting the incidence of mumps in Chongqing based on a SARIMA model. *BMC Public Health* 2021;21:373. doi: 10.1186/s12889-021-10383-x.
- [14] Zheng Y, Zhang L, Zhu X, Guo G. A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China. *PLoS One* 2020;15:e0234660. DOI: 10.1371/journal.pone.0234660.
- [15] Villani M, Earnest A, Nanayakkara N, Smith K, de Courten B, Zoungas S. Time series modelling to forecast prehospital EMS demand for diabetic emergencies. *BMC Health Serv Res* 2017;17:332. doi: 10.1186/s12913-017-2280-6.
- [16] Zhang YQ, Li XX, Li WB, et al. Analysis and predication of tuberculosis registration rates in Henan Province, China: an exponential smoothing model study. *Infect Dis Poverty* 2020;9: 123. . DOI: 10.1186/s40249-020-00742-y.
- [17] Yang X, Zou J, Kong D, Jiang G. The analysis of GM (1, 1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China. *Medicine (Baltimore)* 2018;97:e11787. DOI: 10.1097/MD.00000000000011787.
- [18] Rath S, Tripathy A, Tripathy AR. Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes Metab Syndr* 2020;14:1467–74.
- [19] Singh RK, Rani M, Bhagavathula AS, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (ARIMA) model. *JMIR Public Health Surveill* 2020;6:e19115. DOI: 10.2196/19115.
- [20] Wang H, Tian CW, Wang WM, Luo XM. Time-series analysis of tuberculosis from 2005 to 2017 in China. *Epidemiol Infect* 2018; 146:935–9.
- [21] 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 (Baltimore)* 2016;95:e3929. DOI: 10.1097/MD.0000000000003929.
- [22] Martinez EZ, Silva EA, Fabbro AL. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Rev Soc Bras Med Trop* 2011;44:436–40.
- [23] Zheng YL, Zhang LP, Zhang XL, Wang K, Zheng YJ. Forecast model analysis for the morbidity of tuberculosis in Xinjiang, China. *PLoS One* 2015;10:e0116832. DOI: 10.1371/journal.pone.0116832.
- [24] Fang L, Wang D, Pan G. Analysis and Estimation of COVID-19 Spreading in Russia Based on ARIMA Model. *SN Compr Clin Med* 2020;9:1–7.
- [25] Fang X, Liu W, Ai J, et al. Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infect Dis* 2020;20:222. doi: 10.1186/s12879-020-4930-2.
- [26] Yang E, Park HW, Choi YH, et al. A Simulation-based study on the comparison of statistical and time series forecasting methods for early detection of infectious disease outbreaks. *Int J Environ Res Public Health* 2018;15:966. doi: 10.3390/ijerph15050966.
- [27] Leng A, Li Y, Wangen KR, Nicholas S, Maitland E, Wang J. Hepatitis B discrimination in everyday life by rural migrant workers in Beijing. *Hum Vaccin Immunother* 2016;12:1164–71.
- [28] Yang Y, Yan M, Yue M, et al. Prevalence of hepatitis B and knowledge about hepatitis B among migrant workers in Shandong province, China: a cross-sectional study. *Iran Red Crescent Med J* 2015;17:e26725. DOI: 10.5812/ircmj.17(4)2015.26725.
- [29] Wu R, Du XF, Wang K, Guo XQ, Liu XH. Analysis of the epidemiological characteristics of viral hepatitis from 2010 to 2015 in Yaodu District, Linfen City. *Chin J Gerontol* 2017;37:5446–8.
- [30] Zheng Y, Zhang L, Wang C, et al. Predictive analysis of the number of human brucellosis cases in Xinjiang, China. *Sci Rep* 2021;11:11513. doi: 10.1038/s41598-021-91176-5.
- [31] Moftakhar L, Seif M, Safe MS. Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iran J Public Health* 2020;49:92–100.
- [32] Yin Y, Xue M, Shi L, et al. A Noninvasive prediction model for hepatitis B virus disease in patients with HIV: based on the population of Jiangsu, China. *Biomed Res Int* 2021;2021:6696041. doi: 10.1155/2021/6696041.
- [33] Zheng Y, Zhang L, Zhu X, Guo G. A comparative study of two methods to predict the incidence of hepatitis B in Guangxi, China. *PLoS One* 2020;15:e0234660. doi: 10.1371/journal.pone.0234660. PMID: 32579598; PMCID: PMC7314421.
- [34] Wei W, Jiang J, Liang H, et al. Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng County, China. *PLoS One* 2016;11:e0156768. DOI: 10.1371/journal.pone.0156768.
- [35] Qi C, Zhang D, Zhu Y, et al. SARFIMA model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome and comparing with SARIMA. *BMC Med Res Methodol* 2020;20:243. DOI: 10.1186/s12874-020-01130-L 1138.
- [36] Cong J, Ren M, Xie S, Wang P. Predicting seasonal influenza based on SARIMA Model, in mainland China from 2005 to 2018. *Int J Environ Res Public Health* 2019;16:4760. doi: 10.3390/ijerph16234760.
- [37] Xu Q, Li R, Liu Y, et al. Forecasting the incidence of mumps in Zibo City based on a SARIMA model. *Int J Environ Res Public Health* 2017;14:925. doi: 10.3390/ijerph14080925.
- [38] Li HL, Zheng RJ, Zheng Q, et al. Predicting the number of visceral leishmaniasis cases in Kashgar, Xinjiang, China using the ARIMA-EGARCH model. *Asian Pac J Trop Med* 2020;13: 81-90. . DOI: 10.4103/1995-7645.275416.
- [39] Yu G, Feng H, Feng S, Zhao J, Xu J. Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA-NNAR hybrid model. *PLoS One* 2021;16:e0246673. DOI: 10.1371/journal.pone.0246673.
- [40] Zhao C, Yang Y, Wu S, et al. Search trends and prediction of human brucellosis using Baidu index data from 2011 to 2018 in China. *Sci Rep* 2020;10:5896. doi: 10.1038/s41598-020-62517-7.
- [41] Wang Y, Xu C, Wang Z, Yuan J. Seasonality and trend prediction of scarlet fever incidence in mainland China from 2004 to 2018 using a hybrid SARIMA-NARX model. *PeerJ* 2019;7:e6165. DOI: 10.7717/peerj.6165.
- [42] Meffre C, Le Strat Y, Delarocque-Astagneau E, et al. Prevalence of hepatitis B and hepatitis C virus infections in France in 2004: social factors are important predictors after adjusting for known risk factors. *J Med Virol* 2010;82:546–55.
- [43] Ma GX, Lee S, Wang M, et al. Role of sociocultural factors in hepatitis B screening among Asian Americans. *South Med J* 2011; 104:466–72.
- [44] Tosun S, Aygün O, Özdemir HÖ, Korkmaz E, Özdemir D. The impact of economic and social factors on the prevalence of hepatitis B in Turkey. *BMC Public Health* 2018;18:649. doi: 10.1186/s12889-018-5575-6.

- [45] Wang Y, Xu C, Zhang S, et al. Temporal trends analysis of tuberculosis morbidity in mainland China from 1997 to 2025 using a new SARIMA-NARNNX hybrid model. *BMJ Open* 2019;9:e024409. DOI: 10.1136/bmjopen-2018-024409.
- [46] Wang Y, Xu C, Wang Z, Zhang S, Zhu Y, Yuan J. Time series modeling of pertussis incidence in China from 2004 to 2018 with a novel wavelet based SARIMA-NAR hybrid model. *PLoS One* 2018;13:e0208404. DOI: 10.1371/journal.pone.0208404.
- [47] Song Z, Guo Y, Wu Y, Ma J. Short-term traffic speed prediction under different data collection time intervals using a SARIMA-SDGM hybrid prediction model. *PLoS One* 2019;14:e0218626. DOI: 10.1371/journal.pone.0218626.