

Original Article

Trends of the incidence rate of Hepatitis C in Changsha, China, 2005 to 2023 and Its Prediction to 2025Shuilian Chen^{1,2}, Ping Zeng², Kewei Yang², Xixing Zhang¹, Xuewen Yang¹, Yinzhu Zhou¹¹ Changsha Center for Disease Control and Prevention, Changsha, Hunan Province, China² College of Systems Engineering, National University of Defense Technology, Changsha, Hunan Province, China**Abstract**

Introduction: Hepatitis C is a globally distributed hepatic disease with a high chronicity rate, posing significant global public health concerns. This research aimed to predict the incidence trend of hepatitis C to guide disease prevention and control.

Method: The quarterly incidence rates of hepatitis C in Changsha from the first quarter of 2005 to the third quarter of 2023 were collected. Curve-fitting method, grey model (GM (1,1)), Autoregressive Integrated Moving Average (ARIMA) model, and Back Propagation Neural Network (BPNN) model, were applied to simulate the incidence trend of hepatitis C.

Results: The annual average reported incidence rate in Changsha was 12.518 per 100,000. The incidence rate of hepatitis C showed an upward trend from 1.102 per 100,000 in 2005 to 17.172 per 100,000 in 2013 and remained at a high level from 2014 to 2019. It has been steadily declining since 2020. The BPNN model exhibited the best forecasting performance ($MAE = 0.394$, $MRE = 0.203$). The prediction results from the BPNN model estimated the incidence rate of hepatitis C to be 9.991 per 100,000 in 2024 and 11.920 per 100,000 in 2025.

Conclusions: The incidence of hepatitis C in the next two years is projected to be slightly higher than that during the COVID-19 epidemic. It is imperative to strengthen various measures to achieve the goal of hepatitis C elimination in China.

Key words: Hepatitis C; incidence rate; trend; model; prediction.

J Infect Dev Ctries 2025; 19(1):140-145. doi:10.3855/jidc.19790

(Received 31 December 2023 – Accepted 29 June 2024)

Copyright © 2025 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

Hepatitis C virus (HCV), a major cause of liver cirrhosis and cancer, is a serious public health issue, with approximately 58 million chronic cases worldwide in 2019 [1]. According to the data released by the National Health Commission of the People's Republic of China [2], the reported incidence rate of hepatitis C has increased year by year prior to the COVID-19 outbreak. The incidence rate was 1.6 per 100,000 in 2003 and increased to 15.3 per 100,000 in 2012, marking an increase of about 8.5 times. HCV not only poses a significant threat to public health but also presents a substantial challenge to the demand for national health resources. Research on the prediction of hepatitis C is an emerging field of study. Daw *et al.* used time series models to predict the trend of hepatitis C infection rates among blood donors in Libya [3]. Zheng *et al.* employed an Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) combined model to predict the incidence rate of hepatitis C [4]. Xia *et al.* presented three neural network models, but the results indicated that no single

model for predicting the incidence cases of hepatitis C could be completely superior to others [5].

Accurate prediction is a prerequisite for formulating short-term or long-term response strategies for preventing and controlling infectious diseases. Current studies contribute to improving model accuracy and provide valuable references and guidance for future research. However, there is generally a lack of systematic comparison of the predictive accuracy of various models for hepatitis C. Our study takes Changsha city as an example to illustrate the overall epidemic trend of hepatitis C. Changsha, located in the southeast region of China and the capital city of Hunan Province, has a population of over 10 million. We used the curve fitting method, grey model (GM (1,1)), ARIMA model, and Back Propagation Neural Network (BPNN) model to simulate the incidence trend of hepatitis C. The optimal model would be selected through comprehensive comparison and evaluation to predict the short-term trend in HCV incidence, providing an important resource for planning and evaluating control measures in response to hepatitis C.

Methodology

Data

Hepatitis C cases include clinically diagnosed cases and confirmed cases. A clinically diagnosed case is defined as the presence of anti-HCV antibodies, along with a suspicious exposure history, clinical manifestation, or abnormal biochemical examination. A confirmed case is defined as the presence of HCV RNA. The quarterly data of new hepatitis C cases in Changsha from the first quarter of 2005 to the third quarter of 2023 were retrieved from the Communicable Disease Surveillance System in the China Disease Prevention and Control Information System, and population data in the same period were extracted from the Changsha Statistical Yearbook.

Statistical analysis

Curve-Fitting Method

Curve-fitting method is a fundamental technique in data modeling that aims to find a function which best fits a given set of data points. Commonly utilized functions for curve fitting include polynomials, exponential, power functions, and logarithmic functions [6]. The process of curve fitting involves several steps: cleaning and organizing the data, selecting an appropriate model, estimating the model parameters, and evaluating the goodness of fit. The optimal curve is determined based on the *F*-statistic value of the regression model, the corresponding *p* value, and the coefficient of determination (R^2).

ARIMA Model

ARIMA model is a widely used approach for time series forecasting. The model consists of three components: the autoregressive (AR), integration (I), and moving average (MA) parts, which correspond to the parameters *p*, *d*, and *q* respectively [7]. The parameter *p* represents the lag order, indicating the number of lag observations included in the model. The parameter *d* is the degree of differencing, denoting the number of times raw observations undergo differencing. The parameter *q* is the order of the moving average, specifying the size of the moving average window. To select the appropriate model, the randomness, stationarity, and seasonality of time series data were analyzed using time series diagrams, autocorrelation functions (ACF), and partial autocorrelation functions (PACF). The optimal model is chosen based on the Bayesian information criterion (BIC).

GM (1,1) Model

GM (1,1) model is the fundamental and most widely used grey system model for forecasting with limited and short-term data. The time response equation is given as follows [8].

$$x_1(k + 1) = \left(x_0(1) - \frac{b}{a}\right) e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n$$

Where, *-a* is the development factor, reflecting the development momentum of x_0 and x_1 , *b* is the grey action, reflecting the data changes in the data. The GM (1,1) model is established using the hepatitis C incidence rate, and the parameter vectors are estimated by the least squares method.

BPNN Model

BPNN is a feedforward network consisting of three layers of neurons, or nodes: an input layer, a hidden layer, and an output layer [9]. Each layer contains several neurons, and the hidden layer can be either single or multiple. Neurons communicate with each other through connections with different weights, and each neuron in the hidden layer and output layer has its threshold. The training process involves forward propagation and backward propagation, during which the network's link weights are constantly adjusted under the influence of external input samples, steering the network's output closer to the expected output. The modeling steps are as follows: First, a BPNN model is designed based on the incidence data of hepatitis C; second, the BPNN model is optimized using a training set, with training repeated until the network converges (i.e., reaches the expected training error); finally, the established model is used to predict the incidence rate.

Model Evaluation

R^2 , mean absolute error (MAE), and mean Relative Error (MRE) were adopted to evaluate the model fitting effect. The formulas are as follows:

$$R^2 = 1 - \frac{SS_R}{SS_T}$$

$$SS_R = \sum_{i=1}^n (y_i - x_i)^2$$

$$SS_T = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right|$$

Where: SS_R is the sum of residual errors, SS_T is the total variation of the points, n is the number of samples, x_i is the actual value for observation i , y_i is the predicted value of the model for observation i . The value of R^2 ranges from 0% to 100%, with a higher R^2 value indicating a better fit of the model. Smaller values of MAE and MRE indicate a better fitting effect of the model.

Excel 2016 was used for data collation and plotting. Statistical analysis was performed using R (version 4.2.3), and $p < 0.05$ represented statistical significance. All incidence rates were reported per 100,000 person-years.

Results

Epidemic Trend of Hepatitis C

A total of 17,513 cases of hepatitis C were reported in Changsha from the first quarter of 2005 to the third quarter of 2023. The annual average reported incidence rate was 12.518 per 100,000. The incidence rate of hepatitis C in Changsha exhibited an upward trend, increasing from 1.102 per 100,000 in 2005 to 17.172 per 100,000 in 2013, and then remained at a high level from 2014 to 2019. It has been steadily declining since 2020. The incidence data of hepatitis C from the first quarter of 2005 to the third quarter of 2023 were summarized on a quarterly basis, and the quarterly incidence rates were shown in Figure 1.

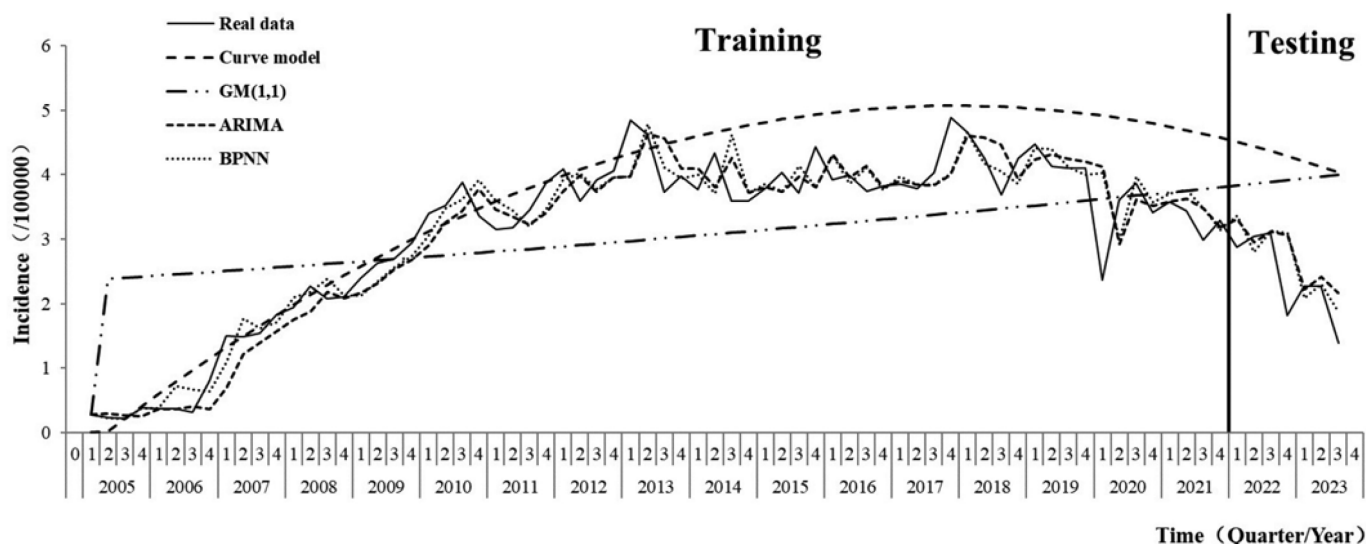
Table 1. Fitting and prediction effect evaluation of the four models.

Model / Stage	MAE	MRE
Curve-fitting (Quadratic curve)		
training	0.627	0.271
testing	1.886	0.911
GM (1,1)		
training	0.826	0.849
testing	1.523	0.770
ARIMA (2,1,2)		
training	0.298	0.117
testing	0.397	0.219
BPNN		
training	0.274	0.114
testing	0.394	0.203

Model Modeling Results

The curve fitting results indicated that the quadratic curve had the best fit ($R^2 = 0.95$). The optimal ARIMA model was the ARIMA (2,1,2) model with the significance of parameters and the minimum value of BIC (BIC = -1.291). During the training process, the curve-fitting method, GM (1,1) model, ARIMA (2,1,2) model, and BPNN model were employed to model the quarterly incidence rates of hepatitis C from 2005 to 2021. The results revealed that the quadratic curve ($R^2 = 0.930$), ARIMA (2,1,2) model ($R^2 = 0.950$), and BPNN model ($R^2 = 0.953$) performed better, while the GM (1,1) model ($R^2 = 0.685$) exhibited poor performance. In testing process, the models were tested with the incidence rates of hepatitis C from the first quarter of 2022 to the third quarter of 2023, and the results demonstrated that the BPNN model had the best forecasting performance ($MAE = 0.394$, $MRE = 0.203$), followed by ARIMA (2,1,2) model ($MAE = 0.397$, $MRE = 0.219$), as detailed in Table 1.

Figure 1. The observed (solid lines) and predicted incidence trend of hepatitis C in Changsha from the first quarter of 2005 to the third quarter of 2023.



Prediction Results

The results above indicated that both the BPNN model and the ARIMA (2,1,2) model were suitable for predicting the reported incidence rate of hepatitis C in Changsha, with the BPNN model showing a slight edge in performance over the ARIMA model. We used the BPNN model to predict the incidence rate of hepatitis C in Changsha, and the prediction results estimated the incidence rate of hepatitis C to be 9.991 per 100,000 in 2024 and 11.920 per 100,000 in 2025, respectively.

Discussion

The incidence rate of hepatitis C in Changsha increased year by year from 2005 to 2013, with an annual average reported incidence rate of 12.22 per 100,000 over the 18-year period. This rate is equivalent to the national annual reported incidence rate, which ranged from 3.03 per 100,000 in 2004 to 14.97 per 100,000 in 2014 as reported by Liu *et al.* [10], but higher than that reported in Shandong by Zheng *et al.* [4]. The incidence rate of hepatitis C in Changsha stabilized between 2013 and 2019, then showed a declining trend in recent years, consistent with research findings from other cities in China [11,12]. Zakrzewska *et al.* also found a significant decrease in the diagnosis rate of HCV infection in Poland in 2020 compared to 2019 [13].

The decline in the incidence rate of hepatitis C over the past four years might be attributed to the COVID-19 pandemic or the efforts made to eliminate hepatitis C. Measures against COVID-19 have played an important role in controlling various infectious diseases, including hepatitis C [14]. In 2016, the World Health Organization (WHO) called for the global elimination of viral hepatitis as a public health threat by 2030 [15], and China has also intensified its efforts to eliminate hepatitis C [16]. Therefore, the significant decrease in the incidence rate of hepatitis C might be primarily due to the comprehensive interventions that have been implemented.

Hepatitis C poses a serious health and economic burden to society, making it crucial to forecast its trends. In this study, the curve-fitting method, GM (1,1) model, ARIMA model, and BPNN model were used to model the reported incidence rate of hepatitis C. The results showed that the curve-fitting method, ARIMA (2,1,2) model, and BPNN model could more accurately simulate the incidence of hepatitis C in Changsha, with the BPNN model demonstrating the highest prediction accuracy.

A forecasting model is a statistical tool designed to predict future trends or outcomes by analyzing patterns

within a given set of input data. Since different types of diseases have unique characteristics, it is impossible to use the same model to simulate them all. Therefore, it is particularly important to select a suitable model for predicting a specific disease.

The curve-fitting method is easy to operate, but it has a limited range of curve types, making it difficult to ensure that the selected curve is the one with the highest degree of fit. We found that the quadratic curve had a high fitting degree, but its performance in the testing stage was not satisfactory. The GM (1,1) can make predictions with limited data and has good predictive effects on data that exhibit certain characteristics [17,18]. It is widely used for predicting population, housing prices, logistics, and other aspects [19-22]. A research showed that the GM (1,1) model had excellent performance in predicting the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China [23]. However, this study found that it was not suitable for predicting the incidence trend of hepatitis C in Changsha City.

ARIMA models are a subset of linear regression models that demonstrate good prediction performance in most applications, especially for stable and linear data [24-26]. The BPNN model is the most commonly used neural network model in nonlinear prediction, with strong self-adaptation ability, strong self-learning ability, strong nonlinear mapping ability, and high fault tolerance [27,28]. The findings of this study indicated that both models could simulate the incidence of hepatitis C in Changsha, but the BPNN model outperformed the ARIMA (2,1,2) model. Nevertheless, these prediction models have certain limitations, and their ability to forecast over the long term remains insufficient.

Conclusions

This study employed the BPNN model to predict the incidence of hepatitis C in 2024 and 2025. The findings indicated that the incidence rate of hepatitis C in Changsha in the next two years would be slightly higher than that observed during the COVID-19 epidemic. Consequently, the situation regarding the prevention and control of hepatitis C remains critical, it is imperative to intensify various measures aimed at achieving the goal of hepatitis C elimination in China.

Authors' contributions

All authors made a significant contribution to the work. Shuilian Chen, Ping Zeng, Xuewen Yang, and Yin Zhu Zhou designed the study. Shuilian Chen, Kewei Yang, and Xixing

Zhang collected and analyzed the data. Shuilian Chen, Ping Zeng, Xuewen Yang, and Yinzhu Zhou drafted the paper. All authors revised the paper.

Funding

This research was supported by the Hunan Provincial Natural Science Foundation Project (No.2024JJ9547).

Corresponding authors

Yinzhu Zhou

Changsha Center for Disease Control and Prevention
No. 509, 2nd section, Wanjieli North Road
Changsha, Hunan province, China
Tel.: +86-0731-84735652
Email: 1047912196@qq.com

Xuewen Yang

Changsha Center for Disease Control and Prevention
No. 509, 2nd section, Wanjieli North Road
Changsha, Hunan province, China
Tel.: +86-0731-84735652
Email: 774751649@qq.com

Ping Zeng

College of Systems Engineering
National University of Defense Technology, Changsha, Hunan Province, China
Email: zengping_2005@126.com

Conflict of interests

No conflict of interests is declared.

References

- World Health Organization (2019) Progress report on HIV, viral hepatitis, and sexually transmitted infections 2019. Accountability for the global health sector strategies, 2016–2021. Available: <https://apo.who.int/publications/i/item/progress-report-on-hiv-viral-hepatitis-and-sexually-transmitted-infections-2019>. Accessed: January 2024.
- National Health Commission of China (2005-2022) China health statistical yearbook. Available: <http://www.nhc.gov.cn/>. Accessed: January 2024.
- Daw MA, Shabash A, El-Bouzedi A, Dau AA, Habas M (2016) Modelling the prevalence of hepatitis C virus amongst blood donors in Libya: an investigation of providing a preventive strategy. *World J Virol* 5: 14-22. doi: 10.5501/wjv.v5.i1.14.
- Zheng W, Li H, Yang X, Wang L, Shi Y, Shan H, He L, Liu J, Chen H, Wang G, Zhao Y, Han C (2023) Trends and prediction in the incidence rate of hepatitis C in Shandong province in China from 2004 to 2030. *Prev Med* 177: 107749. doi: 10.1016/j.ypmed.2023.107749.
- Xia Z, Qin L, Ning Z, Zhang X (2022) Deep learning time series prediction models in surveillance data of hepatitis incidence in China. *PLoS One* 17: e0265660. doi: 10.1371/journal.pone.0265660.
- Dudek K, Kędzia W, Kędzia E, Kędzia A, Derkowski W (2017) Mathematical modelling of the growth of human fetus anatomical structures. *Anat Sci Int* 92: 521-529. doi: 10.1007/s12565-016-0353-y.
- Kaur J, Parmar KS, Singh S (2023) Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environ Sci Pollut Res Int* 30: 19617-19641. doi: 10.1007/s11356-023-25148-9.
- Wang H, Li Y (2024) A novel seasonal grey prediction model with fractional order accumulation for energy forecasting. *Heliyon* 10: e29960. doi: 10.1016/j.heliyon.2024.e29960.
- Lyu J, Zhang J (2019) Bp neural network prediction model for suicide attempt among chinese rural residents. *J Affect Disord* 246: 465-473. doi: 10.1016/j.jad.2018.12.111.
- Liu Z, Yang Q, Shi O, Ye W, Chen X, Zhang T (2018) The epidemiology of hepatitis B and hepatitis C infections in China from 2004 to 2014: an observational population-based study. *J Viral Hepat* 25: 1543-1554. doi: 10.1111/jvh.12938.
- Li Y, Li X, Lan X, Xue C, Zhang B, Wang Y (2023) Impact of COVID-19 on epidemic trend of hepatitis C in Henan province assessed by interrupted time series analysis. *BMC Infect Dis* 23: 691. doi: 10.1186/s12879-023-08635-9.
- Wang L, Ma C, Zhou Y, Wang Y, Zhao N, Chen Y, Miao Z, Yang Y, Liu S (2024) Epidemiological features of hepatitis C in China from 2015 to 2021: insights from national surveillance data. *Asia Pac J Public Health*: 1381269546. doi: 10.1177/10105395241254870.
- Zakrzewska K, Stepień M, Rosińska M (2022) Hepatitis C in Poland in 2020. *Przegl Epidemiol* 76: 233-242. doi: 10.32394/pe.76.23.
- Zhang Y, Feng W (2024) Impact of the coronavirus disease 2019 pandemic on the diversity of notifiable infectious diseases: a case study in Shanghai, China. *PeerJ* 12: e17124. doi: 10.7717/peerj.17124.
- World Health Organization (2016) Global health sector strategy on viral hepatitis 2016–2021. Available: <https://www.afro.who.int/publications/global-health-sector-strategy-viral-hepatitis-2016-2021>.
- Li J, Pang L, Liu Z (2022) Interpretation of the national action plan for eliminating hepatitis C as a public health threat (2021-2030). *China CDC Wkly* 4: 627-630. doi: 10.46234/ccdcw2022.139.
- Hu YC (2017) A genetic-algorithm-based remnant grey prediction model for energy demand forecasting. *PLoS One* 12: e0185478. doi: 10.1371/journal.pone.0185478.
- Yao J, Fan M, Xiao Y, Liang X, Cai C, Wang Y (2023) Spatial-temporal characteristics of corrected total phosphorus pollution loads from agricultural non-point sources in Tuojiang river watershed, Sichuan province of southwestern China. *Environ Sci Pollut Res Int* 30: 42192-42213. doi: 10.1007/s11356-023-25244-w.
- Wang Y (2014) A forecasting research on population size of Jilin province based on grey gm (1,1) model. *Applied Mechanics and Materials* 667: 273-276.
- Cheng Y, Yun S, Xu X, Lv P (2014) A study on teaching and application of grey prediction model-taking commercial housing prices of Huangshi city as a example. *Energy Education Science and Technology Part A: Energy Science and Research* 32: 1821-1828.
- Li Y, Lu S (2015) Research on prediction of regional grain logistics demand based on the grey-regression model. In. Leicester, United kingdom: IEEE Computer Society. 97-101.
- Cheng C, Miao S, Luo B, Sun Y (2017) Forecasting monthly energy production of small hydropower plants in ungauged basins using grey model and improved seasonal index. *J Hydroinform* 19: 993-1008. doi: 10.2166/hydro.2017.062.

23. Yang X, Zou J, Kong D, Jiang G (2018) The analysis of gm (1,1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China. *Medicine (Baltimore)* 97: e11787. doi: 10.1097/MD.00000000000011787.
24. Enders W (2010) *Applied econometric time series*, 3rd edition. Wiley.
25. Norrulashikin MA, Yusof F, Hanafiah N, Norrulashikin SM (2021) Modelling monthly influenza cases in malaysia. *PLoS One* 16: e0254137. doi: 10.1371/journal.pone.0254137.
26. Kissi E, Adjei-Kumi T, Amoah P, Gyimah J (2018) Forecasting construction tender price index in Ghana using autoregressive integrated moving average with exogenous variables model. *Australasian Journal of Construction Economics & Building* 18: 70-82.
27. Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. *Int J Forecast* 14: 35-62.
28. Wang L, Zeng Y, Chen T (2015) Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Syst Appl* 42: 855-863. doi: 10.1016/j.eswa.2014.08.018.