Retrospective Study

Time series analysis-based seasonal autoregressive fractionally integrated moving average to estimate hepatitis B and C epidemics in China

Wang YB et al. SARFIMA for estimating hepatitis epidemics

Abstract

BACKGROUND

Hepatitis B (HB) and hepatitis C (HC) place the largest burden in China, and a goal of eliminating them as a major public health threat by 2030 has been raised. Making more informed and accurate forecasts of the spread is essential for developing effective strategies, heightening the requirement for early warning to deal with such a major public health threat.

AIM

To monitor HB and HC epidemics by the design of paradigm seasonal autoregressive fractionally integrated moving average (SARFIMA) for projections into 2030, and comparing the effectiveness with seasonal autoregressive integrated moving average (SARIMA).

METHODS

Monthly HB and HC incidence cases in China were obtained from January 2004 to June 2023. Descriptive analysis and Hodrick-Prescott were employed to identify trends and seasonality. The two periods (from January 2004 to June 2022 and from January 2004 to
December 2015, respectively) were used as the training sets to develop both models, while the remaining periods served as the test sets to evaluate the forecasting accuracy.

RESULTS
There were incidents of 23,400,874 HB cases and 3,590,867 HC cases from January 2004 to June 2023. Overall, HB remains steady (average annual percentage change (AAPC)=0.44, 95% confidence interval [CI] -0.94 ~ 1.84) while HC was increasing (AAPC=8.91, 95%CI 6.98 ~ 10.88), both had a peak in March and a trough in February. In the 12-step-ahead HB forecast, the mean absolute deviation (15211.94), root mean square error (18762.94), mean absolute percentage error (0.17), mean error rate (0.15), and root mean square percentage error (0.25) under the best SARMFIMA(3,0,0)(0,0.449,2)_{12} were smaller than under the best SARIMA(3,0,0)(0,1,2)_{12} (16867.71, 20775.12, 0.19, 0.17, and 0.27, respectively). Similar results were also observed for the 90-step-ahead HB, 12-step-ahead HC, and 90-step-ahead HC forecasts. The predicted HB incidents totaled 9,865,400 (95%CI 7,508,093 ~ 12,222,709) and HC totaled 1,659,485 (95%CI 856,681 ~ 2,462,290) during 2023-2030.

CONCLUSION
Under current interventions, China faces enormous challenges to eliminate HB and HC epidemics by 2030, and effective strategies must be reinforced. The integration of SARMFIMA into public health for the management of HB and HC epidemics can potentially result in more informed and efficient interventions, surpassing the capabilities of SARIMA.

Key Words: Hepatitis; Seasonal autoregressive fractionally integrated moving average; Seasonal autoregressive integrated moving average; Prediction; Epidemic; Time series analysis

**Core Tip:** This retrospective study utilized seasonal autoregressive fractionally integrated moving average (SARFIMA) to monitor Hepatitis B (HB) and hepatitis C (HC) epidemics, and the forecasting potential was then compared to seasonal autoregressive integrated moving average (SARIMA). The resulting forecast error rates under the SARFIMA were less than under the SARIMA. The integration of SARFIMA into public health decision-making for the management of HB and HC epidemics can result in more informed interventions. The predicted HB totaled 9,865,400 (95% confidence interval [CI] 7,508,093~12,222,709) and HC totaled 1,659,485 (95% CI 856,681~2,462,290) cases in 2030, resulting in major challenges to eliminate hepatitis in China by 2030.

**INTRODUCTION**

Hepatitis is a condition characterized by liver inflammation, which can be caused by various infectious viruses and noninfectious agents, resulting in a variety of health complications[1]. There are five primary strains of the hepatitis virus, namely types A, B, C, D, and E. Types B and C, in particular, can progress into chronic diseases affecting millions of individuals, and they are the leading causes of liver cirrhosis, liver cancer, and mortality related to viral hepatitis[1, 2]. Although significant achievements have been made in the prevention of hepatitis B (HB) and hepatitis C (HC) through the implementation of various strategies, these infections still pose significant health risks globally, affecting millions of people and causing a substantial burden on healthcare systems[3, 4]. According to the World Health Organization (WHO), an estimated 296 and 58 million people were living with HB and HC, respectively, infection in 2019, and HB- and HC-related complications cause 820,000 and 290,000, respectively, deaths each year[1]. These numbers indicate the scale of the issue and emphasize the need for
effective prevention and control strategies. Also, WHO has set a goal of ending HB and HC as major public health threats by 2030 (compared with the 2015 baseline, new infections decreased by 30% by 2020 and 90% by 2030)[2]. China bears the greatest burden of HB virus (HBV) and HC virus (HCV) infections globally and is expected to play a pivotal role in achieving the global goal of eliminating HB and HC diseases by 2030[4]. Accurate prediction aids in anticipating prospective scenarios and making proactive judgments, allowing policymakers to make educated decisions, develop strategies, and prepare for potential challenges and opportunities[5]. It has served as a valuable tool for mitigating risks, forming policies, and optimizing outcomes in different domains.

Time series analysis plays a crucial role in formulating hypotheses to comprehend the epidemics of different diseases and estimating the dynamics of observed phenomena, ultimately contributing to the establishment of a high-quality control system[5]. Various statistical methods, including generalized regression neural network[6], back propagation neural network[6], grey prediction[7], and Elman neural network[8], have been employed to predict the epidemics of infectious diseases. However, epidemics are influenced by diverse factors, resulting in a combination of trends, seasonality, and randomness in the spread. Accordingly, the aforementioned statistical techniques may not adequately capture the complex and dynamic nature of epidemics, and it is challenging to generalize these models[5].

The seasonal autoregressive integrated moving average (SARIMA) has been the most common use of model in different domains thanks to its simplicity, fast applicability, and ability to explain data[5, 2-8]. Previous studies have also demonstrated the successful application of the SARIMA in estimating the prevalence, morbidity, and mortality of hepatitis[2], COVID-19[5], hemorrhagic fever with renal syndrome (HFRS)[9], hand-foot-and-mouth disease (HFMD)[10], and tuberculosis[11]. The SARIMA can effectively capture the dynamic dependency structure by capturing secular trends, periodic fluctuations, and random variations within a time series[5]. Despite the attractive attributes of SARIMA, it was a representative of modelling short fluctuations
in a series instead of the long term\textsuperscript{[12]}. Additionally, using an integer difference in SARIMA for a series exhibiting long memory can cause over-differencing and the removal of valuable features, which harms forecasts\textsuperscript{[13]}. In contrast, the seasonal autoregressive fractionally integrated moving average (SARFIMA) considers not only short memory, but also long memory, and such benefits as irregular fluctuations and complex seasonality by use of fractional differencing\textsuperscript{[12, 14, 15]}. Also, the user-friendly nature of SARFIMA is relatively easy to explain to end-users and does not involve advanced mathematics or statistics. This enhances understanding and enables users to rely on the model for decision-making. The epidemiology of HB and HC is impacted by various factors, contributing to the epidemic patterns of long memory and complex dynamics\textsuperscript{[16]}. Notwithstanding the advantages of SARFIMA, there is still a lack of exploration into how SARFIMA contributes to estimating HB and HC epidemics. Therefore, this study has two aims: (1) To evaluate the usefulness of SARFIMA in monitoring HB and HC epidemics (projection into 2030) in mainland China; and (2) to assess the forecasting potential of SARFIMA compared to SARIMA.

MATERIALS AND METHODS

Data collection

The monthly incidence cases of HB and HC from January 2004 to June 2023 were provided by the Chinese Center for Disease Control and Prevention (https://www.chinacdc.cn/). The population numbers were obtained from the National Statistical Yearbook. All HB and HC cases were confirmed according to the HB and HC diagnosis criteria issued by The National Health Commission of the People's Republic of China (http://www.nhc.gov.cn/wjw/s9491/wsbz.shtml). HB and HC are notifiable diseases in China, and the confirmed cases required to be reported within 24 h. Any duplicate records were removed at the end of the same month. The SARIMA and SARFIMA were constructed using data from January 2004 to June 2022, while the rest was the test set to indicate the predictive ability of both models. Because we required projection into December 2030 from July 2023 (90 data), the models were
developed using the data between January 2004 and December 2015 to project the trends between January 2016 and June 2023 (90 data) to confirm the forecasting reliability.

**SARIMA Development**

ARIMA is a widely used method that incorporates three components to model a time series: including autoregressive (AR), differencing, and moving average (MA) models\cite{12}. SARIMA is an extension of the ARIMA that includes a seasonal component, encompassing all the components of ARIMA and adding seasonal autoregressive (SAR), seasonal differencing, and seasonal moving average (SMA)\cite{18}. The SARIMA is represented as SARIMA(p, d, q)(P, D, Q)s, where (p, d, q) represents the orders of the AR, differencing, and MA, respectively, (P, D, Q) represents the seasonal counterparts, and S is the seasonality. Selecting appropriate values for p, d, q, P, D, and Q in SARIMA involves four procedures. First, HB and HC data must be stationary. If the data is found to be non-stationary through an Augmented Dickey-Fuller (ADF) test, differencing should be applied to obtain stationary series\cite{19}. Second, the initial orders for p, q, P, and Q were identified by doing autocorrelation function (ACF) and partial ACF (PACF) analyses\cite{12}. A series of attempts were then employed to identify the optimal combination by comparing Akaike's information criterion (AIC), corrected AIC (CAIC), Bayesian information criterion (BIC), and log-likelihood (LL)\cite{18}. The model with lower values for AIC, BIC, and CAIC, along with higher LL, was considered the preferred. Third, the errors from the best-fitting SARIMA were analyzed to verify if they behaved like white noise, which was done by the Ljung-Box Q test, ACF, and PACF\cite{12}. Finally, once the model passed all the required statistical diagnoses, it could project into future periods.

**SARIMA Development**

Earlier values often significantly influence subsequent ones present in time series, demonstrating a concept known as hyperbolic decay (HD) time series\cite{13}. By understanding the underlying mechanisms and developing accurate models for HD series, researchers can extract important insights into complex systems and make data-
driven predictions. The SARIMA assumes an exponential decay of ACF, while the SARFIMA considers an HD[9, 13], indicating the presence of long-term memory. For this reason, SARFIMA, with its added fractional differencing (d_f and D_f), has paid considerable attention owing to its potential to capture persistent data[13, 14]. d_f or D_f ranges from -1 to 1[15]. 1) A value < 0.5 shows a stationary series; 2) A value ∈ (-1, -0.5) means an invertible series; 3) A value ∈ (-0.5, 0) shows an anti-persistent series; 4) A value of 0 shows short memory and mean-reverting process in the series; 5) A value ∈ (0, 0.5) illustrates long-term positive dependence in the series; 6) A value ∈ (0.5, 1) suggests mean reverting but not stationary in the series; 7) A value of 1 signifies a unit root process[15]. The SARFIMA is denoted as SARFIMA \((p, d^*, q)(P, D^*, Q)\), where \(d^* = d + d_f\) and \(D^* = D + D_f\), whereby \(d_f\) or \(D_f\) lies within \((-1, 0.5)\) as the fractional term and \(d\) or \(D \geq 0\) is the integer term. The Hurst exponent (H), a measure of long-range dependency, often denotes the fractional differencing as \(d_f\) or \(D_f = H - 0.5\)[15]. The value of \(H\) falls between 0 and 1, with \(H < 0.5\) showing an intermediate-memory series, \(H = 0.5\) showing an uncorrelated series, and \(H > 0.5\) indicating a long-memory series[15]. The \(H\) was computed using the rescaled range (R/S) method in this study[13]. As Veenstra’s finding[14], the SARFIMA uses multiple starting values during its initial modelling phase. Consequently, it results in multiple modes, and the selection of the optimal one becomes a key step. From all possible modes, the one with the highest LL, alongside the lowest AIC and BIC, was the most considered[14]. The remaining steps of constructing the SARFIMA, such as parameter estimation and model diagnostics, were executed as described in SARIMA.

**Statistical analysis**

The Hodrick-Prescott (HP) method was employed to decompose the series into its trend and seasonality components[20]. The seasonal factor (SF) denotes the extent to which the incidence for a specific period deviates from the mean level (SR > 1 indicating a high-risk season; otherwise, a low-risk)[21], which was computed by multiplicative decomposition. The annual percentage change (APC) and average annual percentage change (AAPC) with a 95% confidence interval (CI) were calculated using Joinpoint
(Version 4.8.0.1) to assess the changing trends\cite{23}. The SARIMA and SARDIMA were implemented using the "forecast" and "arfima" packages in R (version 4.2.0, R Development CoreTeam, Vienna, Austria). Additionally, considering the significant impact of the COVID-19 pandemic on disease epidemics, a variable was created that assigned a value of "1" between January 2020 and March 2023 and "0" between January 2004 and December 2019 to control for its impact on forecasting ability.

Evaluating the effectiveness of both models involved comparing the predicted values with the actual values of the test data using various performance metrics such as the mean absolute deviation (MAD), root mean square error (RMSE), mean absolute percentage error (MAPE), mean error rate (MER), and root mean square percentage error (RMSPE)\cite{24}. A lower value on these metrics shows a more accurate forecast.

**RESULTS**

*Data description*

Between January 2004 and June 2023, China reported 23,400,874 HB cases (with incidence rates of 87.92 per 100,000 persons per year and 7.33 per 100,000 persons per month) and 3,590,867 HC cases (with incidence rates of 13.41 per 100,000 persons per year and 1.12 per 100,000 persons per month). Figure 1 illustrates that HB exhibited an overall trend of stabilization (AAPC=0.44, 95%CI -0.94~1.84) but we observed a rise during 2004-2007 (AAPC=9.13, 95%CI 2.02~16.74), a decline during 2007-2014 (AAPC=-3.53, 95%CI 2.02~16.74), and a steady level during 2014-2022 (AAPC=0.86, 95%CI -0.61~2.35) (Figure 1A). HC showed an overall trend of escalation (AAPC=8.91, 95%CI 6.98~10.88) but we noted an unpredictable upsurge during 2004-2007 (AAPC=32.05, 95%CI 20.31~44.94), a considerable increase during 2007-2012 (AAPC=15.48, 95%CI 11.34~19.78), a slight upturn during 2012-2019 (AAPC=1.65, 95%CI 0.08~3.24), and a reduction during 2019-2022 (AAPC=-4.30, 95%CI -8.7~0.32) (Figure 1C). Based on Figure 1B and D, it is evident that HBC and HCV infections could occur year-round. However, there are notable fluctuations in infection rates. March shows the highest infection rates, with an SF of 1.13 for HB and 1.16 for HC, signifying peaks. In contrast, February
records the lowest infection rates, with SF values of 0.87 for HB and 0.81 for HC, indicating troughs. The infection rates for the other months remain relatively stable.

The preferred SARIMA

A stationary test for the HB incidence series between January 2004 and June 2022 generated $ADF=-0.29 \ (P=0.52)$, and thus a non-stationary series was demonstrated. Given seasonality in the series, it was then seasonally differenced once ($ADF=-6.12$, $P<0.001$), showing its stationarity. By observing the ACF and PACF of the differenced series (Supplementary Figure 1), a series of attempts were done, and we finally identified eight possible models with significant parameters (Supplementary Table 1).

To ensure the selection of the optimal model, the "auto.arima" function in R was also run, ultimately leading to the automatic choice of SARIMA$(1,0,2)(2,0,0)_{12}$. By comparing the resulting nine models, it can be seen that SARIMA$(3,0,0)(0,1,2)_{12}$ was preferred as it gave the least values of AIC (4363.97), CAIC (4364.53), and BIC (4387.4), alongside the greatest value of LL (-2174.99). For the diagnoses of residuals, no correlations toughed the significant bounds besides the ones at delays of 11 and 22 in the ACF and PACF plots (Supplementary Figure 2A) and the $P=0.49$ for Ljung-Box Q statistic indicated an unrelated series of residuals, meaning that this preferred model sufficiently fits the data, which could be used to project into the next 12 data (Figure 2A). Likewise, following the modelling methods, SARIMA$(1,0,2)(2,1,0)_{12}$ (AIC=2748.41, CAIC=2749.08, BIC=2765.71, and LL=-1368.21), SARIMA$(3,0,0)(0,1,1)_{12}$ (AIC=3636.75, CAIC=3637.17, BIC=3656.84, and LL=-1812.38), and SARIMA$(3,1,0)(2,1,0)_{12}$ (AIC=2022.09, CAIC=2022.77, BIC=2219.34, and LL=-1095.05) were determined as the optimal for 90-step-ahead forecasts of HB (Figure 2B), 12-step-ahead forecasts of HC (Figure 3A), and 90-step-ahead forecasts of HC (Figure 3B), respectively (statistical checks for the forecast errors are illustrated in Supplementary Figure 2B-D).

The preferred SARMF

The resulting $R/S=0.73$ for the HB incidence series between January 2004 and June 2022 indicated that the series exhibits long-range dependence, and thus it is well suited to build SARMF. Subsequently, based on the fitting steps, the
SARFIMA(3,0,0)(0,0.449,2)_{12} with one mode tended to be indicated the preferred because it yielded the least values of AIC (4014.66) and BIC (4048.69), along with the maximum value of LL (-1997.33). **Supplementary Figure 3A** presents the ACF and PACF analyses for the residuals, showing that all spikes were within the significance bounds and $P=0.48$ for Ljung-Box Q statistic pinpointed that the forecast errors belonged to a white noise series. Accordingly, this preferred model could make a forecast for the 12-holdout data (**Figure 3A**). In the same way, SARFIMA(1,0,2)(2,0.454,0)_{12} (R/S=0.76, AIC=2610.36, BIC=2637.09, and LL=-1296.18), SARFIMA(3,0,0)(0,0.428,1)_{12} (R/S=0.84, AIC=3257.01, BIC=3287.63, and LL=-1619.5), and SARFIMA(3,-0.155,0)(2,-0.274,0)_{12} (R/S=0.83, AIC=2036.61, BIC=2063.34, and LL=-1009.3) were identified as the best model for 90-step-ahead forecasts of HB (**Figure 2B**), 12-step-ahead forecasts of HC (**Figure 3A**), and 90-step-ahead forecasts of HC (**Figure 3B**), respectively (the resulting modes and required diagnoses of the forecast errors are given in **Supplementary Tables 2-4 and Supplementary Figure 3B-D**).

**Assessing predictive accuracy and reliability**

The data in **Table 1** reveals the forecasting accuracy of both SARIMA and SARFIMA. Notably, SARFIMA exhibits smaller values of MAD, MAPE, RMSE, MER, and RMSPE. **Figures 2A-B and 3A-B** demonstrate a comparison between the forecasts generated by both models and the observed values. The SARFIMA depicts a closer resemblance to the actual trends and seasonality in comparison to SARIMA. These results meant that the SARFIMA outperforms SARIMA. Moreover, two sensitivity analyses were conducted to examine the influence of the age of the affected population and the cultural patterns during the spring season on the predictive quality of SARFIMA. These analyses indicated that SARFIMA consistently produced lower forecasting error rates compared to SARIMA (**Supplementary Tables 5 and 6**). This reinforces the robustness of the SARFIMA. Consequently, a projection into 2030 for HB and HC epidemics was done by identifying the optimal SARFIMA(1,0,1)(0,0.438,1)_{12} and SARFIMA(1,0.429,1)(2,0,0)_{12}, respectively, on the whole data. The resulting results indicated that HB would reach a plateau in the upcoming years (**Figure 2C**), and the
forecasts totaled 9,865,400 (95% CI 7,508,093 ~ 12,222,709) incidents, with a yearly average of 1,233,175 (95% CI 938,512 ~ 1,527,839) incidents (Table 2); HC would begin to recede in the next years (Figure 3C), and the forecasts totaled 1,659,485 (95% CI 856,681 ~ 2,462,290) incidents, with an annualized average of 207,436 (95% CI 107,085 ~ 307,786) incidents (Table 2).

DISCUSSION

Hepatitis poses a major threat to public health globally[8,18]. Analyzing and predicting epidemics were of great importance for prevention and control. This study represents an important contribution to the field as it is the first to explore the efficacy of SARFIMA in monitoring HB and HC epidemics, while also comparing its predictive accuracy with SARIMA. The results supported our initial hypothesis, demonstrating SARFIMA as a more comprehensive approach for capturing the epidemic dynamics of HB and HC compared to SARIMA. Importantly, the forecasting robustness was confirmed by our further sensitivity analyses (Supplementary Tables 5 and 6). Also, previous work indicated that SARFIMA showcases a good performance in forecasting costs[13], road fatality rate[15], CO₂ emission[23], temperature[14], and stock markets[24]. These findings provide further validation for the effectiveness of SARFIMA as a promising alternative in monitoring the spread of HB and HC. Moreover, the application of SARFIMA can superbly contribute to guiding the intensity and type of public health measures. For example, if the model clearly shows an upsurge in the midst of receding HB and HC epidemics, suggesting the effectiveness of measures currently in place. Instead, if SARFIMA predicts a decline despite increasing HB and HC epidemics, heightening the need for further or optimized measures. These practical and actionable insights hold great promise for SARFIMA in monitoring and controlling HB and HC epidemics.

SARIMA has gained significant recognition within the fields of economics, finance, meteorology, and healthcare as a reliable method for analyzing and forecasting time series data[11,17,18]. The SARIMA can adapt to different time series patterns by adjusting
the orders of nonseasonal and seasonal terms that incorporate trend, seasonal, and random components into the modelling framework, and can often generate acceptable forecasts\cite{5,7,8}. As evidenced by our study that SARIMA generated a relatively accurate forecast despite its inferiority to SARFIMA in terms of predictive quality. SARIMA has been indicated satisfactory application in predicting the spread of various infectious diseases such as hepatitis\cite{2}, COVID-19\cite{5}, HFRS\cite{9}, HFMD\cite{10}, and tuberculosis\cite{11}. Also, analysts well-versed in the underlying principles and the procedural steps involved with SARIMA have been able to generate informed and precise forecasts, thereby assisting decision-making and planning efforts in containing the spread of diseases. Notwithstanding this, SARIMA has indicated to be effective in capturing regular short-run dynamics and simple seasonality, and it often generates over-differencing\cite{13,18}. By contrast, SARFIMA combining the strengths of SARIMA and fractional integration can capture long-range dependence, handle complex seasonal patterns, accommodate both stationarity and non-stationarity, provide reliable parameter estimates, and offer robustness and stability\cite{14,15,23}, enabling it an invaluable tool for monitoring HB and HC epidemics, contributing to more informed decision-making and improved understanding of complex temporal dynamics. Consider that time series analysis is a crucial aspect of forecasting that combines various factors and the comprehensive effects of uncertain variables into a time variable, which is cost-effective and widely applicable in practice\cite{25}. Promoting the adoption of SARFIMA can contribute to the improved accuracy and reliability of modelling and forecasting other infectious diseases. However, it is crucial to further validate this generalization. It is also worth mentioning that recent studies have unveiled satisfactory applications of alternative models such as Bayesian structural time series and innovation state-space framework for assessing the epidemics of diseases\cite{26,27}. Accordingly, additional studies focus on comparing and confirming the forecasting performance of these models alongside SARFIMA.

Different from the global declined trend in HB and HC incidences\cite{3}, an overall increase at an average rate of 0.44\% for HB and 8.91\% for HC per year was noted in our
study, also consistent with earlier studies in Guangxi\textsuperscript{8}, China\textsuperscript{4}, and Pakistan\textsuperscript{28}. The gradual improvement of the surveillance system and the increased diagnostic capabilities contribute to such a trend in HB and HC\textsuperscript{29}. A significant decline in HB incidence during 2007-2014 can be attributed to the incorporation of HB prevention and treatment into the "Eleventh Five-Year" and "Twelfth Five-Year" plans in China. Comprehensive measures such as strengthening vaccination, enhancing public awareness and education, and conducting training programs were implemented\textsuperscript{14, 30}. Previous studies indicated that there was an upturn in HB cases in China since 2016/2017\textsuperscript{2, 8}, aligning with our results, possibly attributable to the accelerated urbanization process, a significant increase in the migrant population, a rapid rise in co-infections with HB, and the heavy economic burden. HC has insidious onset and nonspecific symptoms, making it difficult to be detected in the early stages. Although there is no effective vaccine for HC so far, with the gradual expansion of monitoring and testing coverage and the reporting of several outbreaks, more people have undergone screening, resulting in a rapid increase in confirmed cases during 2004-2012\textsuperscript{29}, this matched well with our findings. Recent years have witnessed a decline in HC, which may be associated with the strict screening of blood donors and the standardized management of blood products, the comprehensive implementation of monitoring, early warning, intervention and assessment measures for infectious diseases, the improvement in public knowledge-attitude-behavior regarding hepatitis, the continued optimization of policies, and the improvement of medical insurance\textsuperscript{4, 29}. Besides, according to the predicted figures for 2030, HB reached a plateau and HC receded, it can be said that the elimination of HB and HC by 2030 under current interventions faces enormous challenges. Therefore, comprehensive measures should be taken, such as expanding the scope of adult HB vaccination, a breakthrough in vaccination for HC, preventing mother-to-child transmission, investigating high-risk factors, implementing standardized antiviral treatment in rural areas, and enhancing health education and promotion\textsuperscript{2, 4}. 
HB and HC show a seasonal profile in this study, with a trough in February and a peak in March, consistent with prior reports\textsuperscript{[7]}. The seasonal trough during Spring Festival is largely attributed to people's reluctance to seek medical treatment and under-reporting that is more severe than in other months. However, the seasonal peak is associated with large-scaled population movement after Spring Festival and increased participation in various entertainment activities during the holiday. People tend to engage in unhealthy lifestyles such as excessive eating and irregular sleep, which exacerbate the condition of HBC and HCV-infected individuals and lead to an increase in patients after the festival. However, some studies have also indicated that HB and HC incidences follow the epidemiological characteristics of blood-borne and sexually transmitted diseases, with less pronounced seasonality\textsuperscript{[29]}.

Shortcomings also need to be considered. First, data was taken from a passive monitoring system, under-reporting is inevitable. Second, regional heterogeneity varied greatly in HB and HC incidences in China, and the region-specific forecasting ability of SARFIMA may require additional validation. Third, to capture timely information, the model requires integrating new data duly. Fourth, limited data may not exhibit long-term dependencies, and thus a series with 100 or more samples is recommended in application\textsuperscript{[7]}. Fifth, due to the lack of available data pertaining to the nutritional status of the population, comorbidities such as diabetes and hypertension, the specifics of daily water consumption, and immunological resistance related to the genetic traits of the Chinese population, we are unable to provide a more detailed analysis of how these factors may impact the observed results. Lastly, whether the SARFIMA is transferable to monitor other infectious diseases, verification is warranted.

CONCLUSION
Overall, HB remains steady while HC is rising in China, both exhibit a seasonal pattern, a peak in March and a trough in February. Under current interventions, additional feasible and effective control strategies require to be designed to ensure the elimination of HB and HC by 2030. The SARFIMA provides a more sophisticated and adaptable
framework for capturing intricate patterns and interdependencies in monitoring HB and HC epidemics, as opposed to SARIMA. This ultimately leads to enhanced forecasting capabilities and a deeper comprehension of the underlying process. Consequently, the integration of SARFIMA into public health decision-making for the management of HB and HC epidemics can result in more informed and efficacious interventions.

ARTICLE HIGHLIGHTS

Research background
Hepatitis B (HB) and hepatitis C (HC) have the largest burden in China, and a goal of eliminating them as a major public health threat by 2030 has been raised.

Research motivation
Accurate prediction helps to anticipate possible scenarios and make proactive choices, enabling policymakers to make informed decisions, plan strategies, and prepare for potential challenges and opportunities.

Research objectives
This study aimed to evaluate the usefulness of seasonal autoregressive fractionally integrated moving average (SARFIMA) in monitoring HB and HC epidemics (projection into 2030) in mainland China and to assess the forecasting potential of SARFIMA compared to seasonal autoregressive integrated moving average (SARIMA).

Research methods
The monthly incidence cases of HB and HC from January 2004 to June 2023 were obtained. Then, the two periods (from January 2004 to June 2022 and from January 2004 to December 2015, respectively) were used as the training sets to build the SARFIMA and SARIMA models, while the remaining periods served as the test sets to evaluate the forecasting accuracy of both models.
Research results
During the study period, a total of 23,400,874 HB cases and 3,590,867 HC cases were reported. In the 12-step-ahead HB, 90-step-ahead HB, 12-step-ahead HC, and 90-step-ahead HC forecasts, the best SARFIMA generates lower error rates compared with the best SARIMA. The predicted HB incidents totaled 9,865,400 (95% confidence interval [CI] 7,508,093 - 12,222,709) and HC totaled 1,659,485 (95% CI 856,681 - 2,462,290) during 2023-2030.

Research conclusions
The SARFIMA provides a more sophisticated and adaptable framework for capturing intricate patterns and interdependencies in monitoring HB and HC epidemics compared with the SARIMA. This ultimately leads to enhanced forecasting capabilities and a deeper comprehension of the underlying process.

Research perspectives
The integration of SARFIMA into public health decision-making for managing HB and HC epidemics can result in more informed and efficacious interventions.

ACKNOWLEDGEMENTS
We thanked the Chinese CDC for sharing the HB and HC incidence series data in mainland China.
Anna E. Sundström, Camilla Hakelind. "Examining clinical skills and abilities in psychology - implementation and evaluation of an objective structured clinical examination in psychology", The Journal of Mental Health Training, Education and Practice, 2022

Crossref