



## Trends and multi-model prediction of hepatitis B incidence in Xiamen



Ruixin Zhang<sup>a, 2</sup>, Hongfei Mi<sup>b, 2</sup>, Tingjuan He<sup>b</sup>, Shuhao Ren<sup>a</sup>, Renyan Zhang<sup>a</sup>, Liansheng Xu<sup>c</sup>, Mingzhai Wang<sup>d, 1, \*\*</sup>, Chenghao Su<sup>b, 1, \*</sup>

<sup>a</sup> School of Public Health, Xiamen University, Xiamen City, Fujian Province, China

<sup>b</sup> Department of Public Health, Zhongshan Hospital (Xiamen), Fudan University, Xiamen City, Fujian Province, China

<sup>c</sup> Department of Endemic Disease and Chronic Non-communicable Disease Prevention and Control, Xiamen Center for Disease Control and Prevention, Xiamen City, Fujian Province, China

<sup>d</sup> Department of Occupational Health and Poison Control, Xiamen Center for Disease Control and Prevention, Xiamen City, Fujian Province, China

### ARTICLE INFO

#### Article history:

Received 29 December 2023

Received in revised form 30 July 2024

Accepted 5 August 2024

Available online 6 August 2024

Handling Editor: Dr. Raluca Eftimie

#### Keywords:

Hepatitis B

Temporal trends

Prediction

Joinpoint regression model

Age-period-cohort model

Neural network autoregressive model

### ABSTRACT

**Background:** This study aims to analyze the trend of Hepatitis B incidence in Xiamen City from 2004 to 2022, and to select the best-performing model for predicting the number of Hepatitis B cases from 2023 to 2027.

**Methods:** Data were obtained from the China Information System for Disease Control and Prevention (CISDCP). The Joinpoint Regression Model analyzed temporal trends, while the Age-Period-Cohort (APC) model assessed the effects of age, period, and cohort on hepatitis B incidence rates. We also compared the predictive performance of the Neural Network Autoregressive (NNAR) Model, Bayesian Structural Time Series (BSTS) Model, Prophet, Exponential Smoothing (ETS) Model, Seasonal Autoregressive Integrated Moving Average (SARIMA) Model, and Hybrid Model, selecting the model with the highest performance to forecast the number of hepatitis B cases for the next five years.

**Results:** Hepatitis B incidence rates in Xiamen from 2004 to 2022 showed an overall declining trend, with rates higher in men than in women. Higher incidence rates were observed in adults, particularly in the 30–39 age group. Moreover, the period and cohort effects on incidence showed a declining trend. Furthermore, in the best-performing NNAR(10, 1, 6)[12] model, the number of new cases is predicted to be 4271 in 2023, increasing to 5314 by 2027.

**Conclusions:** Hepatitis B remains a significant issue in Xiamen, necessitating further optimization of hepatitis B prevention and control measures. Moreover, targeted interventions are essential for adults with higher incidence rates.

\* Corresponding author. Zhongshan Hospital (Xiamen), Fudan University, No. 668, Jinhua Road, Huli District, Xiamen City, Fujian Province, China.

\*\* Corresponding author. Xiamen Center for Disease Control and Prevention, No 681-685, Shengguang Road, Jimei District, Xiamen City, Fujian Province, China.

E-mail addresses: [zrxzrx1227@163.com](mailto:zrxzrx1227@163.com) (R. Zhang), [mi.hongfei@zsmxhospital.com](mailto:mi.hongfei@zsmxhospital.com) (H. Mi), [he.tingjuan@zsmxhospital.com](mailto:he.tingjuan@zsmxhospital.com) (T. He), [2417929104@qq.com](mailto:2417929104@qq.com) (S. Ren), [1348487646@qq.com](mailto:1348487646@qq.com) (R. Zhang), [xlsxmcdc@163.com](mailto:xlsxmcdc@163.com) (L. Xu), [xmcdcwmz@163.com](mailto:xmcdcwmz@163.com) (M. Wang), [13799752802@163.com](mailto:13799752802@163.com) (C. Su).

Peer review under responsibility of KeAi Communications Co., Ltd.

<sup>1</sup> These authors are joint senior authors and contributed equally to this work.

<sup>2</sup> These authors contributed equally to this study.

## 1. Introduction

Hepatitis B is an infectious disease caused by the Hepatitis B virus (HBV), which can be transmitted through contact with infected blood and semen (Trépo et al., 2014). In high-prevalence areas, mother-to-child transmission is the primary mode of HBV spread, while in low-prevalence areas, sexual transmission predominates (Trépo et al., 2014). The World Health Organization (WHO) has set a goal to eliminate Hepatitis B by 2030; however, achieving this target still faces many challenges (Al-Busafi & Alwassief, 2024). In 2019, approximately 296 million people worldwide were living with chronic HBV infection (World Health Organization, 2021).

China is the country with the highest number of HBV infections, accounting for one-third of the global infected population (Su et al., 2022). Comprehensive hepatitis B serological surveys conducted in 1979 and 2006 in China revealed national positivity rates for Hepatitis B surface antigen (HBsAg) of 8.75% and 7.18%, respectively (Cui et al., 2017; Xu & Zhang, 2017). In contrast, Xiamen, located in southeastern China, consistently exhibits higher prevalence rates than the national average. Surveys conducted in Xiamen during the same years reported HBsAg positivity rates of 14.73% in 1979 and 13.79% in 2006 (Chenghao et al., 2007; Su, 2015).

To effectively prevent and control hepatitis B in Xiamen, understanding the temporal trends in the incidence rate of hepatitis B and exploring the effects of age, period, and cohort on the incidence rate are crucial for developing public health policies. Additionally, predicting future incidence aids in resource allocation planning. However, studies on the temporal trends and the effects of age, period, and cohort on the incidence rates of hepatitis B in Xiamen remain insufficient. Although there have been studies predicting future incidence of hepatitis B, these studies often use a single time series model and lack verification of model performance (Wang et al., 2023; C. Zhang et al., 2022; Zhao et al., 2022).

Recent research has increasingly applied various time series models to predict infectious diseases. For example, the Seasonal Autoregressive Integrated Moving Average (SARIMA) Model, an extension of the ARIMA model designed to capture seasonal patterns in time series data, has been widely applied to forecast the incidence of diseases such as mumps (Qiu et al., 2021), tuberculosis (TB) (Azeez et al., 2016), and acquired immune deficiency syndrome (AIDS) (Luo et al., 2022). The Exponential Smoothing (ETS) Model, which emphasizes level, trend, and seasonal components through exponential weights, has been used to predict TB incidence rates (Kuan, 2022) and acute hemorrhagic conjunctivitis (AHC) cases (Liu et al., 2020). The Prophet model, a forecasting tool created by Facebook that accounts for daily, weekly, and yearly seasonality, has been effectively used in predicting the incidence of AIDS (Luo et al., 2022) as well as hand, foot, and mouth disease (HFMD) (Xie et al., 2021). Additionally, the Bayesian Structural Time Series (BSTS) Model, which decomposes the time series into several components such as trend, seasonality, cyclicity, and regression effects, has been utilized to anticipate malaria cases (Vavilala et al., 2022) and varicella incidence (Bai et al., 2023).

Furthermore, many studies demonstrate that the Neural Network Autoregressive (NNAR) Model, which integrates neural network capabilities with autoregressive techniques, and hybrid models that combine various approaches, outperform traditional methods in forecasting certain infectious diseases. Specifically, NNAR models have shown superior performance in predicting cumulative confirmed daily cases and cumulative confirmed daily deaths of coronavirus disease 2019 (COVID-19) (Daniyal et al., 2022). SARIMA-NNAR hybrid models have demonstrated better performance in forecasting TB incidence (Azeez et al., 2016), HFMD cases (Yu et al., 2021), and human brucellosis cases (Akeremi et al., 2022). Additionally, a SARIMA-ETS hybrid model and a SARIMA-Prophet hybrid model have excelled in predicting the incidence of TB and AIDS, respectively (Kuan, 2022; Luo et al., 2022).

However, research on predicting hepatitis B incidence using various time series models is currently limited. Most prediction studies have utilized a single forecasting model for hepatitis B, such as the ARIMA model (Wang et al., 2020; C. Zhang et al., 2022), SARIMA model (Zhao et al., 2022), and BSTS model (He et al., 2024). Considering that hepatitis B incidence may exhibit unique short-term fluctuations, seasonality, and long-term trends in different regions, employing a wider variety of time series models could better capture the characteristics of the data in Xiamen. Comparing these models would help identify the model with the highest performance for more accurate predictions.

In this study, we utilized the Joinpoint Regression Model to analyze the temporal trends of hepatitis B incidence rates in Xiamen from 2004 to 2022. Additionally, the Age-Period-Cohort (APC) Model was used to evaluate the effects of age, period, and cohort on incidence rates. Moreover, the performance of six distinct time series models, including the NNAR model, BSTS model, Prophet, ETS model, SARIMA model, and Hybrid model, was assessed using Root Mean Square Error (RMSE), R-squared, Mean Absolute Error (MAE), and the Z-normalized composite index. The model with the highest performance was then used to forecast the number of hepatitis B cases from 2023 to 2027. These analyses aim to provide valuable information to health management agencies for the prevention and control of hepatitis B.

## 2. Materials and methods

### 2.1. Data source

We obtained the incidence data of hepatitis B and population statistics from the China Information System for Disease Control and Prevention (CISDCP). This system, designed and constructed by the Chinese Center for Disease Control and Prevention (CCDC), is an internet-based system for reporting 41 types of infectious diseases (Wang et al., 2008). The CISDCP, which has been working since January 1, 2004, had covered over 90% of medical institutions at all levels by 2016 (Chinese Center for Disease Control and Prevention, 2007).

Cases of hepatitis B with residential addresses in Xiamen City and onset dates between 2004 and 2022 were selected for our study. Suspected cases (2592 cases), pathogen carriers (298 cases), and positive test cases (2 cases), which do not fall within the mandated reporting criteria, were excluded. To enhance the accuracy of trend analysis, only clinically diagnosed and confirmed cases were included.

In addition, 6060 cases reported more than once were excluded from the dataset. To identify duplicate entries, for cases reported from 2016 onwards, each entry had a unique and valid identification number. Entries with the same identification number appearing two or more times were considered duplicates. For cases reported before 2016, where identification numbers were unavailable, duplicates were identified by matching entries with the same name, birthdate, and gender. Only the first report was retained, and subsequent duplicates were excluded. The workflow for data processing, model building, and model performance evaluation is shown in Fig. 1.

### 2.2. Statistical analysis

#### 2.2.1. Age-standardized incidence rate

The age-standardized incidence rate was calculated using the direct method, with the standard population structure sourced from the World Health Organization's (WHO) new World Standard Population (<https://seer.cancer.gov/stdpopulations/world.who.html>).

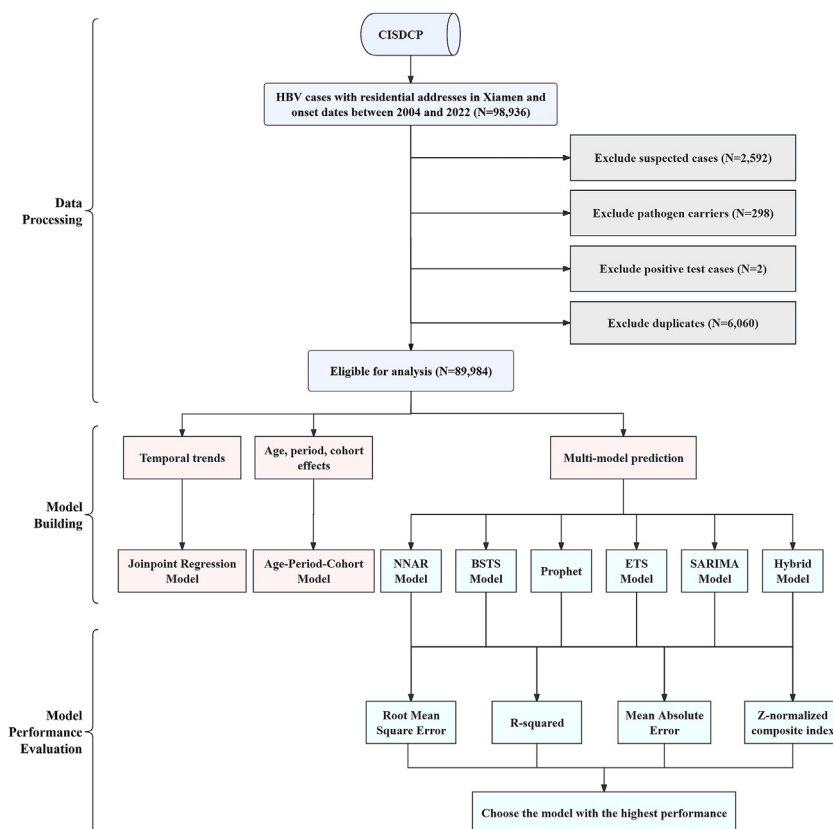


Fig. 1. The flowchart for data processing, model building, and model performance evaluation.

### 2.2.2. Joinpoint Regression Model

The Joinpoint Regression Model was used to analyze the temporal trend of the hepatitis B incidence rate from 2004 to 2022 in Xiamen and identify optimal change points (joinpoints), where a significant change in the trend direction or slope occurs (Kim et al., 2000). The Joinpoint Regression Model, designed to capture shifts in trends within observational data  $(x_1, y_1), \dots, (x_n, y_n)$ , consists of several continuous linear segments. This model can be expressed as (Kim et al., 2000):

$$E[y|x] = \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \dots + \delta_k (x - \tau_k)^+$$

In this formulation,  $E[y|x]$  represents the expected incidence rate of hepatitis B. The parameter  $\beta_0$  is the intercept and  $\beta_1$  is the slope. The coefficient  $\delta_1$  and  $\delta_k$  quantify the changes in the slope of the incidence rate at the joinpoints  $\tau_1$  and  $\tau_k$ , respectively. The function  $(x - \tau_1)^+$  is a piecewise function that equals  $x - \tau_1$  if  $x > \tau_1$  and 0 otherwise.

In the Joinpoint Regression Model, the least squares method is employed to fit linear segments to the data, effectively minimizing the sum of squared errors between observed values and modeled values (Zhang et al., 2022). To optimally identify joinpoints, a grid search algorithm is used (Kim et al., 2000). The significance of these joinpoints is then assessed using the Monte Carlo permutation method (Kim et al., 2000).

In this model, the Annual Percent Change (APC) and Average Annual Percentage Change (AAPC) were calculated. APC and AAPC are used to describe the trends in disease burden over specific time periods and across the entire observation period, respectively (National Cancer Institute). Positive values of AAPC and APC indicate an increasing trend in the disease burden indicators, while negative values indicate a decreasing trend (Huang et al., 2022).

### 2.2.3. Age-Period-Cohort model

The APC model was used to assess age, period, and cohort effects on the incidence rate of hepatitis B in Xiamen (Land, 2013). The APC model can be expressed as follows:

$$Y_j = \mu + \alpha \text{age}_j + \beta \text{period}_j + \gamma \text{cohort}_j + \varepsilon_j$$

In this formulation,  $Y_j$  denotes the net effect on the incidence of hepatitis B for group  $j$ . The coefficients  $\alpha$ ,  $\beta$  and  $\gamma$  quantify the influence of age, period, and cohort, respectively.  $\mu$  serves as the model's intercept, and  $\varepsilon_j$  denotes the residuals of the model. The significance of the parameters was assessed using Wald Chi-Square tests.

We primarily focus on three parameters in the APC model: the longitudinal age curve, the period rate ratios (RR), and the cohort rate ratios. The longitudinal age curve reflects the impact of age on disease burden, while the period rate ratios typically reflect the influence of policy changes, public health interventions, and advancements in medical technology on disease burden. The cohort rate ratios reflect the impact of factors such as economic development and harmful exposures of different birth cohorts on disease burden (Rosenberg et al., 2014).

We utilized the Age Period Cohort Analysis Web Tool developed by the National Cancer Institute of the United States (<https://analysistools.nci.nih.gov/apc/>), which could address the issue of identifiability (Cao et al., 2021; Ma et al., 2022). In the APC model, the number of cases and population data need to be arranged into 5-year periods from 2008 to 2022, and into 5-year age intervals from 5 to 85+ years old.

The Joinpoint Regression Model and the APC Model both analyze time series data, but they differ significantly in their analytical objectives, underlying formulas, and key parameters. A comparative table of these two models specific to the context of this study is available in Appendix A, Table S1.

### 2.2.4. Time series models

To better grasp the data characteristics, we compared six different time series models: the NNAR model, BSTS model, Prophet, ETS model, SARIMA model, and a Hybrid model combining the SARIMA model, ETS model, Seasonal and Trend Decomposition using Loess (STL), and NNAR model. The specifics of each model are detailed below and summarized in a comparative table (Appendix A, Table S2).

The NNAR model is a time series forecasting model that integrates autoregressive techniques with neural network technology (Perone, 2022). The NNAR model is characterized by two principal parameters:  $p$  and  $k$ . The parameter  $p$  denotes the number of input lags. Meanwhile,  $k$  signifies the count of hidden neurons. When addressing seasonal data, the model evolves into the NNAR( $p, P, k$ )[ $m$ ] format, where  $P$  represents the seasonal lags and  $m$  indicates the seasonal cycle (Perone, 2022; Talkhi et al., 2021). The equation of the NNAR model is as follows (Yu et al., 2021):

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-n}) + \varepsilon_t$$

Where  $y_t$  represents the predicted values at time  $t$ .  $(y_{t-1}, y_{t-2}, \dots, y_{t-n})$  are the lagged observations used as input to the neural network to forecast the value at time  $t$ .  $f$  denotes the neural network function.  $\varepsilon_t$  is the prediction error at time  $t$ .

The BSTS Model, a stochastic state space model, decomposes the time series into several components such as trend, seasonality, cyclicity, and regression effects (Feroze, 2020). Integrating prior knowledge with original data updates the Bayesian model to its final form, the posterior distribution, which is simulated using the Markov Chain Monte Carlo (MCMC) algorithm (Katarina & Gunardi, 2023). This model is characterized by a pair of fundamental equations (Punyapornwithaya

et al., 2023).

**Observation equation**

$$y_t = Z_t^T a_t + \varepsilon_t$$

This equation links the observed data  $y_t$  to the latent state vector  $a_t$ , with  $\varepsilon_t$  representing Gaussian observation noise.

**State transition equation**

$$a_{t+1} = T_t a_t + R_t \eta_t$$

This equation describes the evolution of the state vector, where  $T_t$  is the transition matrix and  $R_t$  is the control matrix.  $\eta_t$ , the process noise, is also assumed to be Gaussian and independent of the observation noise.

Prophet is a predicting procedure based on an additive regression model. It can fit non-linear trends with yearly, weekly, daily seasonality, and holiday effects (Ning et al., 2022). The key components of the Prophet model can be summarized as (Luo et al., 2022):

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t$$

Where  $y_t$  represents the predicted value at time,  $g(t)$  is the trend function,  $s(t)$  is the seasonal component and  $h(t)$  represents the effects of holidays or events.

The ETS model applies exponential weights to historical data, where more recent observations are given higher weights (Yang et al., 2023). The model consists of Error (E), Trend (T), and Seasonal components (S), each of which can be added (A), multiplied (M), or omitted (N), leading to different variations of the model such as ETS(A, A, A) and ETS(M, A, N) (Xian et al., 2023). Taking ETS (M, A, N) as an example, its formula is as follows (Athanasopoulos, 2018):

$$y_t = (l_{t-1} + b_{t-1}) \times (1 + \varepsilon_t)$$

Where,  $y_t$  denotes the forecasted value of the series at time t,  $l_{t-1}$  represents the level component at time t-1,  $b_{t-1}$  is the trend component and  $\varepsilon_t$  is the error term.

The SARIMA Model, an advanced version of the ARIMA framework, integrates seasonal elements. Defined by the parameters (p, d, q) for the non-seasonal components and (P, D, Q)<sub>S</sub> for the seasonal aspects. In this context, 'p' and 'P' are the autoregressive (AR) terms. 'd' and 'D' denote the differencing orders. 'q' and 'Q' correspond to the moving average (MA) terms. Lastly, 's' signifies the seasonality period (Yang et al., 2023). This model requires data to be stationary, a condition confirmed in our study through the Augmented Dickey-Fuller Test (Jalil & Rao, 2019). The overall formula for the SARIMA model can be expressed as (Qiu et al., 2021):

$$\Delta^d \Delta_S^D x_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)} \varepsilon_t$$

where:

$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  is the non-seasonal moving average polynomial.

$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  is the non-seasonal autoregressive polynomial.

$\Theta_S(B) = 1 - \theta_1 B^S - \dots - \theta_Q B^{QS}$  is the seasonal moving average polynomial.

$\Phi_S(B) = 1 - \phi_1 B^S - \dots - \phi_P B^{PS}$  is the seasonal autoregressive polynomial.

$\Delta^d$  and  $\Delta_S^D$  are the non-seasonal and seasonal differencing operators, respectively.  $x_t$  denotes the observed value at time t and  $\varepsilon_t$  represents the random error or residuals at time t.

Hybrid models in time series forecasting combine elements from different statistical and machine learning models to leverage their individual strengths (Shaub, 2020). In this study, the hybrid model combines the SARIMA model, ETS model, STL, and NNAR model. The simplified formula for the hybrid model can be expressed as:

$$\hat{y}_t = w_{SARIMA} \times \hat{y}_t^{SARIMA} + w_{ETS} \times \hat{y}_t^{ETS} + w_{STL} \times \hat{y}_t^{STL} + w_{NNAR} \times \hat{y}_t^{NNAR}$$

where  $\hat{y}_t$  is the combined predicted value at time t.  $\hat{y}_t^{SARIMA}$ ,  $\hat{y}_t^{ETS}$ ,  $\hat{y}_t^{STL}$ ,  $\hat{y}_t^{NNAR}$  represent the forecasts at time t from the SARIMA, ETS, STL, and NNAR models, respectively.  $w_{SARIMA}$ ,  $w_{ETS}$ ,  $w_{STL}$ ,  $w_{NNAR}$  are the respective weights for these models, with the constraint that  $w_{SARIMA} + w_{ETS} + w_{STL} + w_{NNAR} = 1$ .

Before constructing predictive models, we used the STL to examine the seasonality of incidence in Xiamen from 2004 to 2022 and discovered regular seasonal patterns in the data (Appendix A, Fig. S1). Subsequently, six time series models were

implemented using specific packages and functions in R. The configuration details of each model in R are outlined below and presented in a summary table (Appendix A, Table S3).

Specifically, in the implementation of the NNAR model using the `nnetar` function from the `forecast` package in R, the optimal number of non-seasonal lags  $p$  was determined based on the Akaike Information Criterion (AIC). To capture seasonality, the number of seasonal lags  $P$  was set to 1 (Yu et al., 2021; Zenia et al., 2023). The optimal number of hidden neurons  $k$  was calculated using the formula  $k = \frac{(p+P+1)}{2}$ . In the training set, the NNAR(10, 1, 6)[12] model provided the most efficient fit.

The BSTS was executed through the `bsts` package, employing functions including `AddLocalLinearTrend`, `AddSeasonal` and `bsts`. Initially, the model was specified by integrating a local linear trend through the `AddLocalLinearTrend` function to capture the intrinsic trend of the time series. To account for regular seasonal fluctuations, a seasonal component of 12 periods was incorporated using the `AddSeasonal` function. The model was then fitted using the `bsts` function, where 500 MCMC iterations were performed to ensure the convergence of Bayesian inference.

For the Prophet model, the function `prophet` in the `prophet` package was used. Within the `prophet` function, the model was configured to capture yearly seasonality patterns, where seasonality effects were assumed to be additive. The seasonality prior scale, holidays prior scale, and changepoint prior scale were set to 10, 10, and 0.05, respectively.

The ETS models were implemented using the `ets` function from the `forecast` package. In this function, the model parameter was set to "ZZZ". This configuration automatically selected the optimal error type, trend type, and season type for the ETS model. The selection was based on information criteria to identify the best-fitting model. In the training set, the ETS(M, N, M) model provided the most efficient fit.

The SARIMA model was implemented using the `auto.arima` function from the `forecast` package. This function autonomously identified the optimal seasonal and non-seasonal differencing orders by conducting unit root tests. Furthermore, it employed a stepwise search algorithm to systematically evaluate various combinations of the seasonal and non-seasonal AR and MA parameters. The model that exhibited the highest performance was selected based on the lowest values of the information criteria (Kuan, 2022). In the training set, the SARIMA(1,0,0)(0,1,2)<sub>12</sub> model exhibited the lowest information criteria. The model passed the Ljung-Box Q test ( $P > 0.05$ ).

Lastly, the hybrid model was implemented using the `hybridModel` function from the `forecastHybrid` package. The model incorporated SARIMA, ETS, STL, and NNAR models, specified by the `models` parameter as "aesn", with the seasonal component included as indicated by the `a.args` parameter. For weighting the forecasts from the various contributing models, we adopted a more robust "equal weights" method.

In evaluating model performance, we used the monthly number of hepatitis B cases in Xiamen from 2004 to 2019 as the training dataset, and the monthly number of cases from 2020 to 2022 as the test dataset. The models' predictive performance was assessed using RMSE, R-squared, MAE, and Z-normalized composite index (Kuhn, 2019). RMSE measures the average magnitude of the error, R-squared indicates the proportion of variance in the dependent variable that is predictable from the independent variables, and MAE represents the average of absolute errors. The specific formulas and ranges of RMSE, R-squared, and MAE are provided in Appendix A, Table S4.

Considering the different scales and ranges of these metrics, the Z-normalized composite index was calculated to achieve a comparable and fair assessment by standardizing different assessment metrics (Andrade, 2021). On one hand, the index rescales the data to have a mean of 0 and a standard deviation of 1, which helps to compare different datasets on a common scale. On the other hand, it ensures that each metric contributes equally to the model evaluation, preventing any single metric from disproportionately influencing the outcome. The formula for calculating the Z-normalized composite index is as follows:

$$Z_m = -\frac{X_{RMSE,m} - \mu_{RMSE}}{\sigma_{RMSE}} + \frac{X_{R^2,m} - \mu_{R^2}}{\sigma_{R^2}} - \frac{X_{MAE,m} - \mu_{MAE}}{\sigma_{MAE}}$$

Where  $x$  is the original value,  $\mu$  is the mean of the variable, and  $\sigma$  is the standard deviation.  $m$  means a specific model.

The model with the highest Z-normalized composite index will be selected as the best-performing model to forecast the number of new cases in Xiamen from 2023 to 2027 using data from 2004 to 2022.

The above analyses were realized by Joinpoint (version 4.9.1.0), Age Period Cohort Analysis Web Tool (Posted: November 10, 2014), and R (version 4.3.0). P values < 0.05 were considered statistically significant.

### 3. Results

The number of hepatitis B cases in Xiamen has generally remained high and showed multiple fluctuations during the period from 2004 to 2022 (Fig. 2). The number of reported cases rose from 1127 in 2004 to 4311 in 2005. In 2010, there was an increase to 7842 cases, which then declined to 4529 in 2011. There was also a noticeable increase in 2020 and 2021, rising to 6761 and 6397 respectively (Fig. 2A). Additionally, hepatitis B cases in Xiamen exhibited seasonal fluctuations, with a significant increase in cases every January (Fig. 2B). The number of hepatitis B cases was significantly higher in males compared to females (Fig. 2C–D). Additionally, the reported cases were highest in the 30–39 age group for both males and females.



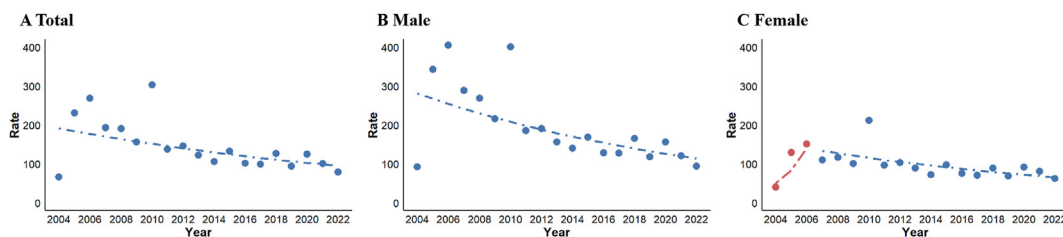
**Fig. 2.** Temporal variation in the number of hepatitis B cases in Xiamen from 2004 to 2022. (A) Annual number of Hepatitis B cases. (B) Monthly number of Hepatitis B cases. (C) Monthly number of hepatitis B cases among male age groups. (D) Monthly number of hepatitis B cases among female age groups.

### 3.1. Joinpoint regression model

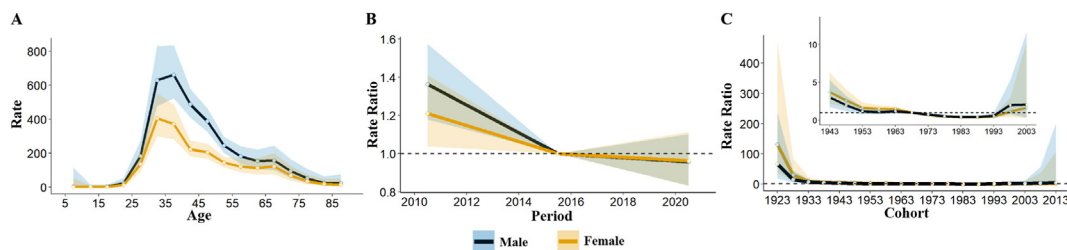
From 2004 to 2022, the age-standardized incidence rate of hepatitis B in Xiamen generally decreased (AAPC =  $-3.83\%$ ,  $P = 0.017$ ) (Fig. 3A). In addition, the incidence rate for males averaged 198.26 per 100,000, whereas for females, it averaged 97.54 per 100,000. Moreover, the incidence rate among males exhibited a more pronounced downward trend (AAPC =  $-4.89\%$ ,  $P = 0.004$ ) (Fig. 3B). The AAPC value for females was 1.47, which did not show statistical significance ( $P = 0.691$ ) (Fig. 3C). There was a turning point for females in 2006. Prior to 2006, the incidence rate in females was increasing (APC = 66.39,  $P = 0.126$ ), and after 2006, it declined (APC =  $-4.66$ ,  $P = 0.001$ ). The specific AAPC values are available in Appendix A (Table S5).

### 3.2. Age-period-cohort model

In the APC model, the longitudinal age curves of the HBV incidence are peaking in the 30–39 age group for both male and female patients. Specifically, in the 30–39 age group, the incidence in males exceeded 600 per 100,000, while in females, it surpassed 350 per 100,000 (Fig. 4A). Regarding the period effect, RR of the incidence among both men and women showed a declining trend with the progression of periods. Prior to the reference period (2015.5), the RR was significant with values greater than 1, while after this period, the RR was not statistically significant, with values less than 1 (Fig. 4B). As for the cohort effect, RR of the incidence also showed a downward trend, with the cohort RR for both males and females being higher before the reference cohort (1968) and lower after it (Fig. 4C). The Wald  $\chi^2$  test results indicated that the period and cohort effects on the incidence rates of hepatitis B in males and females were statistically significant (Appendix A, Table S6).



**Fig. 3.** Joinpoint regression model of hepatitis B incidence rates in Xiamen from 2004 to 2022. (A) Joinpoint regression model of the total population. (B) Joinpoint regression model of males. (C) Joinpoint regression model of females.



**Fig. 4.** Age-period-cohort model of hepatitis B incidence in Xiamen from 2008 to 2022. (A) Longitudinal age curve of the incidence. (B) Period rate ratios of the incidence. (C) Cohort rate ratios of the incidence.

### 3.3. Time series models

After training the NNAR model, Prophet, ETS model, SARIMA model, hybrid model, and BSTS model using the monthly number of hepatitis B cases in Xiamen from 2004 to 2019 and testing them on data from 2020 to 2022 (Fig. 5A–F), their performance was evaluated using RMSE, MAE, R-squared, and the Z-normalized composite index (Fig. 5G–J). The NNAR(10, 1, 6)[12] model achieved the highest Z-normalized composite index of 4.01 (Fig. 5J). The Hybrid model also performed well, with an index of 1.57, followed by the BSTS model with a score of 0.82. In contrast, the SARIMA(1,0,0)(0,1,2)<sub>12</sub>, ETS(M, N, M), and Prophet models exhibited poorer performance, with indexes of  $-0.38$ ,  $-2.63$ , and  $-3.38$ , respectively.

The monthly hepatitis B case data from 2004 to 2022 in Xiamen were used as the training dataset for the NNAR model, Prophet, ETS model, SARIMA model, hybrid model, and BSTS model to forecast new cases from 2023 to 2027 (Appendix A, Fig. S2). The NNAR(10, 1, 6)[12] Model, which exhibited the highest performance, predicts that the number of cases for 2023 will be 4271, with an expected increase to 5314 by 2027 (Fig. 6A). Forecasts from other models suggest that the number of new hepatitis B cases in Xiamen will remain high over the next five years, except for the BSTS model, which predicts a declining trend in the number of cases (Fig. 6B).

## 4. Discussion

This research for the first time analyzed the temporal trends of hepatitis B incidence in Xiamen from 2004 to 2022, and assessed the impact of age, period and cohort effects on the incidence rates. Additionally, we trained six time series models and utilized a multi-metric evaluation approach, including a Z-normalized composite index to select the model with the best performance to forecast the number of hepatitis B cases in Xiamen from 2023 to 2027. These methods enhance the accuracy of prediction. Our analysis of temporal trends, effects of age, period and cohort, along with multi-model forecasting, provides vital support for developing public health measures and optimizing resource allocation.

### 4.1. Joinpoint regression model

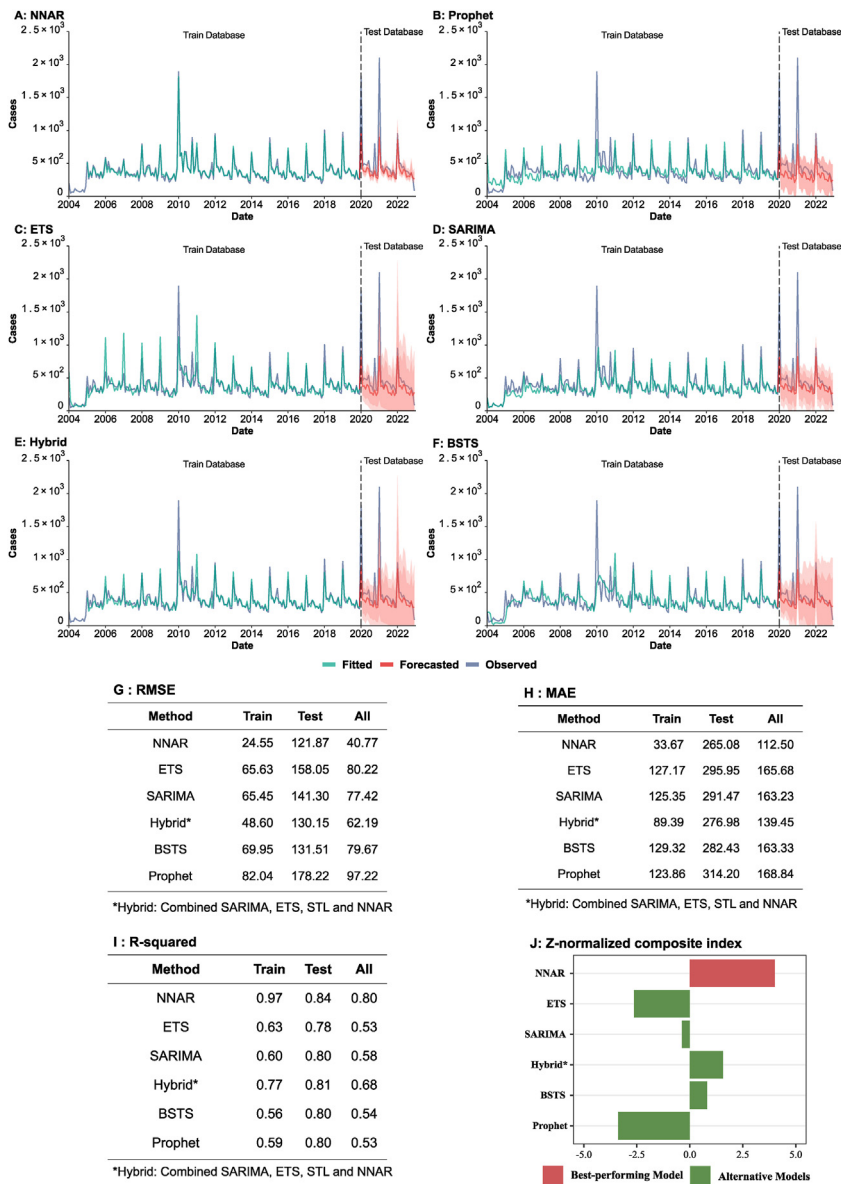
The Joinpoint regression analysis revealed that, from 2004 to 2022, the reported incidence of hepatitis B in Xiamen generally showed a downward trend. This could be attributed to the implementation of free hepatitis B vaccination for newborns since 2003 and the catch-up vaccination campaign for children under 15 from 2007 to 2010 in Xiamen, which effectively interrupted the virus transmission pathway. Moreover, the government and relevant health organizations have improved medical and health conditions and organized educational programs regarding hepatitis B, which are also associated with a decline in incidence rates (Ji et al., 2019; Qin et al., 2022).

We also observed that the incidence rate among men was significantly higher than that in women, aligning with the conclusions of existing research (Hogan et al., 2023; Qian et al., 2023; Wang et al., 2019). The gender disparity can be related to biological characteristics. Research shows that the HBV X protein enhances the transcriptional activity of androgen receptors (ARs) in a manner dependent on androgen concentration (Yang et al., 2009). Another study suggests higher HBV replication and expression in male livers than in female ones (Lee et al., 2018). Furthermore, males may be more inclined to engage in high-risk behaviors, such as having multiple sexual partners, homosexual intercourse, and sharing injection needles, so they may face higher infection risks (Fang et al., 2023; Wang et al., 2019). Emphasizing health education tailored for men may reduce these risks.

### 4.2. Age-period-cohort model

The APC model analysis shows that the adult population in Xiamen has higher hepatitis B incidence rates. This may be related to the late implementation of the vaccination policy and the declining antibody levels with increasing age (Poorolajal & Hooshmand, 2016). The free hepatitis B vaccination policy for newborns began in Xiamen in 2003. Consequently, people born before 2003 may not have been protected by the vaccine and thus have a higher risk of infection. Notably, the incidence



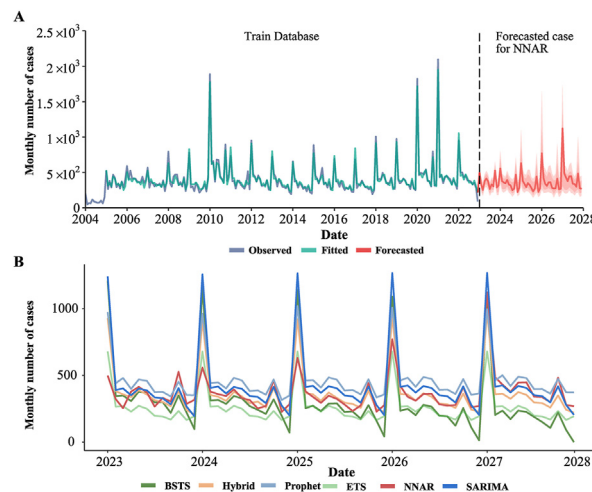


**Fig. 5.** Models training and performance comparison of six time series models. (A) NNAR Model. (B) Prophet. (C) ETS Model. (D) SARIMA Model. (E) Hybrid Model. (F) BSTS Model. (G) RMSE. (H) MAE. (I) R-squared. (J) Z-normalized composite index.

rate in the 30–39 age group is significantly higher than in other age groups. As this age group forms a vital part of the workforce for families and society, it is essential to guide them through health education to undertake vaccine catch-up, early screening, or treatment, thereby reducing the disease burden effectively.

The period RRs of the incidence showed a declining trend over time. This decline suggests improvements in public health policies, such as widespread vaccination and the prevention of mother-to-child transmission, have been effective (Fang, Cao, et al., 2023; Ji et al., 2019). Additionally, health education and public health promotion may have enhanced public understanding of hepatitis B transmission routes, contributing to the reduction in the risk of hepatitis B incidence (Qin et al., 2022). It's worth noting that although the period RR was less than 1 after mid-2015, it is not significant, indicating substantial room for improvement in optimizing prevention and control measures.

As for the cohort effect, the RR was higher before 1968 and lower after it. The higher RR before 1968 might be associated with Xiamen's economic backwardness, social unrest, poor public health conditions, and inadequate public awareness of hepatitis B during that time (Fang et al., 2023). Subsequently, China's societal development after the reform and opening-up in 1978, the implementation of the Technical Operation Procedures in 1997 and the enhancement of educational level and health awareness contributed to the decline in the cohort RR (Gao et al., 2020; Liu et al., 2019; Lu et al., 2019).



**Fig. 6.** Historical and projected monthly hepatitis B cases in Xiamen. (A) Historical trends from 2004 to 2022 and forecasts using the NNAR model from 2023 to 2027. (B) Comparative forecasts from the BSTS, Hybrid, Prophet, ETS, NNAR, and SARIMA models.

#### 4.3. Time series models

Through the comparison of six time series forecasting models, we found that the NNAR(10, 1, 6)[12] model excelled in predicting the hepatitis B cases in Xiamen. The good performance of the NNAR model in predicting other infectious diseases, such as Foot and Mouth Disease and Tuberculosis, has also been corroborated in existing research (Punyapornwithaya et al., 2022; Yadav et al., 2023). Our study applied the NNAR model to predict hepatitis B cases for the first time, revealing its superior performance over other selected models in handling monthly hepatitis B case data characterized by seasonality and non-linearity. Given this finding, the NNAR model may be considered a viable option for forecasting the number of hepatitis B cases in future studies.

In addition, we found that the Hybrid model's performance was second only to the NNAR model. The Hybrid model, by combining the SARIMA model, ETS model, STL, and NNAR model, allows for improved forecast accuracy by effectively managing linear and nonlinear elements, as well as different components like errors, trends, and seasonal variations (Perone, 2022). Although the Hybrid model includes the NNAR model, during the modeling process, we employed a more robust and computationally efficient "equal weights" method, which may have weakened the overall performance of the Hybrid model compared to the standalone NNAR model. Furthermore, the BSTS model also demonstrated good predictive performance, which is related to its effectiveness in handling uncertainty and stochastic behavior in the data (Almarashi & Khan, 2020).

As one of the most frequently utilized predictive models in hepatitis B forecasting research, the SARIMA model effectively captures seasonal fluctuations (Fang, Cao, et al., 2023; Zhao et al., 2022). However, we found that the SARIMA model performed slightly below the average when compared to all other models considered in the study. This suggests that we should be cautious when using the SARIMA model for forecasting monthly hepatitis B cases, as it may not effectively handle nonlinear data and is more suitable for short-term forecasting (Zhao et al., 2022). In past research, the ETS and Prophet models have not been used for forecasting hepatitis B data. However, our findings indicate that both the ETS and Prophet models performed poorly in this context. This may be related to the Prophet model's emphasis on ease of use and tuning at the expense of model performance, and the ETS model's inadequacy in handling nonlinear data (Adeyinka & Muhajarine, 2020; Ogallo et al., 2023).

The NNAR(10, 1, 6)[12] model, which has demonstrated the highest performance, predicts that the number of hepatitis B cases in Xiamen will remain high from 2023 to 2027. This finding highlights the continued severity of hepatitis B incidence in Xiamen, underscoring the need for timely intervention measures. Therefore, it is crucial to strengthen vaccination among high-risk adult populations (such as family members living with hepatitis B patients, healthcare workers, sex workers, individuals with multiple sexual partners, and men who have same-sex sexual behavior) (WorldHealthOrganization, 2016). Additionally, increasing the accessibility of hepatitis B testing and expanding treatment accessibility through the community are essential to reduce transmission and control the burden of hepatitis B. Furthermore, enhancing public knowledge about the prevention and control of hepatitis B through public health education is imperative to strengthen individuals' health management awareness.

Our studies had some limitations. Firstly, due to the APC model's requirement for the period interval to match the 5-year age interval, data from 2004 to 2007 were not included in the APC model analysis, potentially limiting our full use of this four-year dataset. Secondly, although the study attempted to explore the age, cohort, and period effects on hepatitis B incidence in Xiamen using the APC model, it was unable to quantify specific risk or protective factors, such as the impact of vaccination programs. Thirdly, in our predictive models, we primarily based our forecasts on incidence data without incorporating related

external factors, such as public health policies or socio-economic developments. Future research could benefit from including more external variables and exploring more complex models that can handle multiple inputs and interactions. Lastly, the inherent uncertainty of long-term forecasts, due to potential shifts in epidemiological patterns and public health interventions, poses a challenge. To ensure accuracy and relevance, it is essential to monitor epidemiological patterns regularly and update the model with new data.

## 5. Conclusion

Although the incidence rate of hepatitis B in Xiamen has generally shown a declining trend over the past 19 years, the best-performing model, the NNAR model, predicts that the number of hepatitis B cases will remain high in the next five years. Therefore, it is necessary to maintain ongoing efforts in the prevention and control of hepatitis B, particularly by implementing targeted interventions among adults with higher incidence rates.

## Funding

This research was funded by Xiamen Medical and Health Key Project [grant numbers 3502Z20191105].

## Ethics approval

Only broad information (such as the incidence) was collected with no identifying information and therefore the Zhongshan Hospital (Xiamen), Fudan University Research Ethics Committee has confirmed that no ethical approval is required. All methods were carried out in accordance with the relevant guidelines and regulations of the Helsinki Declaration.

## Data availability

The datasets used and analyzed in the current study are available from the corresponding author on reasonable request.

## CRediT authorship contribution statement

**Ruixin Zhang:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Hongfei Mi:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Tingjuan He:** Writing – review & editing, Validation, Data curation. **Shuhao Ren:** Writing – review & editing, Validation, Data curation. **Renyan Zhang:** Writing – review & editing, Data curation. **Liansheng Xu:** Writing – review & editing, Validation. **Mingzhai Wang:** Writing – review & editing, Validation, Resources, Conceptualization. **Chenghao Su:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization.

## Acknowledgments

We are grateful to the Xiamen Center for Disease Control and Prevention for providing data. We would like to thank Dr. Kangguo Li for his guidance in the coding of the predictive model.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.idm.2024.08.001>.

## References

- Adeyinka, D. A., & Muhajarine, N. (2020). Time series prediction of under-five mortality rates for Nigeria: Comparative analysis of artificial neural networks, holt-winters exponential smoothing and autoregressive integrated moving average models. *BMC Medical Research Methodology*, 20, 292. <https://doi.org/10.1186/s12874-020-01159-9>
- Akermi, S. E., L'Hadj, M., & Selmane, S. (2022). Epidemiology and time series analysis of human brucellosis in Tebessa province, Algeria, from 2000 to 2020. *Journal of Research in Health Sciences*, 22, Article e00544. <https://doi.org/10.34172/jrhs.2022.79>
- Al-Busafi, S. A., & Alwassief, A. (2024). Global perspectives on the hepatitis B vaccination: Challenges, achievements, and the road to elimination by 2030. *Vaccines*, 12. <https://doi.org/10.3390/vaccines12030288>
- Almarashi, A., & Khan, K. (2020). Bayesian structural time series. *Nanoscience and Nanotechnology Letters*, 12, 54–61. <https://doi.org/10.1166/nnl.2020.3083>
- Andrade, C. (2021). Z scores, standard scores, and composite test scores explained. *Indian Journal of Psychological Medicine*, 43, 555–557. <https://doi.org/10.1177/02537176211046525>
- Athanasopoulos, R. J. H. G. (2018). *Forecasting: Principles and practice*. Melbourne, Australia: OTexts [cited 2024 2024-6-19]. Available from: <https://otexts.com/fpp2/ets-forecasting.html>.
- Azeez, A., Obaromi, D., Odeyemi, A., Ndege, J., & Muntabayi, R. (2016). Seasonality and trend forecasting of tuberculosis prevalence data in eastern cape, South Africa, using a hybrid model. *International Journal of Environmental Research and Public Health*, 13. <https://doi.org/10.3390/ijerph13080757>
- Bai, Y., Shen, L., Sun, M., Yang, Z., Chen, Z., Zhai, J., Xue, M., Shao, Z., Liu, K., & Zheng, C. (2023). The short and long-term impact of nonpharmaceutical interventions on the prevalence of varicella in Xi'an during the COVID-19 pandemic. *Journal of Medical Virology*, 95, Article e29020. <https://doi.org/10.1002/jmv.29020>

- Cao, J., Eshak, E. S., Liu, K., Arafa, A., Sheerah, H. A., & Yu, C. (2021). Age-period-cohort analysis of stroke mortality attributable to high systolic blood pressure in China and Japan. *Scientific Reports*, 11, Article 19083. <https://doi.org/10.1038/s41598-021-98072-y>
- Chenghao, S., Jianjun, N., Yizhong, Z., & Guilin, M. (2007). Study on epidemic trend and relationship between hepatitis B and liver cancer in Xiamen City. *Science of Travel Medicine*, 30–31. [https://kns.cnki.net/kcms2/article/abstract?v=a2GKbg6fW\\_YPwBl\\_1SyqCPHdYmHugE76Y7qubm519h\\_gFfQhKmmxhnjgXDf9PRCPjtSFLAxxgwG0cmZnDYJf0Y6kQZ0JISqQGRSXVjQbr3DA5-sDhJlrwU3qaB3BjzUINvYtIMjzVQ=&uniplattm=NZKPT&language=CHS](https://kns.cnki.net/kcms2/article/abstract?v=a2GKbg6fW_YPwBl_1SyqCPHdYmHugE76Y7qubm519h_gFfQhKmmxhnjgXDf9PRCPjtSFLAxxgwG0cmZnDYJf0Y6kQZ0JISqQGRSXVjQbr3DA5-sDhJlrwU3qaB3BjzUINvYtIMjzVQ=&uniplattm=NZKPT&language=CHS)
- Cui, F., Shen, L., Li, L., Wang, H., Wang, F., Bi, S., Liu, J., Zhang, G., Wang, F., Zheng, H., Sun, X., Miao, N., Yin, Z., Feng, Z., Liang, X., & Wang, Y. (2017). Prevention of chronic hepatitis B after 3 decades of escalating vaccination policy, China. *Emerging Infectious Diseases*, 23, 765–772. <https://doi.org/10.3201/eid2305.161477>
- Daniyal, M., Tawiah, K., Muhammadullah, S., & Opoku-Ameyaw, K. (2022). Comparison of conventional modeling techniques with the neural network autoregressive model (NNAR): Application to COVID-19 data. *Journal of healthcare engineering*, 2022, Article 4802743. <https://doi.org/10.1155/2022/4802743>
- Fang, K., Cao, L., Fu, Z., & Li, W. (2023). Prediction of reported monthly incidence of hepatitis B in Hainan Province of China based on SARIMA-BPNN model. *Medicine*, 102, Article e35054. <https://doi.org/10.1097/md.00000000000035054>
- Fang, K., Shi, Y., Zhao, Z., Zhao, Y., Guo, Y., Abudunaibi, B., Qu, H., Liu, Q., Kang, G., Wang, Z., Hu, J., & Chen, T. (2023). Trends in disease burden of hepatitis B infection in Jiangsu Province, China, 1990–2021. *Infectious Disease Modelling*, 8, 832–841. <https://doi.org/10.1016/j.idm.2023.07.007>
- Feroze, N. (2020). Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models. *Chaos, Solitons & Fractals*, 140, Article 110196. <https://doi.org/10.1016/j.chaos.2020.110196>
- Gao, D., Li, H., & Wang, K. (2020). The development of a legal framework for blood donation and blood safety in China over 24 years. *BMC Health Services Research*, 20, 1099. <https://doi.org/10.1186/s12913-020-05944-6>
- He, C. Q., Sun, B. H., Yu, W. T., An, S. Y., Qiao, B. J., & Wu, W. (2024). Evaluating the impact of COVID-19 outbreak on hepatitis B and forecasting the epidemiological trend in mainland China: A causal analysis. *BMC Public Health*, 24, 47. <https://doi.org/10.1186/s12889-023-17587-3>
- Hogan, S., Page, A., Dixit, S., & McBride, K. A. (2023). HBV prevalence in sub-continental countries: A systematic review and meta-analysis. *PLoS One*, 18, Article e0295670. <https://doi.org/10.1371/journal.pone.0295670>
- Huang, J., Deng, Y., Tin, M. S., Lok, V., Ngai, C. H., Zhang, L., Lucero-Priso, D. E., 3rd, Xu, W., Zheng, Z. J., Elcarte, E., Withers, M., & Wong, M. C. S. (2022). Distribution, risk factors, and temporal trends for lung cancer incidence and mortality: A global analysis. *Chest*, 161, 1101–1111. <https://doi.org/10.1016/j.chest.2021.12.655>
- Jalil, A., & Rao, N. H. (2019). Environmental kuznets curve (EKC) <https://doi.org/10.1016/B978-0-12-816797-7.00008-4>
- Ji, W., Xie, N., He, D., Wang, W., Li, H., & Wang, K. (2019). Age-Period-cohort analysis on the time trend of hepatitis B incidence in four prefectures of southern xinjiang, China from 2005 to 2017. *International Journal of Environmental Research and Public Health*, 16. <https://doi.org/10.3390/ijerph16203886>
- Katarina & Gunardi, B. (2023). Optimization of bayesian structural time series (BSTS) applications in forecasting stock prices through state components selection. In A. Mustapha, N. Ibrahim, H. Basri, M. S. Rusiman, & S. Zuhair Haider Rizvi (Eds.), *Proceedings of the 8th international conference on the applications of science and mathematics* (pp. 229–248). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978-981-99-2850-7\\_20](https://doi.org/10.1007/978-981-99-2850-7_20), 2023//.
- Kim, H. J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for jointpoint regression with applications to cancer rates. *Statistics in Medicine*, 19, 335–351. [https://doi.org/10.1002/\(sici\)1097-0258\(20000215\)19:3<335::aid-sim336>3.0.co;2-z](https://doi.org/10.1002/(sici)1097-0258(20000215)19:3<335::aid-sim336>3.0.co;2-z)
- Kuan, M. M. (2022). Applying SARIMA, ETS, and hybrid models for prediction of tuberculosis incidence rate in Taiwan. *PeerJ*, 10, Article e13117. <https://doi.org/10.7717/peerj.13117>
- Kuhn, M. (2019). The caret Package. Retrieved from <https://topepo.github.io/caret/measuring-performance.html>. December 11.
- Land, Y. Y. K. C. (2013). Book Age-Period-Cohort analysis: New models, methods, and empirical applications. In *Boca raton (age-Period-Cohort analysis: New models, methods, and empirical applications)* (1st ed.) Boca Raton; 2013.
- Lee, S. A., Lee, S. Y., Choi, Y. M., Kim, H., & Kim, B. J. (2018). Sex disparity in viral load, inflammation and liver damage in transgenic mice carrying full hepatitis B virus genome with the W4P mutation in the preS1 region. *World Journal of Gastroenterology*, 24, 1084–1092. <https://doi.org/10.3748/wjg.v24.i10.1084>
- Liu, H., Li, C., Shao, Y., Zhang, X., Zhai, Z., Wang, X., Qi, X., Wang, J., Hao, Y., Wu, Q., & Jiao, M. (2020). Forecast of the trend in incidence of acute hemorrhagic conjunctivitis in China from 2011–2019 using the seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing (ETS) models. *Journal of infection and public health*, 13, 287–294. <https://doi.org/10.1016/j.jiph.2019.12.008>
- Liu, J., Liang, W., Jing, W., & Liu, M. (2019). Countdown to 2030: eliminating hepatitis B disease, China. *Bulletin of the World Health Organization*, 97, 230–238. <https://doi.org/10.2471/blt.18.219469>
- Lu, Y., Zhang, Y., Cao, X., Wang, C., Wang, Y., Zhang, M., Ferrier, R. C., Jenkins, A., Yuan, J., Bailey, M. J., Chen, D., Tian, H., Li, H., von Weizsäcker, E. U., & Zhang, Z. (2019). *Forty years of reform and opening up: China's progress toward a sustainable path* (Vol. 5), Article eaa9413. <https://doi.org/10.1126/sciadv.aau9413>
- Luo, Z., Jia, X., Bao, J., Song, Z., Zhu, H., Liu, M., Yang, Y., & Shi, X. (2022). A combined model of SARIMA and prophet models in forecasting AIDS incidence in henan province, China. *International Journal of Environmental Research and Public Health*, 19. <https://doi.org/10.3390/ijerph19105910>
- Ma, Y., Cao, J., Mubarik, S., Bai, J., Yang, D., Zhao, Y., Hu, Q., & Yu, C. (2022). Age-Period-cohort analysis of long trend of mortality for stroke and subtypes attributed to high SBP in Chinese adults. *Frontiers in Neurology*, 13, Article 710744. <https://doi.org/10.3389/fneur.2022.710744>
- Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and prophet. *Computers & Geosciences*, 164, Article 105126. <https://doi.org/10.1016/j.cageo.2022.105126>
- Ogalo, W., Wanyana, I., Tadesse, G. A., Wanjiru, C., Akinwande, V., Kabwama, S., Remy, S. L., Wachira, C., Okwako, S., Kizito, S., Wanyenze, R., Kiwanuka, S., & Walcott-Bryant, A. (2023). Quantifying the impact of COVID-19 on essential health services: A comparison of interrupted time series analysis using prophet and Poisson regression models. *Journal of the American Medical Informatics Association: JAMIA*, 30, 634–642. <https://doi.org/10.1093/jamia/ocac223>
- Perone, G. (2022). Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. *The European Journal of Health Economics: HEPAC: health economics in prevention and care*, 23, 917–940. <https://doi.org/10.1007/s10198-021-01347-4>
- Poorolajal, J., & Hooshmand, E. (2016). Booster dose vaccination for preventing hepatitis B. *Cochrane Database of Systematic Reviews*, 2016, Cd008256. <https://doi.org/10.1002/14651858.CD008256.pub3>
- Chinese Center for Disease Control and Prevention. (2007). Introduction to the China information system for disease control and prevention. Retrieved from November 28 [https://www.chinacdc.cn/ztxm/ggwsj/cjcx/200702/t20070215\\_41341.html](https://www.chinacdc.cn/ztxm/ggwsj/cjcx/200702/t20070215_41341.html)
- Punyapornwithaya, V., Arjkumpa, O., Buamithup, N., Jainonthee, C., Salvador, R., & Jampachaisri, K. (2023). The impact of mass vaccination policy and control measures on lumpy skin disease cases in Thailand: Insights from a bayesian structural time series analysis. *Frontiers in Veterinary Science*, 10, Article 1301546. <https://doi.org/10.3389/fvets.2023.1301546>
- Punyapornwithaya, V., Mishra, P., Sansamur, C., Pfeiffer, D., Arjkumpa, O., Prakotcheo, R., Damrongwatanapokin, T., & Jampachaisri, K. (2022). Time-series analysis for the number of foot and mouth disease outbreak episodes in cattle farms in Thailand using data from 2010–2020. *Viruses*, 14. <https://doi.org/10.3390/v14071367>
- Qian, J., Yue, M., Huang, P., Ai, L., Zhu, C., Wang, C., Luo, Y., Yue, N., Wu, Y., Zhang, Y., Wang, C., & Tan, W. (2023). Spatiotemporal heterogeneity and impact factors of hepatitis B and C in China from 2010 to 2018: Bayesian space-time hierarchy model. *Frontiers in Cellular and Infection Microbiology*, 13, Article 1115087. <https://doi.org/10.3389/fcimb.2023.1115087>
- Qin, H., Qiu, Y., Ying, M., & Ren, J. (2022). Evaluation of the health promotion effect of hepatitis B prevention and treatment in the Zhejiang demonstration area, China. *BMC Public Health*, 22, 2073. <https://doi.org/10.1186/s12889-022-14540-8>

- Qiu, H., Zhao, H., Xiang, H., Ou, R., Yi, J., Hu, L., Zhu, H., & Ye, M. (2021). Forecasting the incidence of mumps in Chongqing based on a SARIMA model. *BMC Public Health*, 21, 373. <https://doi.org/10.1186/s12889-021-10383-x>
- Rosenberg, P. S., Check, D. P., & Anderson, W. F. (2014). A web tool for age-period-cohort analysis of cancer incidence and mortality rates. *Cancer epidemiology, biomarkers & prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 23, 2296–2302. <https://doi.org/10.1158/1055-9965.Epi-14-0300>
- Shaub, D. (2020). Using the "forecastHybrid" package. Retrieved from <https://cran.r-project.org/web/packages/forecastHybrid/vignettes/forecastHybrid.html>. December 12.
- Su, C. (2015). *The association study between HBV infection, environmental factors, polymorphisms and the risk of hepatocellular carcinoma in Xiamen*. Fujian Medical University.
- Su, S., Wong, W. C., Zou, Z., Cheng, D. D., Ong, J. J., Chan, P., Ji, F., Yuen, M. F., Zhuang, G., Seto, W. K., & Zhang, L. (2022). Cost-effectiveness of universal screening for chronic hepatitis B virus infection in China: An economic evaluation. *Lancet Global Health*, 10, e278–e287. [https://doi.org/10.1016/s2214-109x\(21\)00517-9](https://doi.org/10.1016/s2214-109x(21)00517-9)
- Talkhi, N., Akhavan Fatemi, N., Ataei, Z., & Jabbari Nooghabi, M. (2021). Modeling and forecasting number of confirmed and death caused COVID-19 in Iran: A comparison of time series forecasting methods. *Biomedical Signal Processing and Control*, 66, Article 102494. <https://doi.org/10.1016/j.bspc.2021.102494>
- Trépo, C., Chan, H. L. Y., & Lok, A. (2014). Hepatitis B virus infection. *The Lancet*, 384, 2053–2063. [https://doi.org/10.1016/S0140-6736\(14\)60220-8](https://doi.org/10.1016/S0140-6736(14)60220-8)
- Vavilala, H., Yaladanda, N., Krishna Kondeti, P., Rafiq, U., Mopuri, R., Gouda, K. C., Rao Bhimala, K., Rao Kadiri, M., Upadhyayula, S. M., & Rao Mutheneni, S. (2022). Weather integrated malaria prediction system using Bayesian structural time series model for northeast states of India. *Environmental Science and Pollution Research*, 29, 68232–68246. <https://doi.org/10.1007/s11356-022-20642-y>
- Wang, H., Men, P., Xiao, Y., Gao, P., Lv, M., Yuan, Q., Chen, W., Bai, S., & Wu, J. (2019). Hepatitis B infection in the general population of China: A systematic review and meta-analysis. *BMC Infectious Diseases*, 19, 811. <https://doi.org/10.1186/s12879-019-4428-y>
- Wang, Y. B., Qing, S. Y., Liang, Z. Y., Ma, C., Bai, Y. C., & Xu, C. J. (2023). Time series analysis-based seasonal autoregressive fractionally integrated moving average to estimate hepatitis B and C epidemics in China. *World Journal of Gastroenterology*, 29, 5716–5727. <https://doi.org/10.3748/wjg.v29.i42.5716>
- Wang, L., Wang, Y., Jin, S., Wu, Z., Chin, D. P., Koplan, J. P., & Wilson, M. E. (2008). Emergence and control of infectious diseases in China. *Lancet (London, England)*, 372, 1598–1605. [https://doi.org/10.1016/s0140-6736\(08\)61365-3](https://doi.org/10.1016/s0140-6736(08)61365-3)
- Wang, Y. S., Wang, S. N., Pan, J. H., & Wang, W. B. (2020). [Trend analysis and prediction of viral hepatitis incidence in China, 2009–2018]. *Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi*, 41, 1460–1464. <https://doi.org/10.3760/cma.j.cn112338-20191024-00761>
- WorldHealthOrganization. (2016). Global health sector strategy on viral hepatitis 2016–2021. Retrieved from <https://www.who.int/publications/i/item/WHO-HIV-2016.06>. December 11.
- WorldHealthOrganization. (2021). *Global progress report on HIV, viral hepatitis and sexually transmitted infections*, Article 2021. Retrieved from <https://www.who.int/publications/i/item/9789240027077>. December 12.
- Xian, X., Wang, L., Wu, X., Tang, X., Zhai, X., Yu, R., Qu, L., & Ye, M. (2023). Comparison of SARIMA model, Holt-winters model and ETS model in predicting the incidence of foodborne disease. *BMC Infectious Diseases*, 23, 803. <https://doi.org/10.1186/s12879-023-08799-4>
- Xie, C., Wen, H., Yang, W., Cai, J., Zhang, P., Wu, R., Li, M., & Huang, S. (2021). Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. *Scientific Reports*, 11, 1445. <https://doi.org/10.1038/s41598-021-81100-2>
- Xu, A., & Zhang, L. (2017). The review and significance of national seroepidemiological surveys on viral hepatitis in China. *Chinese Journal of Preventive Medicine*, 51, 5. <https://doi.org/10.16168/j.issn.1007-1226.2017.01.005>
- Yadav, B. K., Srivastava, S. K., Arasu, P. T., & Singh, P. (2023). Time series modeling of tuberculosis cases in India from 2017 to 2022 based on the SARIMA-NNAR hybrid model. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale*, 2023, Article 5934552. <https://doi.org/10.1155/2023/5934552>
- Yang, W. J., Chang, C. J., Yeh, S. H., Lin, W. H., Wang, S. H., Tsai, T. F., Chen, D. S., & Chen, P. J. (2009). Hepatitis B virus X protein enhances the transcriptional activity of the androgen receptor through c-Src and glycogen synthase kinase-3beta kinase pathways. *Hepatology*, 49, 1515–1524. <https://doi.org/10.1002/hep.22833>
- Yang, W., Su, A., & Ding, L. (2023). Application of exponential smoothing method and SARIMA model in predicting the number of admissions in a third-class hospital in Zhejiang Province. *BMC Public Health*, 23, 2309. <https://doi.org/10.1186/s12889-023-17218-x>
- Yu, G., Feng, H., Feng, S., Zhao, J., & Xu, J. (2021). Forecasting hand-foot-and-mouth disease cases using wavelet-based SARIMA-NNAR hybrid model. *PLoS One*, 16, Article e0246673. <https://doi.org/10.1371/journal.pone.0246673>
- Zenia, S., L'Hadj, M., & Selmane, S. (2023). A hybrid approach based on seasonal autoregressive integrated moving average and neural network autoregressive models to predict scorpion sting incidence in el oued province, Algeria, from 2005 to 2020. *Journal of Research in Health Sciences*, 23, Article e00586. <https://doi.org/10.34172/jrhrs.2023.121>
- Zhang, Y., Liu, J., Han, X., Jiang, H., Zhang, L., Hu, J., Shi, L., & Li, J. (2022). Long-term trends in the burden of inflammatory bowel disease in China over three decades: A joinpoint regression and age-period-cohort analysis based on gbd 2019. *Frontiers in Public Health*, 10, Article 994619. <https://doi.org/10.3389/fpubh.2022.994619>
- Zhang, C., Liu, Y., Zhao, H., & Wang, G. (2022). Global patterns and trends in total burden of hepatitis B from 1990 to 2019 and predictions to 2030. *Clinical Epidemiology*, 14, 1519–1533. <https://doi.org/10.2147/clep.S389853>
- Zhao, D., Zhang, H., Cao, Q., Wang, Z., & Zhang, R. (2022). The research of SARIMA model for prediction of hepatitis B in mainland China. *Medicine*, 101, Article e29317. <https://doi.org/10.1097/md.00000000000029317>
- National Cancer Institute. APC/AAPC/Tau confidence intervals Retrieved from. <https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/apc-aapc-tau-confidence-intervals/average-annual-percent-change-aapc>. (Accessed 14 April 2023).