

Comparison of Three Prediction Models for Predicting Chronic Obstructive Pulmonary Disease in China

Yuhan Teng¹, Yining Jian², Xinyue Chen³, Yang Li⁴, Bing Han², Lei Wang³

¹Department of Clinical Medicine, Southwest Medical University, Luzhou, Sichuan, People's Republic of China; ²Department of Public Health, China Medical University, Shenyang, Liaoning, People's Republic of China; ³Department of General Practice, the First Hospital of China Medical University, Shenyang, Liaoning, People's Republic of China; ⁴Department of General Practice, Hunnan Zhujia Community Health Service Center, Shenyang, Liaoning, People's Republic of China

Correspondence: Lei Wang, Department of General Practice, The First Hospital of China Medical University, 122 N Nanjing St., Heping District, Shenyang, 110001, People's Republic of China, Fax +86-24-83283333, Email wanglei_tz@163.com

Purpose: To predict the future number of patients with chronic obstructive pulmonary disease (COPD) in China and compare the three prediction models.

Methods: A generalized additive model (GAM), autoregressive integrated moving average (ARIMA) model, and curve-fitting method were used to fit and predict the number of patients with COPD in China. Data on the number of patients with COPD in China from 1990 to 2019 were obtained from the Global Burden of Disease (GBD) database. The coefficient of determination (R^2), mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), relative error of prediction, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) were used to evaluate and compare the fitting effect, prediction effect, and reliability of the three models.

Results: The GAM, ARIMA, and curve-fitting methods could predict future trends in COPD in China. The performance of the GAM is the best among the three models, whereas the curve fitting method is the worst, and the ARIMA (0,1,2) model is in between. The prediction results of the three models showed that the number of patients with COPD in China is expected to increase from 2020 to 2025.

Conclusion: GAM and AIRMA models are recommended for predicting the future prevalence of COPD in China. The number of patients with COPD in China is expected to increase in the next few years. The prevention and control of COPD in China still needs to be strengthened. Using appropriate models to predict future trends in COPD will provide support for health policymakers.

Keywords: generalized additive model, ARIMA model, curve fitting method, COPD, prediction

Introduction

Chronic obstructive pulmonary disease (COPD) is a common, preventable, and treatable disease that is usually caused by significant exposure to noxious particles or gases.¹ COPD is a heterogeneous lung condition characterized by chronic respiratory symptoms (dyspnea, cough, sputum production) due to abnormalities of the airways (bronchitis, bronchiolitis) and/or alveoli (emphysema) that cause persistent, often progressive, airflow obstruction.² COPD is a worldwide health issue, which causes a heavy burden to the world. From 1990 In 2019, the number of patients with COPD has increased from 115 million to 212 million globally.³ In China, the number of patients with COPD increased from 28 million in 1990 to 45 million in 2019.³ In addition, COPD is one of the leading causes of mortality worldwide. In 2017, the total number of deaths from COPD was approximately 3.2 million worldwide, and the age-standardized death rate was 42.2 per 100,000, ranking third in the world.⁴ In China, with more than one million in 2019.⁵ Prevention and control of COPD has become an important challenge for the world and China. Therefore, predicting the prevalence of COPD and the number of patients with COPD in the future will provide data support for health decision-makers.

Several methods have been used to predict future epidemiological disease trends. For example, the generalized additive model (GAM) was used to predict lung cancer incidence among women aged 25 to 84 years in five countries, including the United States, the United Kingdom and Australia.⁶ In China, the autoregressive integrated moving average (ARIMA) model was used to predict the number of patients and prevalence trends of diabetes and chronic kidney disease.^{7,8} One study used the curve fitting method and long short-term memory (LSTM) to predict the number of Corona Virus Disease 2019 (COVID-19) cases in India.⁹ However, the number of existing studies on predicting the future trend of COPD is small, and few studies have compared the accuracy of COPD prediction methods.

In this study, we used the GAM, ARIMA model, and curve fitting method to predict the number of patients with COPD in China from 2020 to 2025, and compared the fitting and prediction effects of the three models. The number of patients with COPD in China from 1990 to 2018 was used to fit the models and the number of patients with COPD in 2019 was used to test the models. The aim of this study was to explore appropriate methods to predict the prevalence of COPD in China, estimate the future risk and disease burden of COPD, and provide references for the prevention and control of COPD in China, and provide data support for health decision-making departments to invest medical resources in the clinical treatment of COPD.

Materials and Methods

Data

Data on the number of patients with COPD in China from 1990 to 2019 were obtained from the Global Health Data Exchange (GHDx). GBD Results Tool (2023). <https://vizhub.healthdata.org/gbd-results/>. The database was established using Microsoft Excel 2016 and statistical analyses of the data were performed using R 4.2.0.

GAM

Generalized additive model (GAM) is an extension of the linear regression model, which is based on a Generalized linear model and an additive model.¹⁰ GAM is a flexible regression model. Its predictor variables can be linear or nonlinear and can be applied to data with various distributions, including normal, binomial, Gamma, and Poisson. GAM predicts disease trends by non-parametric fitting with relaxed assumptions on the relationship between response and predictor variables and the flexibility of the GAM in setting predictor variables allows it to achieve high predictive accuracy.¹¹ The GAM has the following inequality:¹²

$$\begin{cases} g(\mu(x)) = \alpha + \sum_{j=1}^p f_j(x_j) \\ E[f_j(x_j)] = 0 \end{cases} \quad (1)$$

where $g()$ is the link function, $\mu(x)=E(y/x)$, α is the intercept, $f()$ is an arbitrary univariate function for each predictor variable, x_j If the response variable y_i is independent of a single-parameter exponential distribution, the distribution has a density function or probability density function as follows:

$$p(y_i; \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right\} \quad (2)$$

where θ_i is the natural parameter of this exponential distribution, different y_i corresponds to different θ_i , ϕ is the nuisance or scale parameter, ϕ is the same for all y_i , b and c are the functions used to determine the distribution type.

ARIMA Model

The autoregressive integrated moving average (ARIMA) model is a common method for time series analysis and prediction which can predict future values from past values.¹³ Therefore, the future disease trends can be predicted by using the past number of people with a disease.^{7,8} ARIMA model consists of an autoregressive (AR) model, a moving average (MA) model, and integration (I), which correspond to the three parameters p , q and d respectively. The AR (p) model represents the p -order autoregressive model, which is used to associate the pattern of a one-time period with its previous time periods. The MA (q) model represents a q -order moving average model that uses the errors associated with the forecast at a previous time step to forecast a variable at a later time step. I (d) is the reverse process of differencing to

generate the forecast, which refers to the difference order that converts a non-stationary time series into a stationary time series. Combining these three components yields the ARIMA (p, d, q) model. The generalized p th-order AR model (Eq. (3)), and q th-order MA model (Eq. (4)) can be expressed as

$$\left\{ \begin{array}{l} y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \\ y_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \end{array} \right. \quad (3)$$

$$\left\{ \begin{array}{l} y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \\ y_t = c + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \end{array} \right. \quad (4)$$

where y_t is the current time-series value, c is the intercept, ϕ_i ($i=1,2 \dots p$) is the autoregressive parameter, θ_i ($i=1,2 \dots q$) is the moving average parameter, ε_t is the white noise series with zero mean, $\varepsilon_t = y_t - y_{t-1}$. Based on Eq. (3) and Eq. (4), we obtain the generalized ARIMA model, as follows:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (5)$$

where y'_t is the stationary time series after differencing. Using the backshift operator $B^m y_t = y_{t-m}$ in Eq. (5), the ARIMA (p, d, q) model can be expressed as:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (6)$$

Curve Fitting Method

Curve fitting method is a common technique for data modeling.¹⁴ It is the process of transforming the observed relationship in a particular dataset into a parameterized function and can be used for prediction and parameter estimation.¹⁵ Curve fitting method constructs a curve or mathematical function of the number of patients to predict the future disease trends by best fitting a series of data points of the number of patients in the past.¹⁶ Curve fitting methods include several classes, such as polynomials, exponential functions, power functions and logarithmic functions.¹⁴ Polynomials are a common choice for curve fitting, and their general form is as follows:

$$Y = c_0 + c_1 X + c_2 X^2 + \dots + c_{n-1} X^{n-1} + c_n X^n \quad (7)$$

where c_i ($i=0,1 \dots n$) is a real number; is the autoregressive parameter, which is the parameter or coefficient of X , n is the degree of the polynomial.

Model Evaluation

Model evaluation metrics can intuitively evaluate the performance of the model. By comparing the various evaluation metrics of different models, we can intuitively compare the performance of different prediction models in various aspects. The prediction model that performs better in all aspects will be more reliable in predicting disease trends. The three prediction methods in this study were evaluated and compared using common model-evaluation metrics. The model fitting effect was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), mean absolute percentage error ($MAPE$), and root mean square error ($RMSE$). The predictive effect of the model was evaluated by using the *relative error of prediction*. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to evaluate the reliability of the time-series analysis. The equations for the model evaluation metrics are as follows:

$$\left\{ \begin{array}{l} R^2 = 1 - \frac{SS_r}{SS_t} \\ SS_r = \sum_i (y_i - f_i)^2 \\ SS_t = \sum_i (y_i - \bar{y})^2 \\ \bar{y} = \frac{1}{n} \sum_i y_i \end{array} \right. \quad (8)$$

where SS_r is the sum of the squares of the residuals, SS_t is the total sum of squares, y_i is the actual value, f_i is the fitted value, \bar{y} is the mean of the actual values. The value of R^2 ranged from 0 to 1, with values closer to 1 indicating a better fit of the model.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f_i| \quad (9)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - f_i}{y_i} \right| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2} \quad (11)$$

Smaller values of *MAE*, *MAPE*, and *RMSE* values indicated a better fitting effect of the model.

$$\text{relative error of prediction} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}}{y_i} \right| \quad (12)$$

where \hat{y} is the predicted value. The lower the value of the *relative error of prediction*, the better the prediction effect of the model.

$$AIC = -2\log L(\hat{\theta}) + 2K \quad (13)$$

$$BIC = -2\log L(\hat{\theta}) + K\log N \quad (14)$$

where $\log L(\hat{\theta})$ denotes the likelihood function, K denotes the total number of model parameters, N denotes the number of observations. Smaller values of *AIC* and *BIC* values indicated a more reliable model.

Results

Model Evaluation

The GAM used to predict the number of patients with COPD was modeled using years as a predictor. Using the number of patients with COPD in China from 1990 to 2019 to fit and validate the model, we found that when the basis dimension $k=22$, the GAM had the optimal fitting and predictive effect. The model selection was based on the generalized cross-validation (GCV) principle. The GCV score (1.5749e+09) and deviance explained (100%) of the GAM established by year showed that the model works well.

The ARIMA (0,1,2) model was constructed by fitting the number of patients with COPD in China from 1990 to 2018. After the first difference transformation, the original time series changed from non-stationary to stationary, and the differencing time series passed the augmented Dickey–Fuller (ADF) test (ADF value = -3.8001 , $P < 0.05$). The ARIMA (0,1,2) model passed the white noise test ($P > 0.05$), indicating that it fits well and can be used to predict the number of patients with COPD in the future.

For the curve-fitting method, a curve-fitting equation was established by fitting the number of patients with COPD in China from 1990 to 2018. According to the polynomial curve fitting method, we obtained the quadratic equation $y = -9357.3x^2 + 38034939x - 3.8605e+10$. The coefficient of determination of the model was 0.9866, indicating a good fit ($F = 960.08$, $P < 0.001$).

Model Evaluation results

The specific values of the fitting and prediction effect evaluation metrics of the GAM, ARIMA (0,1,2) model, and curve-fitting method are listed in Table 1. The R^2 values of the three models were greater than 0.9, which indicates that the three models can fit the data well. The GAM had the smallest R^2 , *MAE*, *MAPE*, and *RMSE*, indicating that the GAM has the best fitting effect among the three models. In addition, the *relative error of prediction* of the three models was small, which shows that these models also have good prediction effects. Moreover, GAM has the smallest *relative error of prediction*, which indicates that it also has the best prediction effect among the three models. According to *AIC* and *BIC*, all three prediction models were reliable, and the values of *AIC* and *BIC* for GAM were the smallest among the three models.

Table 1 Fitting and Prediction Effect Evaluation of the Three Models

Model	R ²	MAE	MAPE (%)	RMSE	AIC	BIC	Relative Error of Prediction (%)
GAM	0.9999	6929.9	0.02	9748.8	660.77	692.05	-0.0003
ARIMA (0,1,2)	0.9973	138,430.7	0.36	238,583.1	784.12	789.44	-0.4100
Curve Fitting Method	0.9866	425,550.3	1.13	521,219.8	853.81	859.28	-4.1891

Abbreviations: GAM, generalized additive model; ARIMA, autoregressive integrated moving average; MAE, mean absolute error; MAPE, mean absolute percentage error; RMSE, root mean squared error; AIC, Akaike information criterion; BIC, Bayesian information criterion.

Prediction Results

We used the GAM, ARIMA (0,1,2) model, and curve-fitting method to fit the number of patients with COPD in China from 1990 to 2018 and predicted the number of patients with COPD in China from 2019 to 2025, as shown in Table 2. In

Table 2 Fitting and Prediction Results of the Three Models from 1990 to 2025

Year	Actual Values	Fitted/Predicted* Values of GAM	Fitted/Predicted* Values of ARIMA (0,1,2)	Fitted/Predicted* Values of Curve Fitting Method
1990	27,750,213	27,749,770	27,723,071	28,151,525
1991	28,631,432	28,633,235	28,473,099	28,934,960
1992	29,479,714	29,476,282	29,431,996	29,699,679
1993	30,298,665	30,303,071	30,204,186	30,445,684
1994	31,130,857	31,126,443	31,052,313	31,172,974
1995	31,948,269	31,951,696	31,882,342	31,881,550
1996	32,759,336	32,758,004	32,678,443	32,571,411
1997	33,495,325	33,493,665	33,503,401	33,242,557
1998	34,189,045	34,193,573	34,128,844	33,894,989
1999	34,898,215	34,891,921	34,873,317	34,528,706
2000	35,562,685	35,569,183	35,566,508	35,143,708
2001	36,229,393	36,224,555	36,176,711	35,739,996
2002	36,812,918	36,814,608	36,905,493	36,317,569
2003	37,321,211	37,323,028	37,321,493	36,876,427
2004	37,800,883	37,796,257	37,886,149	37,416,571
2005	38,203,544	38,209,824	38,294,940	37,938,000
2006	38,503,279	38,497,066	38,650,508	38,440,714
2007	38,606,128	38,609,871	38,873,094	38,924,714
2008	38,624,314	38,624,160	38,790,938	39,389,999
2009	38,750,857	38,746,558	38,887,876	39,836,570
2010	39,090,394	39,100,023	39,099,805	40,264,426
2011	39,762,225	39,748,085	39,622,811	40,673,567
2012	40,571,705	40,589,640	40,551,258	41,063,993
2013	41,450,223	41,429,365	41,270,513	41,435,705
2014	42,372,472	42,396,174	42,306,588	41,788,703
2015	43,108,193	43,084,290	43,150,240	42,122,985
2016	42,698,723	42,713,748	43,690,108	42,438,553
2017	42,426,516	42,422,833	41,967,170	42,735,407
2018	43,608,453	43,608,259	43,191,680	43,013,545
2019	45,164,960	45,164,812	44,981,522	43,272,969
2020	—	46,721,365	45,780,166	43,513,679
2021	—	48,277,917	46,387,883	43,735,674
2022	—	49,834,470	46,995,600	43,938,954
2023	—	51,391,023	47,603,317	44,123,519
2024	—	52,947,576	48,211,033	44,289,370
2025	—	54,504,128	48,818,750	44,436,506

Note: *Predicted values: the numbers in bold are the predicted values.

Abbreviations: GAM, generalized additive model; ARIMA, autoregressive integrated moving average.

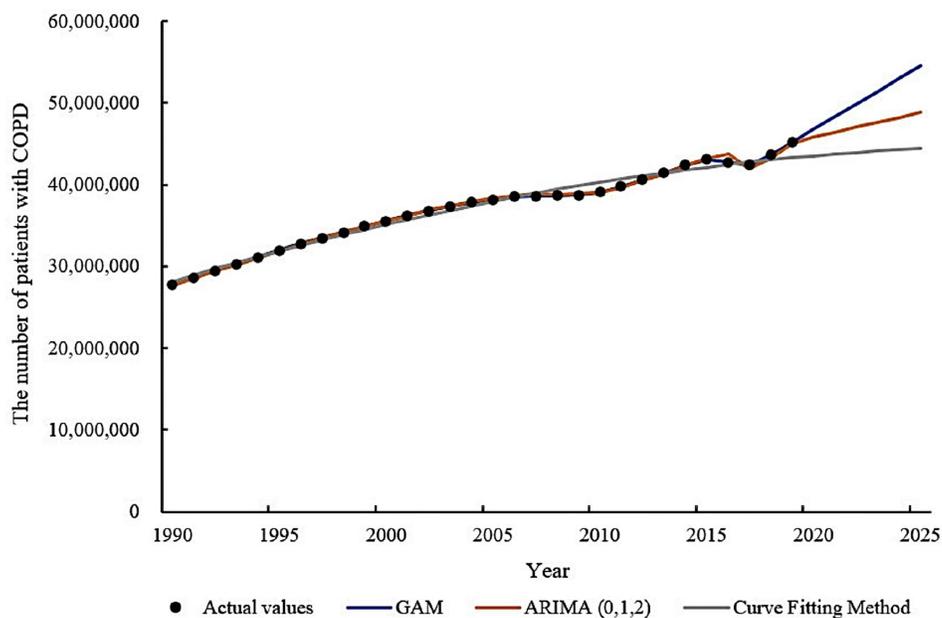


Figure 1 The prevalence trend of COPD in China from 1990 to 2025.

Abbreviations: COPD, chronic obstructive pulmonary disease; GAM, generalized additive model; ARIMA, autoregressive integrated moving average.

addition, the prediction results of the three models showed an overall upward trend in the number of patients with COPD in China from 1990 to 2025, as shown in [Figure 1](#).

Discussion

According to our results, all three models in this study can be used to predict the future number of patients with COPD in China, and all three prediction models showed good fitting and prediction effects. The fitting effect, prediction effect, and model reliability of the GAM are the best among the models, whereas the curve fitting method is the worst among the three models, and the effect of the ARIMA model is in between. In this study, the number of patients with COPD in China from 1990 to 2018 used for model fitting was not normal. For non-normal data, GAM provides a more efficient analysis method than traditional linear models.¹⁷ Similar to the results of this study, GAM also shows some advantages when predicting other diseases. In 2019, a study constructed GAM, ARIMA, and negative binomial regression models (NBM) to predict the number of hand-foot-mouth disease (HFMD) cases in Yiwu City, Zhejiang Province, China. The results showed that the predicted values of the three models were in good agreement with the observed values, and the GAM had the best fitting effect.¹⁸ Based on the number of diagnosed pneumoconiosis cases in China from 2000 to 2019, a study established three prediction models for the incidence trend of pneumoconiosis, including GAM, curve fitting method and GM (1,1) model. This study found that GAM was relatively good among the three prediction models. The coefficient of determination, average fitting relative error, and relative error of prediction of the GAM were better than those of the other two models.¹⁹ The ARIMA model was also suitable for predicting the future number of patients with COPD in China. As a common time-series prediction method, the ARIMA model fully utilizes the temporal information of the original dataset to make accurate predictions. In addition, the modeling process of the ARIMA model is not complex and can be generalized for use.²⁰ At present, ARIMA model has been applied to predict the epidemic trend of various diseases and whether infectious or chronic non-communicable diseases can be well predicted. For example, the ARIMA model was used to predict the estimated confirmed cases and epidemic trends of COVID-19 from April 24 to July 7, 2020 for the top 15 countries with confirmed cases, deaths, and recoveries from COVID-19.²¹ In addition, one study verified the accuracy of using the ARIMA model to predict COVID-19, and the results showed that most of the observed actual values were well within bounds of the ARIMA model prediction at 95% confidence interval, indicating that the prediction accuracy provided by the ARIMA model was appropriate and satisfactory.²² In the United States, the

ARIMA model was used to predict the incidence of cancer from 2015 to 2020.²³ Moreover, the ARIMA model was also been select to predict the trends in the incidence of global bladder cancer from 2017 to 2030.²⁴ The curve fitting method can also be used to predict the future number of patients with COPD in China. However, the fitting effect, prediction effect, and model reliability of the curve-fitting method were the worst of the three methods. Consistent with the results of this study, a study comparing the accuracy of several algorithms used to predict COVID-19 showed that the curve-fitting method had the worst performance in terms of prediction accuracy among the methods considered.²⁵ In this study, we choose the quadratic polynomial model in the curve fitting method, which is a parabolic model and cannot accurately fit the fluctuation part of the data, so the effect of the model may have been affected. However, the sigmoid curve, which is more suitable for volatility data in the curve fitting method, did not show a better fitting and prediction effect in the comparative study of other prediction methods.²⁵ Therefore, compared with the curve fitting method, the GAM or ARIMA model is recommended for analyzing and predicting the future number of patients with COPD in China.

The results indicated that the number of patients with COPD in China showed an overall upward trend from 1990 to 2019 and that the number of patients with COPD in 2019 increased by 62.8% from 1990. Moreover, the predictions of all three models showed that the number of patients with COPD will continue to increase over the next few years. The reasons for the increasing number of patients with COPD in China may be as follows. First, it is related to rapid population aging in China. A meta-analysis of 28 countries worldwide found that the prevalence of COPD increased with age, with a significant increase in people aged 40 years and older. The highest prevalence of COPD among all age groups was in people aged 65 years and older.²⁶ In China, a national cross-sectional study found that the prevalence of COPD increased with age and was as high as 35.5% in people aged 70 years and older in 2015.²⁷ From 1990 to 2021, the number of people aged 65 years and older in China increased from 64 million to 201 million, accounting for the proportion of the total Chinese population increased from 5.6% to 14.2%.²⁸ By 2050, it is expected that there will be 365 million elderly people aged 65 years and older in China, accounting for 26.1% of the total population.²⁹ Secondly, there are a large number of smokers in China. Smoking is a major risk factor for COPD.³⁰ A meta-analysis of 162 population-based articles from 65 countries showed that globally, the risk of COPD in current smokers is 3.2 times higher than that in non-smokers.³¹ In China, a national COPD prevalence study found that the risk of COPD in smokers was 1.87 times higher than that in non-smokers among people aged 40 years and older.³² China is the world's largest consumer of tobacco, accounting for more than one-third of the world's tobacco consumption; in 2019, 341 million (30%) of the world's 1.14 billion tobacco smokers lived in China.³³ A study based on data from five consecutive large nationally representative surveys from 2007 to 2018 showed that the smoking prevalence among Chinese adults aged 18 to 69 years was greater than 25% in all five surveys, and the smoking prevalence among adult males in China was the highest in the world.³⁴ Based on the above reasons, we put forward the following suggestions. First, it is recommended that COPD screening should be strengthened in people aged 40 years and above, especially people aged 65 years and above, so as to achieve early detection, early diagnosis, and early treatment of patients with COPD. Second, health education regarding the dangers of smoking should be strengthened, and smokers should be encouraged to quit smoking. In addition, such a large number of patients with COPD who will continue to increase in the future not only brings challenges to the prevention and control of COPD in China, but also brings some problems to the clinical treatment of COPD. Therefore, health policy makers should increase the resource investment for the clinical treatment of COPD in China to promote the development of clinical treatment of COPD, so that it can better adapt to the increasing number of COPD patients in China.

Conclusion

The GAM, ARIMA model, and curve-fitting method could be used to predict the future number of patients with COPD in China, and the GAM had the best effect among the three models. The prediction results of the three models showed that the number of patients with COPD in China will increase in the future. Therefore, we should continue to strengthen the prevention and control of COPD in China. In addition, it also suggests that the health policy makers in China should invest more medical resources in the clinical treatment of COPD to better face the challenge of the increasing number of patients with COPD in the future.

Abbreviations

COPD, chronic obstructive pulmonary disease; GAM, generalized additive model; ARIMA model, autoregressive integrated moving average model; MAPE, mean absolute percentage error; MAE, mean absolute error; RMSE, root mean squared error; AIC, akaike information criterion; BIC, bayesian information criterion; LSTM, long short-term memory; COVID-19, corona virus disease 2019; GBD, global burden of disease; AR, autoregression; MA, moving average; I, integration; GCV, generalized cross validation; ADF test, augmented dickey fuller test; NBM, negative binomial regression model; HFMD, hand-foot-mouth disease.

Data Sharing Statement

The data used in this study were obtained from the Global Health Data Exchange (GHDx). GBD Results Tool (2022). <https://vizhub.healthdata.org/gbd-results/>.

Ethics Approval and Informed Consent

We applied for ethical approval of China Medical University. The data in the research included no identifiable information, and China Medical University approved the research plan. We have attached the original copy of the ethical approval as an attachment.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. *Am J Respir Crit Care Med.* 2017;195(5):557–582. doi:10.1164/rccm.201701-0218PP
2. Celli B, Fabbri L, Criner G, et al. Definition and Nomenclature of Chronic Obstructive Pulmonary Disease: time for Its Revision. *Am J Respir Crit Care Med.* 2022;206(11):1317–1325. doi:10.1164/rccm.202204-0671PP
3. Global Health Data Exchange (GHDx) [homepage on the Internet]. GBD Results Tool; 2023. Available from: <https://vizhub.healthdata.org/gbd-results/>. Accessed February 3, 2023.
4. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2018;392(10159):1736–1788. doi:10.1016/S0140-6736(18)32203-7
5. Yin P, Wu J, Wang L, et al. The Burden of COPD in China and Its Provinces: findings From the Global Burden of Disease Study 2019. *Front Public Health.* 2022;10:859499. doi:10.3389/fpubh.2022.859499
6. Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics.* 2005;6(4):576–589. doi:10.1093/biostatistics/kxi028
7. Zhu D, Zhou D, Li N, Han B. Predicting Diabetes and Estimating Its Economic Burden in China Using Autoregressive Integrated Moving Average Model. *Int J Public Health.* 2022;66:1604449. doi:10.3389/ijph.2021.1604449
8. Jian Y, Zhu D, Zhou D, et al. ARIMA model for predicting chronic kidney disease and estimating its economic burden in China. *BMC Public Health.* 2022;22(1):2456. doi:10.1186/s12889-022-14959-z
9. Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ.* 2020;728:138762. doi:10.1016/j.scitotenv.2020.138762
10. Hastie T, Tibshirani R. *Generalized Additive Models*. Boca Raton: CRC Press; 1990.
11. Xiao D, Wu K, Tan X, et al. Modeling and predicting hemorrhagic fever with renal syndrome trends based on meteorological factors in Hu County, China. *PLoS One.* 2015;10(4):56.
12. Hastie T, Tibshirani R. Generalized additive models: some applications. *J Am Stat Assoc.* 1987;82(398):371–386. doi:10.1080/01621459.1987.10478440

13. Newbold P. ARIMA model building and the time series analysis approach to forecasting. *J Forecasting*. 1983;2(1):23–35. doi:10.1002/for.3980020104
14. Sit V, Poulin-Costello M. *Catalogue of Curves for Curve Fitting*. Columbia: Forest Science Research Branch, Ministry of Forests; 1994.
15. Li M, Li LD. A novel method of curve fitting based on optimized extreme learning machine. *Appl Artif Intell*. 2020;34(12):849–865. doi:10.1080/08839514.2020.1787677
16. Wang J, Wang C, Zhang W. Data analysis and forecasting of tuberculosis prevalence rates for smart healthcare based on a novel combination model. *Appl Sci*. 2018;8(9):1693. doi:10.3390/app8091693
17. Ravindra K, Rattan P, Mor S, Aggarwal AN. Generalized additive models: building evidence of air pollution, climate change and human health. *Environ Int*. 2019;132:104987. doi:10.1016/j.envint.2019.104987
18. Fu T, Chen T, Dong ZB, et al. Development and comparison of forecast models of hand-foot-mouth disease with meteorological factors. *Sci Rep*. 2019;9(1):15691. doi:10.1038/s41598-019-52044-5
19. Zhou D, Zhu D, Li N, Han B. Exploration of Three Incidence Trend Prediction Models Based on the Number of Diagnosed Pneumoconiosis Cases in China From 2000 to 2019. *J Occup Environ Med*. 2021;63(7):e440–e444. doi:10.1097/JOM.0000000000002258
20. Zhu Z, Zhu X, Zhan Y, Gu L, Chen L, Li X. Development and comparison of predictive models for sexually transmitted diseases-AIDS, gonorrhea, and syphilis in China, 2011–2021. *Front Public Health*. 2022;10:966813. doi:10.3389/fpubh.2022.966813
21. Singh RK, Rani M, Bhagavathula AS, et al. Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: advanced Autoregressive Integrated Moving Average (ARIMA) Model. *JMIR Public Health Surveill*. 2020;6(2):e19115. doi:10.2196/19115
22. Alabdulrazzaq H, Alenezi MN, Rawajfih Y, Alghannam BA, Al-Hassan AA, Al-Anzi FS. On the accuracy of ARIMA based prediction of COVID-19 spread. *Results Phys*. 2021;27:104509. doi:10.1016/j.rinp.2021.104509
23. Xie L. Time series analysis and prediction on cancer incidence rates. *J Med Discovery*. 2017;2(3):1–10. doi:10.24262/jmd.2.3.17030
24. Cai Q, Chen Y, Xin S, et al. Temporal trends of bladder cancer incidence and mortality from 1990 to 2016 and projections to 2030. *Translational Androl Urol*. 2020;9(2):153–165. doi:10.21037/tau.2020.02.24
25. Achterberg MA, Prasse B, Ma L, Trajanovski S, Kitsak M, Van Mieghem P. Comparing the accuracy of several network-based COVID-19 prediction algorithms. *Int J Forecast*. 2022;38(2):489–504. doi:10.1016/j.ijforecast.2020.10.001
26. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: systematic review and meta-analysis. *Eur Respir J*. 2006;28(3):523–532. doi:10.1183/09031936.06.00124605
27. Wang C, Xu J, Yang L, et al.; China Pulmonary Health Study Group. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. *Lancet*. 2018;391(10131):1706–1717. doi:10.1016/S0140-6736(18)30841-9
28. National Bureau of Statistics of China [homepage on the Internet]. Age Composition and Dependency Ratio of Population; 2022. Available from: <https://data.stats.gov.cn/english/easyquery.htm?cn=C01>. Accessed February 15, 2023.
29. Fang EF, Xie C, Schenkel JA, et al. A research agenda for ageing in China in the 21st century (2nd edition): focusing on basic and translational research, long-term care, policy and social networks. *Ageing Res Rev*. 2020;64:101174. doi:10.1016/j.arr.2020.101174
30. Bartal M. COPD and tobacco smoke. *Monaldi Arch Chest Dis*. 2005;63(4):213–225. doi:10.4081/monaldi.2005.623
31. Adeloje D, Song P, Zhu Y, Campbell H, Sheikh A, Rudan I. NIHR RESPIRE Global Respiratory Health Unit. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. *Lancet Respir Med*. 2022;10(5):447–458. doi:10.1016/S2213-2600(21)00511-7
32. Fang L, Gao P, Bao H, et al. Chronic obstructive pulmonary disease in China: a nationwide prevalence study. *Lancet Respir Med*. 2018;6(6):421–430. doi:10.1016/S2213-2600(18)30103-6
33. Reitsma MB, Kendrick PJ, Ababneh E; GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet*. 2021;397(10292):2337–2360. doi:10.1016/S0140-6736(21)01169-7
34. Zhang M, Yang L, Wang L, et al. Trends in smoking prevalence in urban and rural China, 2007 to 2018: findings from 5 consecutive nationally representative cross-sectional surveys. *PLoS Med*. 2022;19(8):e1004064. doi:10.1371/journal.pmed.1004064

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>