European Journal of Preventive Medicine
2023; 11(3): 32-36
http://www.sciencepublishinggroup.com/j/ejpm
doi: 10.11648/j.ejpm.20231103.11
ISSN: 2330-8222 (Print); ISSN: 2330-8230 (Online)

Accumulatively Confirmed Cases Used for Grey Modeling Prediction of the Medium and Long Term Future Epidemic Trend of Infectious Diseases

Li Ming Quan

Nanchong City Committee Office of the Chinese People's Political Consultative Conference, Nanchong, China

Email address:
617738477@qq.com

To cite this article:
Li Ming Quan. Accumulatively Confirmed Cases Used for Grey Modeling Prediction of the Medium and Long Term Future Epidemic Trend of Infectious Diseases. *European Journal of Preventive Medicine*. Vol. 11, No. 3, 2023, pp. 32-36. doi: 10.11648/j.ejpm.20231103.11

Received: April 16, 2023; **Accepted**: May 4, 2023; **Published**: May 17, 2023

Abstract: The author used the actual data obtained from the prevention and control of COVID-19 in China and the cumulative number of confirmed cases obtained at different time intervals to predict the medium and long term (20 days, 40 days and 60 days) future epidemic trend by grey modeling. Research objectives: Grey modeling theory is applied to the modeling and prediction of infectious diseases, appropriate data obtained from the prevention and control of infectious diseases are selected for the simulation and prediction of the medium and long term future epidemic trend of infectious diseases, and an effective method for the prediction of the future epidemic trend of infectious diseases is sought. Research Methods: The author used the actual data obtained from the prevention and control of COVID-19 in China. The trend curve was drawn by statistical data, the trend of epidemic was visually analyzed and observed, and the best series for grey modeling prediction was determined. Then GM(1,1) grey modeling was carried out on the selected series, and the error and accuracy of the built model were tested. Finally, the predicted value of the model was actually verified. Research results: According to the series graph, we selected the cumulative number of confirmed cases with time intervals of 20 days, 40 days and 60 days to model and forecast the future medium and long term epidemic trend of COVID-19 in China, and built the prediction models of cumulative confirmed cases respectively. The average error of the GM(1,1) prediction model established by the cumulative number of confirmed cases at the time node with a time interval of 40 days is too high, reaching 0.6422, and the simulation accuracy is only 37%. It has no practical significance for forecasting. The prediction model of GM(1,1), established by the cumulative number of confirmed cases with a time interval of 20 days, has a large average simulation error of 0.3336 and a simulation accuracy of 67%. Through practical verification, the prediction accuracy of GM(1,1) can reach 99.54%, which has a certain practical value for prediction. The prediction model of GM (1,1) based on the cumulative number of confirmed cases at time nodes with a time interval of 60 days, the average simulation error of GM (1,1) prediction model was 0.01167, and the simulation accuracy was 98.83%. Multiple parameters in the accuracy analysis reached the index of the first-level model. The actual verification of the model showed that the cumulative number of confirmed cases at the predicted time node was 102271. In practice, 107094 cases were recorded, and the predicted number was 4823 cases less than the actual number. The relative error was 0.045, and the prediction accuracy reached 95.49%. Satisfactory gray modeling prediction effect was obtained.

Keywords: Cumulative Confirmed Cases, Gray Modeling, Epidemic Trend, Prediction

1. Introduction

The prediction of the future epidemic trend of infectious diseases is an important work in the prevention and control of infectious diseases [1]. The prediction can increase the initiative and accuracy of the prevention and control of infectious diseases. According to the forecast, the preparation of medical treatment, medicine, equipment, patient beds and other aspects can be made in advance, so as to improve the ability to cope calmly and reduce the phenomenon of the crowding of medical resources. Minimize the harm caused by infectious diseases to human beings. However, it is not easy to predict the future epidemic trend of infectious diseases [2]. Therefore, many experts and scholars in the industry have conducted a lot of research on the future epidemic trend forecast of infectious diseases, especially in the middle and long term during the epidemic period, in order to achieve satisfactory results on the future epidemic trend forecast of infectious diseases. The global epidemic has lasted for three years, and now the global epidemic is coming to an end. China has won a decisive victory in the prevention and control of the epidemic [3]. In order to better apply the grey modeling prediction theory to the future trend of the epidemic in the middle and long term, and find the prediction method of the medium and long term future trend of the epidemic, By using the actual data obtained from the prevention and control of COVID-19 in China and the accumulated data of the confirmed cases, we conducted a gray modeling prediction study on the medium and long term future epidemic trend of COVID-19, in order to provide a more efficient prediction method for the medium and long term prediction of the future epidemic trend of infectious diseases [4]. The research situation is summarized and reported below.

2. Material Sources and Research Methods

2.1. Source and Organization of Material

The original data used in this study were all excerpted from the "Epidemic Bulletin" on the official website of Health Commission, PRC. The time interval of the data was from January 1, 2021 to December 31, 2021. In this study the predictions with a time interval of 20, 40 and 60 days were collectively referred to as the medium and long term predictions of epidemic trends of infectious diseases [5]. The cumulative number of confirmed cases at each time point is shown in Table 1.

Table 1. *Cumulative number of confirmed cases at different time intervals.*

Cumulative number of confirmed cases at each node at an interval of 20 days	88701 89734 89923 90106 90410 90686
Cumulative number of confirmed cases at each node at an interval of 40 days	89734 90106 90686 91337 92342 94819
Cumulative number of confirmed cases at each node at an interval of 60days	89923 90686 91780 94819 96938 100076

2.2. Research Methods

The trend curve was drawn by statistical data, the trend of epidemic was visually analyzed and observed, and the best series for grey modeling and prediction was determined. Then GM(1,1) grey modeling was carried out on the selected series, and the accuracy of the model was tested, and finally the predicted value of the model was actually verified [6].

3. Series Analysis and Modeling

3.1. Graph of Sequence Trend

In order to observe the state of sequence trend formed by cumulative confirmed cases collected at time nodes at different time intervals and select appropriate modeling methods, scatter plots were drawn for cumulative confirmed cases obtained at different time intervals respectively [7]. Set the original series of time points at 20-day intervals, 40-day intervals, and 60-day intervals respectively

$X^{(0)}_{20,}$ $X^{(0)}_{40}$ and $X^{(0)}_{60}$,

$X^{(0)}_{20}$ =[88701,89734,89923,90106,90410,90686];

$X^{(0)}_{40}$ =[89734,90106,90686,91337, 92342,94819];

$X^{(0)}_{60}$ =[89923,90686,91780,94819,96938;100076]. On the X-axis, the accumulatively confirmed cases at this point In time are shown Is the Y-axis, and scatter curves (Figure 1, Figure 2 and Figure 3) are drawn respectively [8]. It can be seen from Figure 1, Figure 2 and Figure 3 that, with the extension of the epidemic time, the cumulative confirmed cases increase along with the extension of the epidemic time, and the increase is in a certain curve mode.

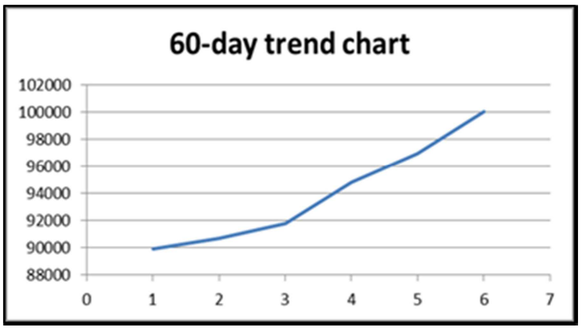

Figure 1. *Scatter plot of cumulative confirmed cases 60 days apart.*

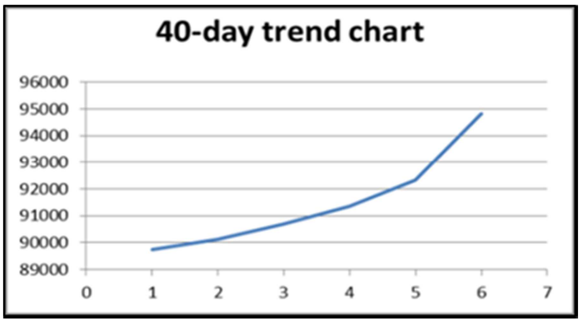

Figure 2. *Scatter plot of cumulative confirmed cases 40days apart.*

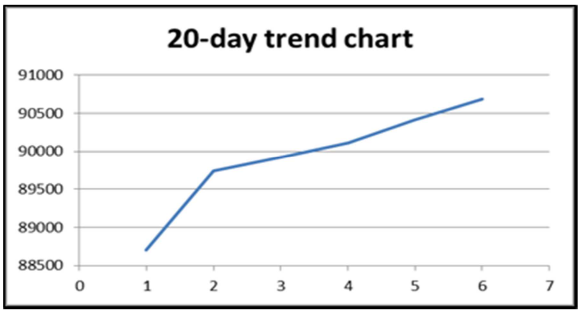

Figure 3. *Scatter plot of cumulative confirmed cases after 20 days.*

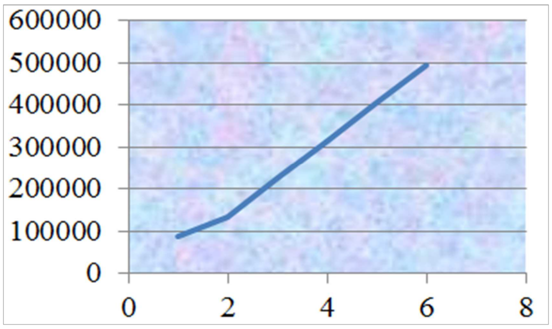

Figure 4. *Scatter plot of the cumulative sequence of confirmed cases at 20-day intervals.*

The time node of the cumulative confirmed cases at a 20-day interval shows an irregular oscillating upward pattern in the scatter plot of the cumulative confirmed cases [9]. To reinforce its actual trend, We used the cumulative sequence $X^{(1)}_{20}$=[88701 178435 268358 358464 448874 539560] of $X^{(0)}_{20}$ to draw the XY scatter curve again (see Figure 4), which is similar to Figure 1 and Figure 2. The curve graph shows that the epidemic trend of COVID-19 at intervals of 20 days, 40 days and 60 days is the same, so the scatter curve graph shows that the original number of cumulative confirmed cases is listed as a monotonously increasing sequence [10].

3.2. The Establishment of Prediction Model

In this study, grey GM(1,1) models were established for cumulative confirmed cases obtained at time nodes with intervals of 20, 40 and 60 days. Here we describe in detail the process of building a GM(1.1) model with a cumulative series of confirmed cases at 60-day intervals. First of all, Form the cumulative sequence $X^{(1)}_{60}(1)=[X^{(1)}_{60}(1)), X^{(1)}_{60}(2), X^{(1)}_{60}(3), X^{(1)}_{60}(4), X^{(1)}_{60}(5), X^{(1)}_{60}(6)]$, whose cumulative number is listed as $X^{(1)}_{60}$ =[89923 180609 272389 367208 464164 564240], Close to the mean sequence $Z^{(1)}_{60} = [Z^{(1)}_{60}(2), Z^{(1)}_{60}(3), Z^{(1)}_{60}(4), Z^{(1)}_{60}(5), Z^{(1)}_{60}(6)]$, the adjacent mean sequence value is $Z^{(1)}_{60}$ = (135266 226499 319798.5 415686 514202). GM (1, 1) model of the basic model for $X^{(0)}_{60}(K) + a_{60}z^{(1)}_{60}(K) = b_{60}$, $\hat{a}_{60} = [a_{60}, b_{60}]^T$ Is the parameter column, $X^{(0)}_{60}(K) + a_{60}z^{(1)}_{60}(K) = b_{60}$ The least squares of the estimated parameter column is $\hat{a}_{60}=(B^T_{60} B_{60})^{-1}B^T_{60}Y_{60}$, Among them:

$$B^T_{60} = \begin{bmatrix} -Z^{(1)}_{60}(2) & -Z^{(1)}_{60}(3) & -Z^{(1)}_{60}(4) & -Z^{(1)}_{60}(5) & -Z^{(1)}_{60}(6) \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -135266 & -226499 & -319798.5 & -415686 & -514202 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$Y_{60} = \begin{bmatrix} 90686 \\ 91780 \\ 94819 \\ 96938 \\ 100076 \end{bmatrix}$$

B^T_{60} is B_{60} transpose. Calculated $\hat{a}_{60}$ = $[-0.025313, 86701.8162]^T$, namely

a_{60} = -0.025313, b_{60} = 86701.8162. Because the time response function of the model

$\hat{X}^{(1)}_{60}(K+1)= [X^{(0)}_{60}(1)-b_{60}/a_{60}]e^{-ak} + b_{60}/a_{60}$ (k=1,2,3,...,n,), To get 60 days with communicable diseases during the interval of the cumulative confirmed cases of infectious diseases in the future the medium and long trend forecast simulation value calculation model of time response function

$$\begin{cases} \hat{X}^{(1)}_{60}(K+1)=3515167.3522e^{0.02531k}-3425244.352 \\ \hat{X}^{(0)}_{60}K+1)= \hat{X}^{(1)}_{60}(K+1) - \hat{X}^{(1)}_{60}(K) \end{cases}$$

In the same way, get 40 with communicable diseases during the interval and 20 days of accumulated cases confirmed the forecast of the future trend of infectious diseases analog value calculation model, the calculation model, respectively, as follows: infectious diseases during the 40 days to predict the future trend of infectious diseases between simulation value calculation model of time response function is:

$$\begin{cases} \hat{X}^{(1)}_{40}(K+1)=802706.6594e^{0.01234k}-712972.6594 \\ \hat{X}^{(0)}_{40}(K+1)= \hat{X}^{(1)}_{40}(K+1)- \hat{X}^{(1)}_{40}(K) \end{cases}$$

The time response function of the model for predicting the future epidemic trend of infectious diseases with a 20-day interval between epidemics:

$$\begin{cases} \hat{X}^{(1)}_{20}(K+1)=182848.7793e^{0.18161k}-93947.7793 \\ \hat{X}^{(0)}_{20}(K+1)=\hat{X}^{(1)}_{20}(K+1)- \hat{X}^{(1)}_{20}(K) \end{cases}$$

3.3. Accuracy Check of Model

The simulated values of confirmed cases were calculated, and their absolute relative errors, sum of relative errors and average errors were shown in Table 2.

Table 2. *Statistical table of simulation errors of confirmed cases at different time intervals.*

Statistical item	statistics	Sum of relative errors	Mean relative error
20-day interval	6	2.0016	0.3336
40-day interval	6	3.8531	0.6422
60-day interval	6	0.100682	0.01167

It can be seen from the data in the table that the average relative error of the simulated values with an interval of 60 days is 0.01167, and the average simulation accuracy is 98.8%. The average relative error of the simulated values with an interval of 40 days is 0.6442, and the average simulation accuracy is 35.78%. The average error of the simulation value after 20 days is 0.3336, and the average simulation degree is 66.64%. Gray modeling requires that the accuracy of the model must reach more than 60% to have the significance of model simulation [11]. Therefore, the prediction model established for the cumulative confirmed cases series at time nodes with an interval of 40 days is not of practical significance for prediction, and the simulation degree of the prediction model established for the cumulative confirmed cases series at time nodes with an interval of 20 days is not high at 66.64%. Only the cumulative number of confirmed cases at time nodes with 60 days interval can the simulation accuracy of the prediction model reached 98.8%, which is a good prediction model [12]. In order to evaluate the grade and prediction accuracy of the available models, we calculated the grade and accuracy of the prediction model based on the cumulative confirmed cases series at time nodes with time intervals of 20 and 60 days. The difference parameters are shown in Table 3. The data obtained from the table and comparative analysis The relative error, mean square ratio, small error rate and 0.67451 value of the prediction model established by the cumulative number of confirmed cases at the time nod.

Table 3. *Comparison table of model accuracy and grade.*

Model grade and model real value	Relative error	Mean square ratio	Small error probability
Level 1 model critical values	0.01	0.35	0.95
Level 2 model critical value	0.02	0.50	0.80
Model real values 60 days apart	0.01167	4.2×10^{-9}	0.99
Model real values 20 days apart	0.3336	9907.5008	0.67

With a time interval of 20 days did not meet the parameter requirements of model level 1-4. The relative error, mean square ratio, small error rate and the value of $0.6745S_1$ of the prediction model established by the cumulative confirmed cases series with a time interval of 60 days meet the requirements of the model level 1 index. The prediction model based on the cumulative confirmed case series at time nodes with 60 days interval reached the optimal level.

3.4. Example Verification of the Mode

In order to verify the actual effect of the above model, we use time nodes with a time interval of 20 days, the prediction model built by the cumulative confirmed cases series is applied to the cumulative confirmed cases at the next (K+1) time point on May 20, 2021. Has carried on the forecast, according to the $\hat{X}^{(1)}_{20}$ (k+1) the time response function prediction calculation, its indeed confirmed that day the number of cases is 90521, through the firm according to international observation records, 90,944 confirmed cases were reported on that day, 423 cases less than the predicted number. With the error of 0.0047, the prediction accuracy reaches 99.54%. We used time nodes with 60 day intervals to accumulate cases the prediction model for the $\hat{X}^{(1)}_{60}$ (K+1) time node for the probable cumulative confirmed cases on February 14, 2022 the simulated prediction data was 102271 cases, while the actual observation recorded a cumulative model of 107094 cases the predicted number was 4823 cases less than the actual data. The relative error was 0.04504, and the prediction accuracy reached 95.50%. The simulation prediction model established with the cumulative number of confirmed cases at the time interval of 20 and 60 days achieved good prediction effect.

4. Discuss

In the process of epidemic prevention and control, in order to analyze and understand the occurrence, development and epidemic law of the epidemic, many data indicators need to be collected and recorded in the work, and different data indicators reveal the characteristics of some aspects of the epidemic situation. Therefore, before we use relevant data for gray modeling, we should carefully analyze the data [13]. Only by selecting scientific data columns can we achieve satisfactory results in modeling. The error and prediction accuracy of the model built by different data column categories and data column obtained at different time points are not the same. Therefore, the error and accuracy of the grey model GM(1,1) built by the original series of nodes at different time points in this study are completely different. In this study, we believe that in the prediction of medium and long-term epidemic trend of infectious diseases in the future, the cumulative series of confirmed cases formed by selecting time nodes at different time intervals will have different internal oscillation and progress modes and strengths, and different internal consistency of data. Therefore, The simulation errors and accuracy of models built with cumulative confirmed case sequences obtained at different time intervals are different, and the effectiveness of models for prediction varies greatly. The cumulative number of confirmed cases is one of the statistical data in the epidemic of infectious diseases. It is the data indicator that the government announces to the society the trend of the epidemic harm of a certain infectious disease. It includes the total number of sick people in a certain period of time, excluding the number of cured and discharged, the number of dead people, that is, the number of people who should be treated in the hospital at present. Therefore, as long as a high-precision prediction model is established to accurately predict the cumulative number of confirmed cases of infectious diseases at a certain point in the future, we can prepare a response plan for the prevention and control of infectious diseases in the future and lay a good foundation for the prevention and control of infectious diseases [14]. In the prediction of the trend of epidemic and development of infectious diseases in the future, it is necessary to combine a variety of forecasting methods and models, and to choose scientific and appropriate original series for modeling and forecasting to make up for the defects and deficiencies of a single forecasting method, so as to make the prediction more perfect and achieve the desired effect.

5. Conclusion

In this study, using the real-time data obtained from the prevention and control of COVID-19 in China since January 1, 2021, the grey GM (1.1) model was established to study the series formed by the cumulative number of confirmed cases at time points with intervals of 20 days and 40 days and 60 days. In this study, XY scatter plot was made for the first few days to graphically describe the trend of data. As can be seen from the graph, the cumulative number of confirmed cases at time nodes with time intervals of 20 days, 40 days and 60 days, and the XY scatter graph drawn by the cumulative number of confirmed cases of COVID-19 in the order of time nodes showed an increase in the cumulative number of confirmed cases with oscillations as time went by. The graph shows monotonous growth. According to the series graph, we selected the cumulative number of confirmed cases with time intervals of 20 days, 40 days and 60 days to model and forecast the future medium and long-term epidemic trend of COVID-19 in China, and built the prediction models of cumulative confirmed cases respectively. The simulation error and accuracy of the models were tested. The average error of the prediction model established by the cumulative number of confirmed cases at the time node with a time interval of 40 days is too high, reaching 0.6422, and the simulation accuracy is only 37%. It has no practical significance for forecasting. The average simulation error of the prediction model established by the cumulative number of confirmed cases with a time interval of 20 days is up to 0.3336, the simulation accuracy is 67%, and the accuracy of the actual verification is 99.54%, which has certain practical value for prediction. The prediction model of GM (1,1) based on the cumulative number of confirmed cases at time nodes with a time interval of 60 days, the average simulation error of the GM (1.1) prediction model was 0.01167, and the simulation accuracy was 98.83%. Multiple parameters in the accuracy analysis reached the index of the first-level model. The actual verification of the model showed that the cumulative number of confirmed cases at the predicted time node was 102271. In practice, 107094 cases were recorded, and the simulated number was 4823 cases less than the actual number [15]. The relative error was 0.045, and the prediction accuracy reached 95.49%, which obtained satisfactory gray modeling prediction effect.

References

[1] Li Ming Quan. Prevention and Control Effects of Different Prevention and Control Mechanisms on the Epidemic of COVID-19 [J]. European Journal of Preventive Medicine, 2022, 10 (6).

[2] Li Ming Quan. Research on the Application of Grey Modeling Theory in the Prediction of Future Epidemic Trend of Infectious Diseases [J]. European Journal of Preventive Medicine, 2023, 11 (1).

[3] Deng Julong, Foundation of Grey Theory (M) Huazhong University of Science and Technology Press, 2002.02.

[4] Liu Sifeng Xie Naiming et al. Grey System Theory and Its Implications (M), Science Press, 2008.

[5] Bernard Rosner (United States) Basis of biostatistics (M) Press: Science Press Time: 2004.

[6] Zhao X F. (2011). Overview of grey system theory (J). Journal of Jilin Provincial Institute of Education (03), 152-154. doi: 10.16083/j.cnki.1671-1580.2011.03.002.

[7] Hu Kun. Research on Grey Prediction Evaluation Method and Application [D]. Nanjing University of Aeronautics and Astronautics, 2004.

[8] Zhang Hu. Research and Application of Time Series Short-term prediction model [D]. Huazhong University of Science and Technology, 2013.

[9] Jia L Y. Grey increment model of population prediction and its application (J). Nanjing University of Information Science & Technology, 2006.

[10] Sun Jing. Research on Product Modeling Design Method Based on Image [D]. Wuhan University of Technology, 2007.

[11] Ma Le. Research on Grey Theory Modeling Method [D]. Dongbei University of Finance and Economics, 2005.

[12] Zhai Xidong. Research on Prediction Model of Port Container Throughput [D]. Dalian University of Technology, 2006.

[13] Rong Luqing, Huang Peihua. Prediction of Cold chain Logistics Demand of Fruit and vegetable in Guangxi and Its Influencing Factors Based on Grey Theory [J]. China, Agricultural Resources and Regional Planning, 2017, 38 (12): 227-234.

[14] Guo Deyong, Li Nian-you, PEI Da-wen, ZHENG Deng-feng. Coal and gas outburst prediction grey theory - neural network [J]. Journal of Beijing university of science and technology, 2007, No. 156 (4): 354-357. The DOI: 10.13374 / j.i ssn1001-053 - x. 2007.04.002.

[15] Wang Zhiming. Research and Application of Multivariate Grey Prediction Model [D]. China University of Mining and Technology2022. DOI: 10.27623/d.cnki.gzkyu.2022.000458.

Biography

Li Ming Quan, male, was born in 1957 with a postgraduate degree in economics, senior animal husbandry division. Long - term engaged in agricultural and animal husbandry economic research work.