COMPARISON OF GM (1,1) AND IMPROVED GM (1,1) MODELS FOR PREDICTION OF HUMAN IMMUNODEFICIENCY VIRUS AND ACQUIRED IMMUNE DEFICIENCY SYNDROME INCIDENCE IN PR CHINA

Daren Zhao[1], Huiwu Zhang[1] and Ruihua Zhang[2]

[1]Administration Office, Sichuan Provincial Orthopedics Hospital; [2]School of Management, Chengdu University of Traditional Chinese Medicine, Chengdu, PR China

Abstract. Human immunodeficiency virus and acquired immune deficiency syndrome (AIDS) continues to be a serious global public health problem negatively impacting development of society and economy, and morbidity and mortality from AIDS are increasing worldwide. Since 1985, when China reported AIDS cases for the first time, incidence of this disease has been continually increasing making prevention and control of AIDS a top public health priority in the country. From AIDS incidence data of the China Health Statistics Yearbook 2010 - 2018, we constructed a standard GM (1,1) model and an improved version to predict AIDS incidence in PR China. Accuracy grade of GM (1,1) and improved GM (1,1) model was at Level 1 (excellent), with mean absolute error of 2.00 and 1.98 and root mean squared error of 2.45 and 2.43 respectively. The improved GM (1,1) model had a higher prediction accuracy performance and simulated well AIDS incidence in PR China from 2010 to 2018. These findings should provide a theoretical basis for science-based guidance on AIDS control and prevention in PR China.

Keywords: AIDS, improved GM (1,1) model, incidence, prediction, PR China

---

Correspondence: Ruihua Zhang, School of Management, Chengdu University of Traditional Chinese Medicine, Chengdu 610041, PR China
Tel: +86 (028) 61800216 E-mail: cdzhangrh@126.com

INTRODUCTION

Human immunodeficiency virus and acquired immune deficiency syndrome (AIDS), has become a serious threat globally, owing to a lack of affordable curative therapy, effective drugs for management and vaccines for prevention

(Aalen *et al*, 1997; Parashar *et al*, 2016; Pongpech *et al*, 2018). Since the first case was reported in USA in 1981, according to the 2020 Global AIDS Prevention and Control Progress Report, there were 38 million HIV-infected patients globally in 2019, with approximately 1.7 million new HIV infections and 690,000 deaths from AIDS-related diseases (UNAIDS, 2020). There are 25.4 million HIV-infected patients on the waiting list for the treatment. In addition, according to a Modelling conducted by UNAIDS and the World Health Organization has shown that if the medical supplies are disrupted for six months for AIDS patients, this could result in 500,000 AIDS-related deaths in sub-Saharan Africa by the end of 2021 (UNAIDS, 2020).

The China Health Statistics Yearbook reported during 2018 64,170 AIDS cases and 18,800 deaths, representing a year-on-year increase of 12.2% and 23.14% respectively (NHCPRC, 2019). This increasing trend in AIDS morbidity and mortality indicates deficiency in control and prevention measures of this infectious disease and poses one of the most important health issues in PR China.

Scientific and reliable prediction of AIDS cases for long-term prevention has proven suitable to monitor and predict the changes in trend and morbidity of AIDS in PR China (Peng *et al*, 2009). However, currently there are several mathematical methods for prediction of AIDS incidence, the most common method being quantitative forecasting methods, viz ARIMA time series forecasting model (Zhou *et al*, 2018), artificial neural network (Renganathan, 2019), exponential smoothing model (Zhang *et al*, 2020), GM (1,1) model (Yang *et al*, 2018), Markov model (Simpson *et al*, 2009), and regression analysis method (Wagner *et al*, 2002). As each of prediction method has different requirements for data characteristics and sample size, choosing a suitable prediction method to obtain an accurate result is the key for accurate prediction of AIDS incidence in PR China.

In 1982, JL Deng at the Huazhong Institute of Technology published a Grey System Theory (Javed and Liu, 2018; Javed and Liu, 2019), which describes a novel mathematical prediction system that can be used to process data where information is partially known, unknown and/or incomplete (Wang *et al*, 2018). It is based on a “small sample”, which refers to the fact that constructing this model requires only a small sample size. Provided that there are four sets of data, the model can be constructed to fit the data set. It is also called a “poor information” model, owing to the fact that some samples provide less information than others. Here the GM (1,1) model and an improved version were constructed and used to simulate AIDS incidence in mainland China from 2010 to 2018. Should the GM (1,1) model (improved or original) prove accurate, it should assist in providing a validated tool for predicting trends in

AIDS incidence, which will be of benefit in developing appropriate AIDS control and prevention for the country.

MATERIALS AND METHODS

GM (1,1) model construction

One advantage of the GM (1,1) model is that it is not necessary to know the data distribution type (Wang *et al*, 2016). The model can be used so long as the data are non-negative monotonic (Ceylan *et al*, 2020). It is based on a random original time series. By accumulation, a new time series is generated. Then, according to the inherent laws of the new time series, a first-order linear differential equation for an approximation was applied (Hu, 2017).

There are six main steps in constructing the GM (1,1) model (Shen *et al*, 2013; Duan *et al*, 2017; Gao *et al*, 2020):

Step 1: Assume the original sequence as

$$x^{(0)} = [\ x^{(0)}(1), x^{(0)}(2), x^{(0)}(3),\dots x^{(0)}(n)\ ] \quad (1)$$

Step 2: Use the formula

$$x^{(1)}(t) = \sum_{i=1}^{t} x^{(0)}(i),\ t = 1,2,3,\dots,\ n\ ,$$

and carry out the original sequence of a first-order accumulation (1-AGO) so as to generate a new sequence:

$$x^{(1)} = [\ x^{(1)}(1), x^{(1)}(2), x^{(1)}(3),\dots x^{(1)}(n)\ ] \quad (2)$$

Step 3: Solve the mean sequence $Z^{(x)}(t)$ as follows:

$$Z^{(1)}(t) = \frac{[\ x^{(1)}(t) + x^{(1)}(t+1)\ ]}{2} \qquad t = 1,2,3,\dots n$$

Step 4: Establish the first order differential equation:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \quad (3)$$

and assuming $\hat{a} = (a,u)^T$ then $\hat{a}$ can be calculated by the following formula:

$$\hat{a} = (a,u)^T = (B^T B)^{-1} B^T Y \quad (4)$$

where $Y = [\ x^{(0)}(2), x^{(0)}(3),\dots x^{(0)}(n)\ ]^T$

$$B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix} \quad (5)$$

$$D = \left\{ (n-1)\left[\sum_{k=2}^{n} Z^2(t)\right] - \left[\sum_{k=2}^{n} Z(k)^2\right]\right\} \quad (6)$$

$$a = \frac{1}{D}\left\{(n-1)\left[-\sum_{k=2}^{n} x(k)Z(k)\right] + \left[\sum_{k=2}^{n} x(k)\right]\left[\sum_{k=2}^{n} Z(k)\right]\right\} \quad (7)$$

$$u = \frac{1}{D}\left\{\left[\sum_{k=2}^{n} Z(k)\right]\left[-\sum_{k=2}^{n} X(k)Z(k)\right] + \left[\sum_{k=2}^{n} Z^2(k)\right]\left[\sum_{k=2}^{n} X(k)\right]\right\} \quad (8)$$

Step 5: GM (1,1) model can be expressed as:

$$\hat{x}^{(1)}(k+1) = \frac{u}{a} + \left[x^{(0)}(1) - \frac{u}{a}\right]e^{-ak} \quad k = 1,2,3,\dots n \quad (9)$$

Step 6: Reduce values using the formula

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}k,$$

and then we obtained the predictive values.

Improved GM (1,1) model construction

First, we assume that the original sequence as:

$$x^{(0)}=\left[\ x^{(0)}(1),x^{(0)}(2),x^{(0)}(3),\dots x^{(0)}(n)\ \right],$$

then the square root transformation of original sequence is performed to generate the new sequence:

$$x_1^{(0)}=\left[\ x_1^{(0)}(1),x_1^{(0)}(2),x_1^{(0)}(3),\dots x_1^{(0)}(n)\ \right]$$

Steps 2-6 are then carried out as described above for GM (1,1) model. Square transformation is conducted when the initial predictive values are calculated values.

Model testing

An *a posteriori* error detection method was employed to test the model accuracy using the formula $C=\frac{S_e}{S_x}$, where C is the mean square deviation, S_e standard deviation of the residual sequence and S_x the standard deviation of the original sequence.

$$S_e=\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[e^{(0)}(k)-e\right]^2},\quad S_x=\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[x^{(0)}(k)-\hat{x}\right]^2}$$

where $e^{(0)}$ is the residual sequence, $x^{(0)}$ original sequence and $\hat{x}$ the predicted sequence value.

The probability of a small error P is calculated using the formula $P=P\{\ e(k)-\bar{e}<0.6745\ S_x\ \}$ where $\bar{e}=\frac{1}{n}e(k)$. The larger the numerical value, the closer it is to .

Performance measures

Performance measures were calculated using the following formulae:

relative error

$$=\left|X_t-\overline{X_t}\right|/X_t\times 100\,\%\,, \qquad (10)$$

mean absolute error (MAE) =

$$=\frac{\sum_{i=1}^{n}\left|X_t-\overline{X_t}\right|}{n} \qquad (11)$$

and root mean squared error (RMSE)

$$=\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(X_t-\overline{X_t}\right)^2} \qquad (12)$$

where X_t is the actual value and $\overline{X_t}$ the predicted value.

Data source

Data were obtained from the China Health Statistics Yearbook for the period 2010 - 2018.

Data analysis

Analysis was carried out using Matlab software (Version R2012b, MathWorks, Natick, MA).

RESULTS

GM (1,1) model

Parameters of GM (1,1) model were a = -0.0672 and u = 2.5539. The equation of GM (1,1) model was $x(k+1) = 40.5449e^{(0.067k)} - 37.9849$, with mean square deviation ratio C ≤0.35 and probability of small error $P \geq 95\%$. Hence, the prediction accuracy grade was at Level 1 (excellent) (Table 1).

Improved GM (1,1) model

Parameters of the improved GM (1,1) model were a = -0.0332 and u = 1.6008. The equation of the improved GM (1,1) model was $x(k+1) = 49.7727e^{(0.033k)} - 48.1727$, with mean square deviation ratio C ≤0.35 and probability of small error $P \geq 95\%$. Hence, the prediction accuracy grade was at Level 1 (excellent) (Table 1).

Comparison between GM (1,1) model and improved version in predicting AIDS incidence in PR China from 2010 - 2018.

Relative error value range of GM (1,1) model was 0-3.60%, whereas the improved GM (1,1) model was 0-3.68% (Table 2). MAE of the two models was 2.00 and 1.98 and RMSE 2.45 and 2.43 respectively.

DISCUSSION

In order to control and monitor the prevalence and spread of AIDS recourse to a mathematical prediction model is a useful strategy (Li and Li, 2020). The model provides a reference for formulation of AIDS preventive measures and optimal allocation of medical resources for AIDS treatment (Martinez *et al*, 2019). Although there are several methods for quantitative prediction of AIDS incidence, it is essential to select the most suitable prediction model validated by comparison with past recorded data.

The GM (1,1) model and an improved version (Liu *et al*, 2019) were constructed and evaluated for their predictive abilities by comparing with

Table 1

Parameters for determination prediction accuracy grade

Prediction accuracy grade	Mean square deviation ratio (C)	Probability of small error (P)
Level 1 (excellent)	C≤0.35	95%≤P
Level 2 (qualified)	0.35<C≤0.50	80%≤P<95%
Level 3 (barely qualified)	0.5<C≤0.65	70%≤P<80%
Level 4 (unqualified)	C>0.65	P<70%

Table 2

Predicted results from GM (1,1) model and improved version of AIDS incidence in PR China (2010 - 2018)

Year	Actual incidence*†	GM (1,1) model predicted incidence*	Relative error (%)	Improved GM (1,1) model predicted incidence*	Relative error (%)
2010	2.56	2.56	0.00	2.56	0.00
2011	2.92	2.82	3.43	2.83	3.14
2012	3.11	3.02	3.03	3.02	2.81
2013	3.12	3.23	3.38	3.23	3.54
2014	3.33	3.45	3.60	3.45	3.68
2015	3.69	3.69	0.01	3.69	0.01
2016	3.97	3.95	0.60	3.94	0.67
2017	4.15	4.22	1.71	4.21	1.55
2018	4.62	4.51	2.29	4.50	2.51

*incidence was reported as per 100,000 population

†Actual incidence was obtained from NHCPRC, 2019

known AIDS incidence in PR China over a nine-year period. The data set is small with a non-normal distribution, meeting the characteristics requirements of the GM (1,1) model. Although an additional square root transformation step is required in the improved GM (1,1) model, sample size and data characteristics remain unchanged.

Surprisingly, there is no significant difference between the models' equation and the a posteriori difference test for two models. The accuracy grade of GM (1,1) model and the improved version were both Level 1 (excellent) indicating that the fitting of the predictive performance was both high. However, MAE and RMSE of the improved GM (1,1) model were (slightly) lower than those of the standard GM (1,1) model, which suggests that the predictive performance of the improved GM (1,1) model was higher than GM (1,1) model, making the prediction of improved GM (1,1) model closer to reality.

In conclusion, the study compared GM (1,1) model and an improved version in their ability to simulate AIDS incidence in China from 2010 to 2018, with the results that the accuracy performance of the improved GM (1,1) model was higher than the standard GM (1,1) model, and could be used as a predictive tool for future AIDS incidence. However, certain precautions should be taken into consideration; for example, the mathematical model does not take into account social, cultural, economic and other unexpected (such as COVID-19 pandemic) factors, and future studies will have to factor in these parameters to make the model more suitable and accurate if it is to be useful in assisting future formulation of AIDS control and prevention policies for PR China.

ACKNOWLEDGEMENTS

The study and publication cost were supported by the Sichuan Provincial Primary Health Service Development Research Center (grant no. SWFZ21-Q-59).

CONFLICTS OF INTEREST DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

Aalen OO, Farewell VT, De Angelis D, Day NE, Gill ON. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Stat Med* 1997; 16: 2191-210.

Ceylan Z, Bulkan S, Elevli S. Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul. *J Environ Health Sci Eng* 2020; 18: 687-97.

Duan J, Jiao F, Zhang Q, Lin Z. Predicting urban medical services demand in China: an improved Grey Markov Chain Model by Taylor Approximation. *Int J Environ Res Public Health* 2017; 14:883.

Gao J, Li J, Wang M. Time series analysis of cumulative incidences of typhoid and paratyphoid fevers in China using both Grey and SARIMA models. *PLoS One* 2020; 15: e0241217.

Hu YC. A genetic-algorithm-based remnant grey prediction model for energy demand forecasting. *PLoS One* 2017; 12: e0185478.

Javed SA, Liu S. Predicting the research output/growth of selected countries: application of Even GM (1, 1) and NDGM models. *Scientometrics* 2018; 115: 395-413.

Javed SA, Liu S. Correction to: Predicting the research output/growth of selected countries: application of Even GM (1, 1) and NDGM models. *Scientometrics* 2019; 120: 1505.

Li Z, Li Y. A comparative study on the prediction of the BP artificial neural network model and the ARIMA model in the incidence of AIDS. *BMC Med Inform Decis Mak* 2020; 20: 143.

Liu YL, Cao WJ, Li F, et al. Application of improved GM (1,1) model in predicting the incidence of tuberculosis, 2019 [cited 2021 May 18]. Available from: URL: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2019&filename=HXYF201903038&uniplatform=NZKPT&v=fmxrIdLhu6G-zxA8w3wFXHWQM4sMZAkbWR4N-9NfFwSjtt55AeHNs3gzxrDkPNmJl6 [in Chinese]

Martinez EZ, Zucoloto ML, Galdino G, Nunes AA, Lizzi EADS. Spatiotemporal distribution of acquired immunodeficiency syndrome incidence in Brazil between 2012 and 2016. *Rev Soc Bras Med Trop* 2019; 53: e20190086.

National Health Commission of the People's Republic of China (NHCPRC). 2019 China Health Statistics Yearbook, 2019 [cited 2021 May 18]. Available from: URL: http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm

Parashar S, Collins AB, Montaner JS, Hogg RS, Milloy MJ. Reducing rates of preventable HIV/AIDS-associated mortality among people living with HIV who inject drugs. *Curr Opin HIV AIDS* 2016; 11: 507-13.

Peng ZX, Wang N, Wang L. Development of methods for estimation and prediction on epidemic situation of HIV/AIDS. *Zhonghua Liu Xing Bing Xue Za Zhi* 2009; 30: 294-7. [in Chinese]

Pongpech N, Avihingsanon A, Chaiwarith R, *et al*. Prediction model of pre-treatment HIV RNA levels in naïve Thai HIV-infected patients: application for resource-limited settings. *Southeast Asian J Trop Med Public Health* 2018; 49: 965-74.

Renganathan V. Overview of artificial neural network models in the biomedical domain. *Bratisl Lek Listy* 2019; 120: 536-40.

Shen X, Ou L, Chen X, Zhang X, Tan X. The application of the grey disaster model to forecast epidemic peaks of typhoid and paratyphoid fever in China. *PLoS One* 2013; 8: e60601.

Simpson KN, Strassburger A, Jones WJ, Dietz B, Rajagopalan R. Comparison of Markov model and discrete-event simulation techniques for HIV. *Pharmacoeconomics* 2009; 27: 159-65.

The Joint United Nations Programme on HIV/AIDS (UNAIDS). UNAIDS data 2020, 2020 [cited 2021 May 18]. Available from: URL: https://www.unaids.org/sites/default/files/media_asset/2020_aids-data-book_en.pdf

Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002; 27: 299-309.

Wang Y, Wei F, Sun C, Li Q. The research of improved grey GM (1,1) model to predict the postprandial glucose in type 2 diabetes. *Biomed Res Int* 2016; 2016: 6837052.

Wang YW, Shen ZZ, Jiang Y. Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China. *PLoS One* 2018; 13: e0201987.

Yang X, Zou J, Kong D, Jiang G. The analysis of GM (1, 1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China. *Medicine (Baltimore)* 2018; 97: e11787.

Zhou L, Zhao P, Wu D, Cheng C, Huang H. Time series model for forecasting the number of new admission inpatients. *BMC Med Inform Decis Mak* 2018; 18: 39.

Zhang YQ, Li XX, Li WB, et al. Analysis and predication of tuberculosis registration rates in Henan Province, China: an exponential smoothing model study. *Infect Dis Poverty* 2020; 9: 123.