Early warning of body lesion trends using grey prediction models

Bo Zeng, Yingjie Yang, Xiaoyi Gou

Abstract: At present, hundreds of millions of Chinese are facing increasingly serious health risks, and health check has undoubtedly played a significant role in finding health risks. However, the current health check in China mainly judges the quality of physical functions by a single index value without dynamic analysis of the changing trends of the index, which may lead to unreasonable diagnostic conclusions. In this paper, the data characteristics of physical indicators are systematically analyzed, and thegrey system models dedicated to data with the aforementioned characteristics are applied to simulate and predict the changing trends of body indicators. On this basis, the possible pathological changes of body organs are identified. Specifically, this paper analyses the state of human kidney functions by grey prediction models. The results showed that even when the renal function index (serum creatinine) is within the normal range, the human renal function might be in trouble. The grey model analysis of the change trends of serum creatinine can predict the potential health hazards of renal functions. The research results of this paper have significant potential to produce tremendous economic value and social benefits.

Keywords: Early warning of body lesion trends; Data characteristics of physical indicators; Grey prediction models; Serum creatinine and renal functions

1.Introduction

China has won many medals in the World Olympic Games. However, it does not mean that Chinese are in good health. On the contrary, hundreds of millions of Chinese are facing serious health risks. According to the big data of Chinese Health [1], chronic diseases and malignant tumors have become fatal factors threatening Chinese health. At present, 160 million Chinese suffer from dyslipidemia; 270 million Chinese endure hypertension and nearly 100 million Chinese have diabetes. In China, on average, one person suffers from cancer each 10 seconds, at least one person dies from cardiovascular diseases every 30 seconds, and 70% of Chinese are at risk of death from overwork.

Facing with the more and more serious health problems, Chinese begin to attach importance to the positive role of routine physical in timely detection of disease risks and protection of human health. At present, health physical has become a vital way to prevent hidden dangers of physical diseases, and has attracted widespread attention of governments and people all over the world.

However, the existing way of health physical analysis has some limitations. At present, doctors analyze the body health status and potential disease risks through various indicators, and draw the corresponding physical conclusions. However, this judgment of health status is mainly based on the static analysis of only one physical result, and it is difficult to find the dynamic development trends of body indicators at different stages. At the same time, since the physical of a different patient has different body diathesis, an identical body index between different patients may not mean the same physical condition. Therefore, it is possible to draw unreasonable conclusions about the health status and physical condition of a patient only by comparing and statically analyzing the body indicators. It is short of the dynamic comparison and longitudinal analysis of the changing trends of the body indicators of the patient.

For example, Serum Creatinine (SC) is the most common indicator used to test the renal function, and it is also a compulsory item for physical. Normal values of Serum Creatinine in men are usually 53-106 microrubbing/liter. However, although the data of Serum Creatinine of a patient is within the interval (that is [53,106]), his kidney may produce lesion. We use a practical example to explain this phenomenon. The results of six recent physical of Serum Creatinine with a patient were 87, 96, 98, 104, 100 and 105 respectively. The results of physical showed that the patient's Serum Creatinine was completely normal, which led to the conclusion that the renal function of the patient was normal. However, The patient often finds that his urine is red and has a large amount of foam, which has some characteristics of chronic nephritis. Finally, through the analysis of kidney biopsy, it was confirmed that the kidney of the patient had pathological changes.

Hence, normal physical indicators do not mean that a person is healthy.

The limitation of the current health physical analysis is shown in Fig.1, as follows.

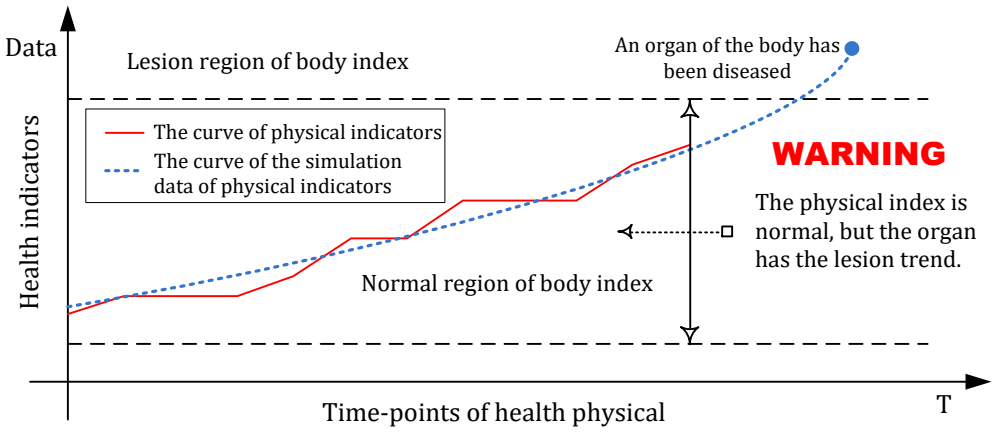

Fig.1 The limitation of current health physical

Therefore, an analysis of the dynamic trends of indicators is better than the current static analysis, which can make a reasonable evaluation of possible future human health status. This paper intends to use prediction modeling technology to describe the law of development and changes of body indicators, so as to accurately predict the changing characteristics and range of body indicators. On this basis, we can implement early warning management for human health. It can help to discover the hidden dangers of human health in time and take preventive therapeutic measures, which is of positive significance to the protection of human health.

The remainder of this paper is organized as follows. In Section 2, we summarize the datacharacteristics of physical indicators. In Section 3, we analyze how to select an appropriate mathematical model to predict the trend of body organ lesions. In Section 4, the deduction process, error test method and modelling steps of GM(1,1) model for health warning are introduced in detail. In Section 5, the validity and feasibility of grey prediction model for early warning of physical health are verified by a real world case study. In Section 6, we analyse four kinds of grey prediciton modeling methods. The last part of the paper discusses the possible economic benefits of the research results.

2. Data Characteristics of Physical Indicators

Any predictive method has certain modeling conditions and application scope. The selection of prediction model is closely related to the data characteristics and sample size of modeling object. For this, the data characteristics of physical indicators will be analyzed firstly .

The index data of a physical at different time points constitute the time-sequence data of the index. The time-sequence data from time point 1 to ***n*** of the ***m***-index is denoted as,

$$X_m = \left(x_m(t_1), x_m(t_2), \cdots, x_m(t_n) \right)$$

Based on an in-dpeth analysis, the following six characteristics of X_m are identified: (A) Small data, (B) Information uncertainty, (C) Non-uniform data intervals, (D) Data type heterogeneity, (E) Complex influencing factors and (F) Data Oscillation. The data characteristics of X_m are described in detail below.

(A) Small data. In China, people normally take 1or 2 physical tests each year. The size of physical data sets is closely related to age structure, living condition, working environment and other factors, which has different significance levels and timeliness. In other words, the old data measured a long time ago is of little reference value in the analysis of current

physical conditions. Normally, the data recorded in the last 6-8 years for indicators of a patient's physical can systematically reflect the evolution process and development trend of body functions. Hence, the number of elements in X_m is usually between 6 and 16. In consideration of the specific situation in which one did not have a physical in some years, the number of elements in X_m might be smaller.

(B) Information uncertainty. Physical indicators are influenced by many factors, such as body and mental state, diet structure, testing equipment, measurement errors, technical level and so on. Hence, it is difficult to obtain accurate values of these indicators usually, which is the main reason why "abnormal indicators" need to be re-examined after a physical. Although we cannot determine the exact indicators, we can make sure their possible values or approximate ranges by multiple physical results. After this, the uncertainty of the interval or discrete set of physical indicators can be evaluated. In Systems Science, the former is named as 'Interval grey number', for exmaple $[\underline{a},\overline{a}]$; and the latter is named as 'discrete grey number', for example $\{a_1,a_2,\cdots,a_t\}$.

(C) Non-uniform data intervals. Let $X_m=(x_m(t_1),x_m(t_2),\cdots,x_m(t_n))$ be a time sequence of the physical indicator ***m***, Δt_k is the time interval between adjacent data in X_m. That is

$$\Delta t_k=t_k-t_{k-1},k=2,3,\cdots,n$$

If Δt_k is not a constant, X_m is named as a non-equidistant sequence. Non-equidistant sequences are often encountered in the process of the data acquisition of physical. If the test is not carried out in some years, the physical indicator of the patient provides an unequal time sequence with data missing in some years.

(D) Data type heterogeneity. The so-called data type heterogeneity refers to the mixed data types among elements in sequence $X_m=(x_m(t_1),x_m(t_2),\cdots,x_m(t_n))$. In other words, the data type is different in X_m. For example, $x_m(t_w)$ in X_m is an interval grey number with a certain range of values, $x_m(t_w)\in[\underline{a},\overline{a}]$; $x_m(t_u)$ is a discrete grey number, $x_m(t_u)\in\{a_1,a_2,\cdots,a_t\}$ and it is impossible to determine which value is true; and $x_m(t_v)$ is a real number, $x_m(t_v)=a$. Because of the uncertainty of the information of the physical indicators, the phenomenon of 'Data type heterogeneity' of the data of physical indicators is brought about.

(E) Complex influencing factors. A body index is an important parameter reflecting the health status of a certain function of the body. For example, Serum Creatinine is commonly used to evaluate the good or bad of the renal function of a patient. The physical indicators are affected by many factors, such as environmental factors (air, water quality, food), genetic factors, psychological factors (anxiety, irritability, fear), lifestyle factors (habits, health

awareness) and health care services. These factors are extremely complex and difficult to be fully identified and quantified.

(F) Data Oscillation. The physical indicators reflecting body health status are always influenced by many uncertain factors, such as internal factors and external environment. As a result, a physical indicator is usually not a constant or meets some monotonous laws of change even for a particular healthy individual. Hence, the physical indicator often shows oscillation characteristics of increasing or decreasing in a certain range.

The six characteristics of time sequence X_m of physical indicators are shown in Fig.2, as follows.

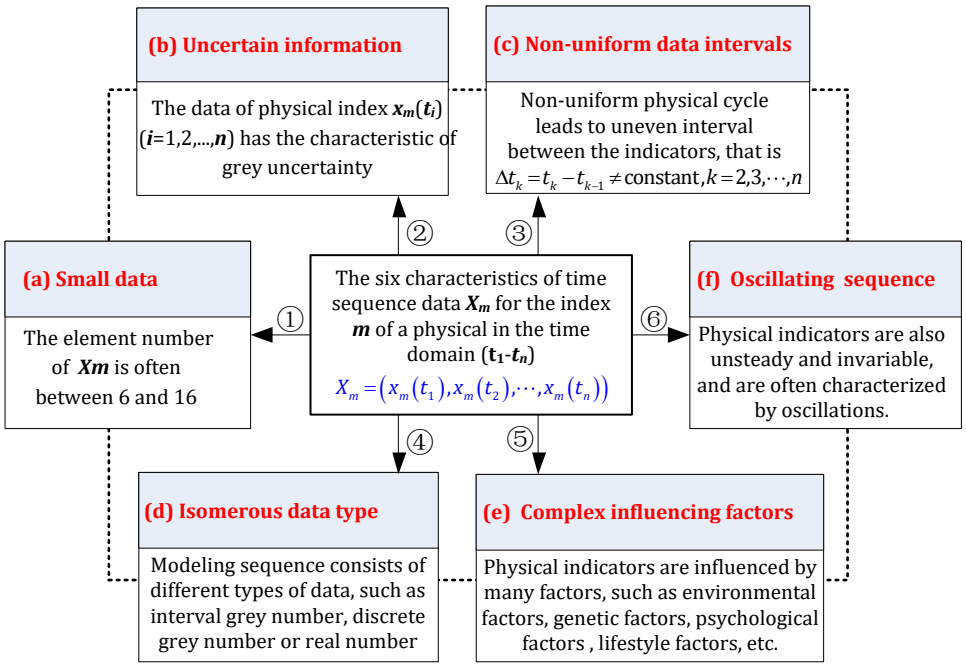

Fig.2 Six data characteristics of physical indicators

3. Selection of health early warning methods

With the rapid development of prediction modelling theory and technology, various prediction methods were proposed one after another. From the number of variables, prediction models can be devided into single variable and multi-variable prediction models. Single variable prediction models simulate and forecast the law of change of a system by analyzing the trend of the characterstic variable of the system. The typical single-variable prediction model includes single-variable neural network models [2], grey prediction models [3], ARIMA models [4], Markov prediction models[5], and their combination models, etc [6-10]. Multi-variable prediction models need to follow with interests on what factors influence system changes, and then the function relation between the dependent variable

and independent variables is built. After this, the dependent variable can be forecasted on the basis of knowing the development trends of independent variables. Multi-variable prediction models mainly include regression analysis models [11], structural equation models [12], multi-variable neural network models [13] and grey prediction models [14], etc.

According to the previous analysis, we can see that the influence factors of physical indicators are complicated (see the characteristic (e) in Fig.2) and we cannot ascertain the composition of independent variables and their degree of influence. Hence, the single-variable prediction model will be employed to forecast the developing trends of physical indicators. As far as single-variable prediction model, such as neural network models and ARIMA models, they are all based on large sample data (the size of sample is not less than 30), otherwise, their simulation and prediction precisions are often unsatisfactory. On the other hand, although the Markov prediction model can be used for small data systems, it is not suitable for medium and long term prediction.

Grey prediction is a new and burgeoning prediction modeling method. It has the advantage of less samples needed and simple model structure [15-16]. In the meantime, grey prediction also has very strong modeling ability for uncertain data (interval grey number, discrete grey number and isomerous data) and non-stationary data (oscillating data, fluctuation data and non-equidistant data) [17-19]. These modeling characteristics of grey prediction models fit with the six data characteristics of physical indicators well. Besides, single grey prediction model does not need to consider the impact of related factors on physical indicators. Hence, grey prediction models with their unique modeling characteristics are suitable to forecast the trends of change of those physical indicators which have the complicated data peculiarities.

Therefore, in this paper, the single grey prediction model will be applied to simulate the rule of change of physical indicators and forecast the range of changes in physical indicators over the next period of time. After this, the early warning management of human health will be facilitated.

Grey System Theory (GST) [20] was first introduced in 1982 by J.L. Deng (1933–2013) at Huazhong University of Science and Technology in China. Grey system theory established a relatively new approach for addressing uncertain systems, or systems of systems, problems having small samples, and low-quality information [21]. The theory enables one to model, analysis, monitor, and control such partially defined systems by generating, excavating, and extracting useful information from what is available. It built on the work of Dr. Lotfi A. Zadeh,

who introduced the concept of fuzzy sets in the 1960s that in turn led to breakthroughs in neural networks and soft computing [22].

Many prominent scholars have commended grey system research [23]. Such scholars include Professor Lotfi A. Zadeh (US), the founder of fuzzy mathematics, Professor Herman Haken (Germany), the founder of synergetics, Professor James M. Tien(US), former vice-president of IEEE and member of the National Academy of Engineering, Professor Robert Valee (France), president of World Organization of Systems and Cybernetics, Professor Alex Andrew (UK), the Secretary General of the World Organization of Systems and Cybernetics and President of the Canadian Royal Academy of Sciences, as well as many Academicians of the Chinese Academy of Sciences and the Chinese Academy of Engineering, including Professor Qian Xuesen, Professor Yang Shuzi, Professor Xiong Youlun, etc.

In 2013, Professor Liu Sifeng, Nanjing University of Aeronautics and Astronautics of China was selected as the Senior Fellow of Maricury International Talent Introduction Program in the Seventh Framework of EU (GS-A-DM-DS FP7-PEOPLE-2013-IIF 629051) with 'excellent' grade. He held a series of academic exchanges on GST in Europe and North America to promote and disseminate Grey System Theory worldwide. In the natural world, uncertain systems with small samples and poor information exist commonly. This fact determines the wide applicability of grey systems theory.

At present, GST has been widely used in human health fields such as disease cause analysis and morbidity prediction. (i) Grey System Theory for Predicting Disease Incidence: Fu,Y used dynamic additional grey catastrophe prediction model to forecast the artificial muscle tremor behavior [24]; Bao, CZ adopted grey model GM(1,1) to forecast all-cause and disease-specific rates of disability adjusted life years (DALYs) in 2015 and 2020 [25]; Shen, X.J applied Grey Disaster Model to forecast Epidemic Peaks of Typhoid and Paratyphoid Fever in China [26]; Gan, R.J implemented a Hybrid Method Combining Grey Model and Back Propagation Artificial Neural Networks to Forecast Hepatitis B in China [27]; Zhang, X.N and Wang, Y.W took ARIMA and grey predictive model to forecast the trend of early childhood caries and hepatitis B in mainland China [28]; Qu, H.M combined grey models and linear regression models to predict cancer mortality [29]. Zhang, L.P optimized Nash nonlinear grey Bernoulli model based on particle swarm optimization and applied it in prediction for the incidence of Hepatitis B in Xinjiang, China [30]. (ii) Grey System Theory for analysing the cause of disease: Han, X applied a comprehensive grey correlation degree to assess the simultaneous and lag associations between the lung cancer morbidity and PM2.5 concentration [31]; Chen, B.X used grey relational analysis to study the relation between

coronary lesion scores and SLS [32]; Zhang, Q.Y employed grey relational analysis to investigate the relation between blood glucose and serum lipids level in type 2 diabetes mellitus (T2DM) patients with secondary malignancy [33].

According to the above literatures, GST is mainly used to analyze the influencing factors of human diseases and predict the incidence of human diseases, while the research on the application of GST to early warning of human health is rare. Therefore, this paper will try to study how to use grey prediction models to analyze and warn one's health status according to his/her physical indicators.

4. Health early warning with grey prediction model

The core idea of applying grey prediction models to study the early warning of human health is to establish a grey prediction model based on the human body characteristic indicators, to simulate and predict the law of change and development trends of the body indicators, and then to analyze and conduct early warning for the health status of a human organ according to the predicted data.

How to establish grey early warning model for human health? The following steps demonstrate the procedure to be taken:

1. To determine which indicators are the core indicators of body function (Step1);
2. To collect and collate historical data with core indicators to prepare for the establishment of grey model (Step2);
3. To establish grey model prediction model to simulate and predict body index (Step3);
4. To analyse and give early warning of human health status based on indicators' predicted data (Step4).

These modeling steps are shown in Fig.3.

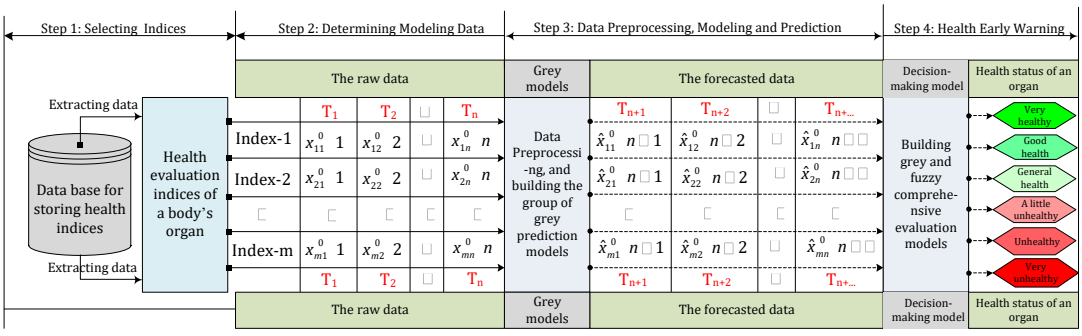

Fig.3 Modeling steps of grey prediction model for human health

In this section, several basic concepts of grey theory are introduced. Then, the most common

grey prediction model GM(1,1) [22, 34], which contains one variable and one derivative, is introduced.

Definition.1 [22] Let $X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right)$ be a sequence of raw data, and D be a sequence operator, satisfying that

$$X^{(0)}D=\left(x^{(0)}(1)d,x^{(0)}(2)d,\cdots,x^{(0)}(n)d\right)$$

Where

$$x^{(0)}(k)d=\sum_{i=1}^{k}x^{(0)}(i),k=1,2,\cdots,n$$

Then D is called the first order accumulating generation operator (1-AGO), and $X^{(0)}D$ is the first order sequence worked on by the 1-AGO. $X^{(0)}D$ can be expressed as $X^{(1)}$, that is

$$X^{(1)}=\left(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\right)$$

Where

$$X^{(1)}=X^{(0)}D \text{ and } x^{(1)}(k)=x^{(0)}(k)d,\ k=1,2,\cdots,n$$

Definition.2 [22] Sequences $X^{(0)}$ and $X^{(1)}$ are shown in Definition.1, then

$$Z^{(1)}=\left(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)\right)$$

Where

$$z^{(1)}(k)=\frac{x^{(1)}(k)+x^{(1)}(k-1)}{2},\ k=2,3,\cdots,n$$

Then $Z^{(1)}$ is the mean sequence of consecutive neighbors generation.

Definition.3 [22] Sequences $X^{(0)}$ and $Z^{(1)}$ are shown in Definition.1 and Definition.2, then the following differential equation

$$x^{(0)}(k)+az^{(1)}(k)=b \tag{1}$$

is called a grey prediction model with the first derivative and one variable, GM(1,1) for short.

Theorem.1 [22] Sequences $X^{(0)}$, $X^{(1)}$ and $Z^{(1)}$ are shown in Definitions.1-3, then the least square estimate sequence of the GM(1,1) model satisfies

$$\hat{a}=\left[a,b\right]^{T}=\left[B^{T}B\right]^{-1}B^{T}Y$$

Where

$$Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \text{ and } B=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

Proof is omitted here.

Definition.4 [22] assume that $X^{(0)}$, $X^{(1)}$, $Z^{(1)}$ and $\hat{a}$ are the same as in **Theorem**.1, then

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b \tag{2}$$

is called a whitenization (or image) equation of the differential equation $x^{(0)}(k)+az^{(1)}(k)=b$, then, the solution (or called a time response function) of the whitenization equation is given by

$$x^{(1)}(t)=\left[x^{(1)}(0)-\frac{b}{a}\right]e^{-at}+\frac{b}{a} \tag{3}$$

the time response sequence of the GM(1,1) model is given by

$$\hat{x}^{(1)}(k+1)=\left[x^{(1)}(0)-\frac{b}{a}\right]e^{-ak}+\frac{b}{a},\ k=1,2,\cdots,n \tag{4}$$

the restored values $\hat{x}^{(0)}(k+1)$ of $\hat{x}^{(1)}(k+1)$ is given by

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k) \tag{5}$$

Normally, only the models that pass various test can be meaningfully employed to make predictions.

Definition 5 [35] Assume that a raw sequence $X^{(0)}$

$$X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right)$$

$X^{(0)}$ is used to build the GM(1,1) model, and the simulation sequence is $\hat{X}^{(0)}$, as follows,

$$\hat{X}^{(0)}=\left(\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n)\right)$$

The error sequence of $X^{(0)}$ is ε, as follows

$$\varepsilon=\left(\varepsilon(1),\varepsilon(2),\cdots,\varepsilon(n)\right)$$

where

$$\varepsilon(u)=x^{(0)}(u)-\hat{x}^{(0)}(u),u=1,2,\cdots,n \tag{6}$$

The relative simulation percentage error (RSPE) of simulation sequence $\hat{X}^{(0)}$ is Δ

$$\Delta=\left(\Delta(1),\Delta(2),\mathrm{L}\ ,\Delta(n)\right)$$

where

$$\Delta(u)=\left|\frac{\varepsilon(u)}{x^{(0)}(u)}\times 100\%\right|,u=1,2,\mathrm{L}\ ,n \tag{7}$$

The mean relative simulation percentage error (MRSPE) of the simulation sequence is $\overline{\Delta}$

$$\overline{\Delta}=\frac{1}{n}\sum_{u=1}^{n}\Delta(u) \tag{8}$$

For giving threshold values α (The threshold is set according to the specific situation of the system), when $\overline{\Delta}<\alpha$ hold true, the grey model is said to be error-satisfactory.

A mathematical model passing the error test can be used for prediction. The purpose of prediction is to provide data support for decision analysis. After applying grey models to predict the change of body index in a future time, the body function reflected by the index values can be analyzed, and the early warning processing based on different situations can be carried out.

5. Case study on early warning for renal failure

. Serum creatinine is the most commonly used indicator of renal function, and is also a compulsory item for health examination. In muscle, creatine slowly forms creatinine through irreversible non-enzymatic dehydration, which is released into the blood and excreted with urine. Therefore, serum creatinine is closely related to the total amount of muscle in the body and is not easily affected by diet.

Creatinine is a small molecule substance, which can be filtered through the glomerulus. It is rarely absorbed in the renal tubules. The creatinine produced in the body every day is almost completely excreted with the urine and is generally not affected by the amount of urine. Clinical detection of serum creatinine is one of the main methods to understand renal function. Usually, the normal value of serum creatinine is 53-106 microrubbing/liter for males, 44-97 microrubbing/liter for females and 24.9-69.7 micromol/L for children. The data of serum creatinine indices of a physical examiner in the last six times are shown in Table 1, as follows.

Table 1 Serum creatinine of a physical examinee in the last six times (unit: micro-friction/liter)

Order	T1	T2	T3	T4	T5	T6
Serum creatinine	87	96	98	104	100	105

(Data from health check, in view of confidentiality and privacy, it is inconvenient to disclose the information

of the patient.)

From Table 1, the raw data sequence $X^{(0)}$ is as follows

$$X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5), x^{(0)}(6)\right)$$
$$= (87, 96, 98, 104, 100, 105)$$

According to Definition 1 and Definition 2, the new sequences $X^{(1)}$ and $Z^{(1)}$ are be obtained from $X^{(0)}$, as follows,

$$X^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), x^{(1)}(4), x^{(1)}(5), x^{(1)}(6)\right)$$
$$= (87, 183, 281, 385, 485, 590)$$

And

$$Z^{(1)} = \left(z^{(1)}(2), z^{(1)}(3), z^{(1)}(4), z^{(1)}(5), z^{(1)}(6)\right)$$
$$= (135.0, 232.0, 333.0, 435.0, 537.5)$$

From sequences $X^{(0)}$ and $Z^{(1)}$, matrices B and Y can be obtained, as follows,

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \\ x^{(0)}(5) \\ x^{(0)}(6) \end{bmatrix} = \begin{bmatrix} 96 \\ 98 \\ 104 \\ 100 \\ 105 \end{bmatrix} \text{ and } B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ -z^{(1)}(4) & 1 \\ -z^{(1)}(5) & 1 \\ -z^{(1)}(6) & 1 \end{bmatrix} = \begin{bmatrix} -135.0 & 1 \\ -232.0 & 1 \\ -333.0 & 1 \\ -435.0 & 1 \\ -537.5 & 1 \end{bmatrix}$$

According to Theorem.1 and matries B and Y, the parameters $\hat{a} = [a, b]^T$ can be computed, as follows

$$\hat{a} = [a, b]^T = \left[B^T B\right]^{-1} B^T Y = [-0.0198, 93.9770]^T$$

From the parameters $\hat{a} = [a, b]^T$ and Definition.4, the GM(1,1) model for the early warning of renal failure can be established, as follows.

$$\hat{x}^{(1)}(k+1) = \left[x^{(1)}(0) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a} = \left[87 + \frac{93.977}{0.0198}\right] e^{0.0198k} + \frac{93.977}{0.0198} \quad (9)$$

From Eq.(9) and **Definition 5**, the simulated sequence $\hat{X}^{(0)}$, the error sequence ε, and the relative simulation percentage error (RSPE) of simulation sequence $\hat{X}^{(0)}$ can be computed, the results are as the follows.

$$\hat{X}^{(0)} = \left(\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \hat{x}^{(0)}(3), \hat{x}^{(0)}(4), \hat{x}^{(0)}(5), \hat{x}^{(0)}(6)\right)$$
$$= (87, 96.653, 98.586, 100.557, 102.568, 104.619)$$

Then

$$\varepsilon = \left(\varepsilon(1), \varepsilon(2), \varepsilon(3), \varepsilon(4), \varepsilon(5), \varepsilon(6)\right)$$
$$= (0, -0.653, -0.586, 3.443, -2.568, 0.381)$$

$$\Delta = \left(\Delta(1), \Delta(2), \Delta(3), \Delta(4), \Delta(5), \Delta(6)\right)$$
$$= \left(0, 0.681\%, 0.598\%, 3.310\%, 2.568\%, 0.363\%\right)$$

And

$$\bar{\Delta} = \frac{1}{n}\sum_{u=1}^{n}\Delta(u) = 1.504\%$$

Since $1\% < \bar{\Delta} < 5\%$, according to the Reference Table of Error Level for Grey Prediction Model [210-211], the error level of the GM(1,1) model is close to the first class standard. Then, the GM(1,1) model can be used to forecast the future trend of renal function in patients undergoing physical examination, the prediciton data are shown as follows.

$$\hat{X}^{(0)} = \left(\hat{x}^{(0)}(7), \hat{x}^{(0)}(8), \hat{x}^{(0)}(9), \hat{x}^{(0)}(10), \hat{x}^{(0)}(11), \hat{x}^{(0)}(12)\right)$$
$$= \left(106.711, 108.845, 111.022, 113.242, 115.507, 117.816\right)$$

According to the prediction data, the renal function of the physical examiner is on the trend of accelerating deterioration. By puncture sampling, the examiner has actually confirmed chronic nephritis. However, from Table 1, the renal function of the physical examiner is still within the reasonable range and everything is normal. Hence, the purpose of this study is to inform the physical examiner of the hidden dangers of diseases in advance, and we think it's a very important job.

6. Discussion

The data sequence in Table 1 has an approximate monotonic increasing law, and the traditional GM (1,1) model is adequate for the modeling of such sequence. However, in fact, physical examination data do not all satisfy the monotonous increasing law. How to use grey prediction model to effectively simulate the trends of change of body indicators at this time?

Generally, according to the trend of data change, time sequence data can be divided into four categories: approximate monotonic increasing (decreasing) sequence, saturated S-shaped sequence, fluctuation sequence and oscillation sequence. For approximate monotonic sequences, GM(1,1) can be used directly to model them, which is not discussed here. For waveform sequence, grey waveform prediction model or GM(1,1|Tan(k-τ)p,sin(k-τ)P) [36] can be applied. For the oscillation sequence, it is slightly complicated. The amplitude of the oscillation sequence is compressed by data transformation to improve the smoothness of the oscillation sequence [37], or the envelope lines [38] are used to simulate the range of the oscillation sequence in order to

realize the interval prediction of the oscillation sequence.

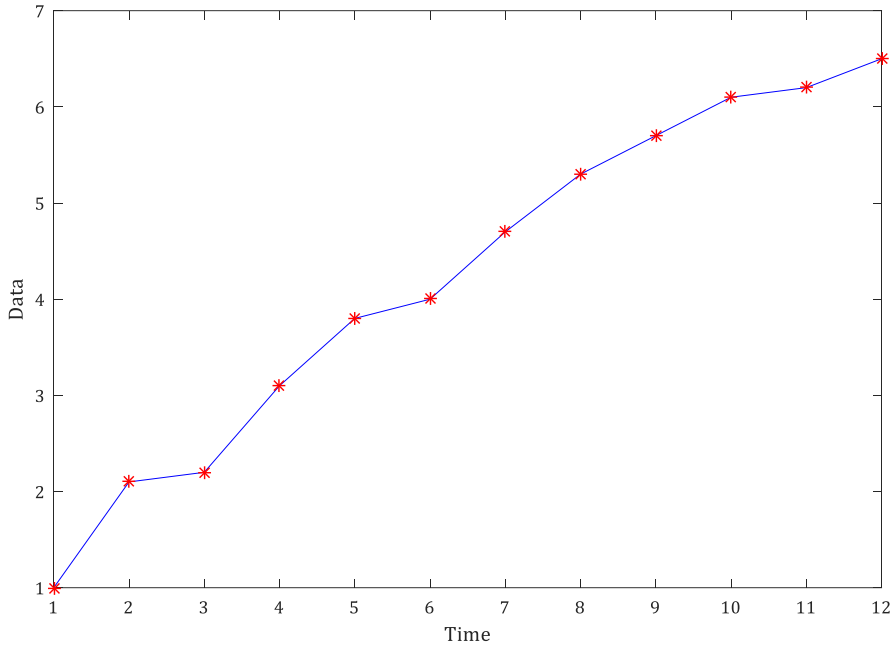

Fig.4.(a) Monotone increasing sequence

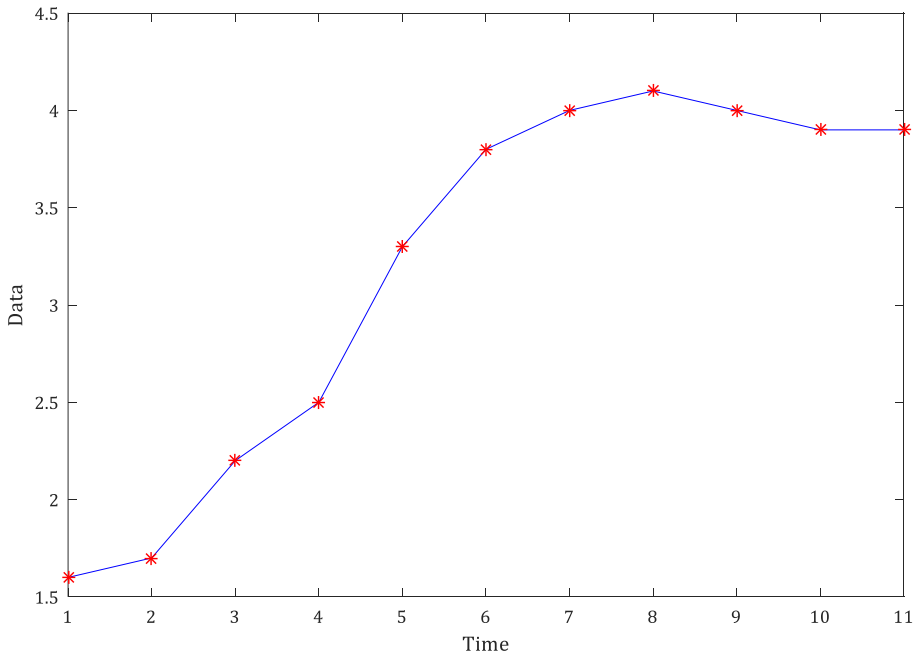

Fig.4.(b) Saturated **S** shape sequence

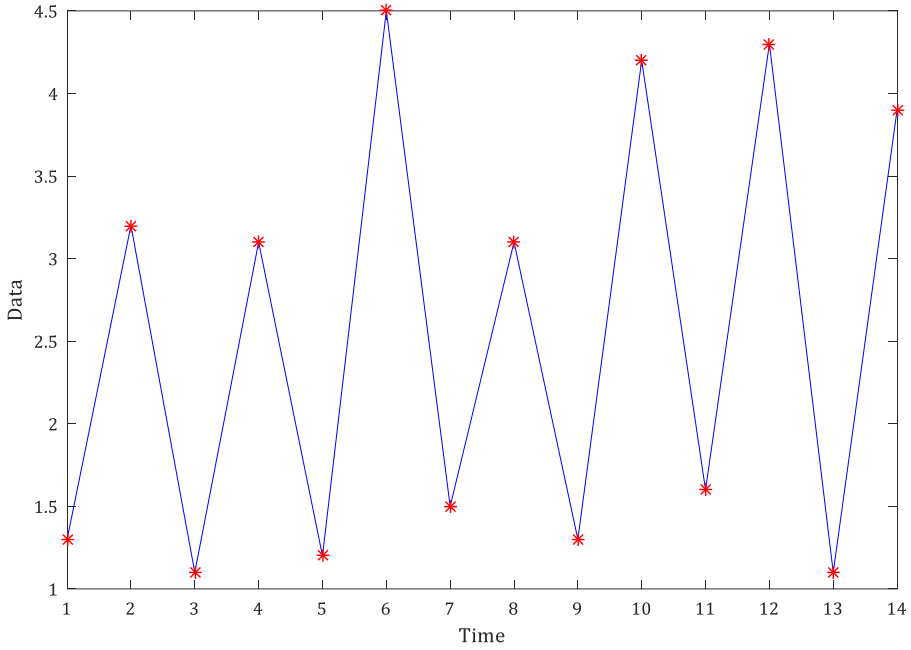

Fig.4.(c) Periodic fluctuation sequence

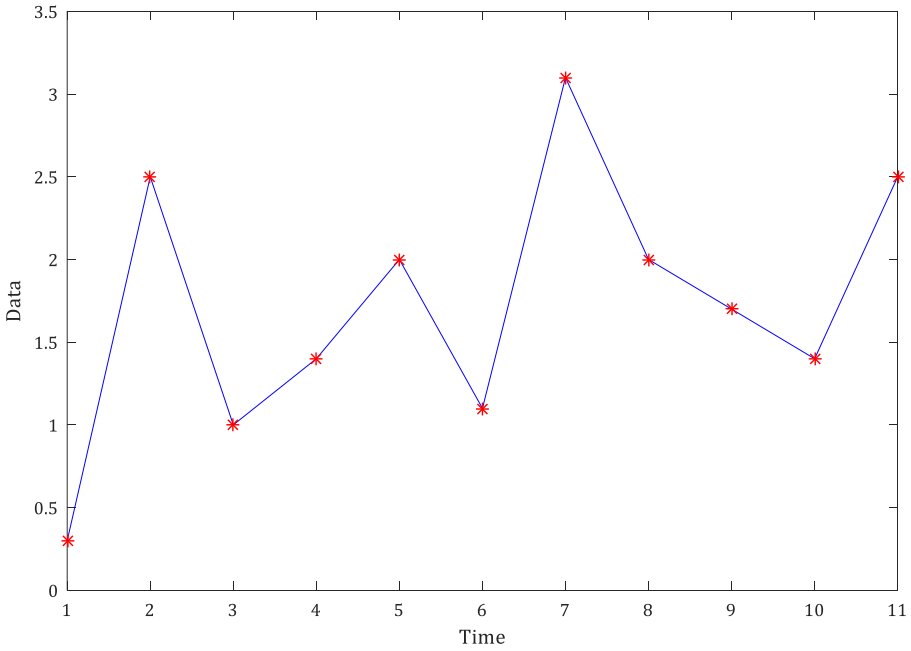

Fig.4.(d) Random oscillation sequence

Fig. 4 Four kinds of common sequences

In short, for different types of sequences, although the difficulty of modeling is different, suitable grey prediction models can always be found to simulate and predict the change trends of these sequences. It can be seen that the grey prediction model provides a set of

effective research means for the study of early warning of human health.

6. Economic Benefit Prospect

Chongqing is the youngest municipality directly under the jurisdiction of the central government of China and the largest industrial and commercial city in Western China. This paper takes Chongqing as an example to analyze the economic prospects of artificial intelligence early warning system. According to official statistics, the permanent population of Chongqing at the end of 2014 is 30.75 million. Considering that some citizens do not take part in physical examination, and some citizens take more than one physical examination every year, we take a median value, i.e. assuming that 15 million people take part in physical examination every year in Chongqing, the cost of each physical examination is mainly distributed in the range of 200-2000 RMB, whichever is the lower limit of 200 yuan, and the rate of intelligence is 1%. For the health analysis fee, the expected annual income is 15*200*1%= 30 million RMB. Considering other uncertain factors, the expected income will be halved, which is 15 million yuan per year. If the research results of this project can be extended to other provinces and regions in China, the benefits will be astonishing.

Acknowledgments

We thank my Doctoral tutor, Professor Liu Sifeng, for encouraging me to contribute to high-level journals. This work is supported by the National Natural Science Foundation of China under grant 71771033 and the joint rsearch grant of Royal Society and National Science Foundation of China IEC\NSFC\170391.

Competing interests

The authors declare no competing interests.

References

[1] Guanghua Bost Consumption Data Center. Big Data on Chinese Health [EB/OL]. http://www.sohu.com/a/252732622_499095 (2018-09-08).

[2] Alunni, V., du Jardin, P., Nogueira, L., Buchet, L., Quatrehomme, G. Comparing discriminant analysis and neural network for the determination of sex using femur head measurements. Forensic Science International. 253, 81-87 (2015).

[3] Zeng,B., Liu, SF. A self-adaptive intelligence gray prediction model with the optimal fractional order accumulating operator and its application. Mathematical Methods in the Applied Sciences. 23, 7843-7857 (2017).

[4] Aasim., Singh, SN., Mohapatra, Abheejeet. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. Renewable Energy. 136, 758-768 (2019).

[5] Wang, YF., Ma, ZM., Xu, CN., Wang, ZK., Yang, XH. Prediction of transfer among multiple states of blood pressure based on Markov model: an 18-year cohort study. Journal of hypertension. 36, 1506-1513 (2018).

[6] Neto, FDN., Baptista, CD., Campelo, CEC. Combining Markov model and Prediction by Partial Matching compression technique for route and destination prediction. Knowledge-based Systems. 154, 81-92 (2018).

[7] Wang, Q., Song, XX., Li, RR. A novel hybridization of nonlinear grey model and linear ARIMA residual correction for forecasting US shale oil production. Energy. 165, 1320-1331 (2018).

[8] Ma, Xin. A brief introduction to the Grey Machine Learning. Journal of Grey System. 31, 1-12 (2019).

[9] Wang, ZX., Li, Q. Modelling the nonlinear relationship between CO2 emissions and economic growth using a PSO algorithm-based grey Verhulst model. Journal of Cleaner Production. 207, 214-224 (2019).

[10] Wang, JZ., Du, P., Lu, HY., Yang, WD., Niu, T. An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting. Applied Soft Computing. 72, 321-337 (2018).

[11] Katayama, S. Computational and statistical analyses for robust non-convex sparse regularized regression problem. Journal of Statistical Planning and Inference. 201,20-31 (2019).

[12] Rothenhausler, D., Buhlmann, P., Meinshausen, N. Causal Dantzig: fast inference in linear structural equation models with hidden variables under additive interventions. Annals of Statistics. 47, 1688-1722 (2019).

[13] Gai, MJ., Cui, SW., Liang, S., Liu, XL. Frequency distributed model of Caputo derivatives and robust stability of a class of multi-variable fractional-order neural networks with uncertainties. Neurocomputing. 202, 91-97 (2016).

[14] Zeng, B., Li, C. Improved multi-variable grey forecasting model with a dynamic background value coefficient and its application. Computers & Industrial Engineering. 118, 278-290 (2018).

[15] Ma, X., Xie, M., Wu, WQ., Zeng, B., Wang, Y., Wu, XX. The novel fractional discrete multivariate grey system model and its applications. Applied Mathematical Modelling. 70, 402-424 (2019).

[16] Ding S. A novel discrete grey multivariable model and its application in forecasting the output value of China's high-tech industries. Computers & Industrial Engineering. 127, 749-760 (2019).

[17] Liu, SF., Lin, Y. An Introduction to Grey Systems. IIGSS Academic Publisher. (1998).

[18] Zeng, B., Chen, G., Liu, SF. A novel interval grey prediction model considering uncertain information. Journal of the Franklin Institute. 350, 3400-3416 (2013).

[19] Xiong, PP., Yin, Y., Shi, J., Gao, H. Nonlinear Multivariable GM (1, N) Model Based on Interval Gray Number Sequence. Journal of Grey System. 30, 33-47 (2018).

[20] Deng, JL. The Control problem of grey systems. System & Control Letters. 1,288-294 (1982).

[21] Liu, SF., Tao, LY., Xie, NM., Yang, YJ. On the New Model System and Framework of Grey System Theory. 28, 1-15 (2016).

[22] Liu, SF., Lin, Y. Grey Systems: Theory and Applications. Springer-Verlag Berlin Heidelberg. (2010).

[23] Liu, SF., Yang, YJ., Forrest, J. Grey Data Analysis Methods, Models and Applications. Springer-Verlag. (2016).

[24] Fu, Y., Yao, JJ., Zhao, HH., Zhao, G., Wan, ZS. Forecast for Artificial Muscle Tremor Behavior Based on Dynamic Additional Grey Catastrophe Prediction. Applied Sciences-basel. 8, 315 (Article Id), (2018).

[25] Bao, CZ., Mayila, M., Ye, ZH., Wang, JB., Jin, MJ., He, WJ., Chen, K. Forecasting and analyzing the disease

burden of aged population in China, based on the 2010 global burden of disease study. International Journal of Environmental Research and Public Health. 12, 7172-7184 (2015).

[26] Shen, XJ., Ou, LM., Chen, XJ., Zhang, X., Tan, XR. The Application of the Grey Disaster Model to Forecast Epidemic Peaks of Typhoid and Paratyphoid Fever in China. Plos One. 8, e60601(Article Id), (2013).

[27] Gan, RJ., Chen, XJ., Yan, Y., Huang, DZ. Application of a Hybrid Method Combining Grey Model and Back Propagation Artificial Neural Networks to Forecast Hepatitis B in China. Computational and Mathematical Methods in Medicine. 328273 (Article Id), (2015).

[28] Zhang, XN., Zhang, L., Zhang, YH., Liao, ZY., Song, JL. Predicting trend of early childhood caries in mainland China: a combined meta-analytic and mathematical modelling approach based on epidemiological surveys. Scientific Reports. 7, 6507(Article Id) (2017).

[29] Qu, HM., Bai, YN., Kui, FR., Hu, XB., Pei, HB., Ren, XW., Shen, XP. Effect of combination model on fitting cancer mortality and prediction. Chinese Journal of Epidemiology. 38, 117-120 (2017).

[30] Zhang, LP., Zheng, YL., Wang, K., Zhang, XL., Zheng, YJ. An optimized Nash nonlinear grey Bernoulli model based on particle swarm optimization and its application in prediction for the incidence of Hepatitis B in Xinjiang, China. Computers in Biology and Medicine. 49, 67-73 (2014).

[31] Han, X., Liu, YQ., Gao, H., Ma, JM., Mao, XX., Wang, YT., Ma, XD. Forecasting PM2.5 induced male lung cancer morbidity in China using satellite retrieved PM2.5 and spatial analysis. Science of the Total Environment. 607,1009-1017 (017).

[32] Chen, BX., Zhu, JX., Tan, XR. Grey Relation between Coronary Lesions and Serum Lipoid Substance. Journal of Grey System. 21, 351-356 (2009).

[33] Zhang, QY., Zheng, XY., Yi, DQ., Ji, WY., Chen, SR. Grey Relation between Glucose and Lipid Metabolism in Type 2 Diabetes Mellitus with Secondary Malignancy. Journal of Grey System. 22, 187-194 (2010).

[34] Zeng, B., Tan, YT., Xu, H., Quan, J., Wang, LY., Zhou, XY. Forecasting the Electricity Consumption of Commercial Sector in Hong Kong Using a Novel Grey Dynamic Prediction Model. Journal of Grey System. 30,157-172 (2018).

[35] Zeng, B., Duan, HM., Bai, Y., Meng, W. Forecasting the output of shale gas in China using an unbiased grey model and weakening buffer operator. Energy. 151, 238-249 (2018)

[36] Deng, J. Undulating grey model GM(1,1|tan(k-τ)p, sin(k-τ)p). Journal of Grey System. 13, 201-205 (2001).

[37] Zeng, B., Meng, W., Liu, SF. Research on prediction model of oscillatory sequence based on GM(1,1) and its application in electricity demand prediction. Journal of Grey System. 25, 31-40 (2013).

[38] Zeng, B., Meng, W. Interval Prediction Modeling Method of Small Sample Oscillation Sequence Based on Grey Theory. Control and Decision. 31, 1311-1316 (2016). In Chinese.