SCIENTIFIC REPORTS

OPEN

Predicting trend of early childhood caries in mainland China: a combined meta-analytic and mathematical modelling approach based on epidemiological surveys

Xiaonan Zhang^{1,2,3}, Lei Zhang^{4,5,6}, Yonghong Zhang⁷, Zhaoying Liao⁸ & Jinlin Song^{1,2,3}

Early childhood caries (ECC) is the most common chronic disease in young children. A reliable predictive model for ECC prevalence is needed in China as a decision supportive tool for planning health resources. In this study, we first established the autoregressive integrated moving average (ARIMA) model and grey predictive model (GM) based on the estimated national prevalence of ECC with meta-analysis from the published articles. The pooled data from 1988 to 2010 were used to establish the model, while the data from 2011 to 2013 were used to validate the models. The fitting and prediction accuracy of the two models were evaluated by mean absolute error (MAE) and mean absolute percentage error (MAPE). Then, we forecasted the annual prevalence from 2014 to 2018, which was 55.8%, 53.5%, 54.0%, 52.9%, 51.2% by ARIMA model and 52.8%, 52.0%, 51.2%, 50.4%, 49.6% by GM. The declining trend in ECC prevalence may be attributed to the socioeconomic developments and improved public health service in China. In conclusion, both ARIMA and GM models can be well applied to forecast and analyze the trend of ECC; the fitting and testing errors generated by the ARIMA model were lower than those obtained from GM.

The early childhood caries (ECC), the tooth decay occurred in any primary tooth in a child 71 months of age or younger¹, has been reported as the most prevalent infectious disease of children. The ECC prevalence in mainland China is comparatively high, 65.5% for 1–6-year-olds and 66.1% for 5-year-olds², far from the target by WHO “half of the 6-year-old children are caries-free”³. It has an adverse impact not only on children’s nutrition intake, speech, and daily routine activities, but also on their physiological health^{4,5}. China is a rapidly developing country with the largest children population in the world⁶. Children’s oral health has become a major public health problem in China. Hence, to better aid an explicit and quantitative direction for the future oral health plan among these population, a reliable prediction method to understand the trend of ECC is needed.

Forecasting techniques, which have been extensively applied to analyze the occurrences, development, and future trends of diseases, such as tuberculosis⁷, malaria⁸, hepatitis⁹, diabetics¹⁰ and influenza¹¹, serve as a policy-supportive tool effectively. The first step to establish the forecasting models was to acquire the time sequence¹². Traditionally, the data used for forecasting came from the regional reports or surveillance data. However, we can hardly get the series data of ECC prevalence by year. Since 1980’s, the Chinese Ministry of Health has invested large human and financial resources to conduct national oral epidemiology survey every decade. The first national survey was conducted in 1982, and the surveyed population were mainly students from primary and secondary schools. In the second and third national oral surveys in 1995 and 2005^{13,14}, 11 and 30

¹College of Stomatology, Chongqing Medical University, Chongqing, China. ²Chongqing key Laboratory of Oral Diseases and Biomedical Sciences, Chongqing, China. ³Chongqing Municipal Key Laboratory of Oral Biomedical Engineering of Higher Education, Chongqing, China. ⁴Melbourne Sexual Health Centre, Alfred Health, Melbourne, VIC, Australia. ⁵Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC, Australia. ⁶Research Center for Public Health, School of Medicine, Tsinghua University, Beijing, China. ⁷Medicine Engineering Research Center, College of Pharmacy, Chongqing Medical University, Chongqing, China. ⁸Children’s Hospital of Chongqing Medical University, Chongqing, China. Correspondence and requests for materials should be addressed to J.S. (email: soongjl@163.com)

provinces were covered respectively, and children aged 5 were chosen as a representative age group. The surveys focused on levels of caries, periodontal disease, mucosal disease and dental fluorosis. Since then, no national surveys on ECC have been performed in China. So far there is no publication devoted to the prediction of ECC, specifically on the national level.

To do the point prediction, the time series analysis is the most commonly used method in statistic. Autoregressive integrated moving average (ARIMA) model¹⁵ is one of the common means of the time series analysis with a complete theoretical basis, which can provide middle-long term forecast analysis. Comparing to other statistical models, the characteristic of the grey predictive model (GM) is outstanding^{16, 17}. It only needs a small sample to establish the model and to predict with a certain precision, which is especially applicable for the system with fuzzy structure or imperfect data. Therefore, these two models are applicable in this study for the prediction.

To establish the optimal model to predicate the trend of ECC in mainland China, we pooled data from existing reports with meta-analysis to calculate the national prevalence of ECC from 1988 to 2013. Then, we forecasted ECC prevalence in mainland China from 2014 to 2018 by the established ARIMA and GM models. The result was expected to provide quantitative basis for allocating medical resources to prevent and control ECC.

Materials and Methods

Data sources. Data used for establishing ARIMA and GM models came from the combined results of a meta-analysis, which was conducted according to the preferred reporting items for systematic review and meta-analyses (PRISMA) checklist. This approach has already been published in the previous literature². No ethical statement was necessary because all data were secondary summary data.

Peer-reviewed articles were searched in the following databases from the date of establishment to March, 2016: PubMed, Embase, Chinese Biomedical Literature database (CBM), Chinese National Knowledge Infrastructure database (CNKI), Chinese Wan Fang database, and Chongqing VIP database, using the key terms 'caries', 'prevalence', 'epidemiology', and 'China'. Two authors screened articles and extracted data independently. Any disagreement was resolved by consensus or the third author. A manual search was also applied to the relevant reference lists of all the eligible articles. Studies were included if they were cross-sectional surveys on ECC using random sampling, at city-level or above in mainland China (except for Hong Kong, Taiwan, and Macao). In order to exclude the effect of age structure, 5-year-olds were chosen as a representative age group. Additionally, studies were based on the general population rather than a specific group. The language of studies was limited to English and Chinese.

To reflect the temporal distribution of ECC, prevalence estimates for ECC in 5-year-olds in each survey year (1988–2013) were calculated by pooling the data from each study, with STATA software 11.1 (Stata, College Station, TX, USA). Statistical heterogeneity was detected by Q-test and I^2 -statistics. A random effects model was adopted in the case of significant heterogeneity ($I^2 > 50\%$ or $P < 0.1$). The quality of the selected studies was assessed using the Reporting of Observational Studies in Epidemiology (STROBE) guideline (Table S2)¹⁸. Potential publication bias was evaluated by funnel plots and Begg's test; $P \leq 0.05$ was considered to be significant.

Combined prevalence rates were divided into two parts to compare the fitting and prediction performances: data from 1988 to 2010 were used to construct the models, while data from 2011 to 2013 were used to test the prediction accuracy of the models. Chow breakpoint test¹⁹ was adopted to identify whether there had been a structural change around 2010; the result was regarded to be significant if $P \leq 0.05$.

ARIMA model construction. ARIMA is a traditional method to study the time series data. Since the sequence of ECC prevalence is a time series and generally have a trend, we chose ARIMA (p, d, q) model to fit it. The following parameters were selected when fitting the ARIMA model: p, the order of auto-regression; d, the degree of difference; q, the order of moving average^{12, 20}.

The sequence of prevalence usually had a trend which was non-stationary, thus augmented dickey-fuller unit root (ADF) test and KPSS test were chosen to test the stationary of the original sequence. If the sequence was non-stationary, differencing was used to transform it to stationary sequence. Under the circumstance, $d = 1$. And the tests were made on the differenced sequence to identify whether the trend still existed. If "yes", $d = 2$, and process went on until the sequence was stationary. Generally, when $d = 2$, the process could stop²¹.

When the differenced sequence was stationary, the variance and covariance of the sequence did not change over time. Then the autocorrelation function (ACF) graph and partial autocorrelation function (PACF) graph were used to identify the order of auto-regression (AR) and the order of moving average (MA) in the ARIMA model²². The model was fitted by the least squares method. The t-statistic was used to test the significance of the parameters and the F-statistic was used to test the significance of the equation. In addition, the Akaike Information Criterion (AIC) was certainly considered to be a comprehensive identifier of the parameters, and the R-squared (R^2) was an important index for model testing²³.

At last, the residual series would be analysed by the Ljung-Box Q-test²⁴ to verify whether it was a white noise time series or not. The white noise series would indicate that the information has been sufficiently extracted, allowing the model to conduct the predictive analysis. Otherwise, the order re-determination and parameter re-estimation were needed. We used the obtained model to forecast the prevalence of ECC from 2011 to 2018. The flow chart to construct ARIMA model was illustrated in Fig. 1.

GM (1,1) construction. If a system is fuzzy in hierarchy relationship, random in dynamic change, and uncertainty in indicator databases, the system is called grey system. The modeling for a grey system is grey model. GM (1,1) is the typical representative for the grey model, which can be used to fit and forecast in the complex system.

Firstly, an accumulative sequence had to be made on the original sequence. Then the new sequence was assumed to be adopted to the differential equation as follows:

Figure 1. Flow chart to construct the ARIMA model.

$$\frac{dy(t)}{dt} + \alpha y(t) = \mu \quad (1)$$

Finally, solving the equation, the GM(1,1) was constructed¹⁶. We also used the established grey model to forecast the future prevalence of ECC from 2011 to 2018.

Performance Statistics Index. The ARIMA model was created with EVIEWS 8 with a significant level of $P < 0.05$; the GM (1,1) was constructed with Matlab 7.0. In order to compare the performance, two statistics indexes were used to evaluate the fitting and prediction accuracy: the mean absolute error (MAE) and the mean absolute percentage error (MAPE). Their calculation formulas were as follows:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \quad (3)$$

where y_t and $\hat{y}_t$ denote the original and the predicted value respectively at time t . The smaller these two indexes were, the better the fitness and prediction performances.

Chow Breakpoint Test: 2010			
Null Hypothesis: No breaks at specified breakpoints			
Equation Sample: 1991 2013			
F-statistic	1.062434	Prob. F(4,15)	0.4090
Log likelihood ratio	5.737285	Prob. Chi-Square(4)	0.2196
WARNING: the MA backcasts differ for the original and test equation. Under the null hypothesis, the impact of this difference vanishes asymptotically.			

Figure 2. The result of chow breakpoint test.

Results

Results from meta-analyses. *Literature search and quality assessment.* A total of 11,776 publications were identified, and 78 eligible articles were included in the meta-analysis (Supplementary Fig. S1). The characteristics of the 78 articles were summarized in Supplementary Table S1. Quality assessment showed that all the studies scored at least 7 out of 10. Publication bias was statistically significant (Begg's test, $P < 0.001$).

Temporal trends in prevalence of ECC. The pooled prevalence of ECC among 5-year-olds was 66.1% (95% CI: 59.0–73.4%, ranging from 81.2% in 1988 to 56.1% in 2013. Figure 2 illustrated the trends in the prevalence at age 5 over time and a decreasing trend in ECC prevalence was observed during the study period.

The result of chow breakpoint test indicated that no structural change occurred in 2010 ($P > 5\%$, Fig. 2). In other words, there was no deviation between the fitting data and the forecasting data.

Simulation Results. *ARIMA.* The prevalence rates fluctuated between 50% and 80% with a downward trend (Fig. 3a). The result of ADF test ($P > 0.05$) indicated that the sequence was non-stationary. After 1-order differencing was used, the differenced sequence tended stationary (Fig. 3b). The result of ADF test ($P < 0.05$) also showed that the sequence was stationary. Then, the figures of ACF and PACF were used to identify the parameter p and q (Fig. 4).

The ARIMA (2,1,3) was chosen. The fitted result showed the significance test of regression was 0.000144, which meant the equation was significant. Moreover, the AIC information was about 6.28 and the R-squared was about 0.71 (Fig. 5). These meant that the effect of the model was good.

Thus, the ARIMA (2,1,3) model on the original sequence has been established.

GM (1,1). Firstly, we generated the accumulative sequence based on the original sequence, shown in Table 1. Then, using the method introduced above, the prediction model was as follows:

$$y(t) = -4988.19e^{-0.01587} + 5069.39 \quad (4)$$

Thus, the GM (1,1) model was established.

Comparison of the results from fitting and prediction. The established ARIMA (2,1,3) model was compared with GM (1,1) from two aspects of fitting and prediction. The comparison in fitting was shown in Table 2. While the average MAE and MAPE of GM (1,1) were 4.81% and 7.34%; the average MAE and MAPE of ARIMA (2,1,3) were 3.63% and 5.74%. Therefore, ARIMA was better than GM in fitting performance.

In addition, the two models were used to predict the prevalence from 2011 to 2013, also showing that the ARIMA was better than GM with lower MAE and MAPE (Table 3). The fitting and prediction curves of two models were compared with actual curve (Fig. 6), indicating that ARIMA (2,1,3) model was more accurate and stable than the GM (1,1).

Discussion

Oral health in children is an important public issue in China and worldwide. Understanding the temporal trend of ECC may facilitate the allocation of oral health resource. As far as we know, this is the first study to forecast the trend of ECC in mainland China based on the data from meta-analysis. It has broadened a new method for the forecasting field, which may provide the base data and theoretical support to establish and evaluate the prevention measures of ECC. The time series data of prevalence of ECC were forecasted by two prediction models, both of which fitted for the data and can be used in forecasting. According to the fitting and prediction accuracy, ARIMA model outperforms the GM (1,1).

This study demonstrates that the prevalence of ECC has declined over the past 30 years and will continue to decrease in the future. The reasons for this declining trend could be the socioeconomic developments and improved public health service in China recently. Since its reform and opening-up, China has experienced rapid socioeconomic changes. The average annual economic growth rate was as high as 9.8%, and per capita gross domestic product increased from 1,112 RMB in 1987 to 38,420 RMB in 2012²⁵. Governmental spending on public health care has grown greatly; the number of dentists has increased 13 times from 1985 to 2008^{26,27}; more and

Figure 3. (a) Temporal trend of early childhood caries prevalence in China during 1988–2010; (b) 1-order differencing of ECC prevalence.

Date: 05/02/17 Time: 22:12						
Sample: 1988 2018						
Included observations: 24						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1 -0.500	-0.500	6.7896	0.009	
		2 -0.166	-0.556	7.5741	0.023	
		3 0.389	-0.035	12.081	0.007	
		4 -0.217	-0.040	13.550	0.009	
		5 -0.136	-0.201	14.155	0.015	
		6 0.328	0.048	17.893	0.007	
		7 -0.251	-0.095	20.205	0.005	
		8 -0.041	-0.139	20.270	0.009	
		9 0.244	-0.031	22.742	0.007	
		10 -0.320	-0.310	27.292	0.002	
		11 0.253	0.090	30.369	0.001	
		12 -0.096	-0.223	30.853	0.002	

Figure 4. ACF, PACF and Q statistic of 1-order differencing sequence of ECC.

more oral health education programmes have been organized across the country²⁸; parental awareness toward oral health and children's oral habits have also been improved greatly according to the two national surveys^{13, 14}. Our observation is consistent with many researches which have reported the inverse relationship between

Dependent Variable: P1 Method: Least Squares Date: 01/23/16 Time: 00:11 Sample (adjusted): 1991 2010 Included observations: 20 after adjustments Convergence achieved after 30 iterations MA Backcast: 1988 1990				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.994961	0.173479	-5.735348	0.0000
AR(1)	-0.986263	0.107712	-9.156478	0.0000
AR(2)	-0.787943	0.098930	-7.964676	0.0000
MA(3)	-0.958830	0.074495	-12.87113	0.0000
R-squared	0.709887	Mean dependent var		-0.915000
Adjusted R-squared	0.655491	S.D. dependent var		8.732653
S.E. of regression	5.125621	Akaike info criterion		6.283237
Sum squared resid	420.3519	Schwarz criterion		6.482383
Log likelihood	-58.83237	Hannan-Quinn criter.		6.322112
F-statistic	13.05032	Durbin-Watson stat		1.416328
Prob(F-statistic)	0.000144			
Inverted AR Roots	-.49-.74i	-.49+.74i		
Inverted MA Roots	.99	-.49+.85i	-.49-.85i	

Figure 5. The result of AMIRA (2,1,3) model.

Year	Original Xt (%)	Accumulative Yt (%)
1988	81.2	81.2
1989	74.2	155.4
1990	73.4	228.8
1991	83	311.8
1992	76.2	388
1993	74.2	462.2
1994	82.2	544.4
1995	66.3	610.7
1996	68.1	678.8
1997	79.5	758.3
1998	71.1	829.4
1999	56.4	885.8
2000	69.4	955.2
2001	55.1	1010.3
2002	60.9	1071.2
2003	56.8	1128
2004	55.3	1183.3
2005	66.4	1249.7
2006	63.6	1313.3
2007	62.0	1375.3
2008	61.0	1436.3
2009	60.1	1496.4
2010	55.1	1551.5

Table 1. Result of accumulative sequence with GM.

socioeconomic status and caries prevalence^{29,30}. If effective interventions are implemented in the near future, the prevalence of ECC may continue to decrease.

The fourth national oral health survey is in progress (2015–2016), and the related data (ECC prevalence at age 5) will be available in the near future. In this study, we have adopted ARIMA (2,1,3) model and GM (1,1) to

Year	Actual (%)	ARIMA (2,1,3)			GM (1,1)		
		Fitted (%)	MAE (%)	MAPE (%)	Fitted (%)	MAE (%)	MAPE (%)
1988	81.2	NA	NA	NA	NA	NA	NA
1989	74.2	NA	NA	NA	78.5	4.3	5.8
1990	73.4	NA	NA	NA	77.3	3.9	5.3
1991	83.0	85.7	2.7	3.3	76.1	6.9	8.3
1992	76.2	71.3	4.9	6.4	74.9	1.3	1.7
1993	74.2	69.7	4.5	6.1	73.7	0.5	0.7
1994	82.2	81.4	0.8	1.0	72.6	9.6	11.7
1995	66.3	68.4	2.1	3.2	71.4	5.1	7.7
1996	68.1	68.6	0.5	0.7	70.3	2.2	3.2
1997	79.5	75.3	4.2	5.3	69.2	10.3	13
1998	71.1	66.1	5.0	7.0	68.1	3.0	4.2
1999	56.4	68.2	11.8	20.9	67.0	10.6	18.8
2000	69.4	70.7	1.3	1.9	66.0	3.4	4.9
2001	55.1	60.6	5.5	10.0	64.9	9.8	17.8
2002	60.9	67.5	6.6	10.8	63.9	3.0	4.9
2003	56.8	64.9	8.1	14.3	62.9	6.1	10.7
2004	55.3	58.8	3.5	6.3	61.9	6.6	11.9
2005	66.4	63.5	2.9	4.4	60.9	5.5	8.3
2006	63.6	61.7	1.9	3.0	60.0	3.6	5.7
2007	62.0	58.2	3.8	6.1	59.0	3.0	4.8
2008	61.0	60.3	0.7	1.1	58.1	2.9	4.8
2009	60.1	58.6	1.5	2.5	57.2	2.9	4.8
2010	55.1	55.4	0.3	0.5	56.3	1.2	2.2
Average			3.63	5.74		4.8	7.4

Table 2. Fitting results of two models.

Year	Actual value (%)	ARIMA (2,1,3)			GM (1,1)		
		Predicted value (%)	MAE (%)	MAPE (%)	Predicted value (%)	MAE (%)	MAPE (%)
2011	63.5	57.3	6.2	9.8	55.4	8.1	12.8
2012	56.1	55.0	1.1	2.0	54.5	1.6	2.9
2013	60.3	54.3	6.0	10.0	53.7	6.6	10.9
2014		55.8			52.8		
2015		53.5			52.0		
2016		54.0			51.2		
2017		52.9			50.4		
2018		51.2			49.6		

Table 3. Prediction results of two models.

forecast the ECC prevalence in 2014–2017, which is 53.5% and 52.0% in 2015, respectively. Accuracy of this prediction method based on data from meta-analyses will be further authenticated, compared with the results of the national survey. If the predicting result is close to the actual data, we can develop and popularize this method in the future research and application.

We have to mention several limitations in this study. First of all, the data used for forecasting were obtained from the pooled results of regional surveys. Error could not be avoided due to publication bias, sampling size and heterogeneity of the included articles³¹. Especially, pooling 1–6 years together included lots of variation and heterogeneities, and choosing year five was the appropriate time marker to demonstrate the epidemic trend for comparisons across studies. For the other age groups, it is not possible to set up the time series sequence due to limited literatures on other ages. Secondly, given the limited information of existing studies, we just conducted time series analysis without considering the risk factors which could affect the occurrence of ECC, such as medical expenditures, GDP, educational levels of the parents, and so on^{4, 32, 33}. Further research can improve the efficacy of these models and provide more clues to explain the variation of the prevalence. Thirdly, there exist obvious economic and population differences among kinds of provinces and cities in China. However, we could not obtain sufficient annual or monthly ECC prevalence from the current publications, to construct the series analysis stratified by geographical and economic differences in different Chinese regions. Fourthly, we have noticed the constrained forecasting to extrapolate in our study, that is, the longer the forecasting duration, the lower the model's

Figure 6. Two models' fitting and prediction curves and the actual data curve.

accuracy^{34,35}. Finally, whether this new prediction method is suitable for other epidemic diseases needs further validation.

In general, the oral health status of children in China has improved over time. We have developed a new prediction method based on data from meta-analyses. Both ARIMA model and GM can be used in fitting and forecasting the prevalence of ECC in mainland China. More precise prediction models may be needed to explain the variation of the ECC trend. We aim to promote general awareness in the local Chinese governments to establish the epidemiological database of ECC on a regional level. Then, we can take the economic or population difference among kinds of provinces and cities in the further research, if we have sufficient information. Developing and applying these prediction models could make us better understand the epidemiological characteristics of ECC and be helpful to prevent and control this disease.

References

1. American Academy of Pediatric, D., American Academy of, P., American Academy of Pediatric Dentistry Council on Clinical, A. Policy on early childhood caries (ECC): classifications, consequences, and preventive strategies. *Pediatr Dent* **27**, 31–33 (2005).
2. Zhang, X. *et al.* Prevalence and care index of early childhood caries in mainland China: evidence from epidemiological surveys during 1987–2013. *Sci Rep* **6**, 18897 (2016).
3. World Health Organization. Oral health global indicators for 2000. Geneva. *World Health Organization* (1988).
4. Petersen, P. E., Bourgeois, D., Ogawa, H., Estupinan-Day, S. & Ndiaye, C. The global burden of oral diseases and risks to oral health. *Bull World Health Organ* **83**, 661–669 (2005).
5. US Department of Health and Human Services. Oral health in America: a report of the Surgeon General. *J Calif Dent Assoc* **28**, 685–695 (2000).
6. National Bureau of Statistics. The 2010 statistical report on the national population. Available at http://www.stats.gov.cn/tjsj/tjgb/rkpcgb/qgrkpcgb/201104/t20110428_30327.html. (Accessed: 21st August 2016) (2011).
7. Moosazadeh, M., Khanjani, N., Nasehi, M. & Bahrampour, A. Predicting the incidence of smear positive tuberculosis cases in Iran using time series analysis. *Iran J Public Health* **44**, 1526–1534 (2015).
8. Medina, D. C., Findley, S. E., Boubacar, G. & Seydou, D. Forecasting non-stationary diarrhea, acute respiratory infection, and malaria time-series in Niono, Mali. *Plos One* **2** (2007).
9. Wei, W. *et al.* Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng county, China. *Plos One* **11**, e0156768 (2016).
10. Lai, Y. W., Toh, M. P. H. S. & Tham, L. W. C. Projection of prediabetes and diabetes population size in Singapore using a dynamic Markov model. *J Diabetes* (2016).
11. Soebiyanto, R. P., Farida, A. & Kiang, R. K. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. *Plos One* **5**, e9450 (2012).
12. Shumway, R. H. & Stoffer, D. S. Time series analysis and its applications [M]. *New York: Springer-Verlag*, 201–202; 289–290 (2000).
13. The National Committee for Oral Health. In *Second national epidemiological survey of oral health* (ed Qi, J.) 132–134 (People's Medical Publishing House, Beijing, 1998).
14. The National Committee for Oral Health. In *Third national epidemiological survey of oral health* (ed Qi, X. Q.) 60–61 (People's Medical Publishing House, Beijing, 2008).
15. Li, Q. *et al.* Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. *Am J Trop Med Hyg* **87**, 364–370 (2012).
16. Liu, D., Dang, Y. & Li, X. Improvement and application of GM(1,1) model. in *Grey Systems and Intelligent Services, 2009. GSIS 2009. IEEE Int Conference on* 344–346 (2009).
17. Dai, T. D. & Huang, X. M. Selection of discrete GM model initial value by designing calculation program. *J Grey System* **15** (2012).
18. VVon Elm, E. *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* **370**, 1453–1457 (2007).
19. Chen, L., Dolado, J. J. & Gonzalo, J. Detecting big structural breaks in large factor models. *J Econometrics* **180**(1), 30–48 (2014).
20. Box, G. E. P. Time series analysis: forecasting and control, 4th edition. *J Marketing Res* **14** (1994).
21. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J Econometrics* **54**, 159–178 (1992).
22. Im, K. S., Pesaran, M. H. & Shin, Y. Testing for unit roots in heterogeneous panels. *J Econometrics* **115**, 53–74 (1995).
23. Mahmood, M., Narges, K., Mahshid, N. & Abbas, B. Predicting the incidence of smear positive tuberculosis cases in Iran using time series analysis. *Iran J Public Health* **44**, 1526–1534 (2015).

24. Kitayama, K., Ohse, K., Shima, N., Kawatsu, K. & Tsukada, H. Regression model analysis of the decreasing trend of cesium-137 concentration in the atmosphere since the Fukushima accident. *J Environ Radioactivity* **164**, 151–157 (2016).
25. National Bureau of Statistics. Economic development of the reform and opening Glory New Chapter -Tremendous changes in China's economic and social development since 1978 in *People's Daily*. Available at: http://www.stats.gov.cn/tjgz/tjdt/201311/t20131106_456188.html. (Accessed: 21st August 2016) (2013).
26. Zhao, L. Y., Sun, C. & Sun, Z. Explore the new way of oral human resource management. *Chin J Hospital Administration* **25**, 512–514 (2010).
27. De-yu, X. & Hong. Oral health in China- trends and challenges. *Int J Oral Sci* **03**, 7–12 (2011).
28. Wang, H. Y., Petersen, P. E., Bian, J. Y. & Zhang, B. X. The second national survey of oral health status of children and adults in China. *Int Dent J* **52**, 283–290 (2002).
29. Sakeenabi, B., Swamy, H. S. & Mohammed, R. N. Association between obesity, dental caries and socioeconomic status in 6- and 13-year-old school children. *Oral Health Prev Dent* **10**, 231–241 (2012).
30. Turner, S. R., Seymour, R. W. & Dombroski, J. R. Socioeconomic status and selected behavioral determinants as risk factors for dental caries. *J Dent Edu* **65**, 1009–1016 (2001).
31. Coory, M. D. Comment on: Heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* **39**, 932–932 (2009).
32. Diehnelt, D. E. & Kiyak, H. A. Socioeconomic factors that affect international caries levels. *Commun Dent Oral Epidemiol* **29**, 226–233 (2001).
33. PhD, G. C. D. *et al.* Early childhood caries and associated determinants: a cross-sectional study on Italian preschool children. *J Public Health Dent* **74**, 147–152 (2014).
34. Alba, E. D. Constrained forecasting in autoregressive time series models: A Bayesian analysis. *Int J Forecasting* **9**, 95–108 (1993).
35. Zhou, L. *et al.* Using a hybrid model to forecast the prevalence of schistosomiasis in Humans. *Int J Environ Res & Public Health* **13**, 355 (2016).

Acknowledgements

This project was supported by “Program for Innovation Team Building at Institutions of Higher Education in Chongqing in 2013”, “Chongqing Municipal Key Laboratory of Oral Biomedical Engineering of Higher Education”, and Research project of Chongqing Municipal Planning Commission “The clinical effect and optimization of LIPUS combined with periodontal treatment”.

Author Contributions

X.N.Z., J.L.S., and L.Z. designed the study. X.N.Z. wrote the manuscript. X.N.Z. and Z.Y.L. collected the data. Y.H.Z. analysed the data. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-06626-w](https://doi.org/10.1038/s41598-017-06626-w)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017