Quantification of Shenzhen's Medical Standards

Qiong Chen

Department of Electrical Engineering, North China Electric Power University, Baoding 130600, China.

1946973833@qq.com

Abstract

In accordance with the development goals of the "Opinions" issued on August 18, 2019, the establishment of pilot demonstration zones, setting benchmarks for the well-being of the people and becoming a pioneer of sustainable development, is an important part of Shenzhen's urban development. In order to realize people's livelihood, it is necessary to establish a high-quality, balanced public service system and a comprehensive and sustainable social security system. Based on principal component analysis and grey prediction, compared with the US level, this article quantitatively describes the goals that Shenzhen's medical insurance needs to achieve in comparison with international standards.

Keywords

Target Quantification; Principal Component Analysis; Grey Forecast.

1. Introduction

The "Opinions" to build a pilot demonstration zone requires the establishment of a high-quality and balanced public service system and a sustainable social security system with full coverage. Shenzhen's urban resources should determine its quantitative targets in order to establish a sustainable society and medical care. And the old-age security system to meet the needs of rapidly developing and changing cities, thereby helping to achieve the development goal of building a national pilot demonstration zone.

2. Analysis of medical level

2.1 Model establishment

2.1.1 Principal component analysis

We collected the population of Shenzhen's residents in recent years x_1, the average monthly disposable income per capita x_2, medical and health services x_3, the number of hospitals x_4, the number of beds per thousand people x_5, and the number of staff per thousand people x_6. The number of technicians per thousand people x_7, the number of doctors per thousand people x_8, and the mortality rate x_9 data.

Steps for evaluation using principal component analysis [1]:

Step1: normalized raw data. Assuming that there are m index variables for principal component analysis, which are $x_1, x_2, ..., x_m$, there are a total of n evaluation objects (referring to the year), and the value of the j-th index in the i-th year is a_{ij}. Convert each index value a_{ij} into a standardized index value $\widetilde{a}_{ij}$, we can get

$$\widetilde{a}_{ij} = \frac{a_{ij} - u_j}{s_j} (i = 1,2,...,n; j = 1,2,...,n)$$

Where $u_j = \frac{1}{n}\sum_{i=1}^{n} a_{ij}, s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a_{ij} - u_j)^2}$, u_j, s_j are the sample mean and sample standard value of the j-th index.

Correspondingly, call $\widetilde{x}_j = \dfrac{x_j - u_j}{s_j}$ is a standardized indicator variable.

Step2: Calculate the correlation coefficient matrix R, the correlation coefficient matrix $R=(r_{ij})_{m\times n}$,where $rij = \dfrac{\sum_{k=1}^{n} \widetilde{a}_{ki}\cdot\widetilde{a}_{kj}}{n-1}, i,j=1,2,...,m$; $r_{ii}=1, r_{ij}=r_{ji}$, r_{ij} is the correlation coefficient between the i index and the j index.

Step3: Calculate eigenvalues and eigenvectors. Calculate the eigenvalues of the correlation matrix R $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{\mathrm{m}} \ge 0$ and the corresponding eigenvectors $u_1, u_2, ..., u_m$,where $u_j = [u_{1j}, u_{2j}, \cdots, u_{mj}]^T$, and m new index vectors are composed of eigenvectors:

$$F_1 = u_{11}\widetilde{x}_1 + u_{21}\widetilde{x}_2 + \cdots + u_{m1}\widetilde{x}_m$$
$$F_2 = u_{12}\widetilde{x}_1 + u_{22}\widetilde{x}_2 + \cdots + u_{m2}\widetilde{x}_m$$
$$\cdots$$
$$F_m = u_{1m}\widetilde{x}_1 + u_{2m}\widetilde{x}_2 + \cdots + u_{mm}\widetilde{x}_m$$

In the formula: F_1 is the first principal component, F_2 is the second principal component, F_m is the m-th principal component.

Step4: Select p(p ⩽m) principal components and calculate the comprehensive evaluation value.

(1) Calculate the information contribution rate and cumulative contribution rate of eigenvalue λ_j .

Where $b_j = \dfrac{\lambda_j}{\sum_{k=1}^{m}\lambda_k}$ is the information contribution rate of the main component F_j, and there $\alpha_p = \dfrac{\sum_{k=1}^{p}\lambda_k}{\sum_{k=1}^{m}\lambda_k}$ is the cumulative contribution rate of the main components F_1, F_2, ..., F_p. When α_p is close to 1, the first p index vectors F_1, F_2, ..., F_p are selected as p principal components instead of the original m index vectors, so that the p principal components are comprehensively analyzed.\

(2) Calculate the comprehensive score, $F = \sum_{j=1}^{p} b_j F_j$ .In the formula, b_j is the information contribution rate of the j-th principal component, which is evaluated according to the comprehensive score value.

2.1.2 GM (1,1) gray forecast

GM(1,1) means that the model is a first-order differential equation and contains only 1 variable gray prediction. The following are the steps of GM(1,1) gray prediction:

Step1: Known reference data column $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n))$, n is the number of data, 1 time Accumulation generates sequence.

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n)) = (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), ..., x^{(0)}(1) + \cdots + x^{(0)}(n))$$

Where $x^{(0)}$ is the original data set of F, $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k=1,2,...,n$ .The mean generating sequence of $x^{(1)}$ is $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), ..., z^{(1)}(n))$ .Where $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2,3,..,n$.

Step2: Establish grey differential equation.

$$x^{(0)}(k) + az^{(1)}(k) = b, k = 2,3,...,n$$

The corresponding white differential equation is $\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b$.

Step3: Use the posterior error test method to test the accuracy of the established model to ensure the feasibility of the result.

According to the $x^{(1)}$ sequence obtained according to the GM(1,1) modeling method, do a cumulative subtraction of $x^{(1)}$ and transform it into $\hat{x}^{(1)}$, $\hat{x}^{(0)} = (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \ldots, \hat{x}^{(0)}(n))$ .Calculate residual $e(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1,2,\ldots,n$ .The variances of the original sequence $x^{(0)}$ and the residual sequence E are S_1^2 and S_2^2:

$$S_1^2 = \frac{1}{n}\sum_{k=1}^{n}[x^{(0)}(k) - \bar{x}]^2,\quad S_2^2 = \frac{1}{n}\sum_{k=1}^{n}[e(k) - \bar{e}]^2$$

Where $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x^{(0)}(k), \bar{e} = \frac{1}{n}\sum_{k=1}^{n} e(k)$. Calculate the posterior difference ratio as $C = \frac{S_1}{S_2}$. Indexes C and p are two important indexes for the posterior error test. The smaller index C is, the better. The following is the evaluation table for the posterior error test.

Table 1 Precision inspection grade reference table

Model accuracy level	Mean square error ratio
1(great)	C≤0.35
2(qualified)	0.35<C≤0.5
3(inadequate)	0.5<C≤0.65
4(unqualified)	C>0.65

2.2 Model solution

2.2.1 Select principal components

According to the literature [2] [3] collated data as shown in the following table:

Table 2 2002-2019 Shenzhen Medical Data

Year	Population /ten thousand	Average monthly disposable income per person/yuan	Medical and health service expenses /10,000 yuan	Number of hospitals	Number of beds per thousand people	Number of staff per thousand people	Number of technicians per thousand people	Number of doctors per thousand people	mortality rate
2002	746.42	2078.39	105817	11808	1.66	3.09	2.49	1.11	0.1
2003	778.27	2161.32	127010	12607	1.75	3.36	2.73	1.21	0.17
2004	800.8	2299.7	115333	14186	1.88	3.57	2.86	1.29	0.21
2005	827.76	1791.2	137817.19	15577	2.03	3.81	3.1	1.4	0.64
2006	871.1	1880.59	164970.23	16193	2.02	6.13	4.97	2.02	0.32
2007	913.37	2025.12	223670.21	16766	1.98	6.49	5.14	2.06	0.51
2008	954.28	2227.44	281265.35	18435	2.09	6.65	5.3	2.11	0.43
2009	995.01	2437.04	306253.61	19872	2.15	6.74	5.4	2.05	0.31
2010	1037.2	2698.41	336483.23	21166	2.2	6.53	5.21	2.05	0.2
2011	1046.74	3042.09	389521.99	22322	2.3	6.88	5.55	2.16	0.32
2012	1054.74	3395.16	438821.44	26124	2.65	7.27	5.87	2.27	0.37
2013	1164.89	3721.09	578426.94	27079	2.75	7.72	6.19	2.39	0.47
2014	1077.89	3412.33	833176.81	28853	2.88	8.03	6.49	2.49	0.36
2015	1137.89	3719.44	1093365.34	35353	3.35	8.15	6.58	2.55	0.44
2016	1190.84	4057.92	1286646.44	38124	3.49	8.14	6.62	2.57	0.23
2017	1252.83	4411.5	1560246.74	39899	3.5	8.33	6.81	2.66	0.18
2018	1302.66	4795.20	1908983.08	43569	3.65	8.82	7.19	2.79	0.26
2019	1343.88	4912.57	3354800	47366	3.83	9.33	7.67	3.01	0.23

With the help of SPSS software, calculate the correlation coefficient matrix R and the eigenvalues of $x_{(1)}, x_{(2)}, \ldots, x_{(9)}$.

Correlation matrix										
		Number of people per 10,000	Average per capita monthly disposable	Medical and health expenses/ten	Number of hospital	Beds per thousand people	Number of staff per 1,000 population	Technical personnel per 1000 population	Doctors per thousand people	Mortality rate (1 in 100,000)
Correlation	Number of people per	1	0.944	0.863	0.971	0.953	0.937	0.945	0.951	0.093
	Average per capita monthly	0.944	1	0.877	0.969	0.96	0.826	0.839	0.843	0.227
	Medical and health	0.863	0.877	1	0.918	0.887	0.724	0.742	0.768	-0.225
	Number of hospital	0.971	0.969	0.918	1	0.994	0.871	0.884	0.896	-0.131
	Beds per thousand people	0.953	0.96	0.887	0.994	1	0.863	0.876	0.887	-0.085
	Number of staff per 1,000	0.937	0.826	0.724	0.871	0.863	1	0.999	0.997	0.082
	Technical personnel per	0.945	0.839	0.742	0.884	0.876	0.999	1	0.998	0.068
	Doctors per thousand people	0.951	0.843	0.768	0.896	0.887	0.997	0.998	1	0.071
	Mortality rate (1 in 100,000)	-0.093	-0.227	-0.225	-0.131	-0.085	0.082	0.068	0.071	1

Figure 1 Correlation matrix

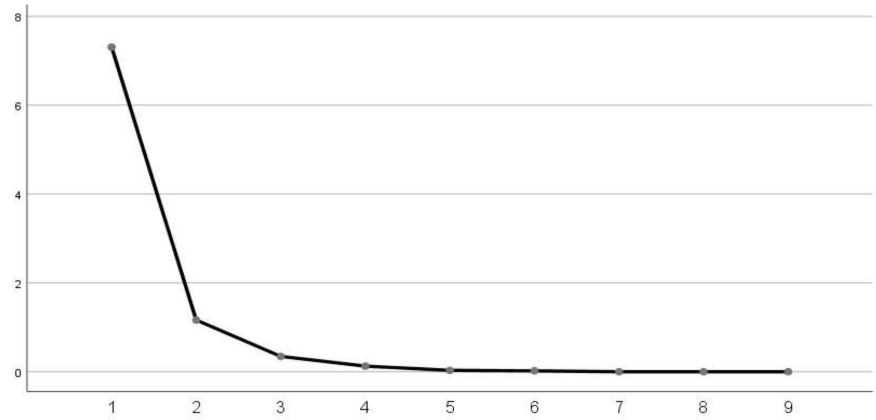


Figure 2 Eigenvalues

Calculate the cumulative contribution rate by eigenvalues.

Total variance interpretation						
composition	Initial eigenvalue			Extract the sum of the squares of the load		
	Total	Percentage of variance	Accumulation%	Total	Percentage of variance	Accumulation%
1	7.308	81.203	81.203	7.308	81.203	81.203
2	1.165	12.946	94.149	1.165	12.946	94.149
3	0.346	3.849	97.997			
4	0.127	1.408	99.405			
5	0.032	0.356	99.761			
6	0.02	0.226	99.987			
7	0.001	0.007	99.994			
8	0	0.005	99.999			
9	9.97E-05	0.001	100			

Figure 3 Cumulative contribution rate

From the total variance explanation graph and the gravel graph, it can be seen that the eigenvalues of the first two eigenvalues are greater than 1 and the cumulative contribution rate reaches more than 90%, and the effect of principal component analysis is very good. Below we select the first two principal components for comprehensive evaluation.

In order to obtain the quantitative relationship between the two principal components and each variable, we calculated the component matrix as shown in the figure:

Component matrix		
	composition	
	1	2
Number of people per 10,000	0.991	-0.007
Average per capita monthly disposable income/yuan	0.887	-0.188
Medical and health expenses/ten thousand yuan	0.887	-2.34
Number of hospital	0.983	-0.09
Beds per thousand people	0.972	-0.049
Number of staff per 1,000 population	0.943	0.227
Technical personnel per 1000 population	0.952	0.208
Doctors per thousand people	0.96	0.203
Mortality rate (1 in 100,000)	-0.081	0.963

Figure 4 Component matrix

Dividing the data in the component matrix by the corresponding characteristic root Kaiping of the principal component is convenient to get the coefficient corresponding to each index in the two principal components [4], the calculation result is:

$$F_1 = 0.367\tilde{x}_1 + 0.352\tilde{x}_2 + 0.328\tilde{x}_3 + 0.364\tilde{x}_4 + 0.36\tilde{x}_5 + 0.35\tilde{x}_6 + 0.352\tilde{x}_7 + 0.355\tilde{x}_8 - 0.03\tilde{x}_9$$

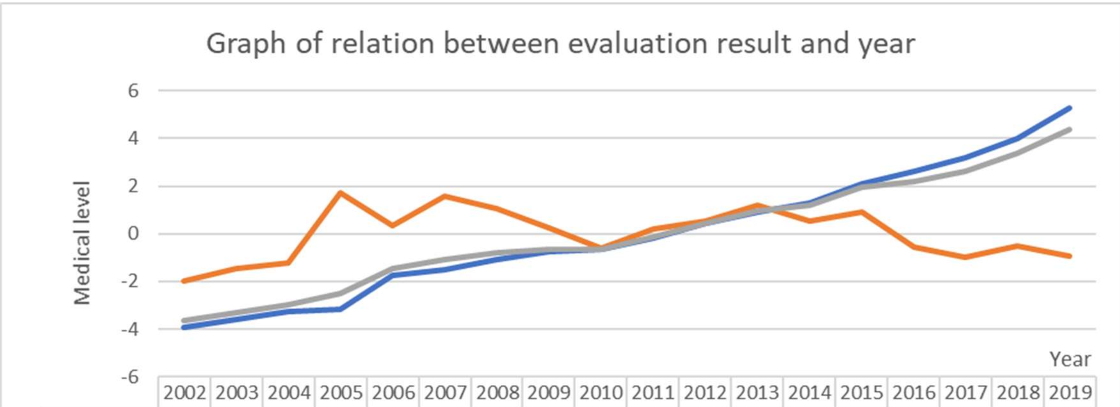
$$F_2 = -0.0065\tilde{x}_1 - 0.174\tilde{x}_2 - 0.2168\tilde{x}_3 - 0.083\tilde{x}_4 - 0.0454\tilde{x}_5 + 0.2103\tilde{x}_6 + 0.1927\tilde{x}_7 + 0.1881\tilde{x}_8 + 0.8922\tilde{x}_9$$

Where $\tilde{x}_i$ is the normalized value of the i-th parameter, and F_i is the expression of the i-th principal component. Take the variance contribution rate of each principal component to construct a comprehensive evaluation function as weight:

$$F = \frac{7.308}{8.473}F_1 + \frac{1.165}{8.473}F_2$$

$$= 0.361\tilde{x}_1 + 0.28\tilde{x}_2 + 0.253\tilde{x}_3 + 0.303\tilde{x}_4 + 0.304\tilde{x}_5 + 0.331\tilde{x}_6 + 0.33\tilde{x}_7 + 0.332\tilde{x}_8 + 0.079\tilde{x}_9$$

According to the expression of the evaluation function [5], the comprehensive evaluation results in recent years can be obtained as shown in the figure.



	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
系列1	-3.91	-3.58	-3.27	-3.18	-1.74	-1.5	-1.1	-0.77	-0.63	-0.2	0.453	0.928	1.28	2.118	2.643	3.204	4.005	5.284
系列2	-1.99	-1.48	-1.2	1.715	0.322	1.572	1.043	0.25	-0.6	0.198	0.514	1.18	0.543	0.902	-0.57	-0.97	-0.5	-0.94
系列3	-3.64	-3.29	-2.98	-2.51	-1.46	-1.07	-0.8	-0.63	-0.63	-0.15	0.461	0.962	1.179	1.951	2.201	2.631	3.385	4.398

Figure 5 Evaluation result graph

2.2.2 Make grey predictions

On the basis of principal component analysis, we get the comprehensive evaluation function F, which is regarded as the medical level value. In order to quantify the value of the objective evaluation function later, the gray GM (1,1) model is used to predict the value of F in the future.

Since the comprehensive evaluation value F has a negative value, it is impossible to directly make gray predictions. Therefore, we add a constant to each data based on the calculated F value, and then use the MATLAB program to gray the comprehensive evaluation value F Forecast, as shown:

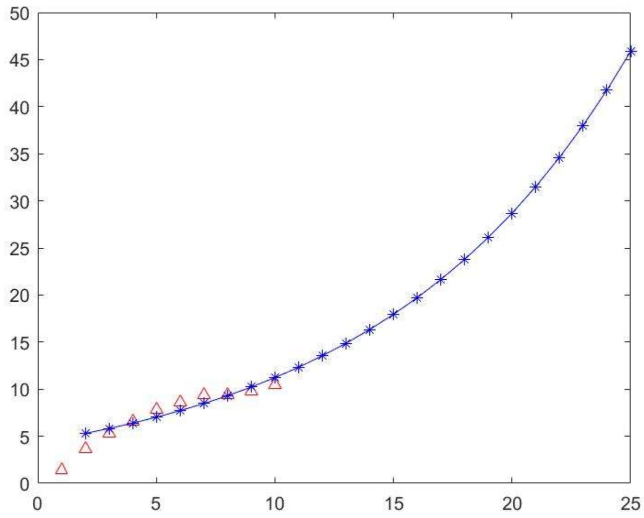


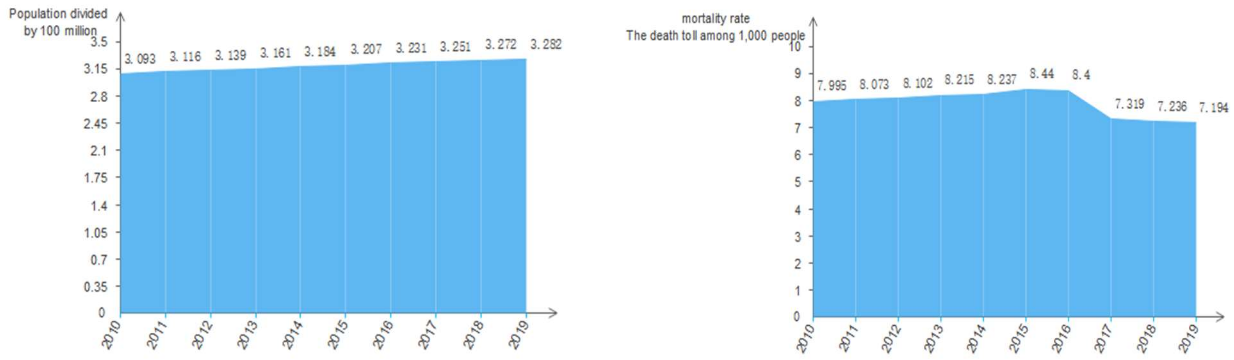
Figure 6 Forecast result

The posterior difference ratio: 0.20869; the system has good prediction accuracy.

The results show that the difference between the calculated value and the actual value obtained by the model is not too discrete. Based on the results, the future data is predicted, the fitting value after 5 years is 10.5191, the fitting value after 10 years is 20.0976, and the fitting value after 15 years is 35.9953. The results show that the overall result of this gray forecast is better. Not only the actual comprehensive evaluation value is calculated, but on this basis, the future development of the F value is predicted, and the result shows that it is similar to what we expected. With the growth of the year, the medical level value has been greatly improved.

3. American medical level

On the website [6], we found data on population and mortality in the United States in recent years, and integrated the data into the Figure 7:



(a) Population (b) Mortality rate

Figure 7 American Medical Data

Based on the same principal component analysis as above, it is $F'=\frac{\widetilde{x}_1}{0.131}\cdot 0.8625+\frac{\widetilde{x}_2}{0.9317}\cdot 0.1375$

It is calculated that the change trend of F′ with years in the United States in the past 10 years is:

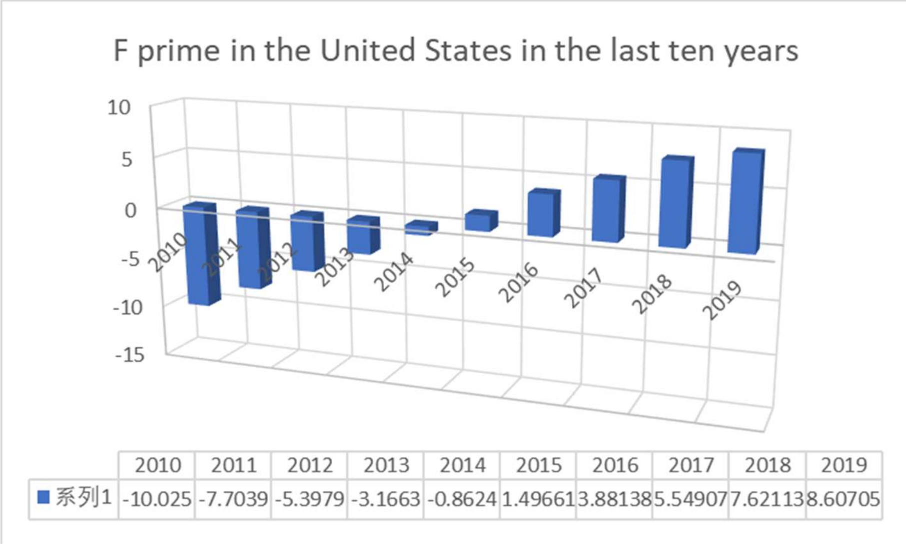


Figure 8 American medical standards

The current international value of F′ is 8.6070451.

4. Conclusions

In view of the F proposed in our model, Shenzhen’s existing medical comprehensive evaluation function value F is 4.397542054, and according to the existing US level value is 8.607045126. In contrast, it can be seen that Shenzhen’s medical treatment has a better value than the international value. There is a gap, and it will take 5 years to reach the current US level. And it is concluded that Shenzhen's F value after 5 years, 10 years, and 15 years is 10.5191, 20.0976, 35.9953, which can further rationally allocate medical resources and formulate matching medical insurance plans based on the data.

References

[1] Mathematical modeling algorithms and applications Si Shoukui, Sun Zhaoliang.

[2] The official website of the Shenzhen Municipal Health Commission. Shenzhen’s health policies, health data, health resources and other data, http://wjw.sz.gov.cn/

[3] Shenzhen Statistical Yearbook, compiled by Shenzhen Bureau of Statistics, National Bureau of Statistics Survey Team.

[4] Huang Zihong, Lin Qiu. Evaluation of urban medical level in Jiangsu Province based on principal component analysis [J]. Science of Normal University. Journal, 2019, 39(10): 33-36.

[5] Xie Ying, Zhao Chunxiang. Application of Grey Relational Analysis in the Evaluation of China's Medical Security System [J]. Chinese Health Statistics. Calculation, 2015, 32(06): 1012-1013 +1016.

[6] https://www.ceicdata.com/zh-hans/united-states/population-and urbanization-statistics/us-death -rate -crude-per-1000-people.