Prediction Analysis of the Prevalence of Alzheimer's Disease in China Based on Meta Analysis

Peilin Zang[1], Zhezhi Jin[2]*

[1]College of Pharmacy, Yanbian University, Yanji, China
[2]College of Economics and Management, Yanbian University, Yanji, China
Email: 1928535062@qq.com, *jinzhezhi@sina.com

Abstract

Objective: To predict the prevalence of Alzheimer's disease (AD) in China's future elderly population, and to provide reference for the formulation and implementation of relevant public health policies in China. **Methods:** A computer search of domestically published literatures with Alzheimer's disease risk factors data was used to conduct a meta-analysis of Alzheimer's disease epidemiological studies that met the inclusion criteria to calculate the Alzheimer's disease of the elderly in different periods over the past 30 years According to the results obtained from the previous Meta analysis, this paper established the ARIMA (1, 0, 1) model and the GM (1, 1) model, which better fit the previously merged prevalence sequence from 1990 to 2018. **Result:** According to the prediction results of the ARIMA (1, 0, 1) model, 2019 to 2023 Alzheimer's disease in China, the prevalence rates are 6.465%, 6.524%, 6.580%, 6.632%, and 6.7955%. **Conclusion:** Without corresponding effective preventive measures, the prevalence of Alzheimer's disease in China will show a continuous growth trend.

Subject Areas

Public Health

Keywords

*Corresponding author: Zhezhi Jin.

1. Introduction

Alzheimer's disease is a chronic disease that occurs in old age and is mainly manifested by progressive cognitive impairment and memory loss. The main clinical features of patients are memory, cognition, language, etc. This kind of physical function is impaired, and the illness lasts for about 3 to 15 years. It is characterized by onset of attack, high damage to the human body, and continuous decline of physical function that cannot be slowed [1]. Since the German psychiatrist ALOIS Alzheimer first discovered the disease and named it Alzheimer in 1906, NINCDS-ADRADA (National Association of Neurology and Traffic Diseases and Stroke and Alzheimer's disease and related Disease Association) Standard [2] defines Alzheimer's disease as a clinical entity and serves as the gold standard. With the expectation of human beings to prolong their lifespan, the aging of the population is showing a gradual upward trend. The current elderly population in the world is extremely increasing. By 2050, this proportion is expected to reach 22% [3]. As a country with a large population, China has more than 200 million elderly people. The aging of the population has made various diseases of the elderly become medically and socially important issues. Dementia in the elderly not only becomes a great problem in the daily life of the elderly; to a large extent, it has brought a huge burden to families and society, and has become a serious public health problem. However, for the chronic disease of Alzheimer's disease, there are currently no ideal clinical treatment drugs and treatments in medicine. Therefore, this article uses the method of Meta analysis to merge the statistics of Alzheimer's disease in mainland China from 1990 to 2018, and establishes the ARIMA model and GM (1, 1) model to predict the prevalence of Alzheimer's disease.

2. Materials and Methods

2.1. Literature Inclusion Criteria

1) Epidemiological survey conducted in mainland China (excluding Hong Kong, Taiwan, and Macao); 2) research method: cross-sectional study; 3) sampling method: random sampling; 4) research object: Chinese elderly People and aged 55 years and over; 5) the study was published from January 1990 to August 2018, and the investigation time was from 1985 to 2017; 6) the selected cases were screened first and then the second step diagnosis. Using the Simple Mental State Examination Scale (MMSE), Hasegawa Dementia Scale (HDS) and other detection methods to carry out the first screening of the surveyed population first, and then a neurologist with a professional physician. Using the ICD-10 Mental Disorder Diagnostic and Statistical Manual (DSM) and other mental disease diagnostic criteria to determine whether it is Alzheimer's disease patient.

2.2. Literature Exclusion Criteria

1) The research did not specify which sampling method was used; 2) The data in the materials were incomplete or did not provide information such as time

and location; 3) The duplicate literature or data were the same;

2.3. Quality Evaluation

Using the 2009 World Alzheimer's Disease Report (WAR) recommended literature quality scoring standard for quality evaluation [4]: 1) When the sample size is less than 500, that is 0.5 points; when the sample size is 500 - 1499, that is 1 point; when the sample size is 1500 - 2999, that is 1.5 points; when the sample size is greater than or equal to 3000, that is 2 points. 2) Research strategy: two-stage design without sampling, that is 0 points; two-stage design with sampling, no weighting, that is 1 point; one-stage or two-stage design with appropriate sampling and weighting, that is 2 points. 3) Diagnostic assessment involves multi-field cognitive tests, including routine disability assessment, information survey, etc. 4) When the response rate is less than 60%, score 1 point; when the response rate is 60% - 79%, score 2 points; when responding When the rate is greater than or equal to 80%, score 3 points.

2.4. Literature Inclusion Results

A preliminary search obtained 12,582 articles, including 317 Pubmed, 231 Embase, 16 The Cochrane Library, 3445 CBM, 4450 CNKI, 3280 Wanfang database, and 843 VIP database. Excluding 4437 articles that were repeatedly published. Then read the abstract and the title to obtain 678 articles that may meet the standards. Finally, through the full text, according to the inclusion and exclusion criteria of the literature, 568 non-compliant documents are excluded, and finally 89 articles are included. It is a cross-sectional study, including 77 Chinese literature and 12 English literature.

2.5. Basic Features of Included Literature

In the included literature, including 13 provincial surveys and 78 municipal surveys, involving 21 provinces, 3 autonomous regions and 4 municipalities in China, 89 articles can obtain the prevalence of Alzheimer's disease According to the data, the prevalence rate ranges from 1.031% in Hainan to 7.81% in Hebei. 89 articles can obtain the specific age of the patient. The minimum age of the patient is 50 years old. All the surveys were conducted in urban communities or rural villages. Among them, 15 One study was conducted from 1990 to 2000, 33 studies were conducted from 2000 to 2009, and 43 studies were conducted from 2010 to 2019. Among them, there were 13,425 patients with Alzheimer's disease, 4569 male patients, and 8856 female patients. The quality evaluation was carried out according to the document quality scoring standard recommended by (WAR), and all the documents were above 6 points, of which 9 points includes 35 articles, 8 points includes 31 articles, 7 points includes 16 articles, 6 points includes 7 articles.

3. Prediction of Alzheimer's Disease in Mainland China

Alzheimer's disease is a common chronic disease of the elderly, and its impact

on the health and life of the elderly cannot be ignored. The International Alzheimer's Disease Society (ADI) pointed out: AD has become a heavy burden and a strategic issue for society [5]. In the prevention and control of diseases, prediction research is a very important topic. Common prediction models include ARIMA (Auto regressive integrated moving average model) model [6], gray model [7], exponential model, etc. The above models use different prediction range. If different models need to be selected for the same disease research materials, this study obtained epidemiological research on the research data of Alzheimer's disease in mainland China from 1990 to 2018, and obtained the past 30 years. Time development trend of prevalence rate, establish ARIMA model according to the temporal distribution characteristics of prevalence rate, predict the development trend of Alzheimer's disease, and evaluate and verify the model to scientifically predict the incidence of Alzheimer's disease. The morbidity provides a theoretical basis.

3.1. Model Establishment

Obtain the published epidemiological survey data on Alzheimer's disease in China, perform meta-analysis on the data, use Revman 5.3 software to merge the prevalence from 1990 to 2018, consider the publication cycle of the literature, and the data in 2019. It is not representative and not included in the final analysis. The database of the combined prevalence will be established in Excel 2007, and the ARIMA model and GM (1, 1) model will be established according to the changing rules of the data Forecast, and evaluate the fit and error of the two models, and compare the advantages and disadvantages of the two prediction methods.

3.2. ARIMA Model

3.2.1. The Establishment Result of ARIMA Model

This paper collates the prevalence data of Alzheimer's disease from 1990 to 2018. Here, the variable is defined as X, and X is used to indicate the prevalence. The following article will use the data from 1990 to 2017 to fit the ARIMA model, and then use the 2018 data to test the degree of model fit. When the model test is passed, it is used to predict the prevalence from 2019 to 2023.

1) Original data X stationarity test

i) Timing diagram: from the timing diagram, it can be seen that X as a whole has a certain upward trend. Among them, the timing diagram is shown in Figure 1.

ii) Unit root test: ADF unit root test is further adopted to verify its stability. The unit root test is shown in Table 1.

From the results of the unit root test on the original sequence, we can see that at the significance level of 0.05, the p value is less than 0.05, so the original hypothesis that there is a unit root is rejected, indicating that the original sequence is stationary. Then you can use the stationary sequence X to build the ARIMA model.
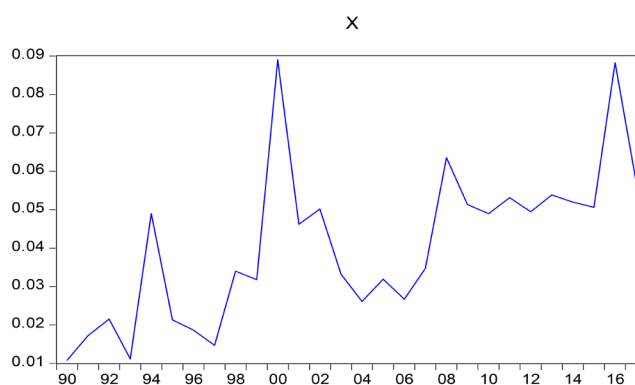
Figure 1. Timing diagram.

Table 1. Unit root test results.

Null Hypothesis: X has a unit root

Exogenous: Constant, Linear Trend

Lag Length: 0 (Automatic—based on SIC, maxlag = 6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		−4.565064	0.0060
Test critical values:	1% level	−4.339330	
	5% level	−3.587527	
	10% level	−3.229230	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(X)

Method: Least Squares

Date: 03/15/20   Time: 16:50

Sample (adjusted): 1991 2017

Included observations: 27 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X (−1)	−0.927363	0.203143	−4.565064	0.0001
C	0.018335	0.007188	2.550742	0.0175
@TREND(1990)	0.001461	0.000527	2.771666	0.0106
R-squared	0.465720	Mean dependent var		0.001715
Adjusted R-squared	0.421197	S.D. dependent var		0.021214
S.E. of regression	0.016140	Akaike info criterion		−5.310618
Sum squared resid	0.006252	Schwarz criterion		−5.166636
Log likelihood	74.69334	Hannan-Quinn criter.		−5.267805
F-statistic	10.46014	Durbin-Watson stat		2.011356
Prob (F-statistic)	0.000541			

2) Model identification and ordering

X autocorrelation graph: from the autocorrelation function graph and partial autocorrelation function graph of X, we can see that the partial autocorrelation coefficient quickly falls within the double standard deviation after one period, which can be regarded as the first-order truncation. Then p is 1, that is, AR (1); the autocorrelation coefficient also enters within twice the standard deviation after one period. If it is regarded as slow entry and tailing, then the AR (1) model is established; at the end, q takes 1, which is MA (1). Therefore, AR (1) and ARMA (1, 1) models can be established for comparative verification. Figure 2 is the X autocorrelation diagram.

Table 2 shows the AR (1) model which is ARIMA (1, 0, 0) model.

Table 3 shows the ARMA (1, 1) model is the ARIMA (1, 0, 1) model.

According to the fitting results of the above two models, when the significance level is 0.05, both model coefficients pass the test. Next, compare the models. The residual variance of the ARMA (1, 1) model is smaller than the AR (1) model, and in the estimation, the R^2 of the ARMA (1, 1) model is greater than the AR (1) model, and the overall fit The degree is better. Because we want to make predictions in this article, it is required that the model fits well, otherwise the predictions do not make much sense. Therefore, here we choose the second model ARMA (1, 1) model, namely ARIMA (1, 0, 1) model.

3) Residual adaptability test

The function values of the partial autocorrelation graph of the residual autocorrelation, as well as the Q-stat and P values show that the residual sequence has no autocorrelation and is white noise, so the model is suitable. Figure 3 is the adaptability test chart of residuals. Figure 4 is model fitting diagram.

Table 2. AR (1) model.

Dependent Variable: X			
Method: Least Squares			
Date: 03/15/20 Time: 17:03			
Sample (adjusted): 1991 2017			
Included observations: 27 after adjustments			
Convergence achieved after 3 iterations			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
---	---	---	---	---
C	0.043050	0.006326	6.804987	0.0000
AR (1)	0.440952	0.172971	2.549285	0.0173
R-squared	0.206320	Mean dependent var		0.041698
Adjusted R-squared	0.174573	S.D. dependent var		0.019998
S.E. of regression	0.018169	Akaike info criterion		−5.106993
Sum squared resid	0.008253	Schwarz criterion		−5.011005
Log likelihood	70.94440	Hannan-Quinn criter.		−5.078451
F-statistic	6.498854	Durbin-Watson stat		2.281929
Prob (F-statistic)	0.017309			
Inverted AR Roots	0.44			

Table 3. ARMA (1, 1) model.

Dependent Variable: X

Method: Least Squares

Date: 03/15/20 Time: 17:05

Sample (adjusted): 1991 2017

Included observations: 27 after adjustments

Convergence achieved after 68 iterations

MA Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.073110	0.040761	1.793613	0.0855
AR (1)	0.929431	0.051595	18.01387	0.0000
MA (1)	−1.548382	0.275332	−5.623700	0.0000
R-squared	0.699211	Mean dependent var		0.041698
Adjusted R-squared	0.674145	S.D. dependent var		0.019998
S.E. of regression	0.011416	Akaike info criterion		−6.003189
Sum squared resid	0.003128	Schwarz criterion		−5.859208
Log likelihood	84.04306	Hannan-Quinn criter.		−5.960376
F-statistic	27.89506	Durbin-Watson stat		2.056898
Prob (F-statistic)	0.000001			
Inverted AR Roots	0.93			
Inverted MA Roots	1.55			

Estimated MA process is noninvertible

Sample: 1990 2017
Included observations: 28

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.428	0.428	5.7071	0.017
		2	0.328	0.177	9.1802	0.010
		3	0.109	-0.104	9.5811	0.022
		4	0.004	-0.079	9.5817	0.048
		5	0.057	0.115	9.7018	0.084
		6	0.183	0.212	10.983	0.089
		7	0.044	-0.161	11.062	0.136
		8	0.194	0.150	12.641	0.125
		9	-0.032	-0.155	12.687	0.177
		10	-0.162	-0.212	13.907	0.177
		11	-0.029	0.200	13.949	0.236
		12	-0.073	-0.022	14.229	0.286

Figure 2. X autocorrelation diagram.

Sample: 1990 2017
Included observations: 27

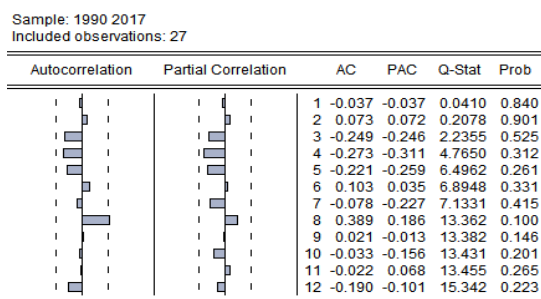| Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.037	-0.037	0.0410	0.840
		2	0.073	0.072	0.2078	0.901
		3	-0.249	-0.246	2.2355	0.525
		4	-0.273	-0.311	4.7650	0.312
		5	-0.221	-0.259	6.4962	0.261
		6	0.103	0.035	6.8948	0.331
		7	-0.078	-0.227	7.1331	0.415
		8	0.389	0.186	13.362	0.100
		9	0.021	-0.013	13.382	0.146
		10	-0.033	-0.156	13.431	0.201
		11	-0.022	0.068	13.455	0.265
		12	-0.190	-0.101	15.342	0.223

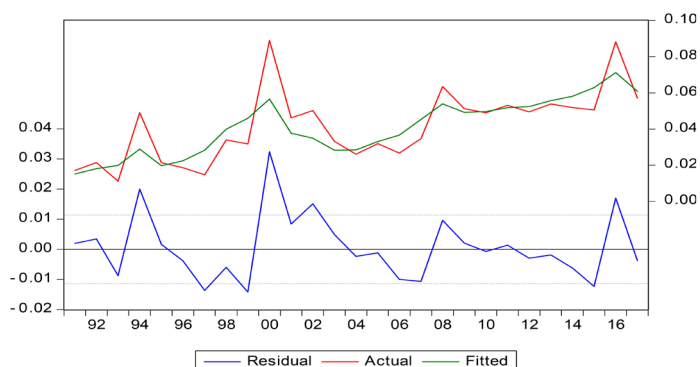**Figure 3.** Residual adaptability test chart.

Figure 4. Model fitting diagram.

3.2.2. Analysis Results of the Model

Static prediction:

First, use static prediction to make a step forward prediction, which is used to compare the static prediction value of 2018 with the original value of the original sequence to test the degree of model fit. The sequence diagram of the group obtained by putting the sequence X and the static prediction sequence XF here in the form of a group is shown in the figure below. It can be seen that the static predicted sequence values in the sample are almost consistent with the original sequence X in trend. From the data point of view, the static prediction value of the prevalence in 2018 is 6.4006%, and the prevalence in 2018 in the original sequence is 5.9360%, which is a slightly higher prediction, but within the acceptable range, and the same increase as the original sequence trend. It shows that the static predicted value and the model estimated value in the sample period are consistent. Among them, the static prediction chart is shown in **Figure 5**, and the group sequence chart is shown in **Figure 6**.

On the whole, the prediction sequence and the original sequence are basically consistent in the change trend. It can be considered that the model fitting is effective, and the prediction result has a certain reference value.

3.3. GM (1, 1) Model Establishment Result

First define the variable as X, and then use X to represent the prevalence rate. The following article will use the data from 1990 to 2017 to fit the GM (1, 1) model, and then use the 2018 data to test the degree of model fit. When the test passes, it is used to predict the prevalence rate from 2019 to 2023. To model the prevalence rate of Alzheimer's disease from 1990 to 2018, first generate a cumulative sequence, the data is shown in **Table 4**.

Using the GM (1, 1) model method for calculation, the following prediction equation can be obtained:

$$Y(t) = -66.742 \times e^{0.0359t} + 67.825$$

Table 5 shows the fitting, testing and prediction results of the GM (1, 1) model. So far we have established the GM (1, 1) model.

Table 4. Generation of diseased cumulative series.

years	Prevalence rate (%)	Cumulative sequence
1990	1.083	1.083
1991	1.72	2.803
1992	2.154	4.957
1993	1.115	6.072
1994	4.9	10.972
1995	2.134	13.106
1996	1.866	14.972
1997	1.47	16.442
1998	3.4	19.842
1999	3.18	23.022
2000	8.9	31.922
2001	4.621	36.543
2002	5.02	41.563
2003	3.322	44.885
2004	2.61	47.495
2005	3.19	50.685
2006	2.67	53.355
2007	3.47	56.825
2008	6.355	63.18
2009	5.136	68.316
2010	4.896	73.212
2011	5.315	78.527
2012	4.95	83.477
2013	5.388	88.865
2014	5.196	94.061
2015	5.063	99.124
2016	8.83	107.954
2017	5.7125	113.6665
2018	5.936	119.6025

Table 5. GM (1, 1) fitted value, test value, predicted value.

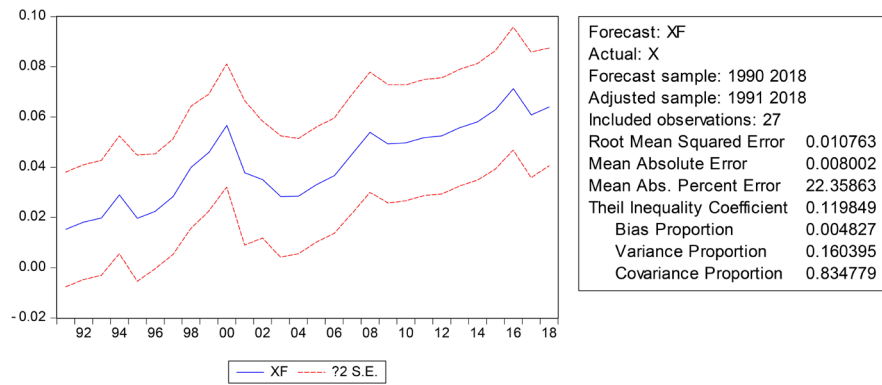| years	Actual value	GM (1, 1)		
		fitted value %	error	error rate %
1990	1.083	NA	0.0005	NA
1991	1.72	2.515705	0.7957	46%
1992	2.154	2.607549	0.4535	21%
1993	1.115	2.702746	1.5877	142%
1994	4.9	2.801419	2.0986	43%
1995	2.134	2.903693	0.7697	36%
1996	1.866	3.009702	1.1437	61%
1997	1.47	3.119581	1.6496	112%
1998	3.4	3.233471	0.1665	5%
1999	3.18	3.35152	0.1715	5%
2000	8.9	3.473878	5.4261	61%
2001	4.621	3.600703	1.0203	22%
2002	5.02	3.732158	1.2878	26%
2003	3.322	3.868413	0.5464	16%
2004	2.61	4.009642	1.3996	54%
2005	3.19	4.156026	0.966	30%
2006	2.67	4.307756	1.6378	61%
2007	3.47	4.465024	0.995	29%
2008	6.355	4.628034	1.727	27%
2009	5.136	4.796996	0.339	7%
2010	4.896	4.972125	0.0761	2%
2011	5.315	5.153649	0.1614	3%
2012	4.95	5.3418	0.3918	8%
2013	5.388	5.536819	0.1488	3%
2014	5.196	5.738959	0.543	10%
2015	5.063	5.948478	0.8855	17%
2016	8.83	6.165646	2.6644	30%
2017	5.7125	6.390743	0.6782	12%
2018	5.936	6.624058	0.6881	12%
2019		6.865890		
2020		7.1166		
2021		7.3764		
2022		7.6457		
2023		7.9248		

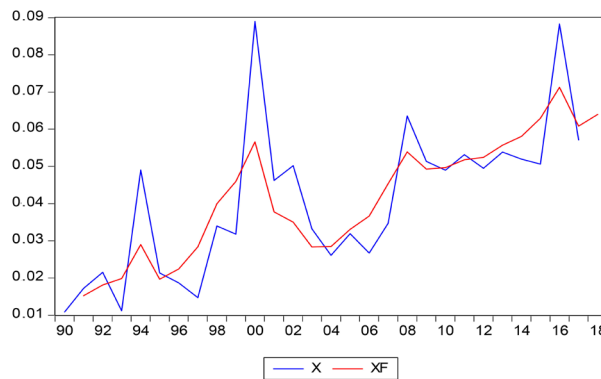**Figure 5.** The static prediction chart.

Figure 6. The group sequence chart.

3.4. Comparison and Evaluation of Models

Finally, the 2018 ARIMA (1, 0, 1) model will be compared with the GM (1, 1) model prevalence fit and prediction. The average error of ARIMA is 0.464, the average error rate is 7.8%, and the GM (1, 1). The average error of the model is 0.6881 and the average error rate is 12%, so the fitting effect of the ARIMA model is more accurate than the GM model. Also the fitting prediction effect of the ARIMA model is more than the GM model is closer to the actual value.

4. Discussion

For the prevention and treatment of Alzheimer's disease, more urgent, accurate and immediate prediction of the prevalence can provide health policymakers with a better understanding of its incidence. With the rapid development of science and technology, more and more models and prediction methods have been discovered. Based on the strong demand for disease prediction in the field of epidemiology, many researchers seek better and better methods for epidemic disease. For prediction and analysis, this article uses the statistical method of Meta analysis to combine the results of the survey and research in various regions of China to obtain the prevalence of Alzheimer's disease in each year in the mainland of China, and then to construct the prevalence prediction model. The innovation of this article is to combine Meta analysis with mathematical models

and apply it to the field of prevalence prediction, reducing the financial and human pressure brought by large-scale epidemiological investigations and establishing the prevalence time. The sequence model predicts Alzheimer's disease, a common disease of the elderly, and evaluates the accuracy of the prediction.

The ARIMA model is a model with typical representative meaning in time series analysis, but the ARIMA model also has certain shortcomings. Its shortcoming is that it requires a high number, and it is very inconvenient to process some sequences that have been extremely missing. Second, because the ARIMA model is built on the basis of the autocorrelation of the sequence, if its autocorrelation is low, it will make the model impossible to build. This study uses the ARIMA model to predict the prevalence of Alzheimer's disease, because there is a certain correlation in the incidence of each year, and the ARIMA model can capture its advantages in this area, and then more accurately predict the prevalence of Alzheimer's disease in the next few years. On the other hand, in the process of model construction, in the selection and selection of parameters, Jin Guo repeatedly hypothesized tests, repeatedly verified and corrected to obtain a statistically reasonable model.

We use the ARIMA (1, 0, 1) model to predict the prevalence of Alzheimer's disease. The static prediction of the prevalence in 2018 is 6.4006%. The predicted value is higher than the real value, there is a certain error, but the error is within an acceptable range. It shows that the static predicted value and the model estimated value in the sample period are consistent. It also predicts the prevalence rate from 2019 to 2023, providing a theoretical basis for the prevention and treatment of Alzheimer's disease.

It is worth noting that a mathematical model was established in this study to predict the prevalence of Alzheimer's disease, but some of its limitations cannot be ignored:

1) Because the source of the prevalence rate is Meta analysis, it is the result of statistically combining the prevalence rates of different provinces, municipalities, and autonomous regions. Due to different sample sizes, the obtained regions are different. The data itself has a certain degree of error.

2) In the process of predicting the prevalence in the next 5 years, only the prevalence data is used to model the autoregressive model, and other possible factors are not considered. Therefore, although the model shows a high accuracy in the prediction process, it cannot provide medical workers with the reasons for the increase and decrease in the prevalence rate. However, in the third chapter of the study, we have already diagnosed Alzheimer's disease. The risk factors have been analyzed, and future researchers can improve the model in this direction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Wang, J. (2011) Prevalence and Preventive Measures of Alzheimer's Disease in China. *Asia-Pacific Traditional Medicine*, No. 2, 55-57.

[2] Fang, M., Zhou, L. and Liu, X. (2012) Advances in Imaging Research of Alzeimer Disease in International Medicine. *Journal of Radiology*, No. 1, 113-118.

[3] Sevcikova, H. (2017) wpp2017: World Population Prospects.

[4] Acosta, D. and Wortmann, M. (2009) Alzheimer's Disease. International World Alzheimer Report.

[5] Sheaff, R., Sherriff, I. and Hennessy, C.H. (2018) Evaluating a Dementia Learning Community: Exploratory Study and Research Implications. *BMC Health Services Research*, **18**, 83. https://doi.org/10.1186/s12913-018-2894-3

[6] Liu, Q., Liu, X., Jiang, B., *et al.* (2011) Forecasting Incidence of Hemorrhagic Fever with Renal Syndrome in China Using ARIMA Model. *BMC Infectious Diseases*, No. 11, 218. https://doi.org/10.1186/1471-2334-11-218

[7] Shao, Z., Wang, C. and Wei, M. (2003) Application of Gray GM (1, 1) Prediction Model in Disease Prediction. *China Hospital Statistics*, No. 10, 146-148.