



Construction of Risk Prediction Model for Alzheimer's Disease Based on Meta-Analysis

Shuang Jiang¹, Zhezhi Jin^{2*}

¹College of Science, Yanbian University, Yanji, China

²School of Economics and Management, Yanbian University, Yanji, China

Email: 1099059878@qq.com, *jinzhezhi@sina.com

How to cite this paper: Jiang, S. and Jin, Z.Z. (2019) Construction of Risk Prediction Model for Alzheimer's Disease Based on Meta-Analysis. *Open Access Library Journal*, 6: e5731.

<https://doi.org/10.4236/oalib.1105731>

Received: August 23, 2019

Accepted: September 13, 2019

Published: September 16, 2019

Copyright © 2019 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objective: Construction of risk prediction model for Alzheimer's disease based on Meta-analysis. **Method:** Documenting such as CNKI, WanFang database and VIP database, which were published in China from 2010 to 2018 and had data on risk factors for Alzheimer's disease. Meta-analysis of standards compliant literature using RevMan5.3 software and then combine the results to obtain a comprehensive risk for each risk factor, the risk prediction model is established based on the natural logarithmic transformation value of the comprehensive risk. **Result:** A total of 28 articles were included in the literature. There are 7 risk factors entering the model through data screening and they were family history of dementia, ApoE, education level, brain trauma, drinking history, family history of dementia and gender and smoking history, a total of 2,229,980 patients were included. **Conclusion:** Ultimately the model will be used to effectively control risk factors and slow the progression of the disease.

Subject Areas

Applied Statistical Mathematics, Nursing

Keywords

Alzheimer's Disease, Risk Prediction Model, Risk Factors, Logistic Regression, Meta-Analysis

1. Introduction

Alzheimer's disease (AD) is the most common type of senile dementia. It is a central nervous system degenerative characterized by progressive cognitive dysfunction and behavioral damage in the elderly and early senile. Clinical manife-

stations are memory impairment, aphasia, abstract thinking and computational power impairing personality and behavioral changes [1]. According to the 2015 World Alzheimer's Disease Report, a new AD patient is diagnosed every 3 seconds. By 2050, the number of patients with Alzheimer's disease worldwide will increase from the current 46 million to 131.5 million, and the increase is more obvious. The main cases are low-income and middle-income families [2]. Therefore, the clinical prevention of Alzheimer's disease, the identification of independent risk factors for patients, and the construction of risk prediction models for Alzheimer's disease are currently very important research topics.

The binary classification logistic regression analysis is a generalized linear regression analysis model, which is often used for automatic diagnosis of diseases, exploring risk factors for causing diseases and predicting the probability of disease occurrence according to risk factors, and the results are accurate and stable [3] [4]. Therefore, this study intends to determine the independent risk factors affecting the condition of Alzheimer's disease patients by logistic regression combined with Meta-analysis, and establish an effective risk prediction model to find the risk factors that affect Alzheimer's disease and to control them so as to improve the quality of life of patients and reduce the burden on society.

2. Data and Method

2.1. Document Retrieval

The literature search uses "Alzheimer's disease, risk factors", as the title or keyword, A total of 1586 articles related to Alzheimer's disease were retrieved from 2010 to 2018, published in three major databases, CNKI, VIP Database and WanFang Database.

2.2. Inclusion Criteria

Patients should be selected to meet the diagnostic criteria for Alzheimer's disease. A parallel randomized controlled clinical trial of the treatment group and the control group was provided; Data on risk factors for Alzheimer's disease can be extracted from the literature.

2.3. Exclusion Criteria

Only the literature of the control group was not set up in the self-experiment; patients with Alzheimer's disease also had other diseases that may cause dementia or congenital cognitive disorders and other types of dementia; Data could not be converted into odds ratio (Odd Ratio) and 95% (CI) confidence interval literature; Repeated inclusion of literature; Poor quality literature, including NOS (Newcastle-Ottawa Scale) scale [5] lower scores or data reporting incomplete literature.

2.4 Literature Inclusion Results

Based on the topic of "Alzheimer's disease, Alzheimer's disease, risk factors", the

related literatures were searched, and 1586 articles were initially searched. According to the inclusion and exclusion criteria of the literature, 1557 articles that did not meet the criteria were excluded. 28 papers were analyzed and the literature screening was shown in **Figure 1**. Among the 28 articles included, there were 2,229,980 patients and 9 risk factors associated with Alzheimer's disease: ApoE (carrier protein E), education level, smoking history, drinking history, family dementia history, diabetes, Negative life events, gender, head trauma. The basic data of the surveyed patients included in the literature are shown in **Table 1**. The literature was evaluated using the NOS scale. The specific evaluation items included: selection of subjects (4 points), comparability between groups (2 points), and measurement of exposure factors (4 points). If the total score is greater than or equal to 5 points, the quality of the literature is indicated. Higher, the total score of less than 5 indicates that the quality of the literature is low. In this study, **Table 2** shows the results of quality evaluation. It can be seen from the table that the 28 documents included are of high quality and can be analyzed.

2.5. Statistical Analysis

Meta-analysis of selected literature using evidence-based medicine software RevMan5.3. Measurement data were expressed as relative risk (OR) and 95% confidence interval (CI). First, the heterogeneity test is carried out on the included documents. When there is no heterogeneity in each of the trials included in the analysis ($p \geq 0.1, I^2 \leq 50\%$), so using a fixed effect model for analysis. When there is heterogeneity between trials included in the analysis ($p < 0.1, I^2 \geq 50\%$), so using a random effects model for analysis.

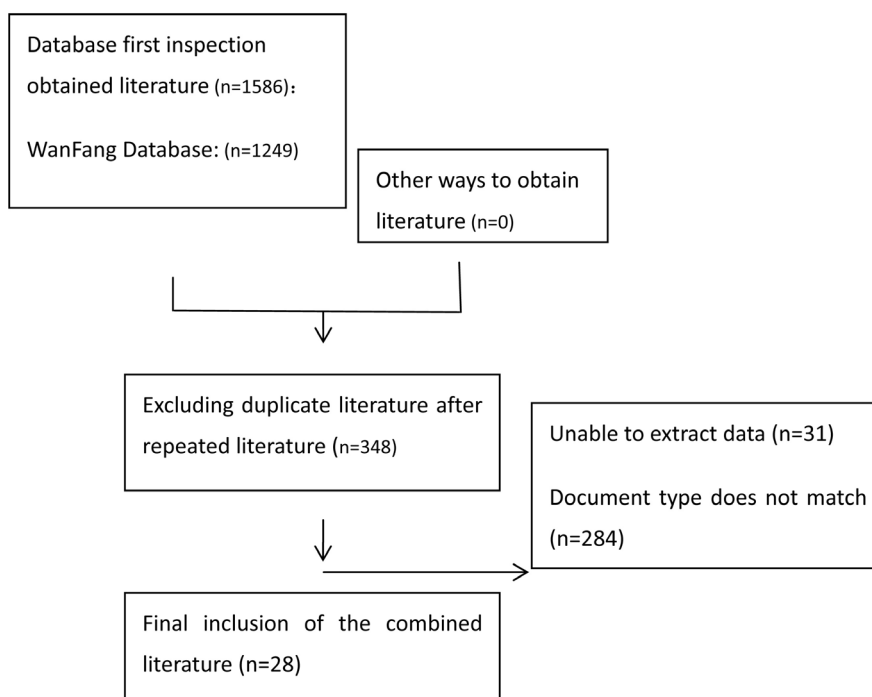


Figure 1. Diagram of searching and screening.

Table 1. Basic data of the surveyed patients.

Risk Factor		Number of Patients with AD	Total People
ApoE	$\epsilon 2, \epsilon 3$	414	1648
	$\epsilon 4$	161	1606
Educational level	Junior high school or above	367	10,522
	Illiterate and junior high school or below	1204	13,415
Gender	Male	647	10,307
	Female	869	10,471
History of smoking	Never smoke	446	2977
	Smoking over three weeks	1240	9278
Drinking history	Never drink	438	1272
	Drinking over three weeks	3680	13,346
Negative life event	No negative life events	25	115
	Negative life event	53	111
Family history of dementia	No history of dementia in the family	138	1232
	Dementia patients in immediate family members	81	10,020
Brain trauma	No brain trauma	13,340	1,970,599
	The head has been hit hard	5909	166,054
Diabetes	No diabetes	418	3588
	Have diabetes	490	3419

Table 2. NOS quality result evaluation form.

The First Author	Research Object Select	Between Groups Comparability	Exposure Factor Measurement	Total Score
Zhang Ting [6]	3	1	3	7
Mai Yicheng [7]	3	1	4	8
Cai Haiyan	3	1	3	7
Dong Xiang	3	2	3	8
Guo Hongmin	3	1	4	7
Zhu Weiming	4	1	2	7
Li Lei	3	1	4	8
Guan Junhua	3	1	4	8
Hu Chengping	4	1	2	7
Huang Heming	3	0	3	6
Liao Jun	4	0	3	7
Sun Zhenghai	3	2	3	8
Ye Guoying	2	1	2	5
Zhan Changqing	4	0	4	8
Cui Bo	2	0	3	5
Zheng Feng	3	2	3	8

3. Meta Analysis

3.1. Heterogeneity Test

By calculating the I^2 value and p value of each risk factor, **Figure 2** shows the heterogeneity test forest map of one of the risk factors for Alzheimer's disease, where $p = 0.29$, $I^2 = 18\%$, so fixed effect model. In this study, except for ApoE, negative life events and educational level, gender, smoking history, drinking history, brain trauma, diabetes and family history of dementia all have heterogeneity. So for ApoE, negative life events, education level, the combined effect model was used to calculate the combined effect values, and the rest were all random effects models.

3.2. Publication Bias

Gender, ApoE, drinking, smoking, family history of dementia, education, diabetes, negative life events, there is no publication bias ($p > 0.05$). There is a certain publication bias in brain trauma ($p < 0.05$). On the whole, the bias of this study included in the literature was small. **Table 3** is the discretionary results of risk factors for Alzheimer's disease.

3.3. Meta-Analysis Results

A total of 9 risk factors were evaluated in this study. Among all the risk factors evaluated, the OR value of diabetes was between 0 and 1 (indicating that the statistical results of this risk factor were not statistically significant), and negative life events were included. There are few literatures, so no risk factors are included in the model construction. Only 7 items are considered as independent risk factors for Alzheimer's disease, including: ApoE, brain trauma, family history of dementia, education level, smoking history, drinking History and gender (**Table 4**).

4. Logistic Regression Model

The logistic regression model refers to the calculation of the OR value of risk factors based on prospective or retrospective data to evaluate the contribution of

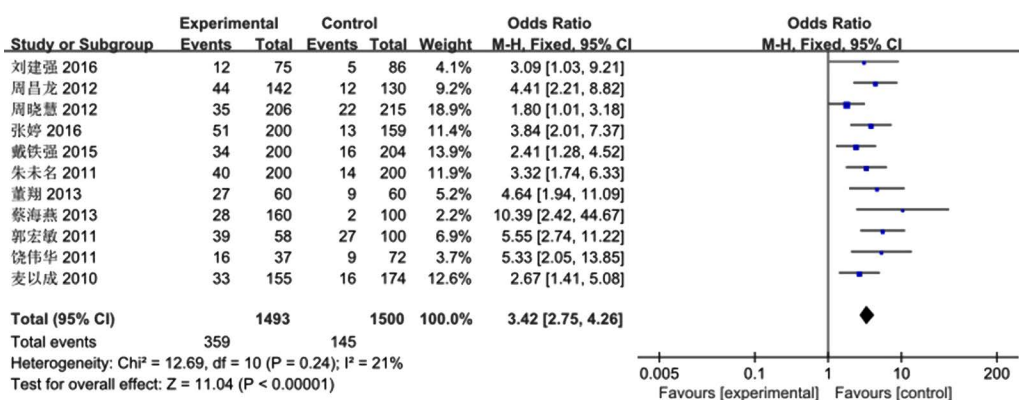


Figure 2. ApoE heterogeneity test forest map.

Table 3. Discretionary results of risk factors for Alzheimer's disease.

Risk Factor	<i>t</i> Value	<i>p</i> Value
Gender	0.15	0.88
ApoE	11.4	0.24
Drinking history	1.62	0.11
Smoking history	0.72	0.98
Brain trauma	5.52	$p < 0.00001$
Family history of dementia	0.17	0.87
Education level	12.49	0.29
Diabetes	0.03	0.98
Negative life events	4.04	0.16

Table 4. Combined results of Meta analysis.

Project	Number of Projects	<i>t</i>	<i>I</i> ² /%	Merge Model	Meta Analysis Results		
					OR	95%CI	<i>p</i>
ApoE	11	11.4	21	fixed	3.42	2.75 - 4.26	0.24
Gender	10	0.15	96	random	2.48	0.54 - 2.04	0.88
Education level	8	12.49	18	fixed	1.66	2.15 - 2.86	0.29
Smoking history	11	0.72	87	random	1.17	0.77 - 1.78	0.98
Drinking history	11	1.62	91	random	1.52	0.92 - 2.53	0.11
Diabetes	7	0.03	82	random	0.99	0.60 - 1.63	0.98
Brain trauma	6	5.52	98	random	2.63	1.86 - 3.70	$p < 0.00001$
Family history of dementia	5	0.17	97	random	3.37	0.13 - 11.54	0.87
Negative life events	2	4.04	50	fixed	3.49	1.90 - 6.42	0.16

risk factors in the risk of specific diseases, and can also be used to predict the risk of disease.

In this study, Meta-analysis was used to calculate the comprehensive risk (OR value) of each risk factor associated with the formation of Alzheimer's disease, and then the corresponding risk prediction model was established based on the natural logarithmic transformation value of the comprehensive risk. This method accumulates the sample size of the current research, overcomes the defects of the single research sample, and ensures the scientificity, reliability and accuracy of the prediction results of the established risk prediction model. The theoretical prediction model for the risk of Alzheimer's disease formation based on the logistic regression model is as follows:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 \quad (1)$$

From the above model, the risk probability of Alzheimer's disease patients can

be calculated:

$$p = \frac{e^{\alpha} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \dots + \beta_n X_n}{1 + e^{\alpha} + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \dots + \beta_n X_n} \quad (2)$$

In the above formula “ $X_1, X_2, \dots, X_i, \dots, X_n$ ” express $1, 2, \dots, i, \dots, n$ risk factors, “ β_i ” is the regression coefficient for each hazard element can be calculated using the natural logarithmic conversion value of the comprehensive risk, calculated as follows:

$$\beta_i = \ln(OR_i) \quad (3)$$

“ α ” is the constant term, from the theoretical model, the constant term can be calculated by the following formula:

$$\alpha = \ln\left(\frac{p_{oj}}{1 - p_{oj}}\right) - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 - \dots - \beta_i \bar{X}_i - \dots - \beta_n \bar{X}_n \quad (4)$$

among them p_o is the prevalence, “ $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_i, \dots, \bar{X}_n$ ” express $1, 2, \dots, i, \dots, n$ average per capita of risk factors, that is, the average exposure rate of the event. This value can be estimated by the prevalence and incidence of Alzheimer’s disease.

5. Establishment of Logistic Risk Prediction Model

Based on the results of the meta-analysis and the basis for the selection of the overall risk, this study first removed the factors that could not be considered to have an impact on the formation of Alzheimer’s disease, and then established a risk prediction model based on the combined results of independent risk factors. In this study, a total of seven risk factors were eventually included in the predictive model. The risk prediction model constructed is as follows:

$$\text{logit}(p) = \alpha + 1.23X_1 + 0.97X_2 + 0.42X_3 + 0.91X_4 + 0.50X_5 + 0.16X_6 + 1.21X_7 \quad (5)$$

However, in this practical application, in addition to estimating the prevalence and incidence of Alzheimer’s disease in the study area, it is also necessary to estimate the risk of each risk factor covered in the risk prediction model of the area. The average value, but the correct calculation of the average of the individual risk levels is extremely difficult, so it is often recommended in practice to ignore the correction of the constant term in the prediction model, and by the formula:

$$\alpha = \ln\left(\frac{p_{oj}}{1 - p_{oj}}\right) \quad (6)$$

to calculate. However, it must be noted in the above formula that although the simplified estimation of the constant term increases the actual operability of the risk prediction model, it also has a higher positive rate or a risk factor due to the risk factors of the target population being assessed. In the case of greater risk, the risk of disease is overestimated [8].

6. Discussion

Meta-analysis is a method of consulting medicine that combines multiple homogeneous studies to provide true and reliable scientific evidence for clinical practice and health decision-making [9]. This study systematically retrieved the current randomized controlled trials evaluating the factors that influence the development of Alzheimer's disease. After rigorous literature screening, 28 articles that met the criteria were finally included. Based on the results of the meta-analysis, a risk prediction model for Alzheimer's disease was established. Through this model, risk factors that affect Alzheimer's disease can be found and controlled to delay the progression of the disease.

An example of a risk prediction model for Alzheimer's disease based on this study is as follows: First, assuming that the incidence rate of Alzheimer's disease patients in the target hospital is 14.91%, and the local prevalence rate is twice that of the incidence rate, then a single calculation is performed. The risk of Alzheimer's disease in patients is as follows: In a local hospital, Alzheimer's disease, the examination report shows that the patient is female, smoking, drinking, and the degree of education is primary school graduation, according to the Alzheimer's disease risk prediction model, Calculate the constant term:

$$\alpha = \ln\left(\frac{14.91 \times 2/100}{1 - 14.91 \times 2/100}\right) = 0.42 \quad (7)$$

Combined with the risk factors of this patient, the risk of the disease is:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = 0.42 + 0.42 \times 1 + 0.91 \times 1 + 0.50 \times 1 + 0.16 \times 1 = 2.41 \quad (8)$$

$$p = e^{2.41} / (1 + e^{2.41}) = 92\% \quad (9)$$

Then the patient's risk is three times that of other patients in the area ($\frac{92\%}{14.91\% \times 2} = 3.09$). Therefore, according to the predicted results, the doctor can appropriately recommend the patient to reduce the frequency of smoking and drinking when the patient's gender cannot be changed and the degree of education cannot be improved for a short period of time.

Although this study based on the meta-analysis, a comprehensive analysis of the included research data to establish a Logistic risk prediction model for Alzheimer's disease provides a predictive tool for the clinical practice of Alzheimer's disease. However, because the Meta-analysis is a secondary analysis based on the original research, the inadequacy of this study cannot be avoided: Because the selected risk factors related to Alzheimer's disease are not analyzed, it may lead to the incomplete meta-analysis, so the constructed risk prediction model has certain defects, so the researcher should find Potential risk factors for Alzheimer's disease and incorporating them into new research to merge to make the risk prediction model more complete; Although this study established a risk prediction model for Alzheimer's disease, in order to ensure that the final constructed risk prediction model is practical, the model should be validated, so the

researcher will continue to verify the model in the follow-up work.

7. Conclusion

Based on a large amount of literature data, this paper establishes a risk prediction model for Alzheimer's disease based on the Meta-analysis combined with the mathematical modeling method of logistic regression model, which can be used to predict the risk of Alzheimer's disease. In addition, in addition to the practical significance of the prediction of Alzheimer's disease, this model can provide new ideas for the next step to establish a risk prediction model for other diseases.

Conflicts of Interest

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

References

- [1] Li, J.C., Huang, J.C., Zhu, Y.F., *et al.* (2017) Evaluation and Analysis of Early Risk Factors for Alzheimer's Disease. *Chinese Journal of Health Statistics*, **34**, 756-760.
- [2] Zhu, Q. and Chen, X.X. (2017) Epidemiological Investigation of Alzheimer's Disease. *Chinese Journal of Physical Medicine and Rehabilitation*, **39**, 866-869.
- [3] Zhang, C.B., Gao, K. and Yang, G.J. (2010) Simulation Comparison between Discriminant Analysis and Logistic Regression. *Journal of Statistics and Information*, **28**, 19-25.
- [4] Yi, S.H., Yi, Y.S., Liu, T.C., *et al.* (2008) Logistic Regression Analysis of Colorectal Cancer Prognosis. *Modern Chinese Medicine*, **14**, 969-971.
- [5] Htenstein, M.J. (1987) Guidelines for Reading Case Control Studies. *Journal of Chronic Diseases*, **40**, 893-903. [https://doi.org/10.1016/0021-9681\(87\)90190-1](https://doi.org/10.1016/0021-9681(87)90190-1)
- [6] Zhang, T., Hu, G.Y. and Lin, Y.L. (2016) Correlation between Apoe Allele Frequency and Lipid Metabolism in Patients with Alzheimer's Disease. *Journal of Practical Medicine*, **32**, 127-129.
- [7] Mai, Y.C. (2010) Correlation between Interleukin-8 and Apolipoprotein E Gene Polymorphisms and Late Onset Alzheimer's Disease. *Journal of Sun Yat-sen University: Medical Science Edition*, **31**, 118-121.
- [8] Ye, C. and Liu, Q. (2010) Exploratory Study on Risk Assessment Model of Colorectal Cancer in Population. *Evidence-Based Medicine*, **10**, 86-91.
- [9] Liu, J.P. (2006) Evidence-Based Chinese Medicine Clinical Research Methodology. People's Medical Publishing House, Beijing.