

Research of Risk Factors for Alzheimer's Disease

Qiqi Feng^{1, *}, Yachen Li² and Zimo Zhang³

¹YALI High School, Changsha, 410007, China

²Knowledge-First Empowerment Academy CEEB Code-440004 9888 Bissonnet, Suite 290,
Houston, Texas 77036, America

³Shanghai No. 2 High School, Shanghai, 200000, China

*Corresponding author: eharris52238@student.napavalley.edu

Abstract. Alzheimer's disease (AD) is the most common type of dementia among the elderly, which has become the focus of global public health and social attention. This study is proposed to explore the relationships between 4 factors (gender, age, years of education and socioeconomic status) with AD by the use of the Binary regression model and correlation regression. It was concluded that female have a dramatic positive impact on AD significantly ($p < 0.01$, $OR = 3.517$), and the years of education will have a remarkable negative impact on the Groups ($p = 0.011$, $OR = 0.784$). However, the Age and socioeconomic status will not have an impact on the AD ($p = 0.748$ and $p = 0.622$). The early prevention is feasibility and importance to delay the onset of AD and to improve patients' quality of life. Accordingly, an effort should be made to improve the professional care and formulate relevant policies for elderly care services, especially for Women with low education.

Keywords: Alzheimer's disease; binary logit regression model; descriptive statistical analysis.

1. Introduction

Alzheimer's disease (AD) is a mental illness which is one of the most frequent types of dementia among the old, accounting for 60-70% of all dementia cases [1]. Not only does this disease phenomenon bring high economic expenditure to the society, but also it has a huge impact on the future life of patients' families [2]. A 2016 report suggested that about 40 million people around the world are suffered from this disease, and the number is predicted to be twice as many as the last 20 years [3]. This disease has become the focus of global public health and social attention [4]. It is generally accepted that non-modifiable risk factors (such as genetics) and modifiable risk factors (such as lifestyle, health) may cooperate to affect the onset and progression of dementia. In contrast to non-modifiable factors that cannot be interfered with (such as age, gender, or genetics), modifiable factors are intervenable before AD occurs (such as smoking, anemia or diabetes). Xu *et al.* proved that about one-third of the risk factors for AD can be modified, reflecting the importance and feasibility of early prevention [5].

Scholars from home and abroad have found many risk factors related to AD, including increasing age, aluminum exposure, diet, head injury, mitochondrial dysfunction, vascular diseases, infectious diseases, etc. [6]. More recently, a growing body of research has shown an increase in identification with some of the factors discussed and a decrease in identification with others, while many new risk factors have also been identified. Wen *et al.* suggested that thyroid disease, physical inactivity, and living alone may also increase the risk of developing AD [7]. Many previous researches have shown that clinical psychiatric symptoms, including depression and anxiety, are related to AD [8, 9].

Therefore, the study is dedicated to discuss the relationship between these factors and AD by focusing on 4 aspects (gender, age, years of education, socioeconomic status) and using the Binary Logit regression model and correlation regression, in order to provide insights into the prevention of dementia, professional care, as well as the management department of elderly care institutions to formulate relevant policies for elderly care services.

2. Methods

2.1. Data Sources

The data for this literature is collected from the Kaggle website. It is collected by Baris Dincer and updated in 2021 for 336 individuals.

2.2. Variable Selection

The data used in this paper count a total of 336 people, including those who have and do not have Alzheimer's disease, of whom 147 are female and 189 are male. The age of patients is boundary from 60 to 98 years. The economic level ranges from 0 to 5, 0 represents the lowest economic level and the 5 means the highest economic level. The years of education ranges from 6 to 23. The data contains 4 variables (gender, age, years of education, socioeconomic status). The abbreviation of the 4 factors is M/F, EDU, Age, and SES, respectively Literature References. Table 1 shows that all the potential factors that can cause the Alzheimer's disease. It also shows the abbreviation of these factors and the meaning of these 4 factors.

Table 1. Description of 4 Factors

Elements	Abbreviation	Meaning
Gender	M/F	Female and male
Years of education	EDU	The time of gain knowledge
Age	Age	The time of alive
Socioeconomic Status	SES	Economic level

Table 1 shows that all the potential factors that can cause the Alzheimer's disease. It also shows the abbreviation of these factors and the meaning of these 4 factors.

2.3. Research Protocol

This article uses the Binary Logit regression model and correlation regression. Correlation analysis is a statistical method that studies whether there is a correlation between things and the strength of the correlation, with the main purpose of studying the closeness of the relationship between variables. And the Binary Logit regression mode studies the dependence of the dependent variable on the explanatory variable. The AD is the dependent variable(Y), and 4 factors are the independent variables (X). The 0 represent demented, the 1 represents non-demented.

3. Results and Discussion

3.1. Descriptive Statistic Analysis

Figure 1 shows that the percentage of number of patients in each the socioeconomic status group. The figure 3 shows that in level 3, the number of patients occupies a relatively big area and in level 5, the number of patients occupies a relatively a small area. This distribution is not increasing in order or decreasing in order, so the SES maybe have no relationship with AD.

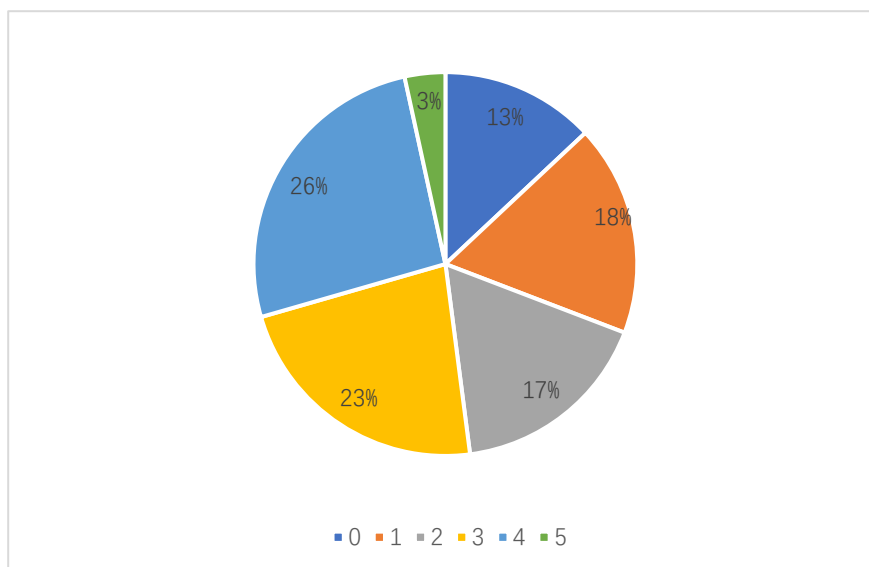


Fig. 1 The percentage of the number of patients in each level

Figure 2 shows that the percentage of number of patients in each years of education group. Here, the time spent on education is divided into four groups. The first group is 6-10, the second group is 11-15, the third group is 16-20, and the last group is 21-23. This figure also shows that in 11-15, the number of patients occupies a large area and in 6-10, the number of patients occupies a small area. Because the number of patients in 21-23 is 0, this graph cannot show that part. From 6-10 to 21-23, it is approximately to say that the longer the years of education is, the less possible the AD appears.

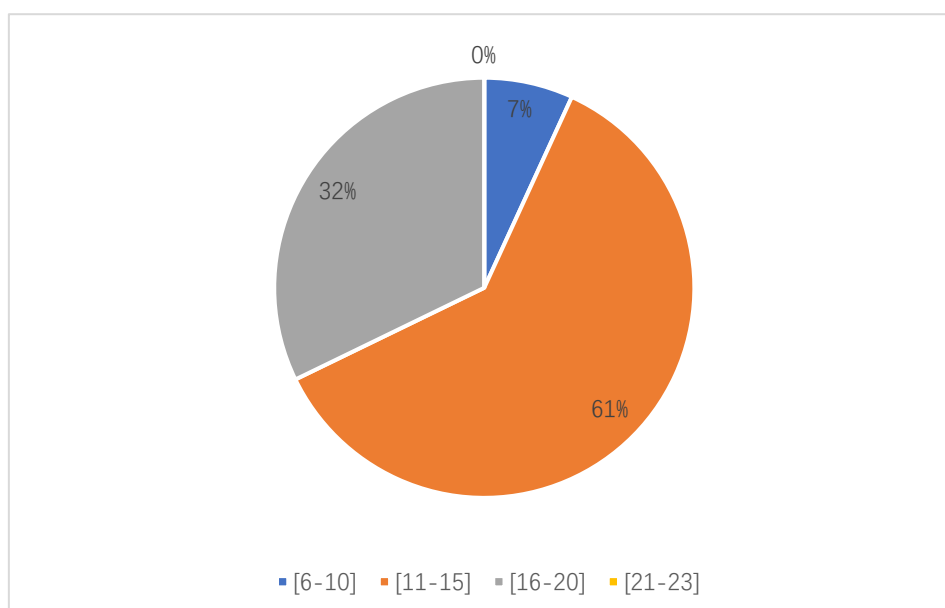


Fig. 2 The percentage of number of patients in each years of education

In the figure 3, age is divided into four groups. The first group is 60-70, the second group is 71-80, the third group is 81-90, and the last group is 91-98. The figure also shows that in age group 71-80, the number of patients occupies a large area and in age group 91-98, the number of patients occupies a small area. From 71-80 to 91-98, the longer age they are, the more patients they have.

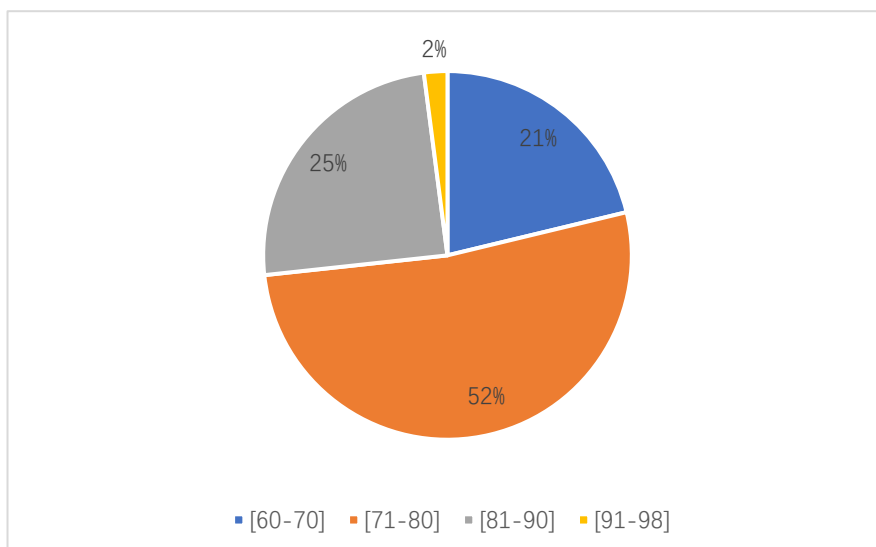


Fig. 3 The percentage of number of patients in each age

In the figure 4, people are divided into 2 groups. The one group is female and the other one is male. The figure also shows that the percentage of number of female of patients is larger than the percentage of number of male of patients. It can say that female will more likely have AD.

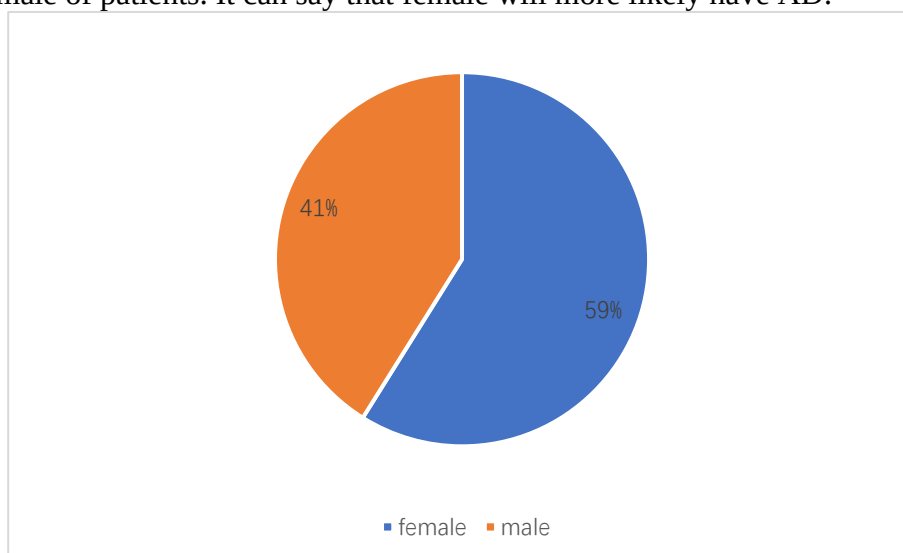


Fig. 4 The percentage of number of patients in each gender

3.2. Correlation Analysis

From table 2 and figure 5, it can be seen that using correlation analysis to study the correlation between AD and M/F, Age, EDUC, SES, Pearson coefficient of correlation is used to represent the intensity of the correlation.

Table 2. Correlation Regression Model Results

	AD	M/F	Age	EDUC	SES
AD	1	-	-	-	-
M/F	0.246**	1	-	-	-
Age	0.018	-0.064	1	-	-
EDUC	-0.263**	0.044	-0.050	1	-
SES	0.187**	-0.027	-0.005	-0.733**	1

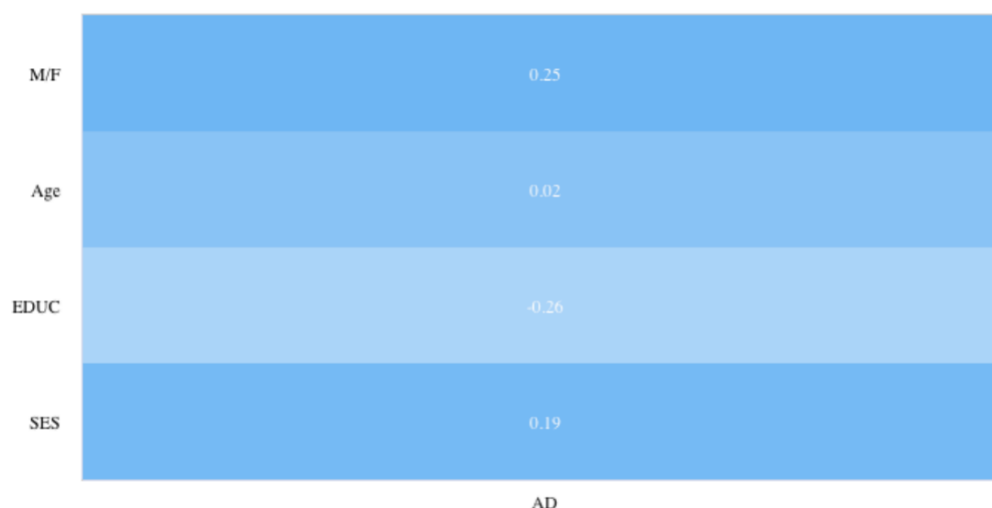


Fig. 5 Pearson Related Visualization

0.246 is the coefficient of correlation among AD and M/F and displays a notable 0.01 level, indicating a notable positive correlation between AD and M/F. 0.018 is the coefficient of correlation among AD and Age, it closes to 0, and the p-value is 0.748>0.05, indicating that there is no correlation among AD and Age. -0.263 is the coefficient of correlation among AD and EDUC and shows a notable 0.01 level, it lights on a notable negative correlation among AD and EDUC. The correlation coefficient among AD and SES is 0.187 and it can be seen a notable 0.01 level, suggesting a significant positive correlation among AD and SES.

3.3. Binary Logit Regression Analysis

This table 3 shows that the accurate data analysis that which factors have a relationship with AD. From the above table, it can be seen that using M/F, Age, Socioeconomic Status (SES), and Years of Education (EDUC) is independent variable, group is dependent variable for Binary logistic regression analysis. It means that M/F, Age, SES, and EDUC can explain the 0.10 change in AD (1 represents demented, 0 represents non-demented).

Table 3. Model results

elements	Regression	S.E.	Z value	Wald	P value	OR	95%CI
M/F	1.258	0.250	5.039	25.389	0.000	3.517	2.156~5.736
Age	0.004	0.016	0.234	0.055	0.815	1.004	0.972~1.036
SES	0.079	0.159	0.492	0.243	0.622	1.082	0.791~1.478
EDUC	-0.165	0.065	-2.534	6.421	0.011	0.848	0.746~0.963
Constant	-0.230	1.889	0.122	0.015	0.903	0.784	0.020~32.234

From the table above, there is a modal formula:

$$\ln(p/1 - p) = 2.109 + 1.268 * M/F - 0.018 * Age - 0.201 * EDUC - 0.029 * SES \quad (1)$$

There is attention that p represents that the probability of Group is 1 and 1-p represents that the probability of Group is 0.

Through the last specific analysis, it can be seen that 1.25 is value of regression coefficient of M/F and at 0.01 level manifest significance ($z=5.039$, $p=0.000<0.01$). It indicates that M/F will have a significance positive effect on the AD. 3.517 is the value of OR, which means that when M/F increases by one unit, the change in AD is 3.517 times. 0.004 is the regression coefficient value of Age, but it doesn't show significance ($z=0.234$, $p=0.815>0.050$). It suggests that Age does not have an impact on the AD. What's more, 0.079 is the value of regression coefficient of SES, but it does not show significance ($z=0.492$, $p=0.622>0.05$). It means that SES does not have an effect on the AD.

And, -0.165 is the regression coefficient value of EDUC. At the 0.05 level, it can show that conspicuousness ($z=-2.534$, $p=0.011<0.05$). It indicates that AD will be influenced by a negative effect by EDUC. EDUC will have a negative effect on the AD. 0.848 is the value of OR, meaning that when EDUC increases by one unit, the change in AD is 0.848 times.

By summarizing all information, it concludes that AD will be influenced by a remarkable positive effect by M/F, and the AD will be influenced by a negative effect by EDUC. However, Age and SES will not have an impact on the AD. The other study, like analysis of the prevalence and related factors of senile dementia in Guandu District, Kunming, has more factors and more samples than this study [10].

4. Conclusion

The current study selects diverse data and focuses on influencing factors that may be associated with developing AD. It was concluded that having AD may be related to age, gender, years of education.

Although the amount of data is limited this model may have errors in addition to the factors, and the sample also did not cover all ages and ethnicities, which caused possible differences and affected the accuracy of the results, this study also has its values and merits. Firstly, the study selects lots of samples from the data set and it contains most of the population so that it works in most cases. On the one hand, a graphical approach was used to visually analyze the differences in the percentage of the number of patients in each level group, in each year of education group, in each age group, and each gender group. This enables visualization of the experiment and makes the results clearer and more intuitive. On the other hand, this paper uses the Binary Logit regression model making it more comprehensive. What is more, it also helps people to know about and treat AD. Besides the age you know is one of the most important factors about AD, this study also tells you that Gender, Years of education may also contribute to AD. Whether these factors are associated with the development of AD requires further medical investigation, which means that these findings will point the way to further relevant research in the future. Once people find a new causative factor, this can help to detect AD earlier, so that people can treat it as soon as possible, improving survival rates and patients' quality of life.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Ferreira D, Perestelo-Perez L, Westman E, et al. Meta-review of CSF Core biomarkers in Alzheimer's disease: the state-of-the-art after the new revised diagnostic criteria. *Front Aging Neurosci*, 2014, 6: 47.
- [2] IDd A, et al. Alzheimer disease: correlation between memory and autonomy. *Rev psiquiatr clín*, 2005, 32(3): 131-136.
- [3] Killin LO, Starr JM, Shiue IJ, Russ TC. Environmental risk factors for dementia: a systematic review. *BMC Geriatrics*, 2016, 16: 175.
- [4] Scheltens P, Blennow K, Breteler MM, et al. Alzheimer's disease. *Lancet*, 2016, 388: 505-517.
- [5] Xu W, Tan L, Wang H F, et al. Meta-analysis of modifiable risk factors for Alzheimer's disease. *J Neurol Neurosurg Psychiatr*, 2015, 86: 1299-1306.
- [6] Armstrong R A. What causes Alzheimer's disease? *Folia Neuropathol*, 2013, 51: 169-188.
- [7] Wen C, Hu H, Ou YN, et al. Risk factors for subjective cognitive decline: the CABLE study. *Transl Psychiatry*, 2021, 11(1): 576.
- [8] Pless A, Ware D, Saggu S, Rehman H, et al. Understanding neuropsychiatric symptoms in Alzheimer's disease: challenges and advances in diagnosis and treatment. *Front Neurosci*, 2023, 17: 1263771.

- [9] Omura JD, et al. Modifiable Risk Factors for Alzheimer's Disease and Related Dementias Among Adults Aged ≥ 45 Years - United States. MMWR Morb Mortal Wkly Rep, 2022, 71(20): 680-685.
- [10] Wang Rujiao, Zhou qiang, Xu Xu, et al. The prevalence and influencing factors of elderly dementia in Guandu District, Kunming City. Alzheimer's disease and related diseases. 2022, 5(3): 189-194.