

The Research of Influence Factors that Possibly Lead to Alzheimer's Disease

Yuhan Jin¹, Chulei Li^{2,*}

¹JuRong Country Garden School, Jurong, 212446, China

²Nanchang No2. high school, Nanchang, 330106, China

*Corresponding author: lcc139168@nspes.ca

Abstract. In known studies it has been shown that the prevalence of Alzheimer's disease is related to factors such as heredity, aluminum intake, lifestyle, and gender, but there are still many factors that have not been considered and studied. Therefore, this study focused on the effect of other factors on the prevalence of Alzheimer's disease. The study used multiple linear regression models to process the data from the Alzheimer's Disease Prediction System, an online web-based system that publishes test data from 373 individuals, and visual graphs to analyze the independent variables. The study concluded that Alzheimer's disease was not significantly associated with age, but was significantly associated with gender, years of education, and socioeconomic status, and positively associated with the latter two. However, there was a significant relationship with gender, level of education and socioeconomic status, and a positive relationship with the latter two. These factors such as years of education and socioeconomic status have not appeared in previous studies, so this study provides a new perspective on Alzheimer's disease medical research and a new direction for predicting the disease.

Keywords: Alzheimer's disease; influence factors; multiple linear regression.

1. Introduction

Since the first officially recorded case of Alzheimer's disease has been reported for 115 years, it has become one of the most devastating diseases in the world which still cannot be cured [1]. The increasing number of patients has a serious negative impact on their families and society. Therefore, it is significant to find out the influence factors in order to reduce financial and psychological costs to families and society and increase the possibility of finding out the therapeutic method. In addition, this paper aims to study the possible factors that generate Alzheimer's disease and help people to be aware of the potential dangers of Alzheimer's diseases for preventing it effectively.

There are incredible factors affecting Alzheimer's disease which involve personal factors and external environmental factors. It is worth nothing that plenty of scientists said that diabetes, genetic inheritance, and the environment had an effect on developing Alzheimer's disease [2]. BBC NEWS have reported that gender also had a relationship with the diseases [3]. Moreover, Zhu mentioned that the Aluminum intake has close relation with the disease, which has proportional relationship with the possibility of Alzheimer's disease [4]. Plenty of patients are caused by these factors which has huge impact. For example, the increasing number of patients has a serious negative impact on their families and society. Therefore, it is significant to find out the influence factors in order to reduce financial and psychological costs to families and society and increase the possibility of finding out the therapeutic method [5]. Liang et al. showed that smoking has a slight relation with smoking [6]. These researches had the features of pertinence and oneness. However, they lack the comprehensiveness of factors to Alzheimer's disease to some extent. As results, this paper demonstrates four influence factors (Gender, Age, Years of Education, Socioeconomic Status which are not common in other papers, if they had relevance with Alzheimer's diseases and how they influence Alzheimer's disease. Then, we choose the proper model and method to explore the relationship.

Some models are used in similar examples. Song et al. used Multivariate Multilevel Models [7]. By estimating the variance at each level and accounting for the effects of differences in explanatory variables, this information can be effectively used to estimate regression coefficients. However, based

on this model, it involves expensive statistical programs and needs substantial sample data, which is not always easy for people to interpret the data. Zhou et al. chose Latent Growth Curve Model (LGCM) [8], the advantage of this model is flexibility, it starts with a single growth trajectory (characterized by an intercept and slope) for a single variable and move to more complex models, but it based on structural equation models (SEM). As time flying, the missing observation needs special treatment in the model [9]. Wang et al. selected the Bayesian Quantile regression joint model [10], which can be observed that the proposed estimator is more stable and efficient especially for the small or modest samples sizes. Nevertheless, it only studied the variables of age, gender, marriage and so on but did not consider the influences by their lifestyles. Also, the size of sample is small, which may include the special cases. Zhang used Multivariate linear stepwise regression model to analyze the factors affecting the quality of life of AD patients [11]. However, they only choose the old people (over 60 years old) to be involved in the sample, which cannot guarantee the diversity of the sample. Besides, they only focused on the lifestyle of the AD patient and ignored age, gender and other factors.

In summary, this paper will verify whether the four factors influence the risk of Alzheimer’s disease completely and objectively, and how the factors affect the possibility of sickening by Alzheimer’s disease. This paper will use multiple linear regression model to analyse the four factors.

2. Methods

2.1. Data Source

The data for this literature is collected from the Kaggle website, which was compiled by Shivani Kharva. The data used in this paper count a total of 373 people, including those who are non-demented, demented and other. The patients' ages range from 60 to 98 years. The data contains 4 variables (Gender, Age, Duration of Years of Education, Level of Socioeconomic Status).

2.2. Variables Selection and Description

Table 1 shows that the total 373 samples and the number of people who suffered by the Alzheimer’s disease. The data where EDUC1 is below the average number of years of education and EDUC2 is above the average number of years of education. The data for level 1 is higher while the data for level 5 is lower. As for convenience, the logogram is shown below.

Table 1. Logogram and Number of four factors

Elements	Logogram	Number1	AD1
Gender	-	373	146
Age	-	373	146
Education Years [6-14.6]	EDUC1	186	89
Education Years [14.6-23]	EDUC2	187	57
Socioeconomic Status [level1-2]	SES1*	191	51
Socioeconomic Status [level 3-5]	SES2**	163	76

*SES1: Socioeconomic Status(level1-2)higher

**SES2: Socioeconomic Status(level3-5)lower

2.3. Introduction to Method

Multiple linear regression (MLR), also known as multiple regression, is a statistical technique that uses multiple explanatory variables to predict the outcome of a response variable. The purpose of multiple linear regression is to model a linear relationship between an explanatory (independent) variable and a response (dependent) variable. In essence, multiple regression is an extension of ordinary least squares (OLS) regression because it includes more than one explanatory variable.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon \tag{1}$$

3. Results and Discussion

The figure 1 shows the number of non-demented, demented and other. As shown in Table 1, The sample of data is 373, of which 39% are demented:

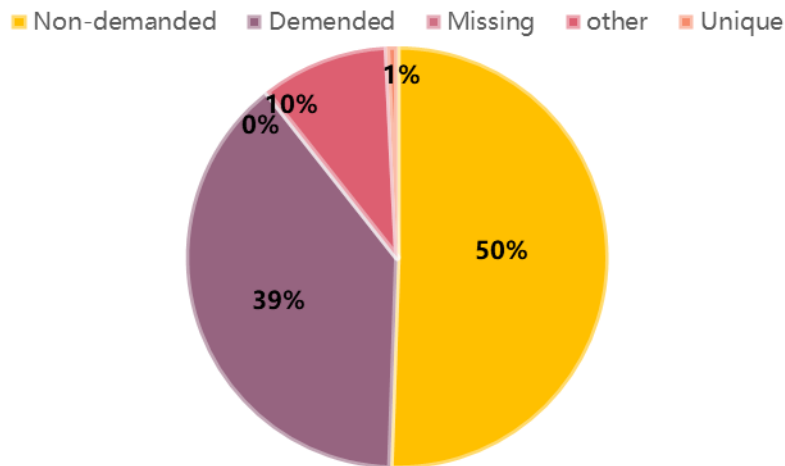


Fig. 1 The proportion of Non-demented people and demented people

The below figure 2 is a histogram of frequency versus age simulated using a normal distribution for a sample of 373 people, which clearly shows the change in frequency with age.

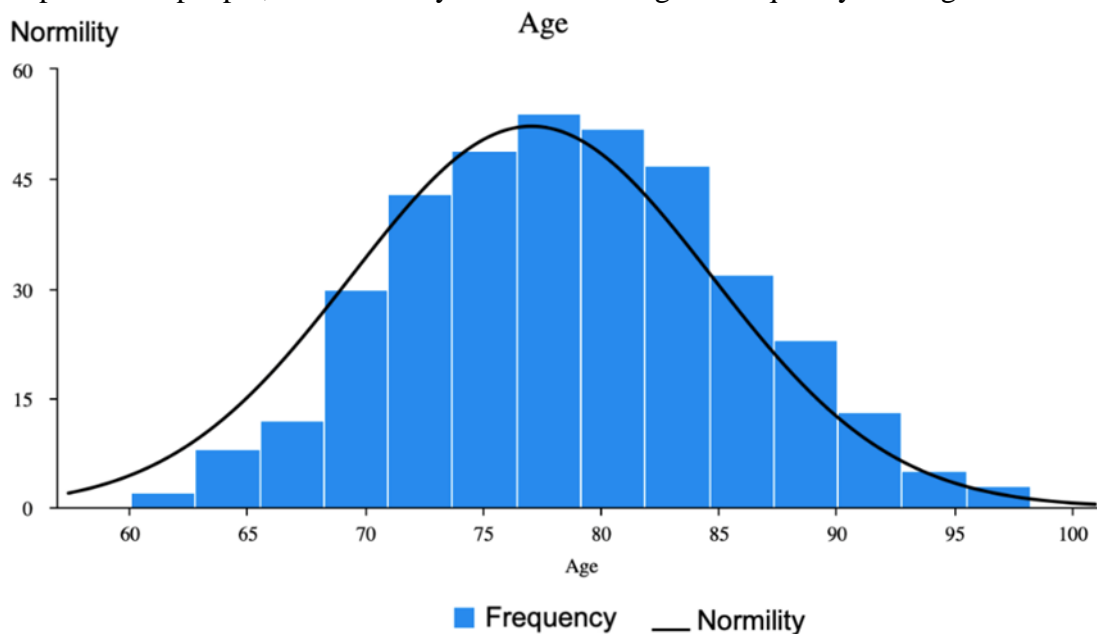


Fig. 2 The frequency of the age

About 53 people in the age group of about 77.5-78.5 account for the largest number of frequencies in the middle, and the frequencies of other age groups decrease in a stepwise manner from the middle to the left and right sides of the population.

Observation shows that the population frequencies of the age groups of about 60-62.5,62.5-66,66-71.5 and 90-92.5,92.5-95.5,95.5-97 on both sides of the picture are small, and the population

frequencies of each subgroup are less than 15, and the values on both sides are relatively close to each other.

In contrast, each of the two populations within the left and right side age stages of about 71.5-77.5 and 78.8-87 have three sub-stages, each with a population frequency of 30 or more, and the right side age stage of about 87-90 has a population frequency of about 20.

Comparison with the modeled normal distribution shows that the right-hand side population frequency is denser and higher than the modeled population frequency at each location, while the left-hand side population frequency is slightly higher than the modeled normal distribution for the population with only three minor stages.

From this we can see that the vast majority of the 373 people in the sample are in sub-stages 5-9, approximately 71.5-84 years of age, with only a small number of people under 71.5 or over 90 years of age.

Table 2. SES distribution by Education years group

Education years	[6-10]	[11-15]	[16-20]	[21-23]
Number	12	191	167	3
SES level 1	0	0	82	3
SES level 2	0	19	50	0
SES level 3	0	24	29	0
SES level 4	5	30	3	0
SES level 5	7	0	0	0
SES no recorded	0	16	3	0
Demented	10	89	47	0

The Table 2 illustrates the SES 1-5 scale on the left hand side of the table represents from top to bottom the social classes from high to low, and the two rows further down represent the number of people with no record of Alzheimer's disease.

It is clear from the Table 2 that of the 12 people with 6-10 years of education, none were in the top three social classes, all were in the bottom social class, and 10 of them had Alzheimer's disease.

Of the 191 people with 11-15 years of education, none were in the top three social classes, but none were at the bottom of the social class either. With the exception of those not recorded, the sample was in strata 2, 3 and 4, with gradually increasing numbers. Of these, 89 had Alzheimer's disease.

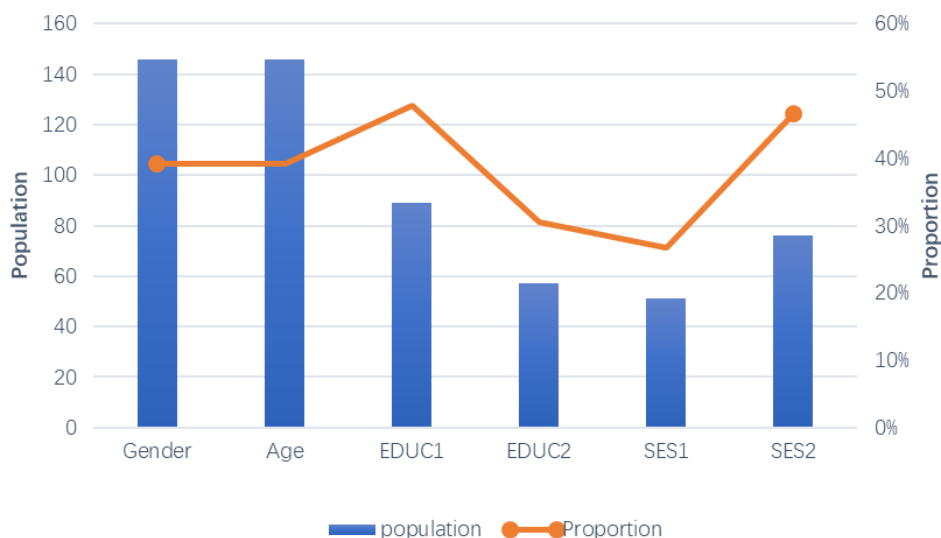


Fig. 3 Variables in the percentage of people with Alzheimer's disease

As for the 167 persons in the sample with 16-20 years of schooling, all of them, with the exception of those who were not recorded, were in the first four strata of the social class, with decreasing

numbers from the largest to the smallest. Of these 167 persons, only 47 suffered from Alzheimer's disease. In the last group, i.e., those with 21-23 years of education, all of them were in the first social class and none of them suffered from Alzheimer's disease.

Figure 3 illustrates several groups of influence factors of Alzheimer's disease in the proportion of people with AD. The factors involve Gender, Age, The duration of Education Years and the Level of Socioeconomic Status. There are a total of 373 samples, to determine how likely each risk factor is to cause Alzheimer's disease, models of Alzheimer's disease were identified and categorized according to the different influencing factors.

According to Figure 2, comparing EDUC1 and EDUC2, we can see that the percentage of EDUC1 (48%) is higher than the percentage of EDUC2 (30%), while the percentage of SES2 (47%) is higher than the percentage of SES1(27%). Thus, it can be seen that short years of education and low socioeconomic status are the main potential risks for Alzheimer's disease.

Table 3. Model Results

	B	Std. Error	Beta	T	P	VIF
Constant	1.745	0.522	-	3.344	0.001**	-
Gender	-0.214	0.071	-0.157	-3.003	0.003**	1.009
EDUC	0.061	0.018	0.263	3.480	0.001**	2.122
Age	-0.003	0.005	-0.038	-0.736	0.462	1.010
SES	0.140	0.045	0.235	3.114	0.002**	2.113
F	F (4,349)=5.387,p=0.000					
D-W	0.797					
Dependent Variable: Alzheimer's disease						
* p<0.05 ** p<0.01						

The table 3 above shows that the model equation is:

$$Y = 1.745 - 0.214\text{gender} - 0.003\text{age} + 0.061\text{EDUC} + 0.140\text{SES} \quad (2)$$

And the R-squared value of the model is 0.058, meaning that gender, age, education and SES are able to explain 5.8% of the variation in the group. The F-test of the model shows that the model passes the F-test (F=5.387, p=0.000<0.05), meaning that at least one of the factors gender, age, education, SES has an influence on the ratio in the group.

Firstly, the regression coefficient for gender is -0.214 (t=-3.003, p=0.003<0.01), which means that gender has a significant negative influence on the group. Secondly, the regression coefficient for age is -0.003 (t=-0.736, p=0.462>0.05), meaning that age has no significant influence on group relationships. Thirdly, the regression coefficient for EDUC is 0.061 (t=3.480, p=0.001<0.01), meaning that EDUC has a significant positive influence on group relationships. Finally, the regression coefficient for SES is 0.140 (t=3.114, p=0.002<0.01), meaning that SES will have a significant positive impact on group relationships.

In summary, the analysis shows that for EDUC, SES has a significant positive effect on the Alzheimer's disease. Gender also has a significant impact on the Alzheimer's disease. Age, on the other hand, has no significant effect on the Alzheimer's disease.

4. Conclusion

A total of four variables were studied in this paper which include Gender, Age, Duration of Years of Education, Level of Socioeconomic Status According to the statistical results Age does not have a significant effect on Alzheimer's disease whereas Gender, Duration of Years of Education, Level of Socioeconomic Status all have a significant effect on Alzheimer's disease. Age has no significant effect on Alzheimer's disease, while Gender, Duration of Years of Education, Level of Socioeconomic Status have significant effect on Alzheimer's disease. It is also worth mentioning that

Duration of Years of Education, Level of Socioeconomic Status have positive effect on Alzheimer's disease.

Obviously, this paper has its limitations. Firstly, due to the small sample size, the modeling results have some limitations. Secondly, more models should be used to analyze the data rather than a single multiple linear regression model. Also, the data in this paper does not cover all age groups and the duration of Years of Education is only six years or more, which may affect the accuracy of the results.

However, it is worth noting that this study is still worth exploring. First of all, the visualization of the charts and graphs directly and concisely analyzes the comparisons between the dependent variables, independent variables, and independent variables. For example, the researcher can clearly understand the percentage of people in the sample who have Alzheimer's disease, the percentage of patients in each independent variable, and so on. Secondly, unlike other studies that utilize a single factor approach, this study involves multiple variables and uses multiple linear regression modeling to screen out not only the influential factors, but also how Duration of Years of Education and Level of Socioeconomic Status affect Alzheimer's Disease. All of these findings can help predict and prevent Alzheimer's disease.

Even though today's technology cannot completely cure Alzheimer's disease, the number of people with Alzheimer's disease will decrease as people's Duration of Years of Education and Level of Socioeconomic Status increase.

Authors Contribution

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

References

- [1] Liu Xia. Can we cure Alzheimer's in the next decade. *Science and Technology Daily*, 2022.
- [2] Xinxin, Wu Jun, Panorama. Brain health, old age health. *Health Home*, 2013, (09): 32-42.
- [3] Li Xiang, Gao Shen. Analysis of lung cancer incidence, prevalence, and mortality trends among Chinese residents from 1990-2019. *Chinese Journal of Prevention and Control of Chronic Diseases*, 2021, 29(11): 821-826.
- [4] Zhu M, Shi H F. Progress in epidemiological studies of aluminum and Alzheimer's disease. *The 8th National Camp on Integrative Medicine and Western Medicine*, 2017.
- [5] Zhang Zhenxin. Topic:Focus on Alzheimer's disease. *Science Today*, 2013, (13): 74-78.
- [6] Liang L, et al. Meta-analysis of the association between smoking and Alzheimer's disease. *Occupation and Health*, 2021, (01): 103-106.
- [7] Song Z J, et al. A study on the relationship between cognition and depression in Alzheimer's disease and its influencing factors based on multivariate multilevel modeling. *China Health Statistics*, 2016, (01): 9-11.
- [8] Zhou LY, et al. Health related quality of life in Alzheimer's disease based on Latent and observable variables growth curve mode State change research. *China Health Statistics*, 2014.
- [9] Andruff H, Carraro N, Thompson A, Gaudreau P. Latent class growth modeling: a tutorial. *Tut Quant Meth Psych*, 2009, 5(1): 11-24.
- [10] Wang Lianyuan, et al.. Bayesian quantile regression joint modeling of risk factors for the development of Alzheimer's disease with different MMSE scores. *Journal of Jilin University (Medical Edition)*, 2023, (02): 395-401.
- [11] Yi Jun, Zhang Daopeng. Analysis of quality of life and its influencing factors in Alzheimer's disease patients. *China Public Health*, 2015, 31(11): 5.