Proceedings of the 2023 International Conference on Mathematical Physics and Computational Simulation
DOI: 10.54254/2753-8818/28/20230426

Exploring the relationship between mortality and potential risk factors in US 50 states: A population perspective using partial least squares and geographically weighted regression models

Chang Guo

Department of Statistical Science, Baylor University, Waco, TX76706, US

chang_guo2@baylor.edu

Abstract. While Alzheimer's disease becomes prevalent in elder population and attracts investment of thousands of billions for its research, its pathogenesis remains unknown. The relationship between multiple potential risk factors (Overall Health, Caregiving, cognitive decline, Nutrition/Physical Activity/Obesity, Screenings and Vaccines, mental health, Smoking and Alcohol Use) and Alzheimer's disease mortality in the 50 US states in 2020 was explored by developing multiple linear regression models, partial least squares regression models, and geographically weighted regression models in this article. In this experiment, through multiple linear regression models, we found eight significant demographic indicator variables, and to solve the covariance problem, we successfully constructed the pls model, and using the regression coefficients in the equation, we screened out the most important variables for the model, TOC06 and sex. In addition, we then explored the characteristics of the spatial distribution of mortality in Alzheimer's disease according to TOC06 and sex variables using the GWR model.

Keywords: Multiple Linear Regression Models, Partial Least Square Regression, Alzheimer's Disease, Risk Factors, Geographically Weighted Regression Model.

1. Introduction

Investigations on Alzheimer's disease (AD) spring up as population aging accelerates. AD is a neurodegenerative disease marked by characteristic neuropathy in the cerebrum, gradually impairing higher cognitive function. As Alzheimer's disease progresses, memory, language, and personality decline. Eventually, the patients are unable to do the simplest tasks. Driven by longer life spans and modern medicine, AD beat chronic diseases (e.g., diabetes, liver cirrhosis, nephritis) and became the US's third leading cause of death among adults over 65 years old in 2019 [1]. In 2006, the global base of AD patients ranged from 11.4 million to 59.4 million, and the global incidence of AD is expected to quadruple by 2050 to a total of between 47.2 and 221.2 million, which is equivalent to one AD patient for every 85 individuals [2]. The increase in the base of AD patients will undoubtedly be a huge burden for society's entire health and welfare system. Although evidence from many observational trials on AD has accumulated in recent years suggesting several potential risk factors [3], more profound research is required to explore the various demographic indicators corresponding to each risk factor theme. This

DOI: 10.54254/2753-8818/28/20230426

study examines the relationship between demographic indicator data and AD mortality in the 50 states of the United States in 2020 under various thematic potential risk factors for AD using partial least squares regression models and geographically weighted regression models. The dataset used in this study includes population-attributable risk proportion of AD risk factors for over-50s in each state, as well as the corresponding sex and age population totals and the number of deaths associated with AD in each state, obtained from the CDC [4].

2. Literature Review

2.1. Classification of Potential Risk Factors for Alzheimer's Disease

According to a 2021 article published by the Alzheimer's Association, it was estimated that 6.2 million Americans over 65 years old have Alzheimer's dementia. Despite numerous risk factors being associated with Alzheimer's disease (AD) in recent literature reviews, the challenge of elucidating how these diverse risk factors interplay in AD and incorporating them into a comprehensive pathogenic explanation persists. A Armstrong classified these risk factors into seven categories: demographic, genetic, lifestyle, medical, psychiatric, environmental, and infection in 2019 [5].

2.2. Age Group 1

Considering the fundamental demographic perspective, it is well-established that age is a significant determinant of sporadic AD. Key factors correlating with AD pathogenesis, such as ribosomal disturbance, phosphorylating oxidation, gene mutation associated with hypothalamic-pituitary-adrenal axis action, and alterations in insulin signaling, are commonly linked to advanced age [5].

2.3. Sex 2

Sex also exhibits notable disparities in AD prevalence, with an estimated 3.3 million women and 2 million men among the 5.3 million individuals over 65 years old diagnosed with AD in the United States [6]. Therefore, both age and gender emerge as critical covariates necessitating consideration in our research, as supported by the existing body of literature.

2.4. Nutrition/Physical Activity/Obesity (TNC03, TNC04) 3

Mid-life suboptimal health conditions have been recognized as a prominent risk factor for AD. Notably, conditions such as mid-life obesity and hyperlipidemia (elevated cholesterol) have been found to exert age-dependent effects on AD development [7]. In addition, regular leisure-time physical activity among middle-aged adults tends to be a protective factor for AD in later life [8]. Risk indicator factors regarding exercise habits at leisure (TNC03) and obesity rates (TNC04) are included in the model.

2.5. Mental Health (TMC01, TMC03) 4

From a mental health and distress perspective, it is essential to note that depression is a common psychiatric comorbidity in dementia [9]. Furthermore, the top 10 percent of the population who experience high distress have a 1-fold higher risk of developing AD than the bottom 10 percent [10]. Exploring the demographic distribution of anxiety and depression within specific age groups can offer valuable insights into the mental dimensions of AD. Although the majority of randomized controlled trials (RCTs) investigating multidomain interventions for AD prevention have shown limited efficacy, it has been emphasized by Crous-Bou et al. that identifying individuals at risk of developing the disease is crucial for successful intervention studies [11]. It will provide solid evidence for mental status examination as a promising tool for early dementia or AD screening, if the relationship between these mental factors and AD can be clarified.

2.6. Overall Health (TOC05, TOC06, TOC07, TOC08, TOC09, TOC10, TOC11, TOC13) 5

The impact of AD on overall health has been investigated in several studies. In a controlled trial involving elderly patients, both mild and moderate Alzheimer's patients were found to have an awareness

DOI: 10.54254/2753-8818/28/20230426

of their declining ability to perform activities of daily living (ADL), despite experiencing a cognitive decline [12]. Additionally, the association between ADL limitation, cognitive impairment, and fall injuries has been established, although the specific contribution of each factor remains unclear [13]. Inflammation has emerged as a potential player in the pathogenesis of AD, as indicated by studies highlighting a higher prevalence of Alzheimer's in patients with arthritic diseases compared to non-arthritic individuals [14]. Sleep disturbances are common among AD patients and various epidemiological studies have demonstrated a strong correlation between cognitive decline and altered sleep patterns [15]. Oral health is another aspect of overall health affected by AD. Using the General Oral Health Assessment Index (GOHAI), researchers found that AD patients exhibited worse oral health than their non-affected counterparts [16]. Furthermore, cognitive performance, as assessed by the Mini-Mental Status Examination (MMSE), emerged as the primary independent predictor of disability. This finding emphasizes the significant impact of cognitive impairment on the development of disability in individuals with AD. As cognitive function declines, the likelihood of experiencing disability in ADL increases [17].

2.7. Screenings and Vaccines (TSC02, TSC08, TSC09) 6.

Numerous studies have investigated the relationship between AD risk and primary prevention strategies like screenings and vaccines. In a study that accounted for confounding variables such as age, sex, and education, it was discovered that individuals who had received vaccines against diphtheria or tetanus, polio, and influenza in the past exhibited a lower risk of developing AD [18]. In the realm of colorectal cancer screening, it was observed that although the overall screening rate for older adults showed an increase from 86.4% to 88.96% between 2015 and 2018, patients with AD were 39% less likely to undergo any form of colorectal cancer screening in comparison to older adults without Alzheimer's disease and related dementias (ADRD) [19]. Furthermore, the findings also strongly suggest that receiving a pneumonia vaccination between the ages of 65 and 75 is linked to a reduced risk of developing AD later in life. Consequently, investigating the pneumococcal vaccination rate within specific demographic groups becomes crucial in comprehending the factors influencing mortality from AD [20].

2.8. Smoking and Alcohol Use (TAC01, TAC03) 7

Since nicotine can compensate for some of the cholinergic deficits observed in AD, moderate smoking plays a protective role in preventing AD to some extent. Combining epidemiological and pharmacological perspectives, although it is uncertain whether alcohol consumption is a risk factor influencing the incidence of AD, the correlation between alcohol consumption and smoking requires that we include both variables in our study [21].

Recent research has demonstrated dynamic changes in the most prominently modifiable risk factors associated with AD over the past decade, suggesting an evolving relationship between AD and these factors [22]. This underscores the importance of regularly updating models that elucidate the complex relationship between multidimensional risk factors and Alzheimer's mortality. Despite identifying these risk factors in the existing literature, AD remains a complex and multifactorial condition with gaps in our understanding of its precise mechanisms. Therefore, investigating trends and associations remains imperative for further comprehending this disease.

3. Methodology

3.1. Dataset Description

The risk factors data analyzed in this study were obtained from a cross-sectional survey conducted by the Centers for Disease Control and Prevention (CDC) from January 2020 to December 2020. The survey was conducted using the Behavioral Risk Factor Surveillance System (BRFSS), the nation's premier health-related cell phone and landline survey system that collects state data about the United States. It was designed to investigate the proportion of different behavioral risk factors (Appendix A)

DOI: 10.54254/2753-8818/28/20230426

associated with Alzheimer's Disease among individuals aged 50 years or older, stratified by age group (50-64 and greater than 65 years), gender, and race.

The mortality data used in this study were sourced from the CDC's comprehensive database, which provided information on the number of deaths due to Alzheimer's Disease by age and gender and the relevant age population of 50 states in the United States. The mortality data included deaths resulting from E85 (Amyloidosis), F01-F09 (Organic, including symptomatic, mental disorders), G30 (Alzheimer's disease), and G31.1 (Senile degeneration of the brain, not elsewhere classified). The mortality data were based on information from all death certificates filed in the fifty states and the District of Columbia. Deaths of nonresidents were not included in the dataset.

3.2. Dataset Processing

Given the accuracy and completeness of the data, we first examined the presence of any vacant values in both datasets. We then utilized the set.seed function to generate random numbers from 0-9 to fill in approximately 14% of the suppressed values (counts <10) in the Alzheimer's death dataset. To investigate the relationship between Alzheimer's mortality and risk factors in the 50 US states, we integrated the two datasets using the software R 4.2.3 version. Since age and gender are known to significantly influence the prevalence of Alzheimer's disease, we used the filter function in R to categorize the proportion of different behavioral risk factors by state, age group (50-64 and >65), and gender in 2020. We then grouped the AD death count dataset by state using the group by function based on the same criteria. Next, we merged the two datasets based on state, age group, and gender using the join function of the dplyr package. By combining the datasets, we were able to analyze the impact of different behavioral risk factors on AD mortality in various age and gender groups across the 50 US states.

3.3. Statistical Analysis

The dataset was partitioned into thematic categories, namely Overall Health, Caregiving, Cognitive Decline, Nutrition/Physical Activity/Obesity, Mental Health, Screenings and Vaccines, and Smoking and Alcohol Use, to facilitate subsequent analysis. We employed multiple linear regression models to comprehensively explore the potential associations between these categories and Alzheimer's Disease (AD) mortality. Within each category, we performed population scaling for each type of refined question to enhance the validity and generalizability of our models. Given the widely recognized role of age as the most significant risk factor for sporadic AD and the notable gender-based discrepancies in AD incidence and prevalence [6], we included age group and gender as covariates in our models. We incorporated all questions related to each thematic category, as well as age group and sex, as covariates. The dependent variable in our models was AD mortality.

Based on the multiple model results and the corresponding scatterplot results, we remove the factors with the proportion of missing values greater than 20% and use the Partial Least Squares (PLS) Regression model to cope with the problem of high co-linearity between the refinement variables corresponding to each topic. PLS regression is a modern method that integrates and extends the principles of principal component analysis and multiple regression. Its purpose is to forecast or analyze a group of response variables based on a group of predictor or independent variables [23]. Through cross-validation of the PLS model, 7 new principal components were identified and used to fit the new regression model. The model used Alzheimer's Disease (AD) mortality as the dependent variable and included age, sex, and various thematic categories as covariates, such as Overall Health (TOC05, TOC06, TOC07, TOC08, TOC09, TOC10, TOC11, TOC13), Mental Health (TMC01, TMC03), Nutrition/Physical Activity/Obesity (TNC03, TNC04), Screenings and Vaccines (TSC02, TSC08, TSC09), Smoking and Alcohol Use (TAC01, TAC03) as independent variables. By using the PLS modeling approach, we were able to observe the association between data on different demographic indicators under each category of risk factors and the corresponding Alzheimer's disease (AD) mortality rates. This approach allowed us to consider a large number of covariates to avoid multicollinearity

DOI: 10.54254/2753-8818/28/20230426

between variables in the same category while screening out data on demographic indicators with a large impact on mortality from Alzheimer's disease.

To further examine the potential spatial heterogeneity in Alzheimer's disease (AD) mortality and its potential demographic associations, we applied a Geographically Weighted Regression (GWR) model. The GWR model is a statistical technique to explore the amount of spatial variation between predictor and outcome variables within a single model [24]. GWR models utilize a differential weighting scheme to generate a separate model for each spatial location in the study area from the same data set, thus allowing for spatial variation in model parameters involving a bandwidth parameter, which is typically determined from the data through a cross-validation procedure [25]. Specifically, we chose the same covariates and dependent variables in the GWR model as in the PLS model. Combined with previous scatter plots of each risk factor grouped by age, we found no significant relationship between mortality and demographic indicator variables for the under-65 age group. Therefore, we focused our GWR analysis on the 65 year and older group to investigate whether there was some relationship between the spatial distribution of demographic indicator data.

4. Result

A demographic and mortality data frame is presented in Table 1 and Table 2.

Table 1. Summary statistics table shows all covariates that selected in the Partial Least Squares (PLS) Regression model.

	Min	1Q	Median	Mean	3Q	Max
Age adjusted mortality	0.064	0.742	5.378	62.96	92.46	601.71
≥ 50 and ≤ 64 yrs rate	0.141	0.184	0.19	0.191	0.199	0.225
≥ 65 yrs rate	0.163	0.172	0.173	0.183	0.216	0.163
Female, ≥ 50 and ≤ 64 yrs rate	0.071	0.093	0.098	0.098	0.103	0.115
Female, ≥ 65 yrs rate	0.062	0.09	0.095	0.095	0.1	0.118
Male, ≥ 50 and ≤ 64 yrs rate	0.07	0.09	0.093	0.093	0.095	0.111
Male, ≥ 65 yrs rate	0.05	0.073	0.076	0.0781	0.084	0.099

Note. The demographic and mortality data for all types of risk variables for the 50 states in the United States in 2020 were analyzed using R version 4.2.3. All data are displayed in Table 1, which contains the means, medians, standard deviations, and Quartiles for all variables.

Table 2. Summary statistics table shows all covariates selected in the Partial Least Squares (PLS) Regression model.

	Categories	Mean±SD	Median (IQR)
Overall Health	TOC05	73.14 ±13.86	74.35 (67.93-82.2)
	TOC06	9.29±3.18	9.35 (7.2-11.6)
	TOC07	19.48±5.9	18.6 (15.35-23.25)
	TOC08	78.56±12.3	81.1 (76.18-84.33)
	TOC09	68.96±11.36	70.7 (65.3-74.9)
	TOC10	34.6±10.32	35.8 (26.675-41.75)
	TOC11	41.54±11.94	42.2 (33.625-50.8)
	TOC13	28.72±7.69	28.25 (24.1-32.7)
Mental Health	TMC01	9.91±4.36	9.05 (7-12.7)
	TMC03	16.63±6.97	15.6 (11.68-20.8)

DOI: 10.54254/2753-8818/28/20230426

Table 2. (continued).

Nutrition/Physical Activity/Obesity	TNC03	27.34±7.23	27.15 (23.05-31.78)
	TNC04	32.83±7.62	32.55 (28.58-37.83)
Screenings and Vaccines	TSC02	73.91±13.95	74.9 (67.08-83.9)
	TSC08	57.05±13.48	59.05 (48.55-67.33)
	TSC09	53.73±19	51.8 (37.1-71.28)
Smoking and Alcohol Use	TAC01	13.42±5.88	12.35 (9.08-17.73)
	TAC03	8.8±5.76	7.5 (4.2-12.63)

4.1. Multiple Linear Regression Model

These regressions from different thematic categories suggested that eight covariates (TCC01, TAC03, TMC01, TNC03, TOC06, TOC11, TSC03, TSC11) from risk factors demonstrate statistical significance (Appendix B). When we performed the diagnosis of the multiple linear regression models, as residual vs fitted plots appeared in clusters of multiple groups evident in almost all linear regression models for different topics, in order to further explore the reasons for this, we plotted two categories reflecting the different variables between the models for each topic according to the age group classification (50 to 64 years, 65+ years) as well as the gender classification of scatter plots (Appendix C). Based on the characteristics of the scatter plots between some of the variables, we conjecture that there was a strong correlation between some of the problem variables under one category of themes and based on the subsequent Variance inflation factor (VIF) values (Appendix D), we determined that there was a serious covariance problem in the multiple linear regression models classified by themes for each risk factor.

4.2. Partial Least Square Model

Based on the severe multicollinearity problem in the multiple linear regression model, we used a partial least squares (PLS) model. Before constructing the model, among all covariates, we removed demographic indicator variables with vacancy values greater than 20%. Combined with the scatter plots according to age in the multiple linear regression model (Appendix C), therefore, in constructing the PLS model, we selected only the population over 65 years old, and the final covariates retained were TAC01, TAC03, TMC01, TMC03, TNC03, TNC04, TOC05, TOC06, TOC07, TOC08, TOC09, TOC10, TOC11, TOC13, TSC02, TSC08, TSC09, and sex, with the dependent variable being Alzheimer's disease mortality. We used 80 percent of the data as a fit and 10 percent as a prediction, and the prediction model was a good fit based on an R^2 value of 0.88 and an RMSE value equal to 0.551. According to all cross-validation and RMSE plot (Figure 1) results, the optimal number of components should be 7. Among them, according to the X^2 explained variance rate plot (Figure 2), we can see that the explanation rate of component 1 is about 32%, the explanation rate of component 2 is about 28%, the explanation rate of component 3 is about 10%, and the remaining components 4 to 7 are less than 5%. The detailed predictor weights of each component can be found in Appendix E. Considering the cross-validation and RMSE image results, we decided to keep 7 components to build the pls regression model. Thus, the final pls regression equation:

$$\begin{aligned} y = {} & 0.042x_{TAC01} - 0.022x_{TAC03} - 0.155x_{TMC01} + 0.205x_{TMC03} + 0.181x_{TNC03} \\ & - 0.021x_{TNC04} + 0.117x_{TOC05} - 0.278x_{TOC06} - 0.025x_{TOC07} \\ & + 0.025x_{TOC08} - 0.059x_{TOC09} + 0.07x_{TOC10} + 0.233x_{TOC11} \\ & + 0.035x_{TOC13} - 0.003x_{TSC02} + 0.098x_{TSC08} + 0.18x_{TSC09} + 0.586x_{sex} \end{aligned}$$

Using the regression coefficients in the equation, regression coefficients plots and the VIP scores plots (Appendix E), we found that the variables TMC01, TMC03, TNC03, TOC06, TOC11, TSC09, and sex had a more significant effect on mortality from Alzheimer's disease. Among them, TMC01 and TMC06 were inversely associated with Alzheimer's disease mortality.

Finally, the comparison of the predicted and observed values is shown in the following Figure3. Although there is some small difference between the predicted and observed values, the model is good

DOI: 10.54254/2753-8818/28/20230426

overall, as evidenced by the coefficient of determination R^2 close to 1 and the relatively small Root Mean Square Error. An informative figure legend and brief title should accompany each graph.

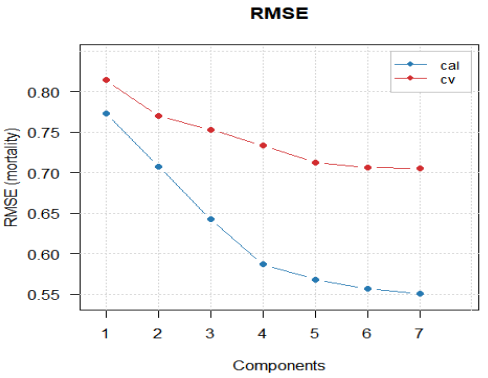


Figure 1. RMSE for PLS Model.

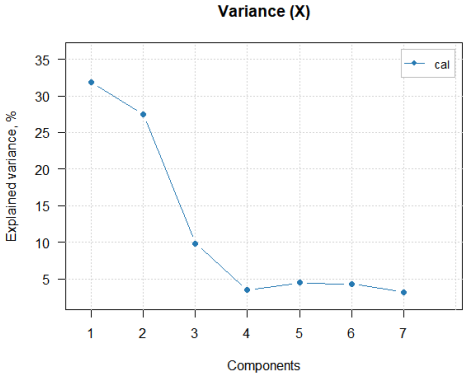


Figure 2. X^2 Explained Variance, %.

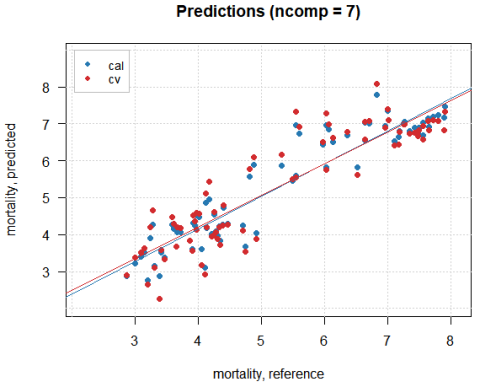


Figure 3. Predictions Plot for the PLS Model.

4.3. Geographically Weighted Regression Model

Based on the results of the pls model and previous studies in the literature, we can identify gender as a very important risk factor influencing mortality from Alzheimer's disease. Combining the regression coefficient plot and the VIP scores plot, we want to explore the association between the demographic indicator variable TOC06 and the sex ratio at the spatial level and mortality from Alzheimer's disease. Therefore, we mapped the demographic indicator variable TOC06 with respect to different sexes as well as different sex ratios based on the GWR, as detailed in Figure 4.

DOI: 10.54254/2753-8818/28/20230426

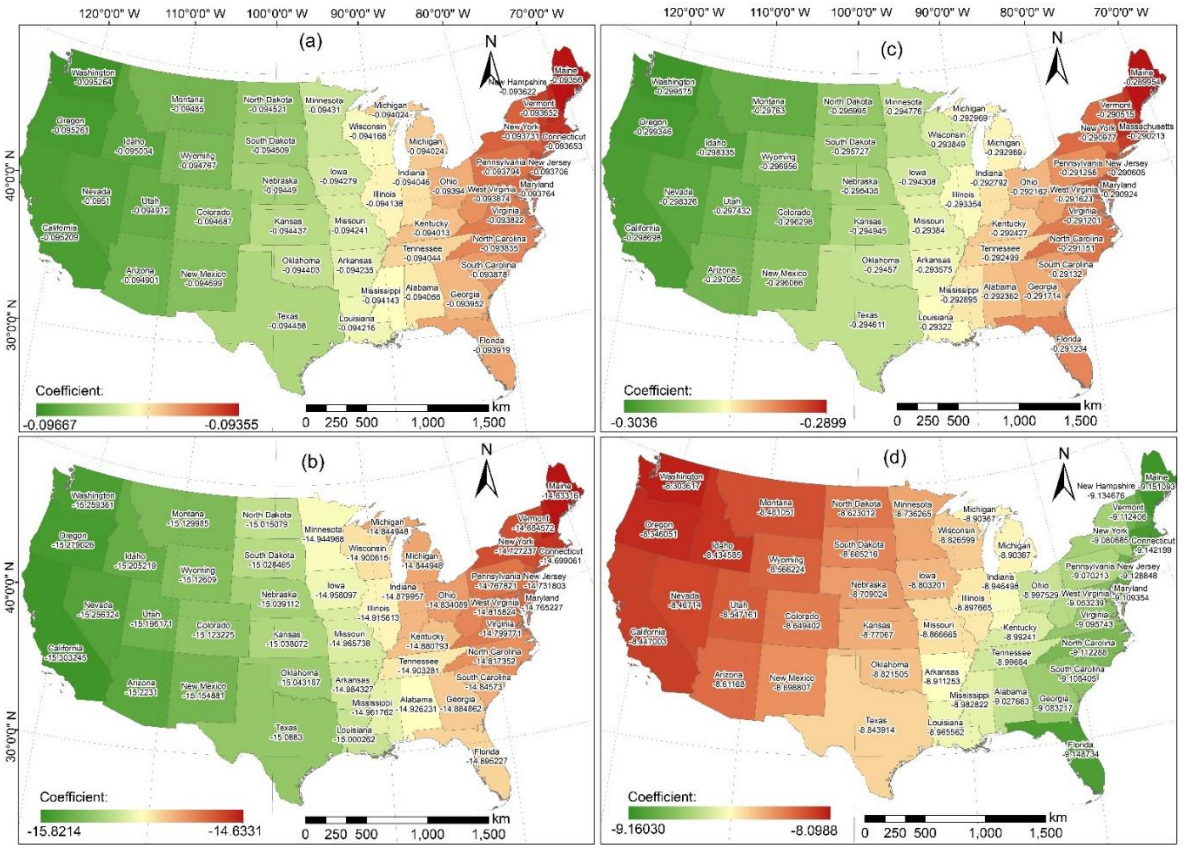


Figure 4. The demographic indicator variable TOC06 with respect to different sexes as well as different sex ratios. (a). Coefficient values of toc06 for the 65+ male population in the 50 US states. (b). Coefficient values for the proportion of males in the 65+ population in the 50 U.S. states. (c). Coefficient values of toc06 for the female population over 65 years of age in the 50 states of the United States. (d). Coefficient values of the proportion of the female population in the 65+ age group in the 50 states of the United States.

From the overall analysis in Figure 4a and Figure 4c, the effect of the tac06 variable on Alzheimer's disease mortality is much smaller in the group of men over 65 years old than in the group of women over 65 years old in all 50 US states. And the larger absolute values of the coefficient of tac06 were mainly concentrated in the western coastal United States, showing an east-west decreasing trend.

From the perspective of sex ratio distribution in Figure 4b and Figure 4d, the larger absolute value coefficient of female sex ratio is mainly concentrated in the eastern United States, and conversely, the larger absolute value coefficient of male sex ratio is mainly concentrated in the west coast of the United States. The absolute values of sex ratio coefficients show a decreasing trend from east to west.

5. Discussion

The results suggest that the demographic indicators TMC01, TMC03, TNC03, TOC06, TOC11, TSC09, and gender have a greater impact on mortality from Alzheimer's disease in the 50 US states in 2020. Based on the regression coefficients of the PLS model, we found that the TOC06 variable and the gender variable were the two variables with the largest absolute values of the regression coefficients. According to the results of the GWR model, we found that both the TOC06 variable and the gender variable played a negative effect on Alzheimer's mortality rate. Compared to the female population over 65 years of age, the toc06 variable had a very weak effect on mortality from Alzheimer's disease in the male population over 65 years of age; combined with the gender differences in injury diagnoses in the elderly population, women were generally injured at a higher rate than men, especially for fractures, which were 2.2 times

DOI: 10.54254/2753-8818/28/20230426

higher in women than in men. Females also had higher injury rates for all body parts, with the most significant rates for leg/foot injuries (2.3), arm/hand injuries (2.0), and lower trunk injuries (2.0) [26]. This, therefore, explains, to some extent, the large differences in the degree of effect of the TOC06 variable on mortality from Alzheimer's disease by gender. Combining the two gender groups, the areas most affected by the toc06 variable were concentrated in the west coast states of the United States, with a decreasing trend of effect from west to east. Although the reported fall rates for the older population in the U.S. states in 2020 did not show a clear pattern of spatial distribution [27], it is possible that the elder individuals in the U.S. West Coast states had higher rates in other injury categories.

In addition, on the map of male and female sex ratio coefficients, we can clearly observe that Alzheimer's disease mortality rates are more influenced by female sex ratios in the eastern United States and by male sex ratios in the western United States. Combined with the Alzheimer's disease mortality distribution map for the female population over 65 years of age in Appendix F, we found that the states with higher Alzheimer's disease mortality rates in the older female population were concentrated in the eastern United States, which is consistent with the trend we observed for the female sex ratio coefficient, but since this variable plays a negative role in Alzheimer's disease mortality, we speculate that there should be other important risk factors that play a positive effect.

We did not exhaustively investigate the impact of several other significant variables on AD mortality. Instead, our focus was solely on two variables, TOC06 and sex, in the Weighted Regression (GWR) model. Consequently, we might have overlooked the influence of additional important factors, such as ethnicity. The influence of ethnicity on AD mortality in each state was not considered in our analysis. Ethnicity is a significant demographic factor that could potentially impact AD mortality rates.

Earlier in 2015, Baumgart and his colleagues affirmed the significance of cognitive levels, such as years of formal education, in reducing the possibility of developing AD [7]. Occupational complexity, level of education, and the number of cognitive leisure activities are all key factors affecting cognitive function in old age and significantly impact AD events. However, in this paper, we did not use data in this area, which may have some impact on the model we constructed.

6. Conclusion

In conclusion, this study aims to investigate the potential risk factors associated with Alzheimer's disease mortality in the 50 US states in 2020 using multiple linear, partial least squares, and geographically weighted regression models. The result indicated eight significant demographic indicator variables. The PLS model was utilized to identify the most critical variables, which were TOC06 and gender factors. Additionally, the part of the analysis of the spatial distribution of mortality in Alzheimer's disease undoubtedly provided new insights into the research and knowledge of this disease. Overall, this study provides valuable information on the potential risk factors associated with Alzheimer's disease mortality. It highlights the importance of considering demographic indicators and spatial distribution in understanding Alzheimer's disease.

This study has several essential implications for public health policy. Identification of significant risk factors such as overall health and caregiving will be helpful to policymakers and healthcare professionals to develop specific and individual interventions to reduce Alzheimer's disease mortality. Moreover, multiple regression models lead to a more comprehensive analysis of the potential risk factors associated with Alzheimer's disease mortality in different geographic locations. This information can be equipped to develop tailored interventions based on the specific needs of other regions.

Overall, this study presented a comprehensive review of mortality and risk factors concerning Alzheimer's disease in 50 states in US with multiple analysis models, allowing further research and medical interventions based on this foundation.

References

[1] Ho, J. Y., & Franco, Y. (2022). The rising burden of Alzheimer’s disease mortality in rural America. *SSM - Population Health, 17*, 101052. https://doi.org/10.1016/j.ssmph.2022.101052

DOI: 10.54254/2753-8818/28/20230426

[2] Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*, *3*(3), 186–191. https://doi.org/10.1016/j.jalz.2007.04.381

[3] Zhang, X.-X., Tian, Y., Wang, Z.-T., Ma, Y.-H., Tan, L., & Yu, J.-T. (2021). The Epidemiology of Alzheimer's Disease Modifiable Risk Factors and Prevention. *The Journal of Prevention of Alzheimer's Disease*, *8*(3), 313–321. https://doi.org/10.14283/jpad.2021.15

[4] Centers for Disease Control and Prevention, National Center for Health Statistics. National Vital Statistics System, Mortality 1999-2020 on CDC WONDER Online Database, released in 2021. Data are from the Multiple Cause of Death Files, 1999-2020, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov/ucd-icd10.html on May 26, 2023 10:55:26 AM

[5] A Armstrong, R. (2019). Risk factors for Alzheimer's disease. *Folia Neuropathologica*, *57*(2), 87–105. https://doi.org/10.5114/fn.2019.85929

[6] Nebel, R. A., Aggarwal, N. T., Barnes, L. L., Gallagher, A., Goldstein, J. M., Kantarci, K., Mallampalli, M. P., Mormino, E. C., Scott, L., Yu, W. H., Maki, P. M., & Mielke, M. M. (2018). Understanding the impact of sex and gender in Alzheimer's disease: A call to action. *Alzheimer's & Dementia*, *14*(9), 1171–1183. https://doi.org/10.1016/j.jalz.2018.04.008

[7] Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H., & Johns, H. (2015). Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*(6), 718–726. https://doi.org/10.1016/j.jalz.2015.05.016

[8] Rovio, S., Kåreholt, I., Helkala, E.-L., Viitanen, M., Winblad, B., Tuomilehto, J., Soininen, H., Nissinen, A., & Kivipelto, M. (2005). Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. *The Lancet Neurology*, *4*(11), 705–711. https://doi.org/10.1016/S1474-4422(05)70198-8

[9] Novais, F., & Starkstein, S. (2015). Phenomenology of Depression in Alzheimer's Disease. *Journal of Alzheimer's Disease*, *47*(4), 845–855. https://doi.org/10.3233/JAD-148004

[10] Wilson, R. S., Evans, D. A., Bienias, J. L., Leon, C. F. M. de, Schneider, J. A., & Bennett, D. A. (2003). Proneness to psychological distress is associated with risk of Alzheimer's disease. *Neurology*, *61*(11), 1479–1485. https://doi.org/10.1212/01.WNL.0000096167.56734.59

[11] Crous-Bou, Marta, Carolina Minguillón, Nina Gramunt, and José Luis Molinuevo. "Alzheimer's Disease Prevention: From Risk Factors to Early Intervention." Alzheimer's Research & Therapy 9, no. 1 (September 12, 2017): 71. https://doi.org/10.1186/s13195-017-0297-z.

[12] Kiyak, H. A., Teri, L., & Borson, S. (1994). Physical and Functional Health Assessment in Normal Aging and in Alzheimer's Disease: Self-Reports vs Family Reports1. *The Gerontologist*, *34*(3), 324–331. https://doi.org/10.1093/geront/34.3.324

[13] Oleske, D. M., Wilson, R. S., Bernard, B. A., Evans, D. A., & Terman, E. W. (1995). Epidemiology of Injury in People with Alzheimer's Disease. *Journal of the American Geriatrics Society*, *43*(7), 741–746. https://doi.org/10.1111/j.1532-5415.1995.tb07042.x

[14] McGeer, P. L., Schulzer, M., & McGeer, E. G. (1996). Arthritis and anti-inflammatory agents as possible protective factors for Alzheimer's disease: A review of 17 epidemiologic studies. *Neurology*, *47*(2), 425–432. https://doi.org/10.1212/WNL.47.2.425

[15] Peter-Derex, L., Yammine, P., Bastuji, H., & Croisile, B. (2015). Sleep and Alzheimer's disease. *Sleep Medicine Reviews*, *19*, 29–38. https://doi.org/10.1016/j.smrv.2014.03.007

[16] Ribeiro, G. R., Costa, J. L. R., Bovi Ambrosano, G. M., & Rodrigues Garcia, R. C. M. (2012). Oral health of the elderly with Alzheimer's disease. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *114*(3), 338–343. https://doi.org/10.1016/j.oooo.2012.03.028

[17] Zanetti, O., Bianchetti, A., Frisoni, G. B., Rozzini, R., & Trabucchi, M. (1993). Determinants of disability in alzheimer's disease. *International Journal of Geriatric Psychiatry*, *8*(7), 581–586. https://doi.org/10.1002/gps.930080709

DOI: 10.54254/2753-8818/28/20230426

[18] Verreault, R., Laurin, D., Lindsay, J., & Serres, G. D. (2001). Past exposure to vaccines and subsequent risk of Alzheimer's disease. *CMAJ*, *165*(11), 1495–1498.
[19] Lv, G., Wang, X., Jiang, X., Li, M., & Lu, K. (2022). Impact of Alzheimer's disease and related dementias on colorectal cancer screening utilization, knowledge, and associated health disparities. *Frontiers in Pharmacology*, *13*, 872702. https://doi.org/10.3389/fphar.2022.872702
[20] Ukraintseva, S., Yashkin, A., Duan, M., Akushevich, I., Arbeev, K., Wu, D., Stallard, E., Tropsha, A., & Yashin, A. (2020). Repurposing of existing vaccines for personalized prevention of Alzheimer's disease: Vaccination against pneumonia may reduce AD risk depending on genotype. *Alzheimer's & Dementia*, *16*(S3), e046751. https://doi.org/10.1002/alz.046751
[21] Tyas, S. (1996). Are tobacco and alcohol use related to Alzheimer's disease? A critical assessment of the evidence and its implications. *Addiction Biology*, *1*(3), 237–254. https://doi.org/10.1080/1355621961000124856
[22] Nianogo, R. A., Rosenwohl-Mack, A., Yaffe, K., Carrasco, A., Hoffmann, C. M., & Barnes, D. E. (2022). Risk Factors Associated With Alzheimer Disease and Related Dementias by Sex and Race and Ethnicity in the US. *JAMA Neurology*, *79*(6), 584–591. https://doi.org/10.1001/jamaneurol.2022.0976
[23] Abdi, H. (n.d.). *Partial Least Square Regression PLS-Regression*.
[24] Matthews, S. A., & Yang, T.-C. (2012). Mapping the results of local statistics: Using geographically weighted regression. *Demographic Research*, *26*, 151–166. https://doi.org/10.4054/DemRes.2012.26.6
[25] Byrne, G., Charlton, M., & Fotheringham, S. (n.d.). *Multiple Dependent Hypothesis Tests in Geographically Weighted Regression*.
[26] Stevens, J. A., & Sogolow, E. D. (2005). Gender differences for non-fatal unintentional fall related injuries among older adults. *Injury Prevention*, *11*(2), 115–119. https://doi.org/10.1136/ip.2004.005835
[27] *Older Adult Falls Reported by State | Fall Prevention | Injury Center | CDC*. (2023, February 28). https://www.cdc.gov/falls/data/falls-by-state.html

Appendix

Appendix A

Category	Variable	Description
Overall Health	TOC01	Duration of physical unhealthy status within last month.
	TOC03	Average duration of limitated movements within last month.
	TOC05	Proportion of elder individuals lost 5 or fewer teeth due to pathologic reason.
	TOC06	Proportion of elder individuals with physical injuries last year.
	TOC07	Proportion of elder individuals who self-reported suboptimal health.
	TOC08	Proportion of elder individuals who self-reported optimal health.
	TOC09	Proportion of elder individuals sleep over six hours per night.
	TOC10	Proportion of elder individuals with mental/physical disability.
	TOC11	Proportion of elder individuals with arthritis.
	TOC13	Elder individuals with suboptimal health diagnosed with arthritis.
Caregiving	TGC01	Proportion of elder individuals taking care of others within last month.

DOI: 10.54254/2753-8818/28/20230426

	TGC02	Proportion of elder individuals planning to take care of others in the following two years.
	TGC03	Proportion of elder individuals having taken care of others for over six months.
	TGC04	Proportion of elder individuals taking care of others over twenty hours per week.
Cognitive Decline	TCC01	Proportion of elder individuals with subjective cognitive decline or memory loss deteriorating within last year.
	TCC02	Proportion of elder individuals with subjective cognitive decline or memory loss impairing daily activity.
	TCC03	Proportion of elder individuals with subjective cognitive decline or memory loss too severe to complete daily activity.
	TCC04	Proportion of elder individuals with subjective cognitive decline or memory loss documented by a health care professional about it.
Mental Health	TMC01	Proportion of elder individuals under frequent mental distress.
	TMC03	Proportion of elder individuals with a lifetime diagnosis of depression.
Nutrition/Physical Activity/Obesity	TNC03	Proportion of elder individuals who have not had any leisure time physical activity in the past month.
Category	Variable	Description
	TNC04	Proportion of elder individuals who are currently obese, with a body mass index (BMI) of 30 or more.
Screenings and Vaccines	TSC01	Proportion of elder female who have received a mammogram within the past 2 years.
	TSC02	Proportion of elder individuals who had either a home blood stool test within the past year or a sigmoidoscopy or colonoscopy within the past 10 years.
	TSC03	Proportion of elder female with an intact cervix who had a Pap test within the past 3 years.
	TSC04	Proportion of elder individuals without diabetes who reported a blood sugar or diabetes test within 3 years.
	TSC08	Proportion of elder individuals injecting influenza vaccine within the past year.
	TSC09	Proportion of individuals with specidic risk factors(have diabetes, asthma, cardiovascular disease or currently smoke) who ever had a pneumococcal vaccine.
	TSC10	Proportion of elder male who are up to date with select clinical preventive services.
	TSC11	Proportion of elder female who are up to date with select clinical preventive services.
Smoking and Alcohol Use	TAC01	Proportion of elder individuals with a smoking history.
	TAC03	Proportion of elder individuals with excessive alcohol consumption within last mouth.

Note. The description of every variable of different categories in risk factors.

DOI: 10.54254/2753-8818/28/20230426

Appendix B

Results of Multiple Regression Models for Variables Corresponding to Demographic Indicators for Each Type of Risk Factor.

variables	beta	95%CI	p_value
TCC01	-0.039	(-0.076,-0.003)	0.0321
TCC02	0.022	(-0.003,0.047)	0.0903
TCC03	-0.015	(-0.042, 0.011)	0.2463
TCC04	-0.003	(-0.0191,0.0133)	0.724
65 years or older	5.316	(5.0621,5.57)	< 2e-16
Male	-1.376	(-1.626,-1.127)	<2e-16
Multiple R-squared: 0.909			
Adjusted R-squared: 0.906			

variables	beta	95%CI	p_value
TAC01	0.012	(-0.022,0.046)	0.479
TAC03	0.101	(0.054,0.147)	3.39E-05
65 years or older	6.129	(5.6,6.678)	< 2e-16
Male	-2.013	(-2.393,-1.634)	< 2e-16
Multiple R-squared: 0.913			
Adjusted R-squared: 0.911			

variables	beta	95%CI	p_value
TGC01	-0.02	(-0.119, 0.059)	0.509
TGC02	-0.003	(-0.102,0.095)	0.947
TGC03	0.016	(-0.007, 0.039)	0.179
TGC04	-0.022	(-0.061, 0.017)	0.266
TGC05	0.005	(-0.044, 0.053)	0.847
65 years or older	5.216	(4.956, 5.476)	< 2e-16
Male	-1.422	(-1.678,-1.166)	< 2e-16
Multiple R-squared: 0.906			
Adjusted R-squared: 0.903			

variables	beta	95%CI	p_value
TMC01	-0.096	(-0.162,-0.030)	0.004
TMC03	-0.014	(-0.057, 0.030)	0.528
65 years or older	4.671	(4.333, 5.010)	< 2e-16
Male	-1.946	(-2.330,-1.562)	< 2e-16
Multiple R-squared: 0.914			
Adjusted R-squared: 0.912			

DOI: 10.54254/2753-8818/28/20230426

variables	beta	95%CI	p_value
TNC03	0.04	(0.010, 0.069)	0.008
TNC04	-0.008	(-0.041, 0.025)	0.651
65 years or older	4.968	(4.532, 5.404)	< 2e-16
Male	-1.197	(-1.488,-0.907)	4.77E-14
Multiple R-squared: 0.909			
Adjusted R-squared: 0.907			

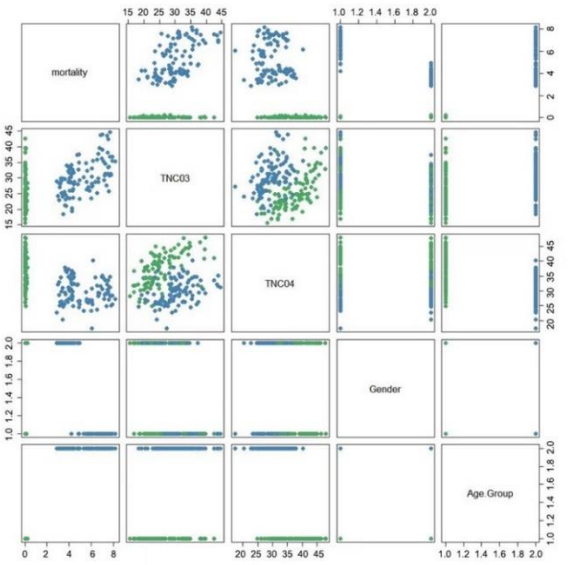
variables	beta	95%CI	p_value
TOC05	-0.008	(-0.043, 0.027)	0.637
TOC06	-0.125	(-0.187, -0.063)	9.35E-05
TOC07	-0.019	(-0.105, 0.068)	0.674
TOC08	NA	NA	NA
TOC09	-0.022	(-0.055, 0.010)	0.182
TOC10	-0.005	(-0.040, 0.030)	0.77
TOC11	0.074	(0.041,0.108)	1.89E-05
TOC13	-0.025	(-0.079, 0.029)	0.366
65 years or older	4.306	(3.626, 4.986)	< 2e-16
Male	-1.274	(-1.722, -0.826)	6.91E-08
Multiple R-squared: 0.924			
Adjusted R-squared: 0.920			

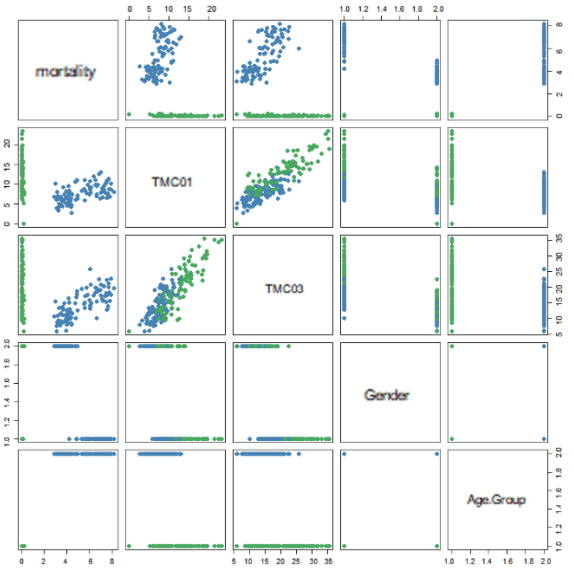
variables	beta	95%CI	p_value
TSC01	0.007	(-0.032, 0.046)	0.717
TSC02	-0.004	(-0.024, 0.017)	0.723
TSC03	-0.028	(-0.049,-0.007)	0.009
TSC04	-0.001	(-0.004,0.002)	0.461
TSC08	-0.009	(-0.036, 0.017)	0.489
TSC09	0.012	(-0.005, 0.029)	0.176
TSC10	0.011	(-0.027, 0.049)	0.574
TSC11	0.071	(0.042, 0.099)	2.79E-06
65 yeas or older	3.6	(2.987, 4.212)	< 2e-16
Male	-0.313	(-2.594, 1.969)	0.787
Multiple R-squared: 0.9668			
Adjusted R-squared: 0.9651			

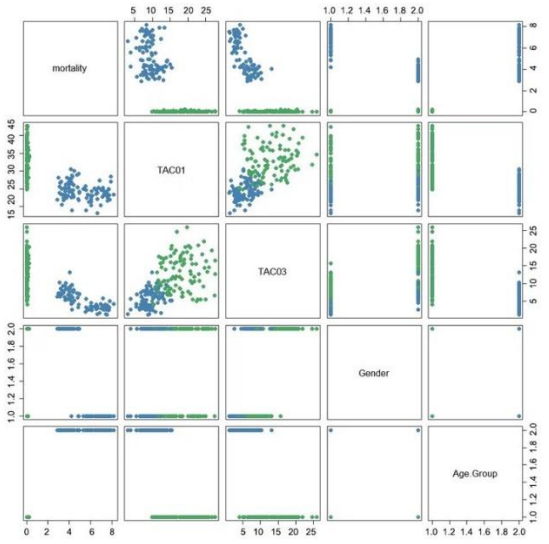
DOI: 10.54254/2753-8818/28/20230426

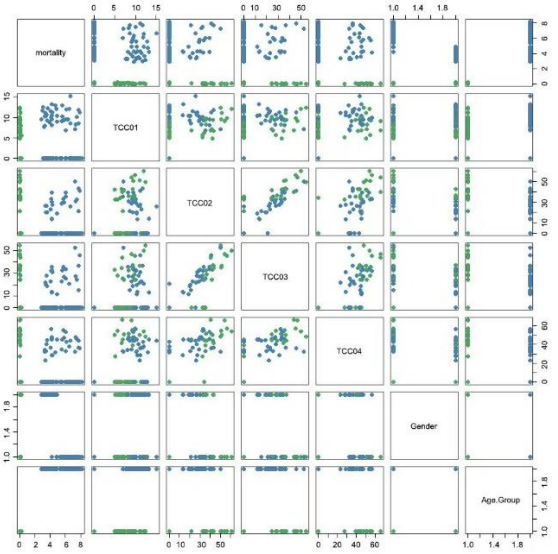
Appendix C

The Scatter Plots for Every Variable for Age Groups.

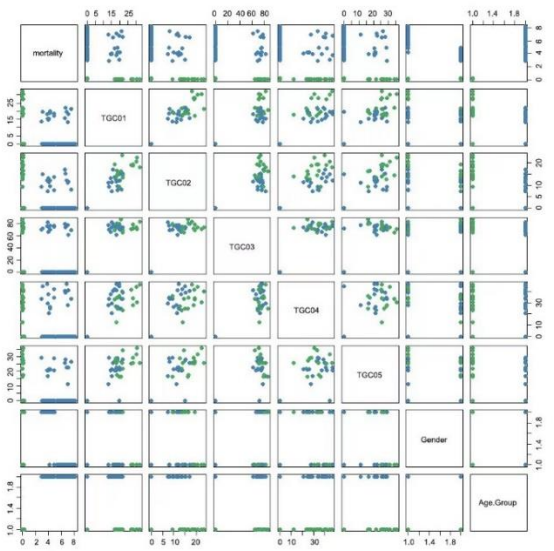


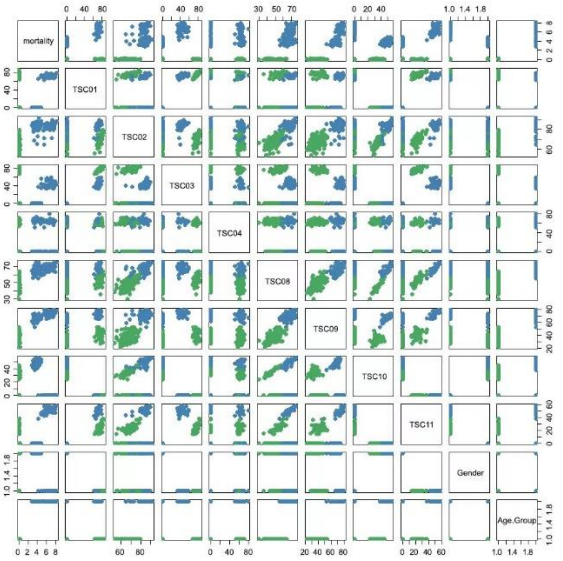


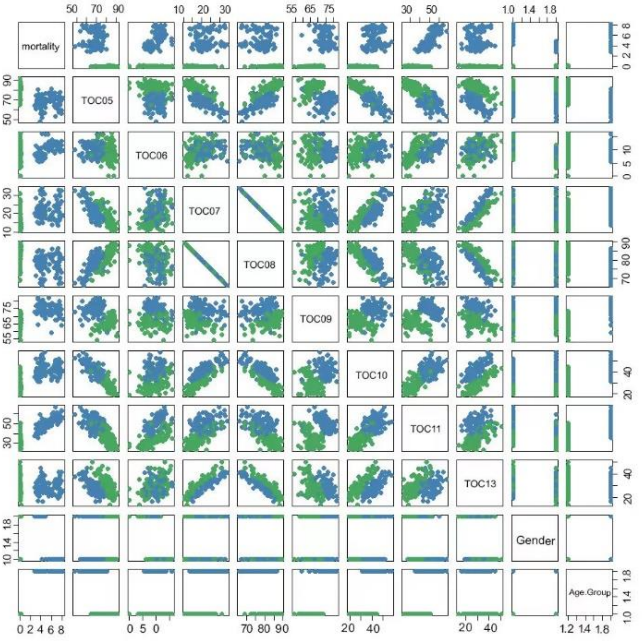




DOI: 10.54254/2753-8818/28/20230426



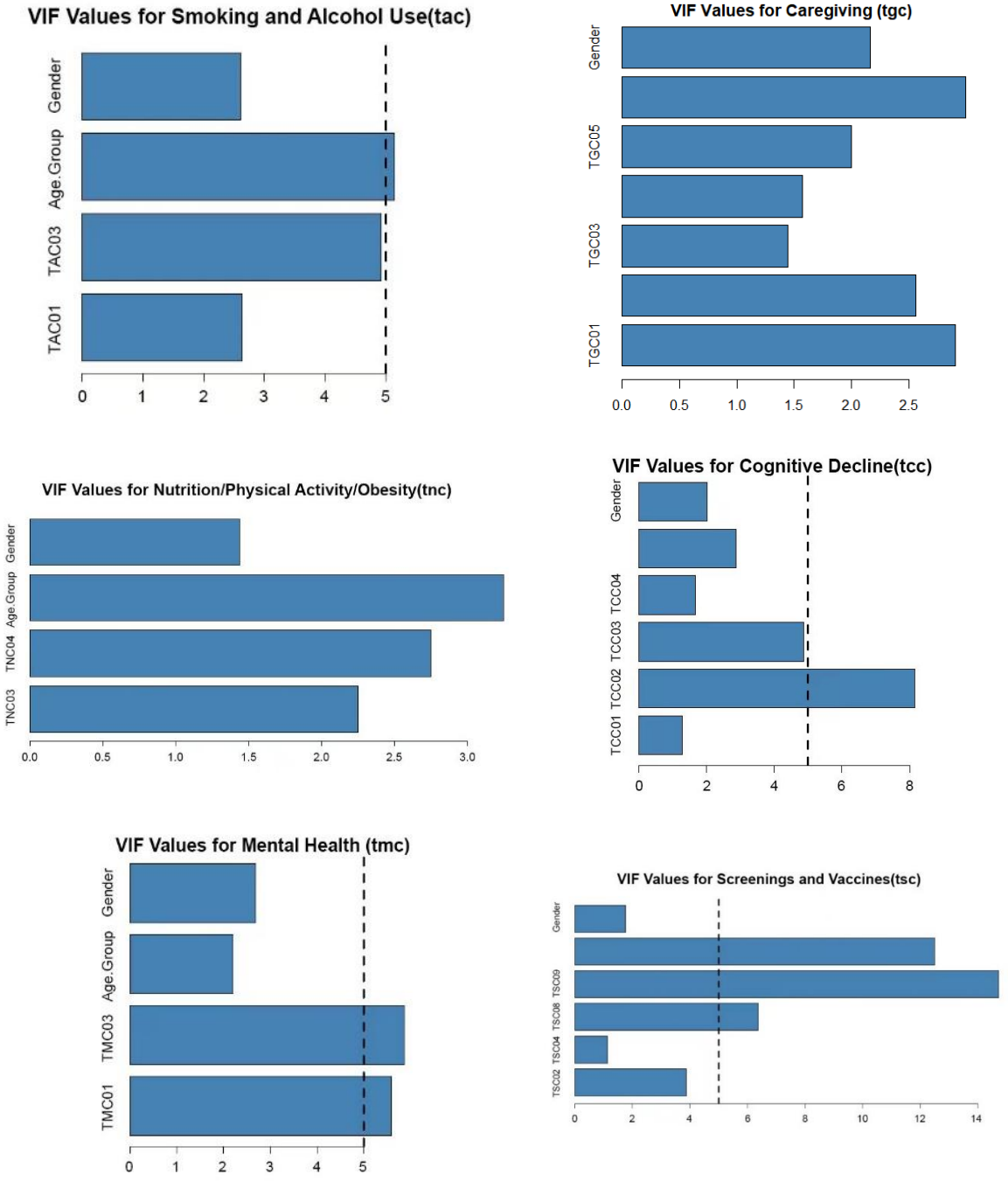




DOI: 10.54254/2753-8818/28/20230426

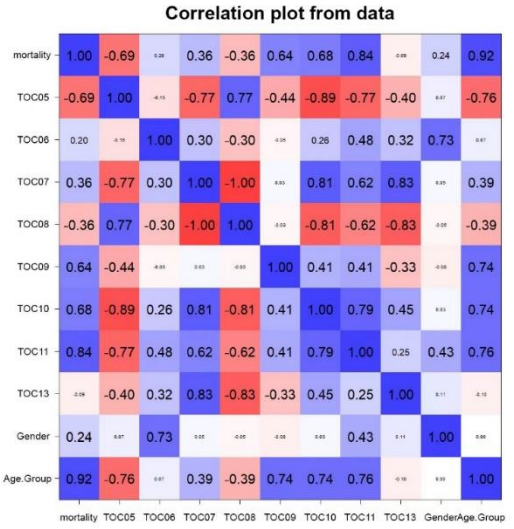
Appendix D

VIF Plots and Correlation Plot for Risk Factors.

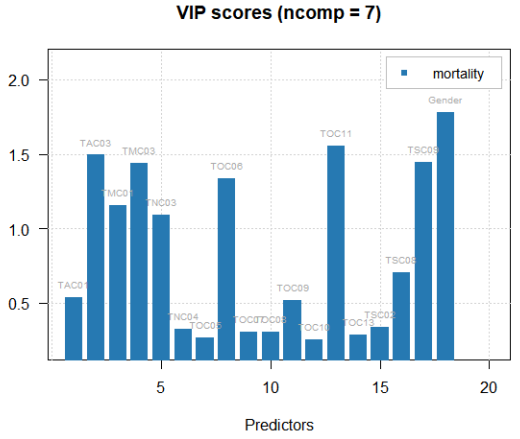


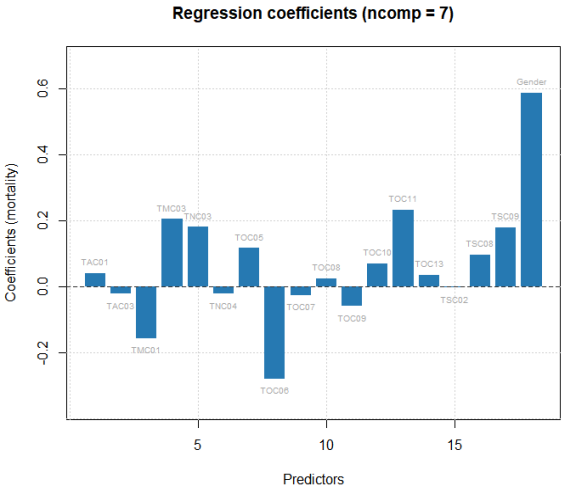
Note. The vif plot could not be shown due to the high level of variable multicollinearity for the overall health theme.

DOI: 10.54254/2753-8818/28/20230426



Appendix E





Appendix F

Map of Alzheimer's Disease Mortality Rates for Women over 65 Years of Age in the 50 States of the United States in 2020