Investigating Physiological and Environmental Impacts on Early Onset Alzheimer's

Sarah Avalle, Anamika Saha, Michelle Osgood, Anthony Masingale

Introduction

Alzheimer's disease (AD) is a significant global health challenge, imposing immense burdens on individuals, caregivers, and healthcare systems worldwide [1]. While traditionally associated with aging, the emergence of early-onset Alzheimer's disease (EOAD) presents a unique subset of cases, characterized by symptom onset before the age of 65 [2,3,4]. Despite its increasing recognition, EOAD remains relatively underexplored compared to late-onset AD, highlighting the need to elucidate the factors driving its onset.

This research endeavors to bridge this critical gap by investigating the physiological and environmental determinants of EOAD. By examining biometric, physiological, and genetic variables alongside demographic, environmental, and geographic features, we aim to provide valuable insights into the mechanisms underlying the early onset of Alzheimer's symptoms. Furthermore, our study seeks to contextualize these findings within the existing scientific literature, building upon and extending the current understanding of EOAD.

Our aspiration is not only to deepen comprehension of EOAD but also to contribute to the development of more effective prevention, diagnosis, and intervention strategies. Through unraveling the complex interplay between biological and environmental factors in EOAD pathogenesis, we aim to inform targeted approaches for risk assessment, early detection, and personalized treatment. Ultimately, our research aims to enhance the quality of life for individuals affected by this debilitating condition.

Alzheimer's disease is frequently perceived as a condition primarily affecting older adults. However, the emergence of cognitive decline in young adults, particularly within the context of EOAD, presents a distinct challenge. EOAD, characterized by symptom onset before age 65, poses unique challenges extending beyond the individual to their families, caregivers, and society at large.

Understanding the factors contributing to early-onset dementia compared to the average age of onset is crucial for comprehending the unique characteristics of EOAD. This includes exploring the roles of biometric, physiological, and genetic variables in predisposing young adults to cognitive decline at an earlier stage in life. Additionally, demographic, environmental, and geographic features may influence the early onset of dementia, further complicating our understanding of EOAD.

While genetic mutations, such as those in the APP and PSEN1/2 genes, have been implicated in some early-onset cases, they do not fully explain the prevalence of EOAD [5]. This underscores the importance of investigating additional genetic and environmental variables. The exposome model, which examines an individual's lifetime exposure to various environmental factors, sheds light on the dynamic relationship between genetic susceptibility and environmental influences.

Environmental factors, including neurotoxic agents like heavy metals and pesticides, as well as lifestyle components such as alcohol consumption and traumatic brain injuries, are under scrutiny for their potential roles in EOAD development [6,7,8]. Socioeconomic

determinants and educational attainment may also play significant roles in shaping dementia risk, emphasizing the importance of considering broader societal factors [9].

Despite the prevalence of EOAD, comprehensive population-based studies on its epidemiology are lacking. This research aims to address these knowledge gaps by investigating a broad spectrum of environmental, occupational, and lifestyle factors associated with EOAD. Through meticulous epidemiological analyses, our goal is to unravel the prevalence and incidence of EOAD, providing crucial insights for medical professionals, policymakers, and society at large.

Ultimately, our research seeks to make a substantial contribution to the understanding of young-onset Alzheimer's disease by exploring the environmental and physiological determinants contributing to cognitive decline in young adulthood. By unraveling the unique determinants that lead to early-onset Alzheimer's, our study aims to advance knowledge and significantly impact the cognitive health of individuals before age 65.

Methods

To identify the environmental influences that can catalyze Alzheimer’s disease and cause onset and progression at a younger age required access to extensive biometric and genetic data. The data from National Alzheimer's Coordinating Center (NACC) was acquired from the website by request. This data included clinical assessments, demographic data, autopsy reports, and biological specimen analysis.

In order to evaluate the data, it was cleaned of null or zero values, organized, and multivariate analysis was performed. All analyses were performed utilizing the Python codling language. Data cleaning utilized the Pandas and NumPy packages. Scatter matrices were produced with the Seaborn package, which helped to determine what variables indicated a correlation that could be further explored and reduced the variables necessary for analysis.

On the qualitative data, principal component analysis (PCA) was performed using Plotly on data that had shown significant correlation from the scatter matrices. PCA (Principal Component Analysis) plot for visualization is paramount due to its ability to effectively capture and illustrate complex multidimensional data in a simplified, two-dimensional format. PCA reduces the dimensionality of the data by identifying the most significant patterns and structures, allowing for visualization of the inherent relationships and variability within the dataset. By plotting the data along the principal components, which represent the directions of maximum variance, PCA facilitates the identification of clusters, patterns, and outliers within the data. This not only aids in data exploration and interpretation but also provides insights into underlying trends and associations that may not be apparent in the original high-dimensional space. Moreover, PCA plots enable researchers to visualize the similarities and dissimilarities between samples or groups, aiding in classification, clustering, and comparative analysis [10]. Once significant variables were identified, box plots were produced using the matplotlib.pyplot package. Box plots are a tool for visualizing categorical data. It is crucial because of the ability to effectively summarize and compare distributions across different categories or groups. Box plots provide a concise and intuitive representation of key summary statistics; including the median, quartiles, and potential outliers. This allows for a quick and easy comparison of the distributional characteristics within each category. This is particularly important when dealing

with categorical variables, as it enables researchers to visually assess variations, central tendencies, and dispersion among groups Thus, facilitating the identification of patterns, trends, and differences in the data. Additionally, box plots are robust against outliers and skewness, providing a clear and standardized method for visualizing categorical data regardless of its distributional properties.

On the quantitative data, a linear regression analysis was performed using Scikits Learn, a machine learning package for python. The Linear regression model for visualization is important due to its simplicity, interpretability, and versatility in capturing linear relationships between variables. Fitting a regression line to the data, helps us visually assess the direction and strength of the relationship, making it easier to discern trends and patterns. In addition to a visual aid in analysis, linear regression provides quantitative metrics such as the slope and intercept of the regression line as well as the coefficient of determination [11]. This offers valuable insights on the extent and significance of the relationship. It also allows for a more comprehensive understanding of the data and facilitates comparison across different datasets or variables.

The accessible clinical assessments provided access to demographic information which was used to provide metrics outside of genetics and physiology. This identified whether there was a variable that progressed young onset dementia from within the patient's ecosystem: location, food availability, socioeconomic status, education levels, climate, or any other common variable often seen that may exacerbate medical conditions. The methods aimed to identify if there was a factor outside of the patient's genetics that increased an individual's predisposition to young-onset dementia against average age-onset dementia counterparts.

Results

Beginning with about 10,000 variables, string data types and data that have no entries were culled, and the remaining variables, about 1000, were reserved for further analysis. From that 1000, the most relevant demographic categorical survey data were chosen based on common demographic variables associated with age. The remaining variables, about 80, were reserved for further analysis. We included all quantitative variables which totaled about 30 variables, mostly from neuropathology autopsy records.

From these variables we produced a collection of scatter matrices that each includes, age and sex, and about 5 variables for each matrix. Producing about 50 matrices, for each of these scatter matrices, we analyzed the relationship of age against any given variable. If the plot output was categorical and visually indicated to share an age difference, we would assign it as such. Conversely, if it produced a scatter plot and visually indicated even slight correlation we would include this, and assign it separately as quantitative.

Once visually indicated significant variables have been identified and grouped as categorical or quantitative, the categorical data was analyzed through Principal component analysis (figure 1), to derive the variables with the highest influence on the principal component. The top three were identified as the Lumbar puncture year (CSFLPYR), T-tau assay method (CSFTTMD) and year of death (NACCYOD). PC1 represents CSFLPYR, which corresponds to the year of lumbar puncture. PC2, represents CSFTTMD, the method used for the T-tau assay. Lastly, PC3 corresponds to EDUC, indicating the level of education for subjects in the dataset.

The results of the box plots (figure 2) of these three variables seem to show little to no difference with the strongest being the lumbar puncture year. Additionally, these are not the limitations of the categorical data, box plot generation does reveal other variables that have age related differences (figure 3). This reveals age related differences in inherited AD mutations (NACCAM), total number of medication use (NACCAMD), and posture status (POSTURE)

Regression analysis was performed on all remaining quantitative variables with the strongest correlations by at least one order of magnitude (figure 4). Plotting of these regressions revealed a slight negative correlation in the right pars orbitalis mean cortical thickness (RPARORBM), a slight negative correlation in right pars triangularis gray matter volume (RPARTRI), a negative correlation in left pericalcarine mean cortical thickness (LPERCALM), and a positive correlation in segmented total third ventricle volume (THIRVENT).

RPARORBM (right pars orbitalis mean cortical thickness) records the mean cortical thickness of the right pars orbitalis for a given MRI. RPARTRI (right pars triangularis gray matter volume) records the gray matter volume of the right pars triangularis for a given MRI. LPERCALM (left pericalcarine mean cortical thickness) indicates the mean cortical thickness of the left pericalcarine region for a given MRI. THIRVENT (segmented total third ventricle volume) records the volume of the segmented total third ventricle for a given MRI.

The dataset provided longitudinal numeric data sourced from the IDeA Lab, encompassing measurements of cortical thickness and gray matter volume, as well as segmented total third ventricle volume from MRI scans. Each variable offered insights into structural brain changes potentially associated with AD progression.

Figure 1.

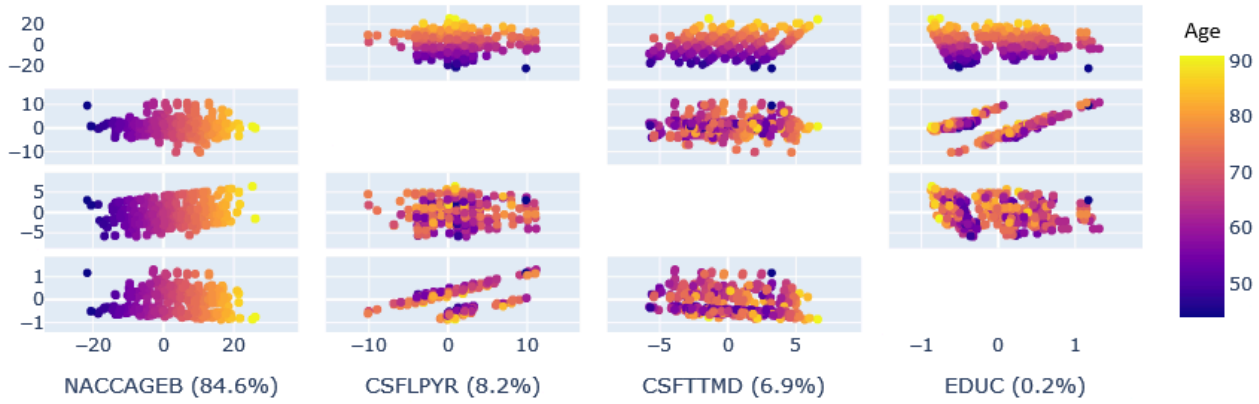


Figure 2.

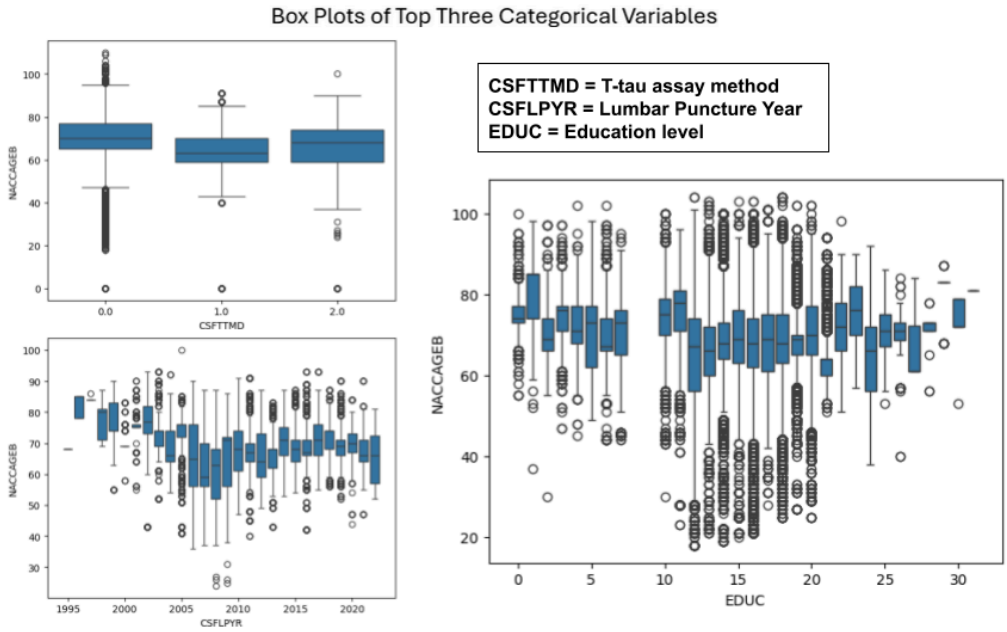


Figure 3

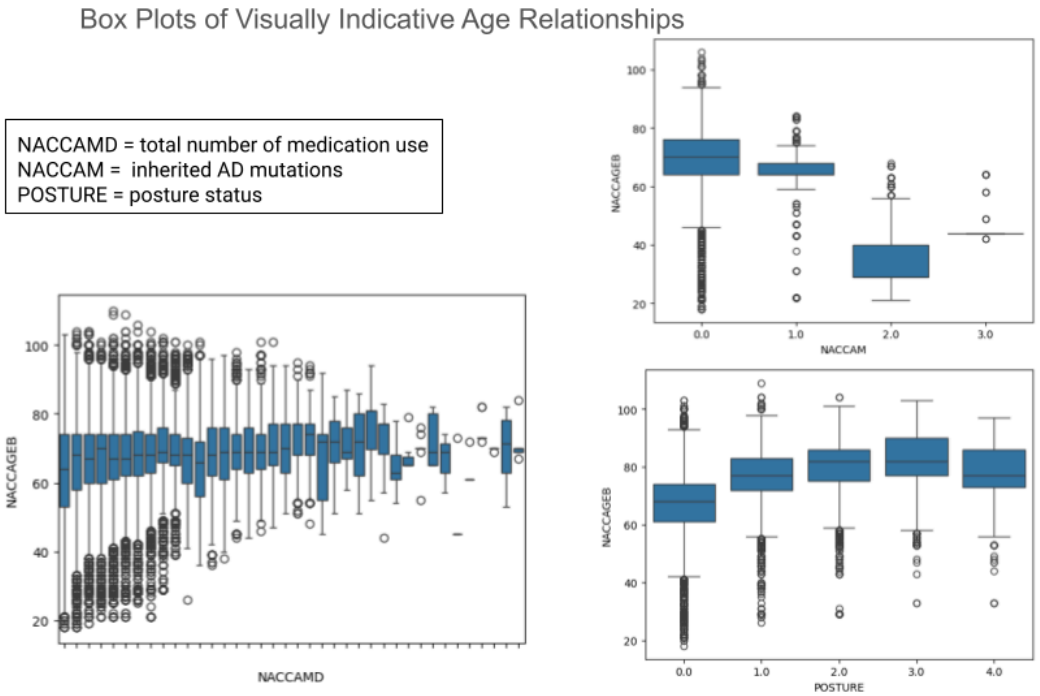
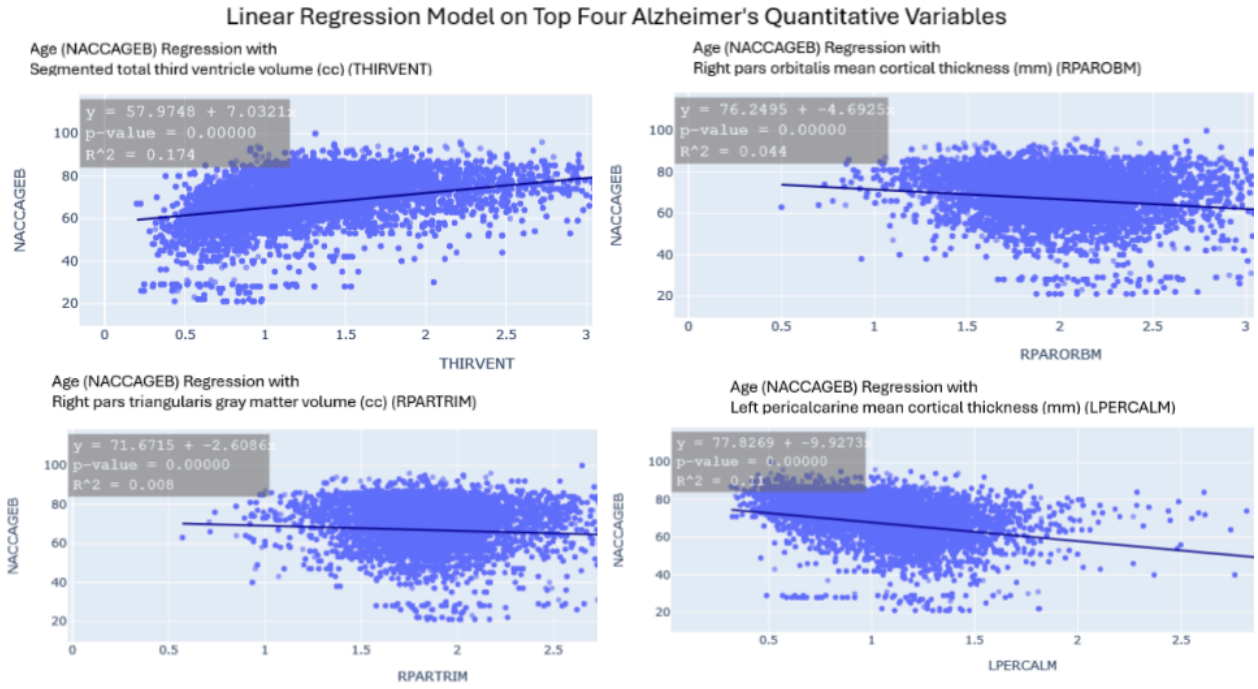


Figure 4.



Discussion

In our analysis, PC1 represents CSFLPYR, which corresponds to the year of lumbar puncture. This variable provides temporal information related to the timing of the lumbar puncture procedure, which could be relevant in understanding disease progression or treatment effects over time. PC2, represented by CSFTTMD, denotes the method used for the T-tau assay. T-tau protein is associated with Alzheimer's disease pathology, and different assay methods may yield variations in measurement accuracy or sensitivity. Understanding the nuances of these assay methods is crucial for interpreting tau protein levels accurately and comprehensively assessing their role in disease mechanisms. Lastly, PC3 corresponds to NACCYOD, indicating the year of death for subjects in the dataset. This variable, derived from neuropathology and milestones data, provides valuable information for analyzing disease progression and mortality outcomes in relation to other variables in the dataset.

The inclusion of PC2, representing the T-tau assay method, underscores the importance of considering different methodological approaches in biomarker measurement, particularly in the context of Alzheimer's disease research. Additionally, the percentages associated with each principal component, such as the 37.8% explained variance for PC2, provide insights into the relative contribution of each variable to the overall variation captured by the PCA plot.

In our exploration of Alzheimer's disease (AD) biomarkers, we've utilized linear regression analysis to examine the relationship between specific MRI metrics and the progression of AD. Initially, we produced scatter plots to explore all the variables of interest within our dataset. The scatter plots assisted in narrowing down the focus to variables that showed a correlation to the age of the Alzheimer's patients within the dataset. Initial variables included MRI data, CSF data, and qualitative data from a survey of Alzheimer's patients completed that covered environmental exposures. Our analysis focused on four key variables derived from MRI data that showed a correlation to age: Right pars orbitalis mean cortical thickness (RPARORBM), Right pars triangularis gray matter volume (RPARTRI), Left pericalcarine mean cortical thickness (LPERCALM), and Segmented total third ventricle volume (THIRVENT).

Our findings revealed significant associations between these MRI metrics and AD progression. Specifically, changes in cortical thickness in regions such as the right pars orbitalis and left pericalcarine, as well as reductions in gray matter volume in the right pars triangularis, were indicative of neurodegenerative processes characteristic of AD. Additionally, enlargement of the segmented total third ventricle, reflecting cerebral atrophy and ventricular expansion, was observed in association with AD progression.

These insights are crucial for understanding the structural changes occurring in the brains of individuals with AD, potentially serving as biomarkers for disease progression. By identifying these MRI biomarkers, we aim to contribute to the development of early diagnostic tools and intervention strategies for Alzheimer's disease.

Our analysis of the box plots reveals distinct patterns in the distribution of AD mutation types among our study participants. The PC2 T-tau assay method may be important to note regarding Alzheimers. Abnormal chemical changes can cause the tau to detach from microtubules and stick to other tau molecules and form threads that eventually lead to joining that causes tangles inside an individual’s neurons. The tangles can then cause a block in the neuron's transport system that essentially harms the synaptic communication between those neurons. The value for PC2 being 37.8% suggests that it explains approximately 37.8% of the relationship. The prevalence of PS-1 and PS-2 mutations suggests a potential genetic predisposition in individuals with young onset Alzheimer's. This observation prompts further investigation into the role of genetic predispositions in the disease's manifestation and progression. By highlighting these patterns, our study initiates a crucial exploration into the correlations between genetics and Alzheimer's disease.

Future Implications and Conclusion

In our analysis, we've identified several key findings that have significant implications for understanding early-onset Alzheimer's disease (EOAD) and its underlying determinants. Our investigation into the physiological and environmental factors associated with cognitive decline in young adults has revealed complex interrelationships between various variables. By examining biometric, physiological, and genetic factors alongside demographic, environmental, and geographic features, we've gained valuable insights into the mechanisms driving the early onset of Alzheimer's symptoms.

One of the significant findings of our study is the identification of specific variables that show correlations with age and disease progression. Through multivariate analysis and regression modeling, we've identified biomarkers and environmental factors that may play a role in the pathogenesis of EOAD. Changes in cortical thickness, gray matter volume, and ventricular volume observed in MRI scans provide insights into structural brain changes associated with AD progression. Additionally, variables such as lumbar puncture year, T-tau assay method, and year of death contribute to our understanding of disease progression and mortality outcomes.

Furthermore, our research highlights the importance of considering both genetic and environmental factors in shaping disease risk and progression. While genetic mutations such as those in the APP and PSEN1/2 genes have been implicated in some EOAD cases, they do not fully explain the prevalence of the disease. Our study emphasizes the need to explore additional genetic and environmental variables, including neurotoxic agents, lifestyle factors, and socioeconomic determinants, to gain a comprehensive understanding of EOAD etiology.

Moving forward, our findings have several implications for clinical practice and future research endeavors. Firstly, our identification of biomarkers and risk factors associated with EOAD could inform the development of targeted intervention strategies and personalized treatment plans for affected individuals. By understanding the relationship between genetic and environmental factors, healthcare professionals can better assess disease risk, facilitate early detection, and implement preventive measures.

Additionally, our research contributes to the growing body of knowledge surrounding EOAD epidemiology and etiology. Through epidemiological analyses via data-driven approaches, we've advanced our understanding of the prevalence, incidence, and risk factors associated with EOAD. This knowledge can inform public health initiatives, policy decisions, and resource allocation efforts aimed at addressing the societal impact of early-onset dementia.

In conclusion, our research represents a significant step forward in the fight against early-onset Alzheimer's disease. By elucidating the complex interplay of physiological and environmental factors contributing to cognitive decline in young adulthood, we aim to inform targeted interventions, raise awareness, and ultimately improve the lives of those affected by this devastating condition. However, further research and validation are imperative to confirm and build upon our findings, promoting a more comprehensive understanding of EOAD and the development of effective prevention and treatment strategies.

References

1. Li, X., Feng, X., Sun, X., Hou, N., Han, F., & Liu, Y (2022). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990-2019. *Frontiers in aging neuroscience*, *14*, 937486. https://doi.org/10.3389/fnagi.2022.937486

2. Harvey, R. J., Skelton-Robinson, M., & Rossor, M. N. (2003). The prevalence and causes of dementia in people under the age of 65 years. *Journal of neurology, neurosurgery, and psychiatry*, *74*(9), 1206–1209. https://doi.org/10.1136/jnnp.74.9.1206

3. Nwadiugwu M. (2021). Early-onset dementia: key issues using a relationship-centred care approach. *Postgraduate medical journal*, *97*(1151), 598–604. https://doi.org/10.1136/postgradmedj-2020-138517

4. Adani, G., Filippini, T., Garuti, C., Malavolti, M., Vinceti, G., Zamboni, G., Tondelli, M., Galli, C., Costa, M., Vinceti, M., & Chiari, A. (2020). Environmental Risk Factors for Early-Onset Alzheimer's Dementia and Frontotemporal Dementia: A Case-Control Study in Northern Italy. International Journal of Environmental Research and Public Health, 17(21), 7941. https://doi.org/10.3390/ijerph17217941

5. Fenoglio, C., Scarpini, E., Serpente, M., & Galimberti, D. (2018). Role of Genetics and Epigenetics in the Pathogenesis of Alzheimer's Disease and Frontotemporal Dementia. Journal of Alzheimer's disease : JAD, 62(3), 913–932. https://doi.org/10.3233/JAD-170702

6. Aloizou, A. M., Siokas, V., Vogiatzi, C., Peristeri, E., Docea, A. O., Petrakis, D., Provatas, A., Folia, V., Chalkia, C., Vinceti, M., Wilks, M., Izotov, B. N., Tsatsakis, A., Bogdanos, D. P., & Dardiotis, E. (2020). Pesticides, cognitive functions and dementia: A review. Toxicology letters, 326, 31–51. https://doi.org/10.1016/j.toxlet.2020.03.005

7. Vinceti, M., Bottecchi, I., Fan, A., Finkelstein, Y., & Mandrioli, J. (2012). Are environmental exposures to selenium, heavy metals, and pesticides risk factors for amyotrophic lateral sclerosis?. Reviews on environmental health, 27(1), 19–41. https://doi.org/10.1515/reveh-2012-0002

8.Power, M. C., Adar, S. D., Yanosky, J. D., & Weuve, J. (2016). Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: A systematic review of epidemiologic research. Neurotoxicology, 56, 235–253. https://doi.org/10.1016/j.neuro.2016.06.004

9. Velilla, L., Acosta-Baena, N., Allen, I., Lopera, F., & Kramer, J. (2022). Analysis of family stigma and socioeconomic factors impact among caregivers of patients with early- and late-onset Alzheimer's disease and frontotemporal dementia. *Scientific reports*, *12*(1), 12663. https://doi.org/10.1038/s41598-022-16400-2

10. Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). International Journal of Livestock Research, 7(5), 60-78.

11. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear regression. In An introduction to statistical learning: With applications in python (pp. 69-134). Cham: Springer International Publishing.