The Impact of Socioeconomic Status on the Incidence of Alzheimer's Disease

Data 621 Winter 2022 Course Project

Kimberley Chiu(ID 00322617), Marc McCoy(ID 30136987), Mark Ly(ID 00504696)

2022-04-16

Abstract

Background: Alzheimer's Disease (AD) is the leading cause of dementia and sixth leading cause of death in the United States. Patients who experience AD are more likely to have difficulties performing activities of daily living in old age, which can result in decreased quality of life for both the patient and family members. Currently, the pathophysiology of AD is not well understood.

Methods: This retrospective cohort study of 150 adults aged 60 to 96 that were selected from a larger database of 400 individuals from the Washington University Alzheimer Disease Research Center (ADRC). The primary and secondary objectives were to estimate the risk of dementia (non-demented, demented) based on socioeconomic status (SES) other confounders and covaraites.

Results: A total of 142 adults were available for study. After screening, 72 patients (50.7%) where identified as non-demented and 70 patients (49.3%) were identified as demented. For our primary outcome, there was a decrease in the odds ratio (OR) as we moved from a high SES (SES=1) to a low SES (SES=4). At the significance level of 0.05, the OR for dementia going from SES=1 to a SES=2 is 0.16 (95% CI: 0.04, 0.58; P=0.008), from SES=1 to a SES=3, we have a OR of 0.39 (95% CI: 0.08, 1.75; P=0.22601) and from SES=1 to SES=4, we have a OR of 0.08 (95% CI: 0.01, 0.54; P=0.011). The OR for age is 0.91 (95% CI: 0.83, 0.99; P=0.021) while the odds ratio for education is 0.76 (95% CI: 0.58, 0.96; P=0.024). For males, the OR is 3.95 (95% CI: 1.14, 15.1; P=0.030) and has the greatest increase in odds developing dementia. In our secondary analysis, we found that sex and education had significant interaction with a p-value from the Wald test of 0.024 at a significance level of 0.05. The likelihood ratio test produced a p-value of 0.020. The sex and education interaction term had a OR of 1.55 (95% CI: 1.07, 2.32; P=0.020).

Conclusion: Using demographic data as explanatory variables produced a simple analysis that produced counterintuitive results. One major drawback of our study design is it did not account for the fact that patients with a lower SES likely received less timely care than patients with higher SES, which is why the OR for higher level SES categories were less than 1.0. Including repeated measures data, such as the change in nWBV, could act as a proxy explanatory variable for brain atrophy. More sophisticated mixed-effects logistic regression models should be explored in a follow up analysis.

Abstract word count: 412 words. **Word count:** 4217. **Figure:** 10 figures. **Tables:** 3 tables. **References:** 7 References.

Contents

1 Introduction:

1.1 Background

1.1.1 Context:

Major neurocognitive disorder (previously called dementia) is an acquired disorder of cognitive function that may result in impairments in memory, speech, reasoning, intellectual function, and spatial awareness. The pathophysiology of dementia is complex, and research has not yet pinpointed a cause.

Alzheimer's disease (AD), which is more prevalent in the female population, is the leading cause of dementia and the sixth leading cause of death in the United States.[2] Due to the underlying pathophysiology, incidence and prevalence increase significantly with age. AD presents enormous social and economic burdens on society and family members due to the patient's potential challenges with activities of daily living (ADL). Early-onset familial AD, which makes up approximately 10% of AD cases, is particularly challenging, as patients experience symptoms before the age of 65 years old.[2,3]

1.1.2 Rational:

Currently, there exists no curative therapy available for AD. However, lifestyle modifications such as adhering to a regular sleep schedule and maintaining a familiar environment may slow the development of symptoms. Physical activity and cognitive rehabilitation exercises such as memory training via puzzles and games have been shown to slow cognitive and functional decline. There is the potential for screening procedures to lower the incidence of AD - as memory training can improve cognitive capabilities through targeted stimulation prior to the patient's AD worsening to a level where ADLs cannot be performed.[4]

The potential impacts for routine screening procedures, prognostic indicators and genetic markers for AD are vast. Studying physiologic changes within the brain assists in the development of research seeking to understand the pathophysiology of AD, which is currently not well understood.

1.1.3 Definitions:

The following definitions for widely used terms are described below. Definitions for variables can be found in Table 1.

- **AD** - Alzheimer's Disease
- **ADL** - Activities of Daily Living
- **CDR** - Clinical Dementia Rating Scale
- **glm** - generalized linear model
- **Mass** - Modern Applied Statistics with S package in RStudio
- **OASIS** - Open Access Series of Imaging Studies

1.2 Research Question and Objectives

1.2.1 Research Question(s):

Primary Analysis: In older adults aged 60-96 years old, is socioeconomic status (SES) associated with AD as operationalized by binary outcome Group (non-demented, demented) while controlling for age, education, sex, MMSE, eTIV, and nWBV as covariates?

Secondary Analysis: In older adults aged 60-96 years old, are covariates sex and education a confounder or effect modifier for binary outcome Group (non-demented, demented)?

1.2.2 Study population:

The study population consists of adults (aged 60 to 96) who are demented and non-demented residing in North America.

1.2.3 Study sample:

The study sample consists of 150 adults aged 60 to 96 were selected from a larger database of 400 individuals from the Washington University Alzheimer Disease Research Center (ADRC). The dataset was originally a longitudinal study centered around a series of imaging studies, however information from only the first visit was used in the analysis.

Subjects with a primary cause of dementia other than AD, active neurological or psychiatric illness, serious head injury, history of clinically meaningful stroke, use of psychoactive drugs, anatomical abnormalities were excluded. Subjects with age-typical brain changes were included.

1.3 Data Source and Description

1.3.1 Data Source

The dataset was taken from the Open Access Series of Imaging Studies (OASIS), which is a compilation of open source distributing MRI datasets. OASIS is made available by the Washington University Alzheimer's Disease Research Center. This dataset is licensed under the CCO 1.0 Universal (CCO 1.0) Public Domain Dedication and is open source where no research ethics approval is required. For the purposes of this project, the assumption is that the measurements were all taken at the end of the study.

1.3.2 Data Description

The longitudinal dataset contains 9 demographic, clinical and derived anatomic values for 150 patients. To satisfy the independence assumption, we will only use information from the first visit in our analysis. Covariates include sex (M/F), age (60 - 96), years of education (6 - 23), social economic status (SES; ranked 1-5), mini-mental state examination (MMSE), clinical dementia rating (CDR), estimated total intracranial volume (eTIV), atlas scaling factor (ASF), and normalized whole brain volume (nWBV).

2 Methods

2.1 Study Design

The study is a retrospective cohort study of 150 adults aged 60 to 96 that were selected from a larger database of individuals from the Washington University Alzheimer Disease Research Center (ADRC). The dataset was originally a longitudinal study centered around series of imaging studies, however, to maintain independence assumption in the data, information from only the first visit will be used in the subsequent analysis.

The sample size is limited to 150 patients, and 8 subjects with missing data were removed. With this given sample size, the study has a power of 31% as seen in Figure 2. Studies on retention in clinical trials show that around 30% of subjects drop out, therefore, to achieve a power of 80% and to account for the dropout rate, a sample size of 712 individuals would be needed as seen in Figure 3.

2.2 Statistical Principals

2.2.1 Level of Significance

A level of significance of 0.05 was assumed for all statistical tests which were performed.

2.2.2 Reporting of Confidence Intervals

A logistic regression via the generalized linear model function (glm) in the MASS library will be used. The log odds ratio, and subsequently the odds ratio of the outcome (group) for categorical exposure variable SES (levels 1 – 4) and the corresponding 95 % confidence interval (CI) will be presented.

The second outcome will be to identify any potential effect modifiers. These potential effect modifiers are sex and education. This will be done by creating an interaction term between each potential effect modifier and the main effect variable of SES. If the interaction term is significant, then that variable can be considered an effect modifier

2.3 Study Outcomes

2.3.1 Primary Outcome

The primary outcome will be to explore in older adults aged 60-96 years old, if socioeconomic status (SES) is associated with AD as operationalized by binary outcome Group (Non-demented, Demented) while controlling for age, education, sex, MMSE, eTIV, and nWBV as covariates. The exponent of the coefficients of covariates to the response outcome Group will be applied to the results as a calculated odds ratio.

2.3.2 Secondary Outcome

The secondary outcome will be to explore covariates sex and education and their impacts as a confounder or effect modifier for binary outcome Group. The exponent of the coefficients of covariates with interaction terms to the response outcome Group will be applied to the results as a calculated odds ratio.

2.4 Analysis Methods

2.4.1 Data Wrangling

The Group outcome variable will be used as it is divided into two categories (Demented and Non-Demented). The exposure variable, SES, is assessed by the Hollingshead Index of Social Position with 5 categories: 1 being the highest and 5 the lowest.

For the purposes of this project, we are assuming that measurements for age, SES, education, and group outcome were taken at the end of the study period.

The dataset was obtained in a csv format that was read into R and analyzed using built-in packages in the R library. Columns for 'Hand', 'MRI.ID', 'Visit', 'Subject_ID' and 'MR.Delay' were excluded from our study. All the patients were right-handed and the first visit only was considered for all the patients, so 'Visit' and 'MR.Delay' columns were excluded. 'MRI.ID' and 'Subject_ID' were also excluded from the analysis as they were not informative.

From our exploratory analysis, it was noticed that the lowest SES level (5) only had 3 patients in total (1 in Non-demented, 2 in Demented). At least 10 observations for each unique level is needed to provide reasonable estimates and standard errors. Therefore, it was decided to collapse the lowest level of SES (5) to the second lowest (4) which would adjust the SES variable to have 4 levels, with 1 being the highest and 4 being the lowest.

There are 14 patients in the dataset who were under the "Converted" group; in order to transform the Group variable to a binary variable, the 14 patients were moved from "Converted" group into "Demented" group. Finally, the categorical columns of 'Group', 'sex' and 'SES' were converted into factors, with 'Group' having two levels (Non-demented, Demented), 'sex' having 2 levels ('F','M').

2.4.2 Statement on Covariates, Confounding, and Effect Modification:

There are potential confounding variables when investigating the association between SES and Alzheimer's Disease, and these include: age, sex, SES, and education. Therefore, in this study, we will consider these variables as potential confounders. Additionally, many health and medical studies show that sex and education are often considered to be effect modifiers. This can be explained by how AD is more prevalent in the female population. Therefore, we will test if sex and education have some interaction with SES on the outcome of dementia, and further determine whether the confounding covariates mentioned above have impacts on the analysis.

2.4.3 Defining Study Outcomes

Non-demented is classified as a CDR = 0. Whereas demented is classified as a CDR > 0. CDR is clinical dementia rating and is calculated through a clinical calculator after assessing a patient's

memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care. A categorical scoring system is used within each category, and the aggregation of categorical scores is used to derive the CDR value which classifies the patient as demented or non-demented.[5] Refer to Figure 1 to clarify the units of measurement for each variable. The outcome is calculated via the CDR clinical calculator, but there are no transformations performed on the outcome.

2.4.4 Assumption Checks:

Our final model was check for multicollinearity, linearity, independence and potential outliers. the constant variance and normality assumptions are not required as the random aspect of the logistic model is not included as an additive term in the regression equation. A goodness of fit test was performed prior to solidifying the primary outcome model. From there, all interaction terms were tested to look for effect modification between variables. Subsequently, interaction terms were tested one by one via likelihood ratio tests. Testing for confounding was then performed and compared to the 10% magnitude rule.

2.4.5 Subgroup Analysis:

No subgroup analysis will be conducted in this project.

2.4.6 Missing Values:

Using the *vis miss* function from the *visdat* package in R, a heatmap was created to determine the percentage of missing data in each column. There are 8 missing values in the SES variable. Since SES cannot be determined on other values, we removed 8 individuals leaving us with 142 left for analysis. See Figure 4 for heatmap.

2.4.7 Software for Analysis

RStudio (Version 1.4.1717) was used to perform the data cleaning and statistical analysis on the relationship between dementia and age, considering potential confounding and effect modifiers. Lucidchart (online visual diagramming software) was used to create the methodology flowchart in Figure 5.

3 Results

3.1 Descriptive Statistics

The descriptive statistics for the potential predictors of dementia (n = 70) and nondemented (n = 72) for individuals between the ages of 60 - 96 years old can be found in the Table 1 using the *gtsummary* package in R. The p-values reported in the figure are calculated for the Person's Chi-Squared test for categorical variables and Wilcoxon rank sum test for numerical variables. We used a significance level of $alpha = 0.05$ to compare our findings. A summary of our findings can be found in Table 1.

The Person's Chi-Squared test determines if there is a statistical significance between categorical variables. Where the null hypothesis is that there is no relationship between categorical variables and the alternative hypothesis is that there is a relationship between the categorical variables. The Wilcoxon Rank sum test is a nonparametric test to determine if two groups are derived from the same population. The null hypothesis for is that the two populations are equal, and the alternative hypothesis is that the two populations are not equal.

There were significantly more female participants in the non-demented group (69%) than the dementia group (49%, p-value = 0.011). The mean age of people who did not have dementia was 75 years old (sd = 8) which is exactly the same as those in the dementia group (75 years old, sd =7; p-value = >0.9). The mean years of education were also the same in both the Non-demented (15 years, sd = 3) and demented groups (14 years, sd = 3; p-value = 0.029).

We see high a higher proportion of participants with higher SES levels in the non-dementia group (SES1 = 21%, SES2 = 38%) than in the dementia group (SES1 = 26%, SES2 = 21%, p-value = 0.3). While more participants in the dementia group were found to be in the lower SES levels (SES4 = 19%) compared to those in the non-dementia group (SES4 = 27%, p-value = 0.3).

Age and SES both have a p-value that is greater than our significance level of $\alpha = 0.05$ (Age = >0.9, SES = 0.2). In both cases will fail to reject the null hypothesis. In terms of Age we would say that the two populations (Non-demented, Demented) are equal and for SES we would say that there is no relationship between the categorical variables.

The Estimated total intracranial volume (eTIV), Atlas scaling factor (ASF) and, Normalized whole brain volume (nWBV) mean values for both the non dementia group (eTIV = 1,480, sd = 184; p-value = 0.8, nWBV = 0.26, sd = 1.04; p-value = 0.002, ASF = 1.20, sd = 0.14; p-value = 0.8) and the dementia group (eTIV = 1,471, sd = 167; p-value = 0.8, nWBV = -0.27, sd = 0.89; p-value = 0.002, ASF = 1.21, sd = 0.13; p-value = 0.8).

3.2 Primary Outcome Results (Logistic Regression)

The main deviation from our SAP was switching our from a cross sectional dataset to a longitudinal dataset. After further investigation, we determined that the majority of the missing data was due to the selection criteria of participants. Only those that were over the age of 60 were select to undergo the full clinical assessment. Initially, we planed on using age as our primary indicator however, that does not align with our research question and after further discussion, our primary indicator changed from age to SES. For our primary analysis, we changed from a uni-variate model to a multivariate model that included all relevant covaraites and confounding terms. Our secondary analysis was modified to explore any interaction effects rather than performing a ordinal logistic regression on based on CDR levels in our SAP.

3.2.1 Variable selection

Variables selection was done using pairwise plot from GGAlly, vif function from the car and the backwards stepwise variable selection from the MASS package. ASF and eTIV both have a VIF value that is greater than 5 and have high collinearity that needs to be addressed. We performed backwards stepwise regression to determine which covariates to keep in our model. Our initial AIC was 141 from the full model and 138.27 after dropping ASF variable. A partial f-test was done at a significance level of 0.05 to determine if we should drop the ASF variable. Our null hypothesis

is that there is a there is no difference on coefficients if we remove them from our model and the alternative is that at least one of the coefficients removed from the model is non-zero. With a f test-statistic with df 1,131 and a p-value of 0.9287 which means we fail to reject the null hypothesis and drop the ASF coefficient because it does not significantly improve the fit of our model.

3.2.2 Confounding

We expect there to be confounders in our dataset as they are linked to multiple risk factors associated with dementia. From our stepwise model, the Wald test p-values for all of of our covariates are less than our significance level of $\alpha = 0.05$. A check for for confounding was not needed. Ethnicity may also have a confounding effect on the incidence of AD, however, we are unable to adjust for ethnicity since it is unmeasured in our dataset.

3.2.3 Regression analysis

From our backwards stepwise selection we have 7 significant predictors to be kept. Our baseline is women who are non-demented with the highest SES (SES=1). Our main outcome is dementia status which is binary (non-demented/demented) and our main exposure is social economic status which is ordinal (1,2,3,4).

Table 2. Primary Outcomes of Dementia and Socioeconomic Status

Characteristic	**OR**	**95% CI**	**p-value**	**GVIF**	**Adjusted GVIF**
SES			**0.011**	2.8	1.2
1					
2	0.16	0.04, 0.58			
3	0.39	0.08, 1.75			
4	0.08	0.01, 0.54			
Age	0.91	0.83, 0.99	**0.021**	2.2	1.5
EDUC	0.76	0.58, 0.96	**0.024**	2.3	1.5
SEX			**0.030**	2.0	1.4
F					
M	3.95	1.14, 15.1			
MMSE	0.45	0.30, 0.61	**<0.001**	1.2	1.1
eTIV	1.00	0.99, 1.00	**0.013**	2.2	1.5
nWBV	0.45	0.22, 0.88	**0.019**	2.2	1.5

From Table 2 above, SES (p-value = 0.011), age (p-value = 0.021), education (p-value = 0.024), sex (p-value = 0.030), MMSE(p-value = <0.001), eTIV (p-value = 0.013) and nWBV (p-value = 0.019) are all significant under 0.05.

At the significance level of 0.05, the OR for dementia going from a high SES level (1) to a slightly lower SES level (2) is 0.16 (95% CI: 0.04, 0.58; p-value=0.008), this means that the patient is 84.0% less likely to develop dementia as they move down from SES 1 to SES 2. Going from a SES of 1 to a SES of 3, we have a OR of 0.39 (95% CI: 0.08, 1.75; p-value=0.22601), which means that the patient is 61% less likely to develop dementia when moving from a SES of 1 to a SES 3, however these value is statistically insignificant as the p-value is greater than our acceptance criteria $\alpha = 0.05$ and the

confidence interval does include 1. Finally, going from SES 1 to lowest SES (4), we have a OR of 0.08 (95% CI: 0.01, 0.54; p-value=0.011) which means that a patient is 92% less likely to develop dementia as they move from SES 1 to SES 4. The OR for age is 0.91 (95% CI: 0.83, 0.99; p-value = 0.021) which means the patient is 0.09% less likely to develop dementia for every one year increase in age. The odds ratio for education is 0.76 (95% CI: 0.58, 0.96; p-value=0.024), where the patient is 24% less likely to develop dementia for every additional year of education they have. The OR for males is 3.95 (95% CI: 1.14, 15.1; p-value = 0.030) which means that males have a 3.95 times greater odds of developing dementia than women. The OR for MMSE is 0.45 (95% CI: 0.30, 0.61; p-value = <0.001), which means that the patient is 55% less likely to develop dementia for every 1 unit increase in MMSE. The OR for eTIV is 1.00 (95% CI: 0.99, 1.00; p-value=0.013), which means that every one unit increase in eTIV does not change the odds of dementia. Finally, the OR for nWBV is 0.45 (95% CI: 0.22, 0.88; p-value=0.019) after standard scaling using the mean and standard deviation. From these odds ratio, the sex of the patient has the greatest increase in odds of a patient developing dementia. The greatest decrease in odds of dementia is from SES level. We expected a higher odds of dementia for every change in SES status.

3.2.4 Goodness of Fit

A goodness of fit hypothesis test was done at the significance level of $\alpha = 0.05$, where our null hypothesis is the the model is correct and the data fits. Our alternative hypothesis is that there is an evidence of a lack of fit. With a Residual deviance of 118.79 on 132 degrees of freedom, we get a probability of 0.788, which is greater than our significance level of $\alpha = 0.05$. We fail to reject our null hypothesis and we can say that our data is a good fit with our selected model.

3.3 Assumptions and Model Diagnostics

3.3.1 Linearity Assumption

The linearity assumption assumes that our continuous variables, X, and the log-odds are linear. Figure 7 contains the plots of the logit versus our continuous predictors.MMSE, Education and nWBV look somewhat linear with the log-odds. Age does not look linear and may need to be transformed to a higher power (cubic). The Box-Tidwell test adds interactions between the continuous variables into the model. Our null hypothesis is that the our continuous variables are linearly related to the log odds and the alternative hypothesis is that our continuous variables are not linearly related to the log odds. At a significant level of $\alpha = 0.05$, the p-values for all continuous variables are greater than our significance level of $\alpha = 0.05$. We will fall to reject our null and say that our continuous x variables are linearly related to the log-odds. Our linearity assumption has been met.

3.3.2 Independence Assumption

As we are only using the first visit for each patient in the longitudinal dataset, each patient is unique and only counted once. Using the data in this format, we can say that our independence assumption has been met.

3.3.3 Multicollinearity

Multicollinearity was assessed during our variable selection. The presence of correlation can cause errors in our results by using redundant information and could reduce our precision in our analysis. We don't see any evidence of collinearity between any of our predictor variables based on vif values.

3.3.4 Influential Points and Outliers

Boxplots and Cook's distance values were used to identify outliers and influential points. In Figure 7, we see that MMSE and eTIV both have outliers however, not all outliers are influential observations. If a point has a Cook's distance that is greater than 1, then the data point is likely to be influential. From Figure 8, we see that the 3 highest values are from point 93 , 135 and 136. However, all 3 points have a Cook's distance that is less than 1 so we can leave them in our model.

3.4 Secondary Outcome Results (Effect modification)

Within the literature, we expect sex to be an effect modifier on all our variables. The p-values from the Wald test show that only sex and education had a significant p-value of 0.024 ($\alpha = 0.05$). For a likelihood ratio test, the null hypothesis is there is no difference between the models (no effect modification) and our alternative hypothesis is that there is a difference between the two models (effect modification). The p-value for testing sex as a effect modifier was 0.020 ($\alpha = 0.05$) which means that we reject our null hypothesis and will include this interaction term in our model. The OR for the interaction between sex and education is 1.55 (95% CI: 1.07, 2.32; p-value = 0.020) which is the odds association between males and education at the highest SES level (SES=1). A summary of the outcomes from the secondary analysis can be found in Table 3.

4 Conclusion

Socioeconomic status is associated with Alzheimer's disease as operationalized by group (non-demented, demented) while controlling for: age, education, sex, MMSE, ETIV, and NWBV as covariates. There were three deviations from our initial expectations that were rooted in the literature when analyzing the results. Recall that the base case was for sex as female with a SES=1. It was found that the odds dementia for males is 3.95 times the odds of dementia for women, this is conflicting with current literature as female prevalence is higher according to medical literature. The lower odds ratio for higher SES was also contradictory to expectations. We observed that the odds of the patient being demented with SES = 4 is 0.08x the odds of a patient with an SES=1. We expect that this OR was larger than one the result of censoring due to death. As mentioned in the descriptive statistics section, the average age of our sample patients was around 75, which is the peak prevalence of AD. Patients with AD are likely to die at an earlier age than patients who do not have AD.

We realized after performing our analysis that we treated the eTIV and nWBV variables inappropriately in our statistical analysis plan and should have considered the delta over time of nWBV as an explanatory variable (or a proxy for brain atrophy), and because eTIV doesn't change over the patient's lifetime, but nWBV may possibly change, taking a ratio of nWBV/eTIV would have been more appropriate for our study.

4.1 Limitations

Major limitations stem from the intuition which underpins our model construction – research has not yet proven the directionality of the relationship between symptoms and pathophysiology – changes in the brain could be the result of AD or they could be the cause, therefore including them as covariates may not be appropriate. Another limitation of the current data is that the sample size was small with 150 subjects. This coincided with a power of 31%. For 80% power, a sample size of 712 individuals is required to have a robust final sample size including the 30% dropout rate factor. For future studies, we would obtain a sample size of 712 to achieve adequate power. There is also a high likelihood of measurement bias because different physicians performed the clinical tests that were used to derive the CDR data and ultimately demented vs. non-demented status of the patient. Another drawback of our study design is it did not account for the fact that patients with a lower SES likely received less timely care than patients with higher SES – which is why our OR for higher level SES categories were less than 1.

4.2 Future plans

Future plans would include revisiting the study design to properly incorporate the change or delta in nWBV as an explanatory variable. This would be a better proxy for brain atrophy, which is considered to be one of the key drivers of the symptoms of AD. Using the ratio of (nWBV/eTIV) over time could also be explored. It can be clearly seen from Figure 9, that the nWBV drops more rapidly over 2-4 year timespans in patients who exhibit clinical symptoms of AD; the slopes with CDR of 0.5 or CDR of 1 are steeper than the slopes with a CDR = 0 on average. This is intuitive as brain atrophy is documented in the literature as being associated with AD. If we were to include the change in (nWBV/eTIV) as a explanatory variable, however, our study would have repeated measures for the same sample, and we would lose the previously met assumption of independence and would need to use a mixed-effects logistic regression model to account for this. This could be done via the *xtmelogit* command in R.

Alternative future plans include placing more emphasis on social determinants of health, as opposed to demographic variables, as explanatory variables. One major drawback of our study design is it did not account for the fact that patients with a lower SES likely received less timely care than patients with higher SES – which is why our OR for higher level SES categories were less than 1. We could look at ways of designing a data collection process that would ensure the data represented this health access disparity (Figure 10).

5 Contribution statement

Mark Ly contributed to model building and assumption checking. General skeleton for R markdown file. Wrote sections for methods and results.

Marc McCoy contributed to the construction of the lucid chart and supported methods. Wrote sections: introduction, conclusions drawn, limitations, and future plans.

Kimberley Chiu contributed to building the report sections: introduction, methods, conclusion, quality checking, and finding relevant research statistics.

6 References

1. Marcus, DS, Wang, TH, Parker, J, Csernansky, JG, Morris, JC, Buckner, RL. Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. Journal of Cognitive Neuroscience, 19, 1498-1507. doi: 10.1162/jocn.2007.19.9.1498

2. The Alzheimer's Association. 2020 Alzheimer's disease facts and figures. Alzheimer's & Dementia .2020; 16(3): p.391-460. doi: 10.1002/alz.12068.| Open in Read by QxMD

3. Kasper DL, Fauci AS, Hauser SL, Longo DL, Lameson JL, Loscalzo J. Harrison's Principles of Internal Medicine. New York, NY: McGraw-Hill Education; 2015

4. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM–5).2013. doi: 10.1176/appi.books.9780890425596.| Open in Read by QxMD

5. Hughes, C., Berg, L., Danziger, W., Coben, L., & Martin, R. (1982). A New Clinical Scale for the Staging of Dementia. British Journal of Psychiatry, 140(6), 566-572. doi:10.1192/bjp.140.6.56.

6. Piedmont Healthcare. The Odds of Developing Alzheimer's Disease. https://www.piedmont.org/living-better/can-you-beat-the-odds-of-developing-alzheimers-disease. Accessed March 3, 2022.

7. The true cost of patient drop-outs in clinical trials. mdgroup. https://mdgroup.com/blog/the-true-cost-of-patient-drop-outs-in-clinical-trials/. Published December 18, 2021. Accessed March 3, 2022.

7 Table and Figures

Figure 1 - Data Source Attributes

Data Category	Attribute	Data Type	Details
Demographic	Gender	Categorical	M=Male (71), F=Female (145)
	Age	Continuous	Years of age (33 – 96)
	Education	Categorical	1 = less than high school, 2 = high school grad, 3 = some college, 4 = college grad, 5 = beyond college
	Socioeconomic Status	Categorical	1 = highest status to 5 = lowest status
Clinical	Mini-Mental State Examination (MMSE)	Continuous (to Categorical)	A maximum of 30 points is possible ≥ 24: No dementia, 20–24: mild dementia,13–20: moderate dementia, <13: advanced dementia
	Clinical Dementia Rating (CDR)	Categorical	0 = nondemented, 0.5 = very mild dementia, 1 = mild dementia, 2 = moderate dementia
Derived Anatomic Values (from MRI)	Estimated total intracranial volume (eTIV)	Continuous	Estimates intracranial brain volume (1123–1992 mm^3).
	Atlas scaling factor (ASF)	Continuous	Scaling factor that allows for comparison of eTIV based on differences in human anatomy (0.88–1.56)
	Normalized whole brain volume (nWBV)	Continuous	This variable measures the volume of the whole brain. 0.64–0.84 mg (observed).

Figure 2 - Power calculation

```
## Power for logistic regression
##
##       p0   p1     beta0     beta1   n alpha    power
##     0.58 0.46 0.3227734 -0.483116 150  0.05 0.311424
##
## URL: http://psychstat.org/logistic
```

Figure 3 - Sample Size calculation

```
## Power for logistic regression
##
##       p0   p1     beta0     beta1        n alpha power
##     0.58 0.46 0.3227734 -0.483116 546.8521  0.05   0.8
##
## URL: http://psychstat.org/logistic
```

Figure 4 - Missing values Heatmap

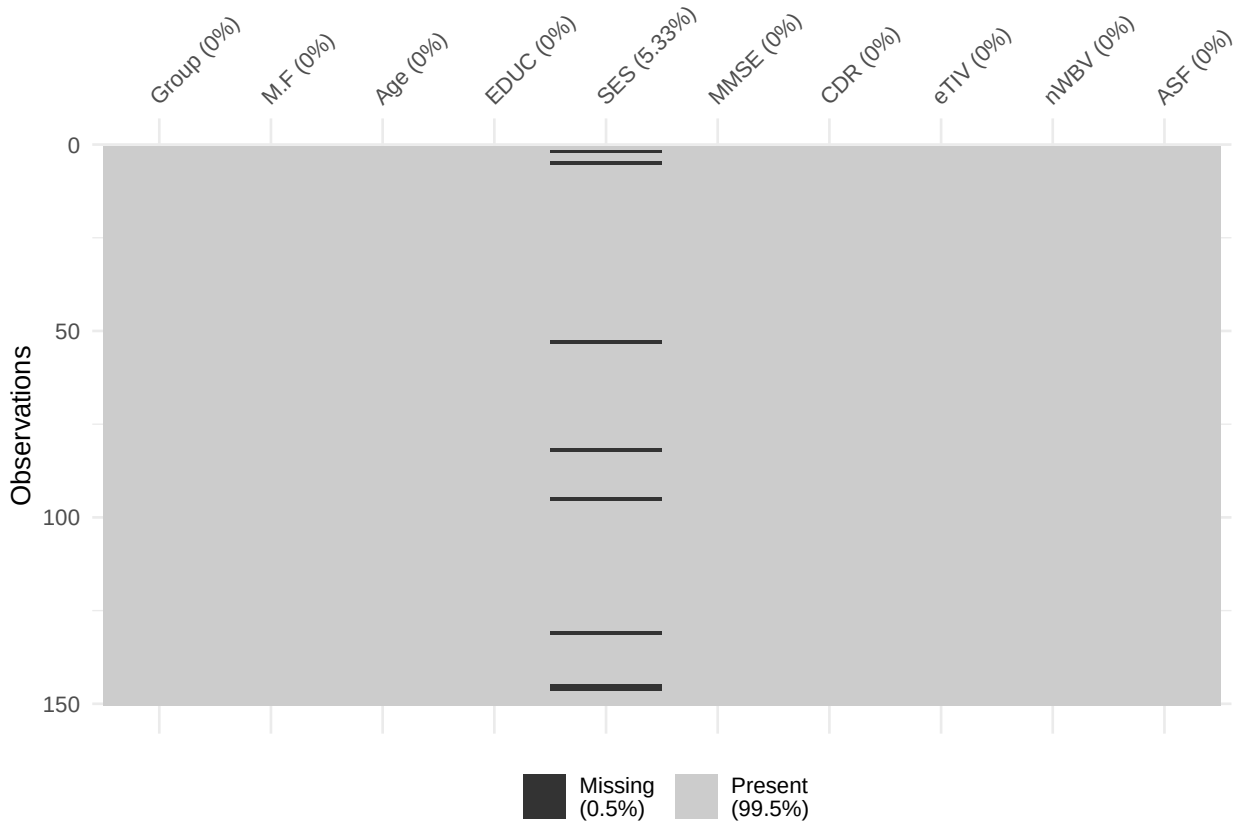

Figure 5 - Methodology Flowchart

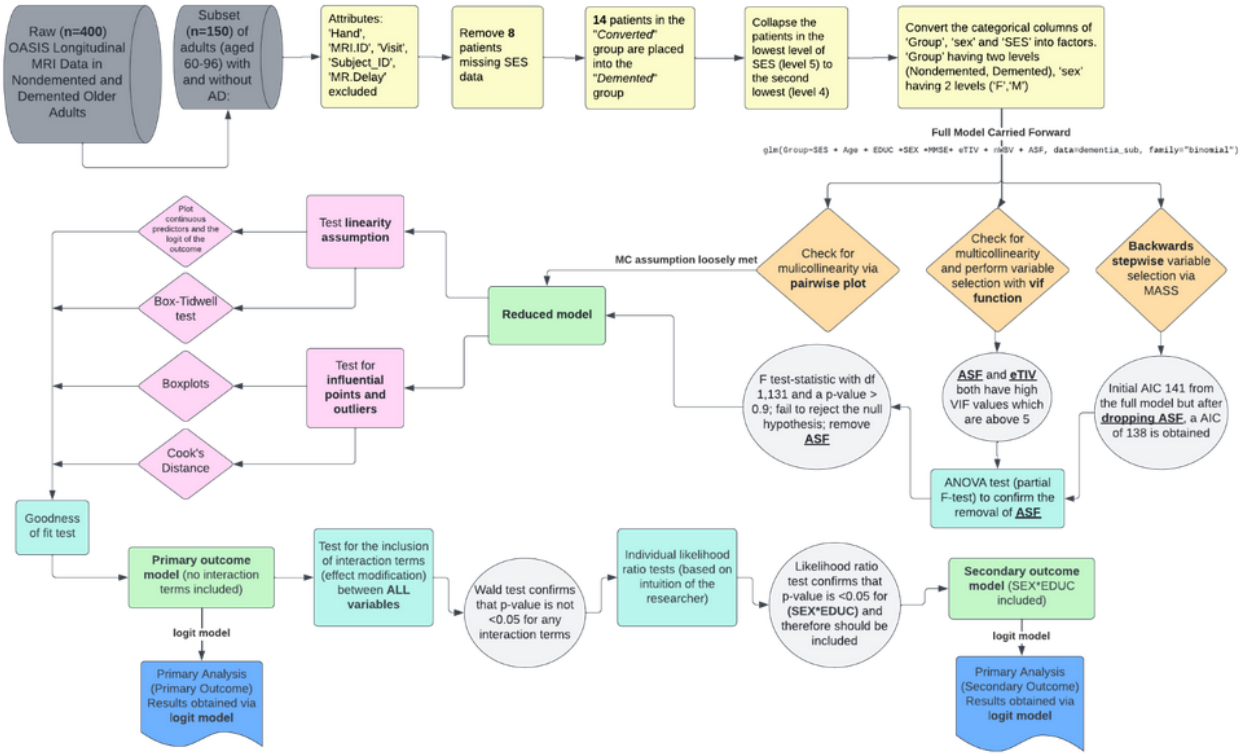

Table 1 - Demographic and Clinical Characteristics of the Study Population:

Variable	**Nondemented**, N = 72	**Demented**, N = 70	**p-value**
SEX			0.011
F	50 (69%)	34 (49%)	
M	22 (31%)	36 (51%)	
Age	75 (8)	75 (7)	>0.9
EDUC	15 (3)	14 (3)	0.029
SES			0.2
1	15 (21%)	18 (26%)	
2	27 (38%)	15 (21%)	
3	16 (22%)	18 (26%)	
4	14 (19%)	19 (27%)	
MMSE	29 (1)	26 (4)	<0.001
eTIV	1,480 (184)	1,471 (167)	0.8
nWBV	0.75 (0.04)	0.73 (0.03)	0.002
ASF	1.20 (0.14)	1.21 (0.13)	0.8

Figure 6 - Scatterplots of Logit vs Continuous Variables

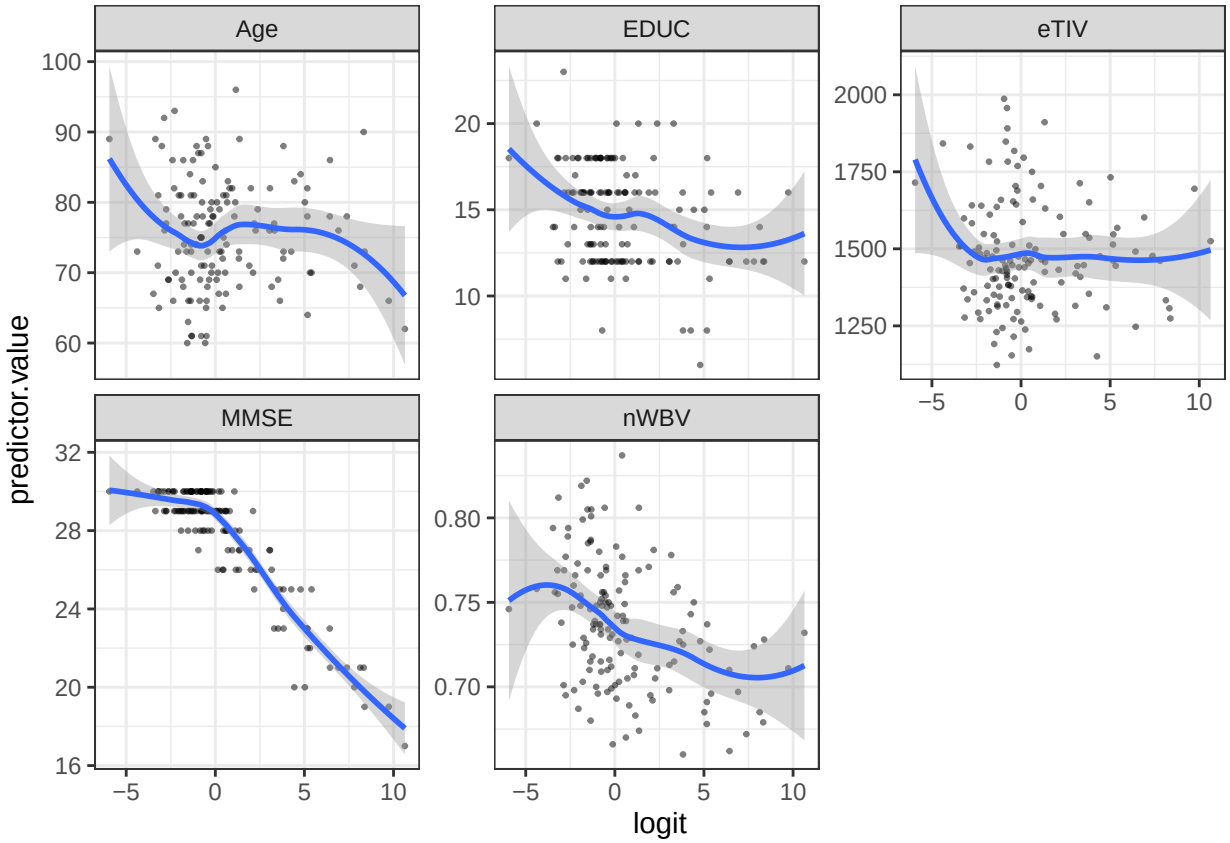

Figure 7 - Boxplots of Continuous Variables

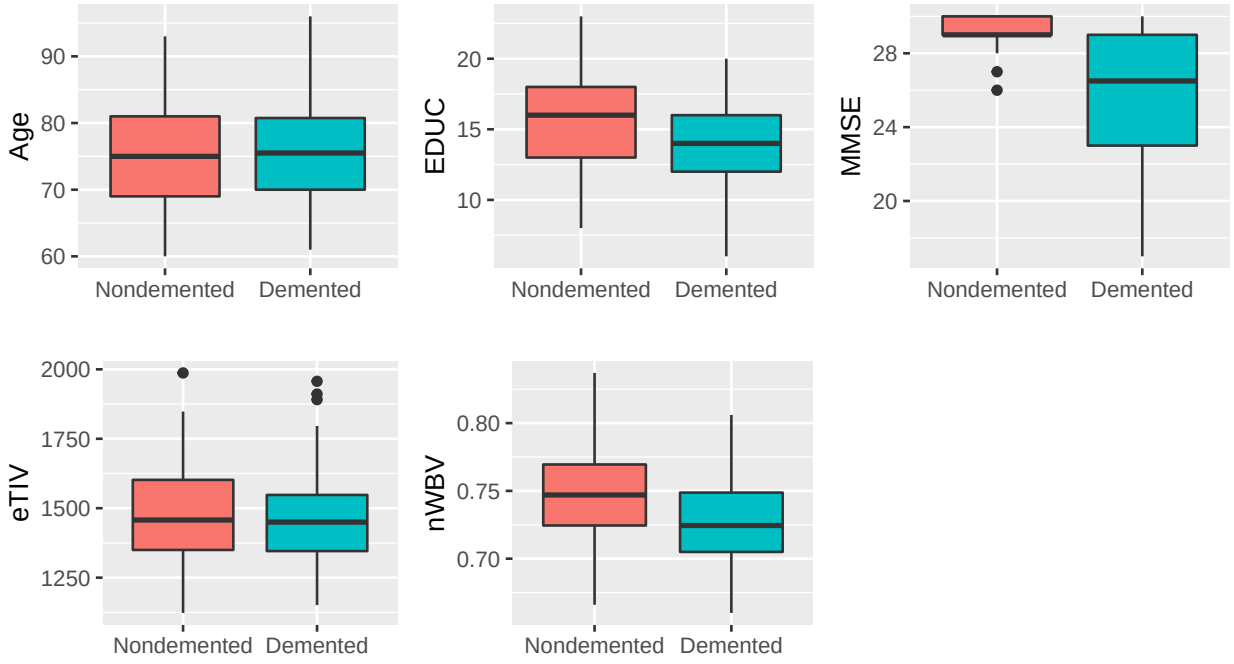

Figure 8 - Cook's Distance:

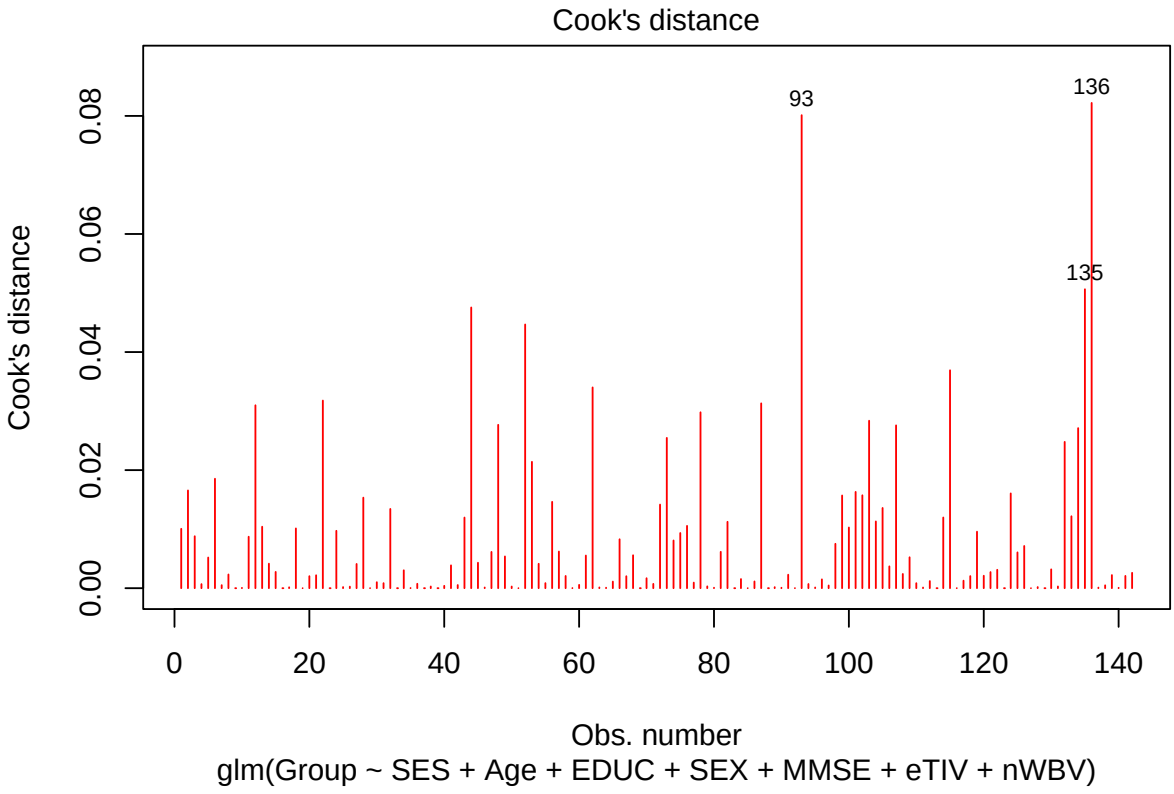

Table 3 Primary Outcomes with Effect Modiciation of Dementia and Socioeconomic Status

Variable	OR	95% CI	p-value	GVIF	Adjusted GVIF
SES			**0.014**	2.7	1.2
1					
2	0.14	0.03, 0.54			
3	0.35	0.07, 1.67			
4	0.09	0.01, 0.61			
Age	0.91	0.83, 0.99	**0.027**	2.1	1.5
EDUC	0.64	0.46, 0.85	**0.002**	3.2	1.8
SEX			0.078	36	6.0
F					
M	0.01	0.00, 1.72			
MMSE	0.43	0.28, 0.60	**<0.001**	1.3	1.1
eTIV	0.99	0.99, 1.00	**0.003**	2.5	1.6
nWBV	0.00	0.00, 0.06	**0.023**	2.1	1.5
EDUC * SEX			**0.020**	41	6.4
EDUC * M	1.55	1.07, 2.32			

Figure 9 - Ratio of CDR overtime

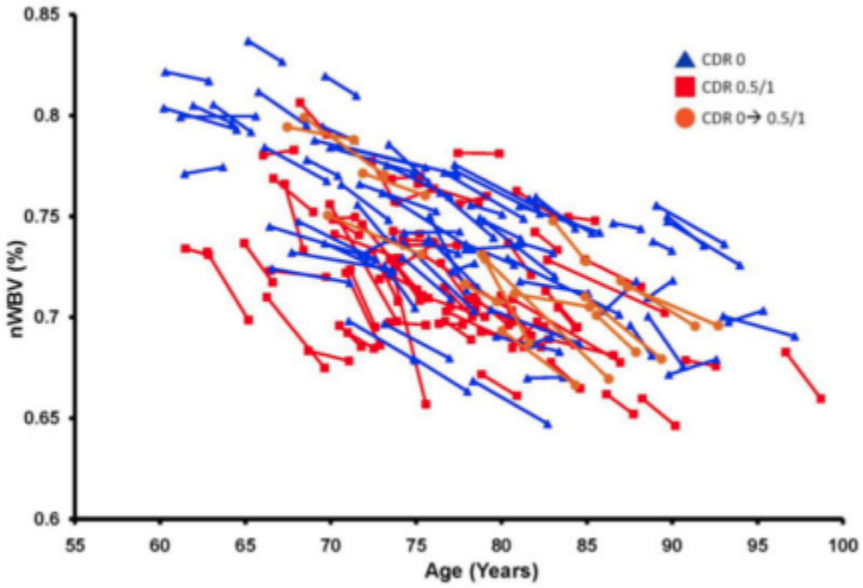

Figure 10 - Health Access Disparity

Economic Stability	Neighborhood and Physical Environment	Education	Food	Community and Social Context	Health Care System
Employment	Housing	Literacy	Hunger	Social integration	Health coverage
Income	Transportation	Language	Access to healthy options	Support systems	Provider availability
Expenses	Safety	Early childhood education		Community engagement	Provider linguistic and cultural competency
Debt	Parks	Vocational training		Discrimination	Quality of care
Medical bills	Playgrounds	Higher education			
Support	Walkability				

Health Outcomes
Mortality, Morbidity, Life Expectancy, Health Care Expenditures, Health Status, Functional Limitations