DATA 621

ALZHEIMER'S DISEASE ANALYSIS

Group 2:
Kimberley Chiu (ID: 00322617)
Marc McCoy (ID: 30136987)
Mark Ly (ID: 00504696)

TABLE OF CONTENTS

Presentation Outline

- Background
- Research Question
- Methods and Analyses
- Results
- Discussion

BACKGROUND

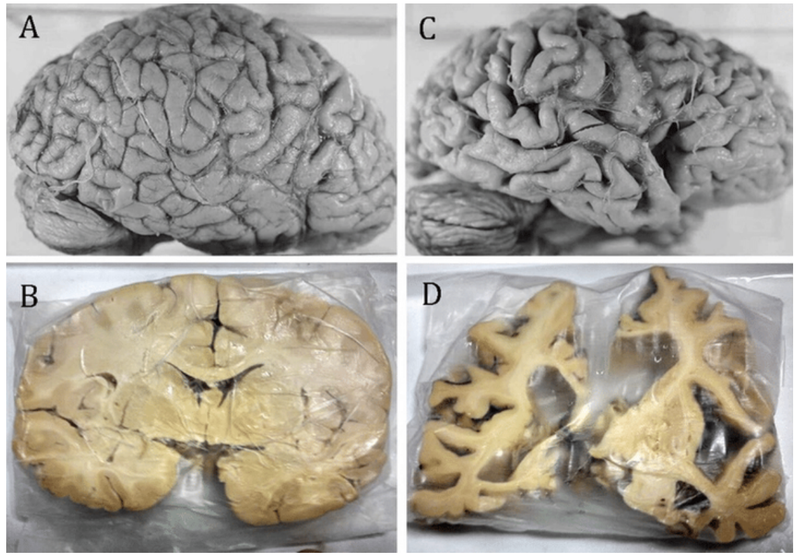
RATIONALE

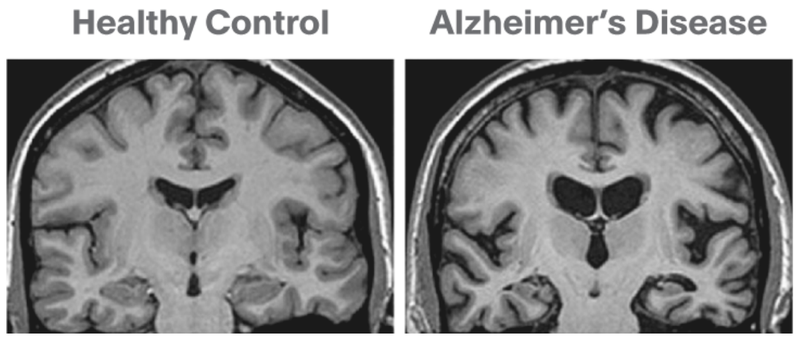
- Dementia is a major neurocognitive disorder that may result in impairments in memory, speech, reasoning, intellectual function, and spatial awareness.
  - Alzheimer's disease (AD) is a subtype of dementia.

- AD is more prevalent in the female population:
  - leading cause of dementia
  - sixth leading cause of death in the United States

- Presents enormous social and economic burdens on society and family members due to potential challenges with activities of daily living.

- 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.

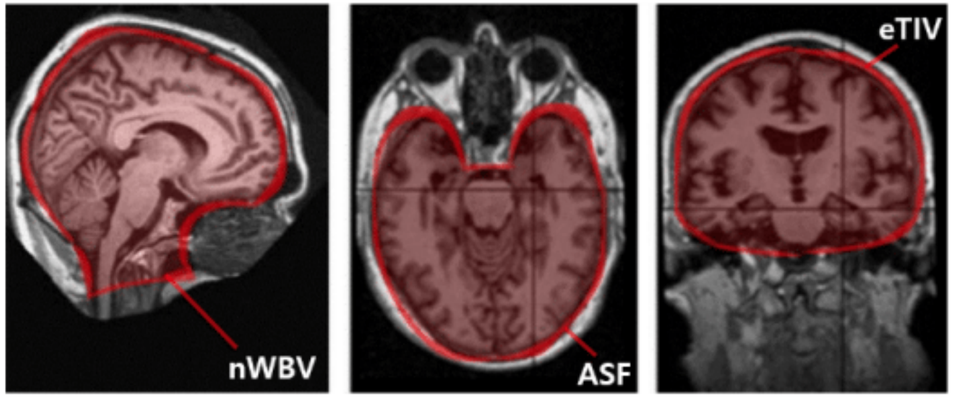




BACKGROUND

DEFINITIONS

- **Outcome (Categorical):**
  - **Group**
    - Non-demented
    - Demented
- **Main Exposure (Categorical):**
  - Socioeconomic Status (**SES**):
    - 1 = highest status to 4 = lowest status
- **Co-variates:**
  - **Education** in years
  - **Age** in years
  - **Sex** (M/F)
  - Mini-Mental State Examination (**MMSE**):
    - ≥24: No dementia, 20–24: mild dementia,13–20: moderate dementia, <13: advanced dementia
  - Estimated total intracranial volume (**eTIV**)
  - Normalized whole brain volume (**nWBV**)
  - Scaling factor that allows for comparison of eTIV based on differences in human anatomy (**ASF**)



MRI outputs nWBV, ASF, and eTIV

RESEARCH QUESTION

- **WHAT** we want to explore:
  - Primary Analysis: In older adults aged 60-96 years old, is **socioeconomic status (SES)** associated with AD as operationalized by binary outcome **Group (nondemented, demented)** while controlling for age, education, sex, MMSE, eTIV, and nWBV as covariates?
  - Secondary Analysis: Are covariates Sex and Education a confounder or effect modifier for binary outcome Group?

- **WHY** we want to explore:
  - 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.

METHODS AND ANALYSES

- *Retrospective cohort study* of adults **(n=150)** aged 60 to 96 that were selected from a larger database of individuals **(n=400)** from the Washington University Alzheimer Disease Research Center (ADRC).
- Originally a longitudinal study centered around series of imaging studies, however we used information from only the **first visit** in our analysis.
- Subjects with a primary cause of dementia other than AD, active neurologic 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**.

- **Source Population:**
  - Demented and nondemented adults (aged 60 to 96) in North America
- **Sample Size:**
  - 150 subjects of adults with and without AD aged 60-96 years from the longitudinal imaging dataset

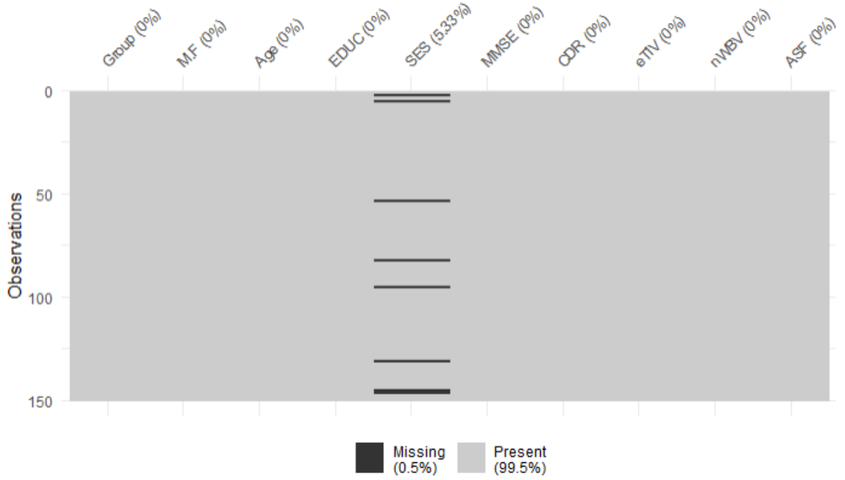


FIGURE 1: Heatmap showing 8 subjects with missing SES data

	CDR 0	*CDR 0.5*	*CDR 1*
Number	86	51	13
Female/male	60/26	21/30	7/6
Age (years)	75.8 ± 8.2 (60–93)	74.8 ± 6.3 (62–90)	75.7 ± 8.7 (61–96)
Education (years)	15.2 ± 2.7 (8–23)	13.6 ± 2.8 (6–20)	14.0 ± 3.2 (8–20)
MMSE	29.1 ± 0.8 (27–30)	26 ± 3.1 (17–30)	23.0 ± 3.3 (19–30)
Prescriptions (*n*)	2.9 ± 2.1 (0–9)	3.2 ± 2.4 (0–11)	2.5 ± 2.4 (0–7)
Systolic BP (mmHg)	135.5 ± 20.3 (98–192)	143.5 ± 19.4 (118–188)	143.4 ± 24.9 (90–188)
Diastolic BP (mmHg)	72.8 ± 10.2 (50–100)	77.1 ± 10.1 (58–98)	76.9 ± 9.2 (60–88)
Reported HBP (%)	54.6	46.0	53.3
Diabetes (%)	9.3	14.0	13.3

FIGURE 2: Descriptive statistics of 150 patient sample (by CDR status)

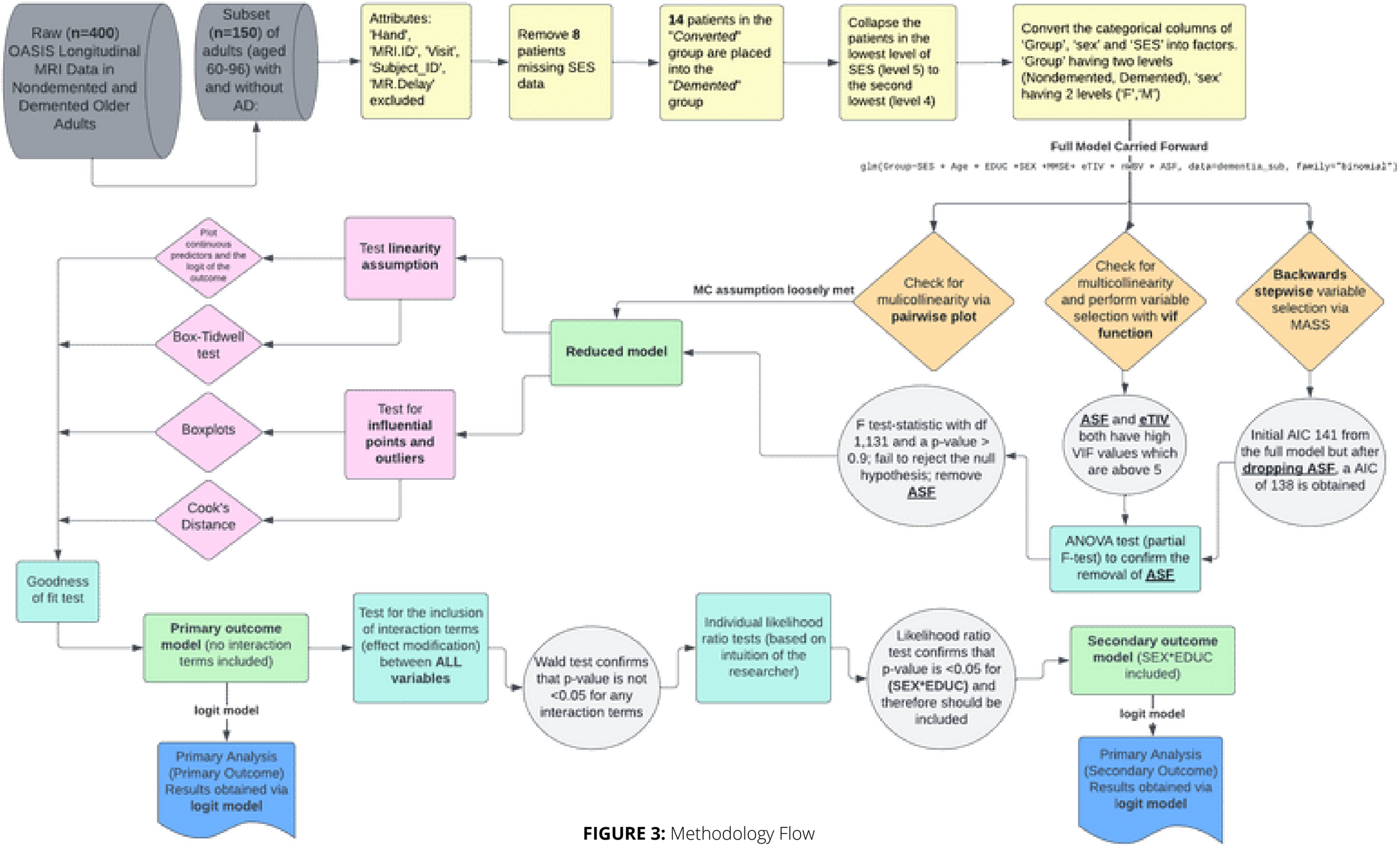


FIGURE 3: Methodology Flow

METHODS AND ANALYSES

- **ANALYSES USED/LEVEL OF SIGNIFICANCE:**
  - *MASS package* was used to run generalized linear models (glm) for **logistic regression**.
  - **Effect modification** and **confounding** was checked via likelihood ratio tests and the 10% rule, respectively
  - Binary outcomes were analyzed with logistic regression and estimated coefficients were exponentiated and reported as adjusted and unadjusted odds ratios with 95% confidence intervals.
  - All p-values are two-sided.
  - **Level of significance of α=0.05**

- **ASSUMPTIONS:**
  - **Linearity:** Scatter plots (Explanatory vs logit of outcome) were plotted and visually inspected.
  - **Independence:** Assumed to be met because raw data is not time series (after transformation)
  - **Multicollinearity:** Variance inflation factors (VIF) were used; the ASF variable was removed from the final model

- **MISSING VALUES:**
  - 8 subjects with missing SES data were removed instead of imputing, to prevent skewing the data

- **OUTLIERS:**
  - Assessed via **residual vs. leverage plots, cook's distance, and leverage points**; a separate model was run with the outliers removed

RESULTS

DESCRIPTIVE STATS

- The gtsummary package in R was used to compare the potential predictors in the "*Demented*" group (n = **70**) and the "*Non-demented*" group (n = **72**) for individuals between the ages of 60 - 96 years old.

FINDINGS:

- Significantly more female participants in the nondementia group (**69**%) than the dementia group (**49**%).
- Mean age of boths groups was **75** years old
- Mean education years same in both groups (~**15**).
- Higher proportion of patients with higher SES levels in non-demented group than in the dementia group.
- More participants in the dementia group in the lower SES levels compared to those in the no dementia group.

METHOD:

- The p-values reported in the figure are derived from:
  - Pearson's Chi-Squared test for categorical variables with expected cell counts ≥ 5
  - Wilcoxon rank sum test for numerical variables
- We used a significance level of alpha = **0.05** to compare our findings.

FIGURE 4: Descriptive Statistics

Variable	Nondemented, N = 72[1]	Demented, N = 70[1]	p-value[2]
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

[1] n (%); Mean (SD)
[2] Pearson's Chi-squared test; Wilcoxon rank sum test

RESULTS

PRIMARY ANALYSIS

$$\widehat{logit(p)} = \widehat{\beta_0} + \widehat{\beta_1} * SES2 + \widehat{\beta_2} * SES3 + \widehat{\beta_3} * SES4 + \widehat{\beta_4} + Age + \widehat{\beta_5} * EDUC + \widehat{\beta_6} * Males + \widehat{\beta_7} * MMSE + \widehat{\beta_8} * eTIV + \widehat{\beta_9} * nWBV$$

$$\widehat{logit(p)} = 54.04 - 1.857 * SES2 - 0.949 * SES3 - 2.477 * SES4 - 0.0968 * Age - 0.2778 * Educ + 1.373 * Males - 0.802 * MMSE - 0.004296 * eTIV - 21.551 * nWBV$$

Reference case is non-demented, female, and SES=1

- **e^β1** = 0.156 = OR of dementia change from SES level of 1 to a SES level of 2
- **e^β2** = 0.387 = OR of dementia change from SES level of 1 to a SES level of 3
- **e^β3** = 0.084 = OR of dementia change from SES level of 1 to a SES level of 4
- **e^β4** = 0.908 = OR of dementia for each 1 year increase in age for females with a SES of 1.
- **e^β5** = 0.757 =OR of dementia for each 1 year increase in education for females with a SES of 1.
- **e^β6** = 3.95 = OR of dementia for men vs. women at a SES of 1.
- **e^β7** = 0.448 = For every one point increase in MMSE, the odds of dementia is 0.488 times higher for women in the highest SES
- **e^β8** = 0.995 = For every 1 unit increase in eTIV, the odds of dementia is 0.995 times higher for women in the highest SES.
- **e^β9** = 0.00 = For every 1 unit increase in eTIV, the odds of dementia is 0 times higher for women in the highest SES.

The odds ratio of dementia for males at a SES level of 1 is **3.95** (95% CI: 1.14 - 15.1; p-value =0.030) compared to females at the same level of SES, holding all other variables constant

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 57.040108  13.783034   4.138 3.50e-05 ***
SES2        -1.857340   0.694932  -2.673  0.00752 **
SES3        -0.948694   0.783591  -1.211  0.22601
SES4        -2.477091   0.979889  -2.528  0.01147 *
Age         -0.096753   0.043975  -2.200  0.02779 *
EDUC        -0.277784   0.128329  -2.165  0.03042 *
SEXM         1.373275   0.652781   2.104  0.03540 *
MMSE        -0.802380   0.185206  -4.332 1.48e-05 ***
eTIV        -0.004296   0.001811  -2.372  0.01771 *
nWBV       -21.550511   9.470697  -2.275  0.02288 *
```

FIGURE 5: Model Result Summary

Characteristic	OR[1]	95% CI[1]	p-value	GVIF[1]	Adjusted GVIF[2,1]
SES			**0.011**	2.8	1.2
1	1.00	1.00			
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	1.00	1.00			
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	0.00	0.00, 0.03	**0.019**	2.2	1.5

[1] OR = Odds Ratio, CI = Confidence Interval, GVIF = Generalized Variance Inflation Factor
[2] GVIF^[1/(2*df)]

RESULTS

GOODNESS OF FIT

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  57.040108  13.783034   4.138 3.50e-05 ***
SES2         -1.857340   0.694932  -2.673  0.00752 **
SES3         -0.948694   0.783591  -1.211  0.22601
SES4         -2.477091   0.979889  -2.528  0.01147 *
Age          -0.096753   0.043975  -2.200  0.02779 *
EDUC         -0.277784   0.128329  -2.165  0.03042 *
SEXM          1.373275   0.652781   2.104  0.03540 *
MMSE         -0.802380   0.185206  -4.332 1.48e-05 ***
eTIV         -0.004296   0.001811  -2.372  0.01771 *
nWBV        -21.550511   9.470697  -2.275  0.02288 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 196.83  on 141  degrees of freedom
Residual deviance: 118.79  on 132  degrees of freedom
AIC: 138.79
```

Cumulative Chi Square density:

- Want to see if the data fits our model
  - **Ho**: The data fits our selected model distribution
  - **Ha**: The data does not fit our selected model distribution
- Significance level
  - **0.05**

```r
1-pchisq(118.79,132)
```

```
[1] 0.7882735
```

- Our model with a residual deviance of 118.79 on 132 df had a p-value of **0.788** which is greater than our significance level of **0.05**

We **fail** to reject the null hypothesis and can say that the data does fit our selected model distribution.

RESULTS

LINEARITY

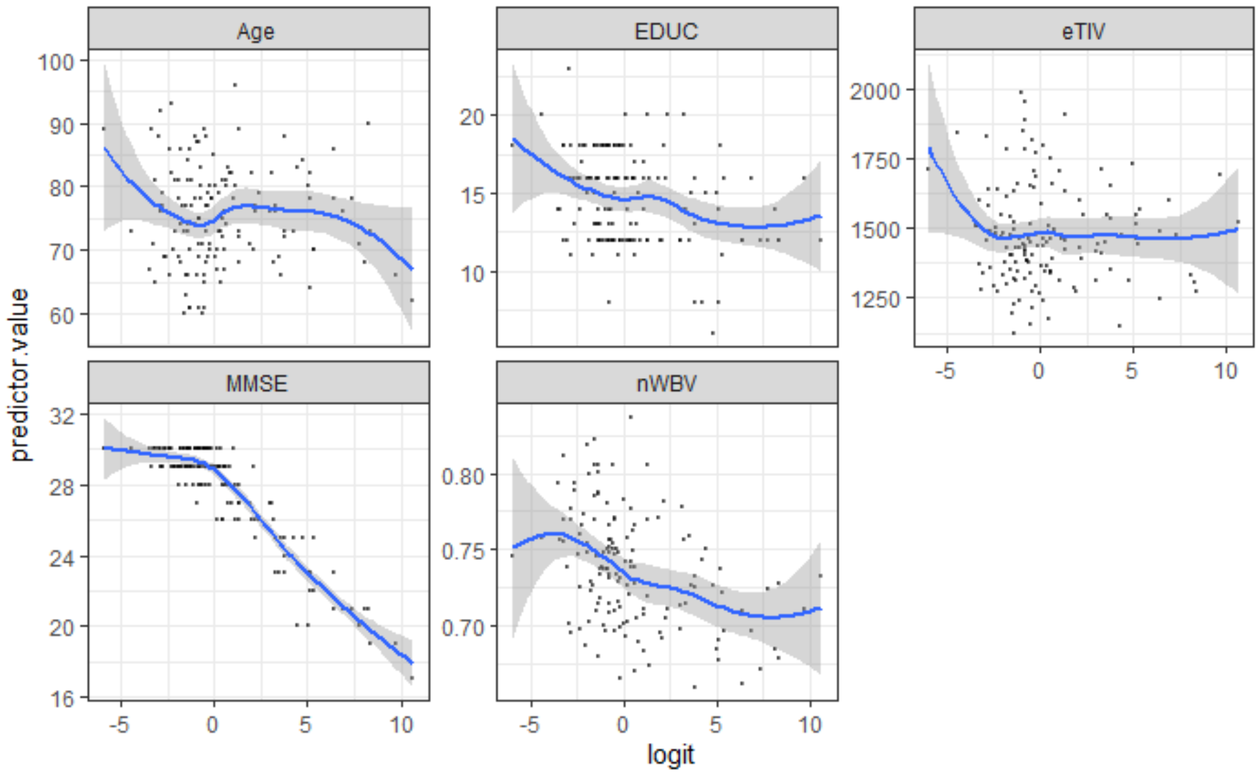


FIGURE 6: Linearity Plots

Box-Tidwell:

- Adds interactions between continuous variables and corresponding natural log into the model.
  - **Ho**: Continuous X variables are linearly related to the the log odds
  - **Ha**: Continuous X variables are not linearly related to the log odds
- Significance level
  - **0.05**

```
                      MLE of lambda Score Statistic (z) Pr(>|z|)
dementia_sub$Age            0.36427              0.2876   0.7736
dementia_sub$EDUC           0.47207              0.1121   0.9108
dementia_sub$MMSE           1.54500             -0.9509   0.3416
dementia_sub$eTIV           1.56662             -0.1761   0.8602
dementia_sub$nWBV          -0.30807              0.5896   0.5554

iterations =  15
```

- All z scores for each of the variables have a p-value greater than our significance level **0.05**

We **fail** to reject the null hypothesis and can say that our continuous x variables are linearly related to the log-odds and the linearity assumption is met

RESULTS

MULTICOLLINEARITY

VIF:

- Was checked during variable selection
- Remove VIF values one by one if they are greater than 5 (**ASF** and **eTIV**)

	GVIF	Df	GVIF^(1/(2*Df))
SES	2.808293	3	1.187792
Age	2.175392	1	1.474921
EDUC	2.338931	1	1.529356
SEX	2.003545	1	1.415466
MMSE	1.206772	1	1.098532
eTIV	45.555634	1	6.749491
nWBV	2.229360	1	1.493104
ASF	44.798171	1	6.693144

- Backwards stepwise regression removed **ASF**

	GVIF	Df	GVIF^(1/(2*Df))
SES	2.769844	3	1.185066
Age	2.151668	1	1.466856
EDUC	2.258286	1	1.502759
SEX	2.008336	1	1.417158
MMSE	1.206625	1	1.098465
eTIV	2.222543	1	1.490820
nWBV	2.226062	1	1.491999

Partial F-Test:

- Determine differences between two models
  - **Ho**: All coefficients removed from the model are zero
  - **Ha**: At least one of the coefficients removed from the model is non-zero
- Significance level
  - **0.05**

```
Analysis of Deviance Table

Model 1: Group ~ SES + Age + EDUC + SEX + MMSE + eTIV + nWBV
Model 2: Group ~ SES + Age + EDUC + SEX + MMSE + eTIV + nWBV + ASF
  Resid. Df Resid. Dev Df  Deviance Pr(>Chi)
1       132     118.79
2       131     118.78  1 0.0080112   0.9287
```

- F-test statistic with df 1,131 and a p-value **0.9287**. This is greater than our significance level of **0.05**.

We **fail** to reject the null hypothesis and we can drop the ASF covariate because it does not significantly improve the fit of the model.

RESULTS

INDEPENDENCE

Study Design:

Original data was longitudinal, we decided to use the first visit so each patient is unique and counted once. The independence assumption is met.

> The data set includes longitudinal MRI data from 150 individuals age 60 to 96 years, including 64 individuals with very mild to moderate AD as diagnosed clinically and characterized using the Clinical Dementia Rating (CDR) scale (Morris et al., 2001; Morris, 1993) at their initial visit. Another 14 of the individuals were characterized as nondemented at the time of one or more scans and then clinically

CONFOUNDING

Covariates:

- Age, Sex, SES, education are expected to have a confounding effect
- Significance level
  - **0.05**

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  57.040108  13.783034   4.138 3.50e-05 ***
SES2         -1.857340   0.694932  -2.673  0.00752 **
SES3         -0.948694   0.783591  -1.211  0.22601
SES4         -2.477091   0.979889  -2.528  0.01147 *
Age          -0.096753   0.043975  -2.200  0.02779 *
EDUC         -0.277784   0.128329  -2.165  0.03042 *
SEXM          1.373275   0.652781   2.104  0.03540 *
MMSE         -0.802380   0.185206  -4.332 1.48e-05 ***
eTIV         -0.004296   0.001811  -2.372  0.01771 *
nWBV        -21.550511   9.470697  -2.275  0.02288 *
---
```

- The Wald test of significance for each of the variables are <**0.05**.
- Magnitude of confounding not required.

RESULTS

OUTLIERS

- MMSE and eTIV both have outliers however, not all outliers are influential observations.
- 3 highest values are from point 93 , 135 and 136.
  - All 3 points have Cook's distance that is less than 1 (leave them in our model).

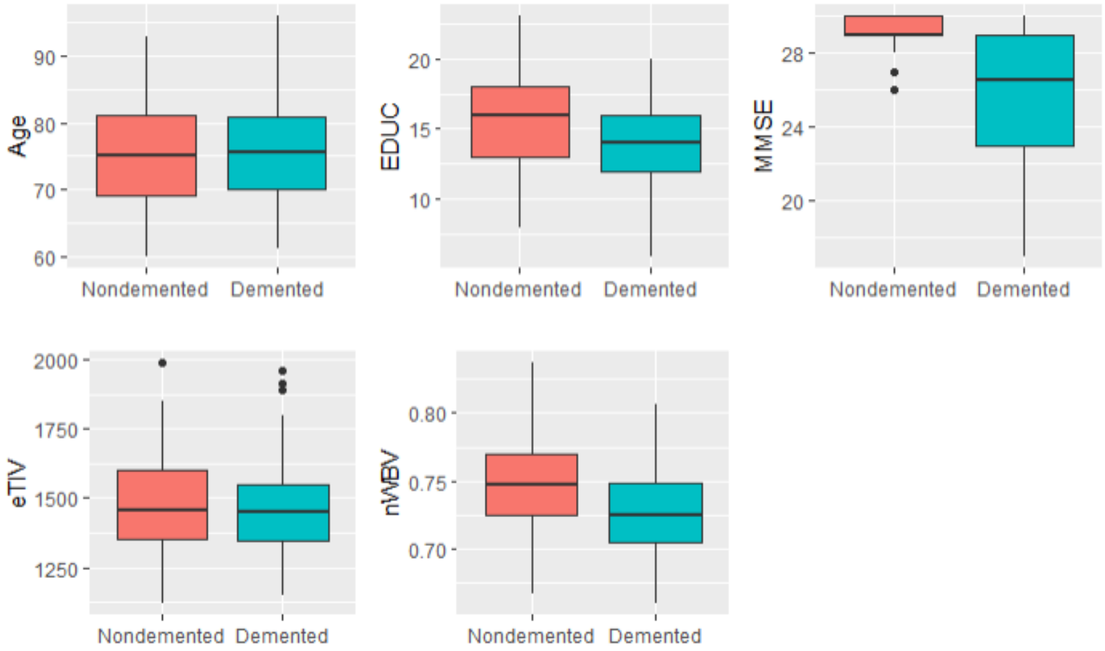


FIGURE 7 - Boxplot to Determine Outliers

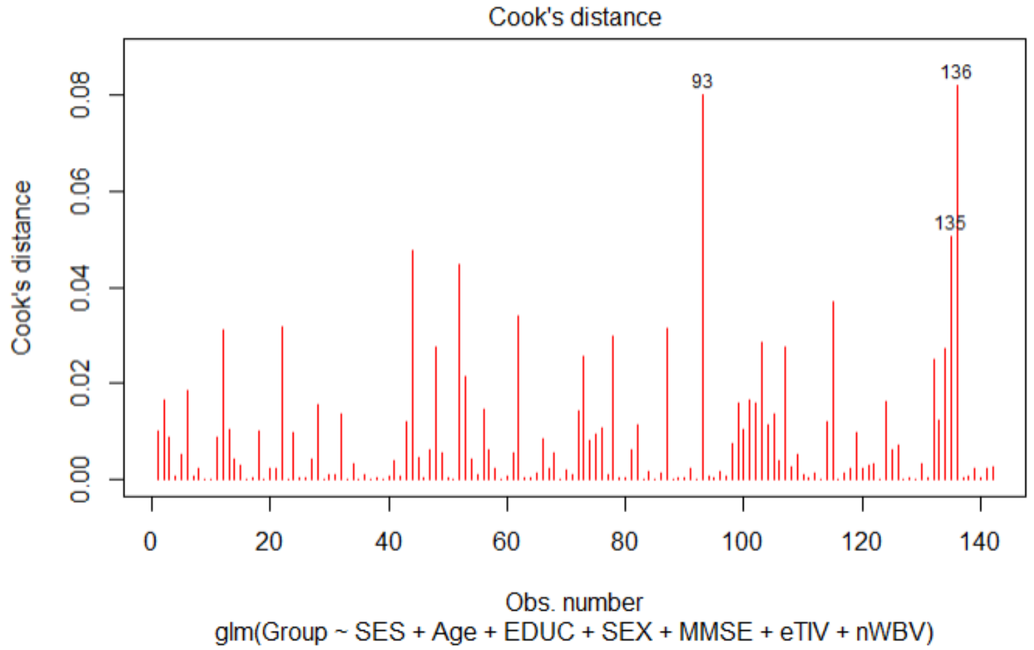


FIGURE 8 - Cook's Distance

RESULTS

SECONDARY ANALYSIS

Might have interaction with the sex and other covariates

- Sex and Education was a significant interaction based on the Wald test (p-value = **0.02437**)
- Sex now has a insignificant Wald test (p-value = **0.08202**)

$$\widehat{logit(p)} = \widehat{\beta_0} + \widehat{\beta_1} * SES2 + \widehat{\beta_2} * SES3 + \widehat{\beta_3} * SES4 + \widehat{\beta_4} + Age + \widehat{\beta_5} * EDUC + \widehat{\beta_6} * Males + \widehat{\beta_7} * MMSE + \widehat{\beta_8} * eTIV + \widehat{\beta_9} * nWBV + \widehat{\beta_{10}} * Males * Education$$

$$\widehat{logit(p)} = 54.04 - 1.857 * SES2 - 0.949 * SES3 - 2.477 * SES4 - 0.0968 * Age - 0.2778 * Educ + 1.373 * Males - 0.802 * MMSE - 0.004296 * eTIV - 21.551 * nWBV + 0.4394 * Males * Education$$

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	61.978805	14.448400	4.290	1.79e-05	***
SES2	-1.985206	0.725372	-2.737	0.00620	**
SES3	-1.058593	0.816948	-1.296	0.19505	
SES4	-2.367873	0.984393	-2.405	0.01615	*
Age	-0.095184	0.044835	-2.123	0.03376	*
EDUC	-0.451744	0.156414	-2.888	0.00388	**
SEXM	-4.889628	2.811603	-1.739	0.08202	.
MMSE	-0.842155	0.197405	-4.266	1.99e-05	***
eTIV	-0.005493	0.002005	-2.740	0.00615	**
nWBV	-21.167812	9.587129	-2.208	0.02725	*
EDUC:SEXM	0.439478	0.195222	2.251	0.02437	*

Likelihood ratio test:

- Determine differences between two models
  - **Ho**: There is no difference in the two models
  - **Ha**: Using the full model is 'better' than reduced
- Significance level
  - **0.05**

```
Analysis of Deviance Table

Model 1: Group ~ SES + Age + EDUC + SEX + MMSE + eTIV + nWBV
Model 2: Group ~ SES + Age + EDUC + SEX + MMSE + eTIV + nWBV + SEX * EDUC
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       132     118.79
2       131     113.36  1   5.4304  0.01979 *
```

- Change in residual deviance is 5.4304 with df 1,131 and a p-value **0.01979**. This is less than our acceptance criteria

We will **reject** the null hypothesis and say that the interaction term is statistically significant and should be left in our model.

RESULTS

SECONDARY ANALYSIS

$$\widehat{logit(p)} = \begin{cases} \widehat{\beta_0} + \widehat{\beta_1} + \widehat{\beta_2} + \widehat{\beta_3} + \widehat{\beta_4} + \widehat{\beta_5} + \widehat{\beta_7} + \widehat{\beta_8} + \widehat{\beta_9} + \epsilon, & \text{if person is female} \\ \widehat{\beta_0} + \widehat{\beta_1} + \widehat{\beta_2} + \widehat{\beta_3} + \widehat{\beta_4} + \widehat{\beta_5} + \widehat{\beta_6} + \widehat{\beta_7} + \widehat{\beta_8} + \widehat{\beta_9} + \widehat{\beta_{10}} + \epsilon, & \text{if person is male} \end{cases}$$

Reference case is non-demented, female, and SES=1

- **e^β1** = 0.137 = OR of dementia change from SES level of 1 to a SES level of 2
- **e^β2** = 0.469 = OR of dementia change from SES level of 1 to a SES level of 3
- **e^β3** = 0.009 = OR of dementia change from SES level of 1 to a SES level of 4
- **e^β4** = 0.902 = OR of dementia for each 1 year increase in age for females with a SES of 1.
- **e^β5** = 0.637 = OR of dementia for each 1 year increase in education for females with a SES of 1.
- **e^β6** = 0.01 = OR of dementia for men vs. women at a SES of 1.
- **e^β7** = 0.431 = For every one point increase in MMSE, the odds of dementia is 0.431 times higher for women in the highest SES
- **e^β8** = 0.995 = For every 1 unit increase in eTIV, the odds of dementia is 0.995 times higher for women in the highest SES.
- **e^β9** = 0.00 = For every 1 unit increase in eTIV, the odds of dementia is 0 times higher for women in the highest SES.
- **e^β10** = 1.55 = The odds ratio for the association between Males and Education at the highest SES level.

FIGURE 9: Model Output Summary

Characteristic	OR[1]	95% CI[1]	p-value	GVIF[1]	Adjusted GVIF[2,1]
SES			**0.014**	2.7	1.2
1	1.00	1.00			
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	1.00	1.00			
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			

[1] OR = Odds Ratio, CI = Confidence Interval, GVIF = Generalized Variance Inflation Factor
[2] GVIF^[1/(2*df)]

DISCUSSION

CONCLUSIONS DRAWN

CONCLUSION:

- **Socioeconomic status (SES) is associated with AD as operationalized by binary outcome group (nondemented, demented) while controlling for age, education, sex, MMSE, eTIV, and nWBV as covariates**
- Deviations from our expectations:
  - Expected the OR would be >1 for lower SES (levels: 2, 3, 4)
  - Expected that the OR for males would be <1
  - Expected that the OR would be >1 for age

INTERPRETATIONS (PRIMARY ANALYSIS):

- The odds of the patient being demented with SES = 4 is 0.08x the odds of a patient with an SES=1 and it is statistically significant.
- The odds of a male patient being demented is 3.95x the odds of a female patient and it is statistically significant.
- A one year increase in education corresponds with a 24% decrease in AD risk and it is statistical significant.
- A one unit score increase in MMSE is corresponds with a 55% decrease in AD risk and it is statistically significant.
- All tested variables showed significance with the exception of ASF (removed), which is used to derive eTIV.
- eTIV, and nWBV are significant predictors but have OR of 1.00 and 0.00 respectively:
  - We treated these variables inappropriately in our statistical analysis plan and should have considered the delta over time of nWBV as an explanatory variable (proxy for brain atrophy)
  - eTIV doesn't change over the patient's lifetime, but nWBV may possibly change

Characteristic	OR[1]	95% CI[1]	p-value	GVIF[1]	Adjusted GVIF[2,1]
SES			**0.011**	2.8	1.2
1	1.00	1.00			
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	1.00	1.00			
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	0.00	0.00, 0.03	**0.019**	2.2	1.5

[1] OR = Odds Ratio, CI = Confidence Interval, GVIF = Generalized Variance Inflation Factor
[2] GVIF^[1/(2*df)]

FIGURE 10: Primary analysis model outputs

Characteristic	OR[1]	95% CI[1]	p-value	GVIF[1]	Adjusted GVIF[2,1]
SES			**0.014**	2.7	1.2
1	1.00	1.00			
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	1.00	1.00			
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			

[1] OR = Odds Ratio, CI = Confidence Interval, GVIF = Generalized Variance Inflation Factor
[2] GVIF^[1/(2*df)]

FIGURE 11: Secondary analysis model outputs

DISCUSSION

LIMITATIONS

PAUCITY OF LITERATURE TO DRAW ON:

- Pathophysiology of AD is still not well understood (are the physical changes driving the symptoms or are the symptoms driving the physical changes… we still do not know); the issue of **bidirectionality.**
- Logistic regression does not require the assumption of normality of errors to be met.
- Up to this point, the open source data has been utilized primarily for **exploratory analysis** to understand the physiologic relationships between MRI findings and the development/presence of AD.

LIMITED SAMPLE SIZE REDUCES POWER:

- The limitation of the current data is that we will have a lower power with **(n = 150)** total subjects. This coincides with a power of 31%.
  - Studies on patient retention in clinical trials indicate 30% in dropout rates.
  - A larger sample size is required **(n = 547)** to achieve adequate power (80%) with a two-tailed significance level of α = 0.05
  - A sample size of **(n = 712)** individuals is required to have a robust final sample size.

To find power with our data's sample size:

```r
wp.logistic(n = 150, p0 = P0, p1 = P1, alpha = Alpha,
                        power = NULL, alternative = c("two.sided"),
                        family = c("Bernoulli"), parameter = px1)
```

```
Power for logistic regression

      p0   p1     beta0     beta1   n alpha    power
    0.58 0.46 0.3227734 -0.483116 150  0.05 0.311424
```

FIGURE 12: Power with (n=150) is only 31%

```r
# list all the parameters
Alpha = 0.05
P0 = 0.58
px1 = 0.5 #default for Bernoulli
Power = 0.80
P1 = 0.46


output <- wp.logistic(n = NULL, p0 = P0, p1 = P1, alpha = Alpha,
                        power = Power, alternative = c("two.sided"),
                        family = c("Bernoulli"), parameter = px1)
output
```

```
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
```

FIGURE 13: Required sample size of (n=712) to obtain 80% power based on an assumed 30% dropout rate

DISCUSSION

FUTURE PLANS

EXPLANATORY VARIABLE ADJUSTMENTS:

- **Change in nWBV (or change in nWBV/eTIV)** instead of first visit nWBV as an explanatory variable
  - Assumption of independence is no longer met
  - Would need to explore the utilization of a mixed-effects logistic regression model via xtmelogit command in R
- More emphasis on exploring **social determinants of health** as potential explanatory variables

FORMALIZED STUDY DESIGN AND DATA COLLECTION PROCESS:

- Unusual result: OR for lower SES status:
  - Literature suggests this could be an indicator of **systemic healthcare access inequity** (patients with higher SES receive access to physician quicker and an therefore receive an earlier diagnosis)
  - Formalize a study design to **combat the imbalance in the data** with regards to SES and other social determinants of health

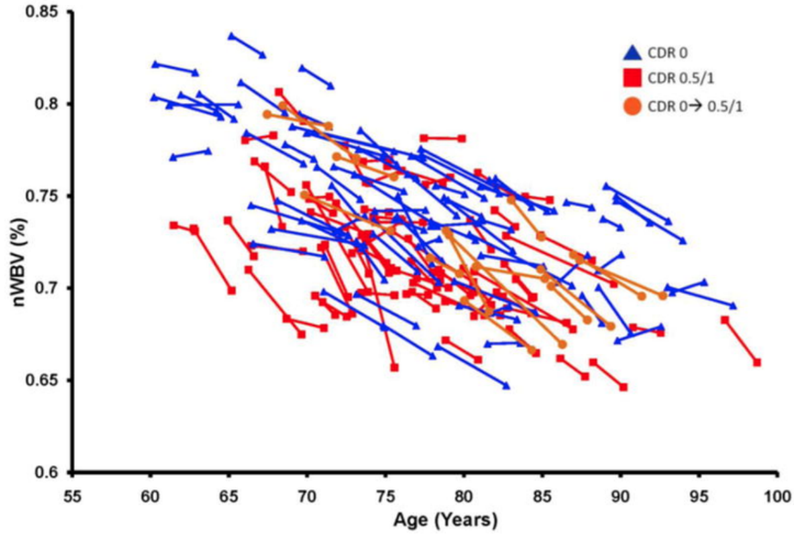


FIGURE 14: Longitudinal plot of the delta in nWBV (atrophy)

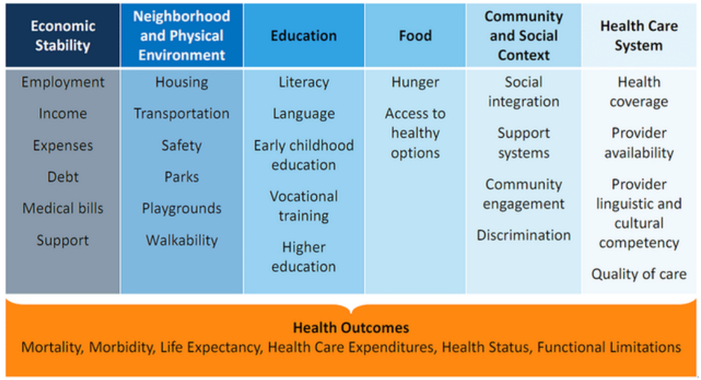
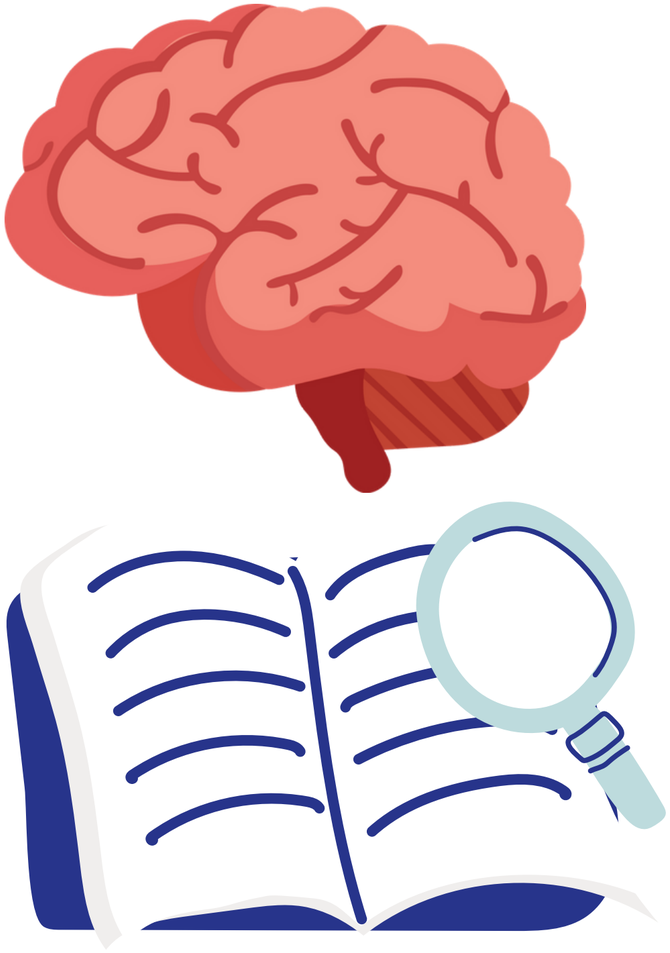


FIGURE 15: Social Determinants of Health defined by WHO



THANK YOU!

REFERENCES

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM–5).2013. doi: 10.1176/appi.books.9780890425596.| Open in Read by QxMD
- Boysen J. MRI and Alzheimer’s. Kaggle. https://www.kaggle.com/jboysen/mri-andalzheimers?select=oasis_longitudinal.csv. Published August 16, 2017. Accessed February 11, 2022.
- 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. MIT Press. https://direct.mit.edu/jocn/article-abstract/19/9/1498/4427/Open-Access-Series-ofImaging-Studies-OASIS-Cross?redirectedFrom=fulltext. Published September 1, 2007. Accessed February 11, 2022.
- 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.
- 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.