Final Project - Alzheimer’s Disease

Michael Paris

March 2, 2021

# Introduction

  Dementia is defined as a decline in mental ability severe enough to interfere with daily life. Alzheimer’s is a degenerative brain disease that is caused by complex brain changes following cell damage. It leads to dementia symptoms that gradually worsen over time. (“Dementia Vs. Alzheimer’s Disease: What Is the Difference?” n.d.) Two of my grandparents died with several of the signs of dementia but never had a diagnosis of Alzheimer’s disease.

  In this analysis, we’ll discuss several research questions and attempt to glean some incites from the data.

* What are the main risk factors for developing this disease?
* Are there any secondary risk factors?
* Is Alzheimer’s disease becoming more common?
* Is it possible to predict who might develop this disease?

  To answer these questions, we’ll perform an analysis of the data to determine the significant correlation between risk factors to determine which factors may be considered major versus minor. If the data is available, a similar approach will be taken to determine if there is a correlation within families. To determine if different ethnicities are more or less at risk, analyzing positive cases as a percent of the total ethnic population will be reviewed.

  To determine if the disease is becoming more common, I plan to plot the positive diagnosis numbers against the general population over time to see if variables are increasing at similar rates or if they are not connected. With the information above, we may be able to show which groups are more at risk of developing the disease.

# Datasets

  The datasets we will analyze are listed below:

* Weekly counts of death by jurisdiction and cause of death \*\* Center for Disease Control \*\* <https://healthdata.gov/dataset/weekly-counts-death-jurisdiction-and-cause-death> \*\* Updated February 17, 2021 \*\* 334K records, 15 columns
* Population, Population Change, and Estimated Components of Population Change: April 1, 2010 to July 1, 2019 (NST-EST2019-alldata) \*\* United States Census Bureau \*\* <https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/national/totals/nst-est2019-alldata.csv>
* ACS Demographic and Housing Estimates \*\* United States Census Bureau \*\* <https://data.census.gov/cedsci/table?q=demographics&tid=ACSDP1Y2019.DP05&hidePreview=false>
* Alzheimer’s Disease and Healthy Aging Data \*\* Center for Disease Control \*\* <https://healthdata.gov/dataset/alzheimers-disease-and-healthy-aging-data> \*\* Updated January 20, 2021 \*\* 144k records, 39 columns
* Oasis MRI Demographics Data \*\* Oasis \*\* <https://www.oasis-brains.org/>

# Data Cleanup

For the weekly count of death by jurisdiction and cause of death, we have performed several operations on the dataset. In the first step in the cleanup, we removed rows that contained estimated deaths for a specific period while keeping the rows of actual death data. In step two, we combined weekly data into a single row representing each year.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Year | Cause.Group | Cause.Subgroup | Number.of.Deaths |
| 1 | 2015 | Alzheimer disease and dementia | Alzheimer disease and dementia | 489139 |
| 14 | 2016 | Alzheimer disease and dementia | Alzheimer disease and dementia | 497457 |
| 27 | 2017 | Alzheimer disease and dementia | Alzheimer disease and dementia | 523662 |
| 40 | 2018 | Alzheimer disease and dementia | Alzheimer disease and dementia | 533988 |
| 53 | 2019 | Alzheimer disease and dementia | Alzheimer disease and dementia | 543350 |
| 66 | 2020 | Alzheimer disease and dementia | Alzheimer disease and dementia | 613544 |

For the population dataset, we reduced the number of columns to match the same years for the death data. Years 2015 through 2019. We removed data for the year 2020 as it was only a partial year’s worth of data.

|  |  |
| --- | --- |
| Year | Population |
| 2015 | 320635163 |
| 2016 | 322941311 |
| 2017 | 324985539 |
| 2018 | 326687501 |
| 2019 | 328239523 |
| 2020 | 331002651 |

For the Oasis data, several of the columns needed to be changed from character over to factors including columns such as Male/Female and Group. Columns that were not pertinent to the analysis were also removed. These columns included hand, subject.id, and MRI.Id. The column hand was removed as all of the patients in the dataset were right-handed.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject.ID | MRI.ID | Group | Visit | M.F | Hand | Age | EDUC | SES | MMSE | CDR | eTIV | nWBV | ASF |
| OAS2\_0001 | OAS2\_0001\_MR1 | Nondemented | 1 | M | R | 87 | 14 | 2 | 27 | 0 | 1987 | 0.696 | 0.883 |
| OAS2\_0001 | OAS2\_0001\_MR2 | Nondemented | 2 | M | R | 88 | 14 | 2 | 30 | 0 | 2004 | 0.681 | 0.876 |
| OAS2\_0002 | OAS2\_0002\_MR1 | Demented | 1 | M | R | 75 | 12 | 2 | 23 | 0.5 | 1678 | 0.736 | 1.046 |
| OAS2\_0002 | OAS2\_0002\_MR2 | Demented | 2 | M | R | 76 | 12 | 2 | 28 | 0.5 | 1738 | 0.713 | 1.010 |
| OAS2\_0002 | OAS2\_0002\_MR3 | Demented | 3 | M | R | 80 | 12 | 2 | 22 | 0.5 | 1698 | 0.701 | 1.034 |
| OAS2\_0004 | OAS2\_0004\_MR1 | Nondemented | 1 | F | R | 88 | 18 | 3 | 28 | 0 | 1215 | 0.710 | 1.444 |

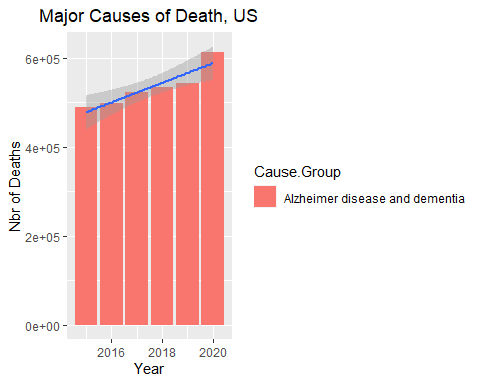
Group Non-Demented, Demented  
EDUC Years of education SES Socioeconomic Status MMSE Mini Mental State Examination CDR Clinical Dementia Rating eTIV Estimated Total Intracranial Volume nWBV Normalize Whole Brain Volume ASF Atlas Scaling Factor

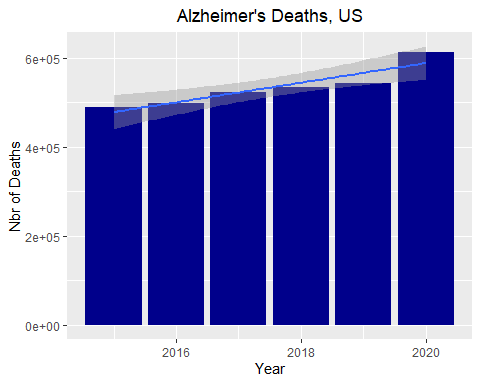
# Technical Libraries

These are the libraries used in the analysis.

* library(bibtex) Package used to create bibliographies
* library(ggplot2) Package used to create graphics
* library(dplyr) Package used for data manipulation
* library(caTools) Package containing utility functions

# Graphs

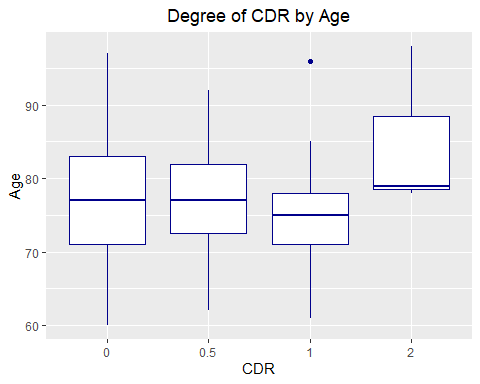




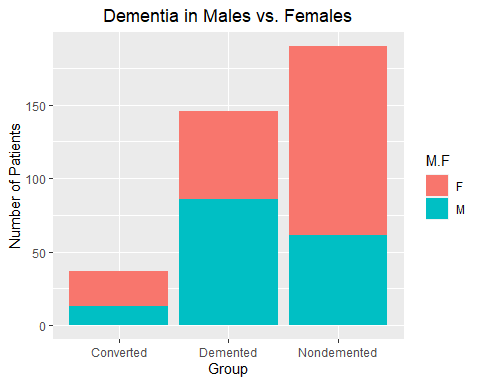
From 2015 through 2019, there has been an increase in deaths from Alzheimer’s disease of just over 25%. Over the same period, the population of the United States grew at a rate of just over 3%. Does this by itself suggest that instances of Alzheimer’s are increasing or could it indicate that doctors can diagnose the disease with more accuracy?

I believe the Oasis MRI dataset is the most interesting of the group, so we’ll spend some time analyzing it.

We can see the average age for each degree of the CDR Scoring table is relatively close, but there is a definite difference in the median age relative to the interquartile range.



From the Oasis MRI dataset, we can see that more men than women were afflicted with the disease.



# Modeling the Oasis Data

Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

Call:  
glm(formula = diagnosis ~ m\_f + age + educ + ses + mmse + CDR +   
 eTIV + nWBV + ASF, family = "binomial")  
  
Deviance Residuals:   
 Min 1Q Median 3Q Max   
-2.98272 0.05529 0.24088 0.43144 1.29609   
  
Coefficients:  
 Estimate Std. Error z value Pr(>|z|)   
(Intercept) -49.69218 36.68907 -1.354 0.175604   
m\_fM 0.02569 0.59303 0.043 0.965444   
age -0.08002 0.03531 -2.266 0.023454 \*   
educ 0.14144 0.10230 1.383 0.166792   
ses2 2.28273 0.61998 3.682 0.000231 \*\*\*  
ses3 2.37507 0.66684 3.562 0.000368 \*\*\*  
ses4 3.66705 1.06637 3.439 0.000584 \*\*\*  
ses5 19.43175 3768.61537 0.005 0.995886   
mmse -0.30330 0.13766 -2.203 0.027573 \*   
CDR0.5 -1.59969 0.51661 -3.096 0.001958 \*\*   
CDR1 15.37247 1378.60065 0.011 0.991103   
CDR2 16.16303 5501.45980 0.003 0.997656   
eTIV 0.02445 0.01244 1.966 0.049300 \*   
nWBV -6.48180 7.96640 -0.814 0.415850   
ASF 26.81757 15.06294 1.780 0.075016 .   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for binomial family taken to be 1)  
  
 Null deviance: 241.19 on 372 degrees of freedom  
Residual deviance: 177.89 on 358 degrees of freedom  
AIC: 207.89  
  
Number of Fisher Scoring iterations: 18

From the summary of the model’s fit, we can see that there are two statistically significant variables: Socioeconomic status (ses) and Mini-Mental State Examination (mmse). Socioeconomic status (ses) was really surprising to me that it had such a high level of significance. What would drive this variable to be significant? This finding would suggest that additional study around this variable would be warranted. Could it be a lack of regular medical care, diet, or other factors such as smoking?

“Dementia Vs. Alzheimer’s Disease: What Is the Difference?” n.d. *Alzheimer’s Disease and Dementia*. Alzheimer’s Association. <https://www.alz.org/alzheimers-dementia/difference-between-dementia-and-alzheimer-s>.