

# Dementia Status and Possible Correlations With Health and Lifestyle Factors

Alyssa Diaz

## Abstract

Worldwide, more than 55 million people have dementia, with 10 million more people being diagnosed every year.<sup>1</sup> Numerous studies have explored how individuals might be able to minimize their chances of developing dementia by identifying risk factors. In this analysis, various lifestyle and health markers are examined for possible correlations with the development of dementia. A total of 6 markers were identified and categorized as either “nature” or “nurture;” the strongest correlations with dementia status were within the “nature” category, although some interesting patterns could be found in the “nurture” category as well. Since so many varied factors exist that could affect the chances of developing dementia, further analysis is needed to take all the complex relationships between these factors into consideration and draw more solid conclusions.

## Introduction

The Dementia Patient Health Dataset contains data on 1,000 individuals aged 60-90 years (mean age 74.9), and includes many health and lifestyle variables with a focus on dementia status, such as age, smoking status, physical activity levels, and family history of dementia, among others.<sup>2</sup>

The objective of this analysis is to investigate possible correlations between dementia status and factors that can be attributed to either “nature” or “nurture.” Although many health and lifestyle factors were included in this dataset, 3 were chosen for initial analysis in each group; positivity for the apolipoprotein E gene (APOE-ε4), family history, and age for the “nature” category, and education level, smoking status, and physical activity for the “nurture” category.

## Methods

Much of the data in this dataset, such as APOE-ε4 Status and Education Level, are categorical rather than continuous or numerical. To render the categorical data more useful for future clustering and regression analysis, each of the categorical columns was converted to ordinal scale by mapping the name of each category to an integer (Figure 1).

Figure 1

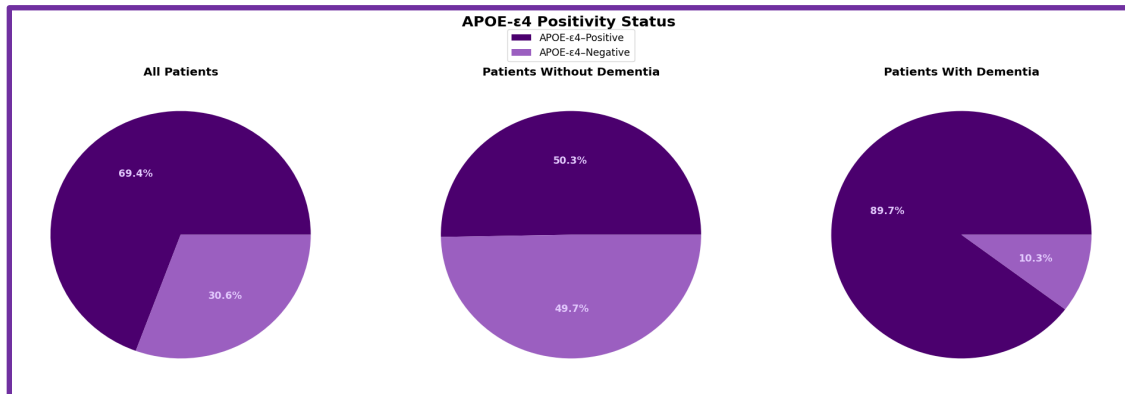
```
# Convert Education Level to ordinal scale
# (1=No School, 2=Primary School, 3=Secondary School, 4=Diploma/Degree)
edu_map = {'No School':1, 'Primary School':2, 'Secondary School':3, 'Diploma/Degree':4}
df['Education Scale'] = df['Education_Level'].replace(edu_map)
df['Education Scale'] = df['Education Scale'].astype(int)
```

As an indicator of dementia, the dataset’s Dementia Status provided a simple binary answer to whether or not an individual was diagnosed with dementia. In addition, the Cognitive Test Scores included in the dataset derive from the 10-point Cognitive Screener (10-CS), which is an aptitude test that is commonly used to diagnose cognitive impairment and dementia. This test classifies patients as having “normal cognition (≥8 points), possible (6–7 points), or probable cognitive impairment (≤5).”<sup>3</sup> This metric was used as an additional marker of individuals’ dementia status in conjunction with the binary Dementia Status factor.

## Results

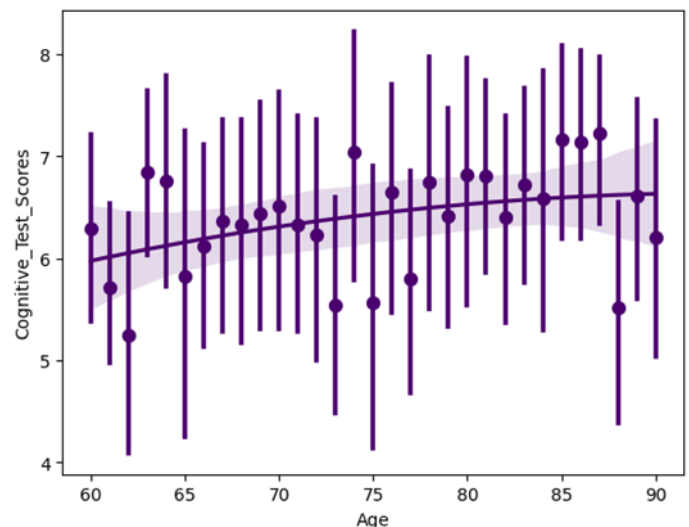
The most striking correlation was found between having dementia and testing positive for the APOE- $\epsilon$ 4 gene. Of those without dementia, 50.3% had the gene, and 89.7% of those diagnosed with dementia had the gene (Figure 2). This is not surprising, as it is well-established that having at least one APOE- $\epsilon$ 4 gene more than doubles the risk of developing Alzheimer's disease.<sup>4</sup> Similarly, patients with dementia were slightly more likely to have a family history of cognitive impairment: 43.7% of those without dementia had a family history, as compared to 47.4% of those with dementia.

Figure 2



Of the “nurture” factors analyzed, smoking status differed the most between those with and without dementia. Of the patients without dementia, 40% and 17.5% were former and current smokers, respectively, while 48% of individuals with dementia were former smokers and none currently smoked. Although interesting, this does not imply causation; for example, those with dementia may be more likely to be under close care and thus not permitted to continue their smoking habit.

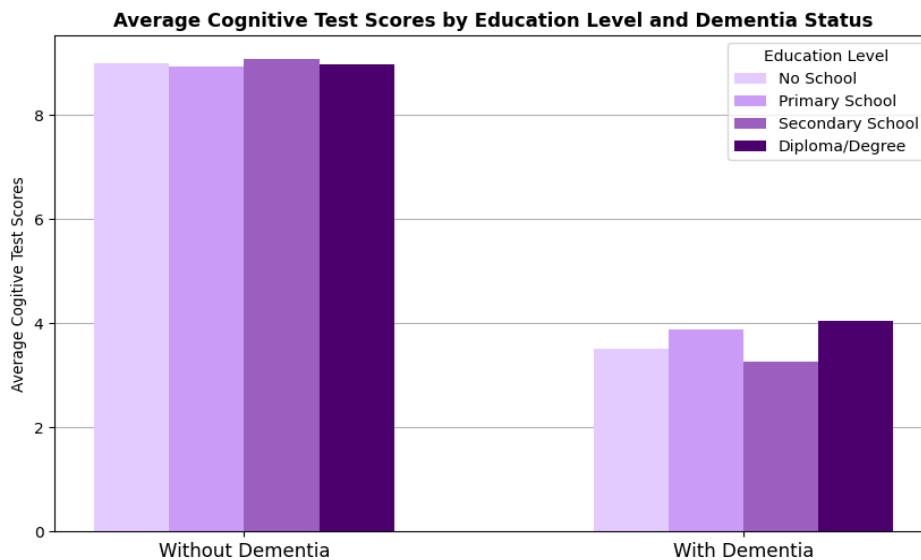
Figure 3



Age was also explored, since higher age is a well-known risk factor for dementia.<sup>1</sup> Testing this possible correlation by performing regression on Cognitive Test Scores versus aggregated distinct ages (Figure 3) resulted in a Mean Squared Error of 4,774.7 and an R-squared value of .568. Although these results do not demonstrate a high degree of correlation, the plot of the regression line over the aggregated age data shows a trend toward slightly higher Cognitive Test Scores in older individuals, which runs counter to what one would expect to see in an aging population.

Education level in relation to cognitive test scores was also analyzed (Figure 4). As expected, the group without dementia scored significantly higher on the test than did those with dementia. When comparing performance within each group, those without dementia performed at similar levels on the test regardless of education level, but more variability in performance was apparent in the group with dementia, with those with the highest level of education scoring better on the test.

Figure 4



## Conclusion

Based solely on the data analyzed here, “nature” stands out as being more influential on dementia diagnosis, with testing positive for the APOE-ε4 gene being closely correlated with dementia status, and a lower level of correlation between family history and dementia status. The “nurture” categories show interesting patterns, notably in Smoking Status, but the correlation is less clear. Factors that may influence developing dementia are numerous and varied, making it difficult to point to any single factor as being the cause of dementia, and warrant further exploration.

To fully account for the complex relationships between all the health and lifestyle factors included in the dataset, a multivariate analysis of the data is needed. Performing regression and clustering on all variables included in each of the “nature” and “nurture” groups in conjunction with Dementia Status and Cognitive Test Scores may reveal that there is a higher correlation within one group versus the other. However, the relatively small size of this dataset may limit the conclusions that can be drawn; it is possible that performing a similar analysis on a more robust dataset with a larger sample size would yield more conclusive results.

## References

- 1) World Health Organization. Dementia fact sheet. March 15, 2023. Accessed May 26, 2024.  
<https://www.who.int/news-room/fact-sheets/detail/dementia>
- 2) Adeyemi T. Dementia patient health dataset. Kaggle. Updated February 2024. Accessed May 26, 2024.  
<https://www.kaggle.com/datasets/timothyadeyemi/dementia-patient-health-dataset>
- 3) Rico BMH, Aliberti MJR, Trajano da Silva NO, et al. Advancing cognitive assessment in telemedicine: validity and reliability of the telephone 10-point cognitive screener. *J Am Geriatr Soc*. 2023;71(3):977-980.  
doi:10.1111/jgs.18091
- 4) Mayo Clinic. Alzheimer's genes: are you at risk? April 29, 2023. Accessed May 26, 2024.  
<https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-genes/art-20046552>