Assessment of Demographic Disparities and Associated Impacts in Dementia Diagnosis

Raffay Ahmed, Sampson Akwafuo, *California State University, Fullerton, CA 92831*

***Abstract*— The continued rise in cognitive ailments during aging calls for global concerted efforts towards curbing this threat. These ailments, such as Dementia are growing public health challenges, The prevalence is influenced by a multitude of demographic factors. This study aims to analyze the relationship between demographic factors and dementia diagnosis by using a dataset of categorical variables compiled from NHIS survey data. This dataset was used with visualization techniques, such as frequency tables and stacked bar charts, and statistical methods, such as Chi-square test of independence and Cramer's V. The analysis revealed significant association of age groups ($X^2(5) = 4425.936, p < 0.001$, *Cramér's* $V = 0.18$) and educational level ($X^2(4) = 492.704, p < 0.001$, *Cramér's* $V = 0.06$), with strong and moderate strength respectively. Additionally, demographic factors such as geographical location, physical attributes, and socioeconomics play a complex role in the statistical significance of dementia diagnosis. These findings highlight the importance of creating public health initiatives to address aging populations, educational disparities, and economic burdens people face to help reduce the impact of dementia in the future.**

I. Introduction

A. Importance of Public Health and Aging

Public health plays a critical role in ensuring the well-being of populations by addressing widespread health issues, promoting preventive care, and mitigating health disparities. By the end of the 20th century, public health advancements such as vaccinations, infectious disease control, and fortified foods scientifically increased life expectancy. Vaccinations eradicated smallpox and controlled diseases like influenza, improved sanitation curbed tuberculosis, and fortified foods help prevent nutritional deficiencies like goiter. Increased knowledge and positive attitude towards health care even allowed for better family planning, improved hygiene, and fewer pregnancy-related deaths [1], [2]. However, the 21st century has introduced new public health challenges, including chronic disease, substance addiction, food insecurity, and an aging population, which is often overlooked despite its significant societal impact [3].

Aging, a natural biological process, has been influenced by advancements in healthcare and living conditions, enabling longer and healthier lives. By 2030, 1 in 6 individuals globally will be aged 60 or older, while the population of individuals aged 80 years or older is projected to triple from 2020 to 2050, reaching 426 million [4]. While aging offers opportunities for individuals and society, it also increases the risk of chronic diseases and cognitive decline. Conditions such as hypertension, high cholesterol, diabetes, and arthritis are common among older adults, with nearly 60% of Americans over 65 managing at least one chronic illness [5]. Cognitive decline, particularly dementia, poses a significant challenge, affecting individuals, families, and caregivers, and highlights the urgent need to address the need to address the health concerns of gaining populations.

B. What is Dementia?

Dementia is a broad term that describes a range of symptoms that affect cognitive ability such as memory, reasoning, communication and the ability to perform everyday tasks [6] While not a disease itself, dementia encompasses various conditions, including Alzheimer's disease, frontotemporal dementia (FTD), Lewy body dementia (LBD), vascular dementia, and mixed dementia. Alzheimer's, the most common form, is caused by protein buildup and nerve cell damage in the brain. Other forms have distinct causes, such as FTD impacting the frontal lobe and vascular dementia resulting from blood vessel damage in the brain. Despite their differences, all forms of dementia share cognitive symptoms like memory loss and confusion, along with phycological symptoms such as anxiety and depression [7].

The effects of dementia extend beyond individuals, significantly impacting caregivers and healthcare systems. In 2024, an estimated 6.9 million Americans aged 65 and older will live with Alzheimer's dementia—approximately 1 in 9 people in this age group [8]. With the aging population, dementia diagnoses and associated costs are projected to rise, with annual healthcare and long-term care costs estimated at $360 billion in 2024, reaching $1 trillion by 2050. This growing burden underscores dementia as a critical public health challenge requiring immediate attention.

C. Objective

The paper strives to investigate the association between dementia and various demographic factors, which include education, age, race/ethnicity, geographic location, and education, using publicly available data. By leveraging data analysis techniques such as the Chi-square test of independence and Cramer's V, the study identifies notable relationships between demographic factors and the prevalence of dementia. This research focused on non-genetic determinants, which are often overlooked in favor of biological factors, to provide a more comprehensive understanding of dementia. By addressing disparities in how dementia impacts different demographic groups, this project can inform public health strategies, reduce healthcare inequities, and assist at-risk populations. These insights contribute to preventative strategies, improved early detection, and the development of effective public health policies.

Additionally, we created simple graphical user interface (GUI) to provide the Chi-square test of independence and Cramer's V as tools for more researchers across various domains.

II. Literature Review

The literature review aims to summarize reports and studies about dementia. This section will encompass reports and studies on the prevalence of dementia, its economic impacts, and the role of demographic disparities in diagnosis and care. This section aims to set the stage for our study by reviewing these topics, which focus on analyzing the correlation between dementia diagnosis and demographic factors to bring data-driven insights to healthcare providers and legislators.

A. Dementia Prevalence

With an aging population, dementia's prevalence and economic impact are commonly studied and reported by researchers and government health organizations. Establishing current and predicting future populations with dementia and the burden it creates on a healthcare provider's and an individual's finances allows individuals and policymakers to make informed decisions. Some studies have used the Health and Retirement Study (HRS) and statistical analysis to estimate the United States population with dementia [9], [10]. The HRS is a survey done on aging adults that identifies dementia through cognitive assessment and reports from proxies. Plassman et al. estimated a dementia prevalence of 13.9% for individuals aged 71 and older had dementia, equating to 3.4 million U.S. residents in 2002. Langa et al. estimated a dementia prevalence of 8.8% for individuals aged 65 and older in 2012. Both studies used the samples from the national surveys and derived weights to calculate how much of the U.S. population an individual signified. Using logistic regression techniques and cognitive assessments, they both determined that age played a role in the prevalence of dementia disease. However, Langa et al. also noted that individuals with higher education had a lower incidence of dementia.

Aside from national prevalence statistics, researchers and government agencies look at global prevalence to identify patterns and understand the impact of dementia on a broader scale. Understanding a disease on a global scale can allow for internationally developed strategies, healthcare, and resource allocations. Nichols et al. [11] looked at statistics internationally for factors such as BMI, smoking, and high plasma glucose to estimate dementia prevalence from 2019 through 2050. The Global Burden of Disease (GBD) framework, and other related studies estimated the prevalence of dementia and their causative agents [12]. using how various factors impact it were estimated, those statistics were statistically modeled using Bayesian meta-regression, and, enabling estimations through 2050. They estimated 57.4 million global dementia incidences in 2019, with a projection of 152.8 million in 2050. North and Sub-Saharan Africa have the most prominent potential increases, and higher education levels are recommended to reduce projected levels. The World Health Organization [13] also synthesized a report by cumulating data from global health agencies, focusing on aging populations with influences by education and cardiovascular health. They estimated 47.47 million global incidences of dementia in 2015, with a projected increase to 75.63 million in 2030, stating that low-medium-income countries (LIMCs) would experience the most substantial gains.

These studies are traditionally done on aging populations; however, Hendriks et al. [14] established a systemic study to determine the global prevalence of young-onset dementia (YOD). By reviewing published data from 1990 through 2020, they estimated the prevalence of YOD for 5-year age bands. The global prevalence was found to be 119 incidences per 100,000 people aged between 30 and 64, equating to 3.9 million cases worldwide in 2019. Hendricks et al. noted in their study that upper-medium-income countries (UMICs) had the highest rates of YOD, with similar incidence between male and female individuals. This study supplements traditional studies done on aging populations to provide a holistic understanding of the prevalence of dementia in all age groups.

B. Economic Impact

Researching the prevalence of dementia is only one part of the picture. Dementia also has a substantial economic impact - placing a financial burden on the individual, their families, and the healthcare system. Due to it being the most prevalent form of dementia, Alzheimer's disease (AD) is often the primary focus of economic research to provide insights into financial costs associated with dementia care [15], [16]. Frech et al. utilized commercial and Medicare data from 2014 to 2019 to calculate U.S. healthcare costs, including inpatient, outpatient, emergency, and pharmacy. For individuals with mild cognitive impairment (MCI), mean costs were $24,541 annually, but progressed Alzheimer's disease and related dementia (ADRD) meant costs were $34,599. Tay et al. systematically reviewed cost-of-illness studies related to AD from significant databases to establish costs for healthcare. In their research, the economic burden of AD varied based on the progression of the disease, where mild AD would cost $468 and severe AD would cost $171,283 annually. Both studies showcase how the economic burden of dementia increases as the disease progresses from mild impairment to Alzheimer's disease, highlighting the need to reduce costs.

While the costs for Alzheimer's disease are substantial, they represent just a portion of the overall economic impact caused by dementia. On a global scale, the financial burden for all forms of dementia is substantial and poses a significant economic challenge worldwide. Researchers have created economic models to provide current estimates for the economic burden dementia has and future projections it is going to develop [17], [18]. Wimo et al. created an economic model using global governmental agencies' data, which estimated global dementia costs in three categories: 1) Medical Costs, 2) Social Sector Costs, and 3) Informal Care Costs. From the model, they estimated dementia care cost $1.3 trillion globally in 2019; however, 61% of dementia incidence occurred in low-middle-income countries, while 74% of global costs occurred in high-income countries. Chen et al. established a health-augmented macroeconomic model to calculate the projected costs of Alzheimer's disease and other dementia (ADODs). This model considered the reduction in healthcare supply due to increased dementia incidence, treatment costs, and loss of productivity from the individual - projecting ADODs would cost $14.5 trillion globally from 2020 to 2050. Both studies highlight that global costs would place a more considerable economic burden on low-income countries, and preventative action should be taken.

C. Demographic Disparities in Diagnosis

While the economic burden of dementia is a significant topic, researchers have also explored how demographic factors affect diagnosis rates. Some studies have revealed how factors such as race, socioeconomic status, and education influence dementia prevalence and health outcomes [19], [20]. Kornblith et al. examined dementia incidence across five racial and ethnic groups using U.S. Veterans Health Administration (VHA) dataset of over 1.8 million individuals. Adjusting for age, dementia incidence was the highest among Hispanic and Black participants, followed by AIAN (American Indian and Alaka native), Asian, and finally White. This study highlights how race and ethnicity plays a complex role in dementia prevalence. Balls-Berry and Babulal provided a narrative report of previous studies on Alzheimer's disease and related dementias (ADRDs). They highlight how racial and ethnic minority groups have a higher prevalence of dementia and have lower access to quality healthcare while experiencing discrimination and misunderstanding in the healthcare system.

While demographic factors can contribute to disparities in diagnosis incidence, studies have found they can also lead to delayed or missed diagnosis [21], [22]. Tsoy et al. conducted a cross-sectional study using Medicare data for a population aged 65 and older. They identified individuals who were diagnosed with mild cognitive impairment (MCI) or dementia and assessed how fast they were able to receive specific healthcare services within six months. This study concluded that ethnic minorities and disadvantaged neighborhoods were less likely to receive a timely diagnosis and evaluation. Lin et al. used the HRS national survey data for individuals aged 70 and older with dementia and employed logistic regression to estimate the chances of missed or delayed diagnosis across demographic factors. The study found that Black and Hispanic individuals had diagnosis delays of two years, which typically caused diagnosis at more advanced stages of dementia. These studies illustrate how demographics play a factor in dementia diagnosis and highlight the necessity of further research and intervention.

III. METHODOLOGY

A. Data Set Surveying and Collections

The success of the project is dependent on the dataset used to find correlations between dementia diagnosis and the demographic background of an individual. Due to financial limitations and time constraints for this project, we must seek out a dataset that is free to use without restrictions and readily available for download. The dataset must have observations that represent unique individuals, their diagnosis status of dementia, and demographic information.

The National Health Interview Survey (NHIS) dataset provides data from a comprehensive survey conducted by the U.S. Census Bureau and managed by the National Center for Health Statistics (NCHS) [23]. The NHIS survey is a vital source of information for tracking health and demographic trends in the United States. The survey data is acquired by sampling adults and children through in-person interviews with established questions. The questions cover health status, healthcare access, regional location, and demographic factors. Some notable variables within the survey data include age, sex, race, education level, employment, and income. Additionally, this survey provides health conditions, behaviors, and lifestyles, making it a comprehensive source of health data.

NHIS dataset provides a breadth of demographic and health-related data, making it ideal for the project's scope. It offers standardized and clean data from randomly surveyed individuals across the United States from different regional and demographic backgrounds. It provides demographic data, health, behavior, and numerous other variables that can be identified from survey codebooks and summaries. Additionally, it offers individuals if they've been explicitly diagnosed with Alzheimer's by a doctor. However, due to the amount of data the NHIS dataset provides, we must preprocess the data to a workable form for statistical analysis.

B. Dataset Creation and Preprocessing

The NHIS dataset website provides a CSV of all the data collected from multiple years of surveys from 2019 to 2023. Additionally, it provides a survey summary and codebook to help identify the variables in which they collected data and summarize the incidences for each variable. By exploring the 2019 survey summary, we identified critical demographic variables for our study: urban-rural classification, geographical region, sex, race/ethnicity, income-to-poverty ratio, BMI, weight, height, age, education level, and dementia diagnosis. From the 2019 variable list, we cross-matched the variables in the 2020 to 2023 surveys and adjusted the variable names if needed. Combining the survey data, we have a dataset of 12 columns: three are continuous, and eight are categorical values. There are also 150,000 unique samples, which makes this dataset suitable for the scope of our project.

The responses in the survey data are represented as integer values and are converted to strings using a data dictionary provided by the NHIS. Additionally, we converted the continuous age, weight, and height data into binned categorical data and grouped education-level data into broader categories. Age was binned into age ranges representing an age group, education level into more straightforward groups, and height and weight into evenly distributed ranges. Converting continuous data into categorical helps improve interpretability, facilitate the use of specific statistical methods, and simplifies the visualization of relationships. Additionally, binning improves interpretability while addressing data sparsity, making patterns and trends withing the data more apparent.

The NHIS survey responses allow sample adults to respond with 'Refused,' 'Not Ascertained,' 'Don't Know,' 'Unknown,' 'Not Available,' 'I Don't Know the Answer,' and 'I Don't Know.' Additionally, for the continuous data responses for age, weight, and height, specific numerical values represent unknown, refusal to answer, or out of range, like 99 and 999. Consequently, we must clean any responses that are not useful for statistical analysis. After cleaning the data instances with non-usable entries, we still have around 136,000 sampled adults viable.

C. Data Visualization

Graphical representation visually explores the relationship between dementia diagnosis and a demographic factor. Creating charts provides an insightful way to examine the distribution of dementia diagnosis across demographic categories, such as education, age, and race/ethnicity. The methods to graphically represent the categorical data in the dataset for this study are frequency tables, with quantity and percentage, and stacked bar charts, both simple and 100%. Since we are comparing the categories only to dementia diagnosis, these methods will provide clear visualizations to identify trends.

Creating a frequency table with percentages is an essential step in data exploration, as it provides a simple summary of the distributions of the variables. Creating a visualization for the counts of each category, along with their relative proportions, can identify patterns, trends, and data imbalances. Using percentages to quantify the proportions enables comparisons across varying groups of each demographic factor, allowing for easier interpretation. This step is essential to validate and explore large datasets as it enables us to ensure quality and generate insights for subsequent analysis.

Stacked bar charts visualize the distribution of categorical variables across various groups. Each bar represents a group of the categorical variable and is divided into segments that represent another categorical variable. Stacked bar charts allow us to visualize each demographic category's composition and a subcategory's contributions while displaying the overall total. This allows us to identify data disparities and trends in the proportion. For this project, we created a stacked bar chart for each categorical variable in the dataset, created a bar for each answer of the category, and then subdivided the bar for the subcategory we are comparing with dementia diagnosis.

The 100% stacked bar chart works similarly; however, that quantifies the number of responses for a regular stacked bar chart is a percentage instead. Seeing the total percentage of each group in a category makes it easier to see relative differences between the amounts of each group. Using this method provides a more meaningful comparison than absolute counts to see trends and correlations. Additionally, if this visualization method is used on ordinal categorical variables, we can potentially see trends based on original ranks, such as educational levels or income-to-poverty ratios.

D. Statistical Analysis

Our study intends to use the prepared dataset for statistical analysis to find correlations between dementia diagnosis and demographic factors. Statistical analysis will utilize the Chi-square test for independence and Cramér's V for bivariate categorical data [24], [25]. To prepare our data for statistical analysis, we will create contingency tables from the dataset that contain dementia diagnosis frequency with a demographic factor. Each contingency table has a column with the available categories for each variable along with the corresponding frequency of dementia diagnosis.

The contingency tables allow for the calculation of an expected frequency as denoted as 'Expected Count'. Expected frequency is a value we would anticipate if there was no association between the variables and is calculated using the totals of the table, as seen in Equation 1. Each expected frequency must be greater or equal to five.

$$E_{ij} = \frac{\left(R_i \cdot C_j\right)}{N} \tag{1}$$

Where E_{ij} is the expected frequency for a cell in row i and column j, R_i is the row total for row i, C_j is the column total for column j, and N is the total sample size (sum of all frequencies). All tables show values greater than this threshold, confirming Chi-square test of independence is a valid method for this dataset.

Chi-square test of independence calculates whether there is a significant association between two categorical variables by using the observed and expected frequencies in the contingency tables. It outputs a Chi-square value and an asymptotic significance to determine whether the observed frequencies were statistically significant, using Equation 2.

$$x^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{2}$$

Where x^2 is the Chi-square test value and O_{ij} is the observed frequency for a cell in row i and column j .

The Chi-Square value is the difference between the observed and expected frequencies. A larger value means a greater difference, suggesting a stronger association. The Chi-Square asymptotic significance value is derived using the Chi-square distribution table using the value and the degrees of freedom, which helps accept or reject the null hypothesis.

For this study, the null (H_o) and alternative (H_a) hypotheses for the Chi-square test of independence are:

- H_o: There is no association between the demographic factor and dementia diagnosis.
- H_a: There is an association between the demographic factor and dementia diagnosis.

Commonly if asymptotic significance is <0.05, we reject the null hypothesis and accept the alternative, indicating an association.

Cramer's V, as shown in Equation 3, was done subsequently after the Chi-square test of independence to provide a standardized value between 0 and 1, where 0 indicates no association and 1 indicates perfect association.

$$V = \sqrt{\frac{x^2/N}{min\left(C_j - 1, R_i - 1\right)}} \tag{3}$$

Where V is the Cramer's V score. The magnitude of the value defines the strength of the correlation found using the Chi-square test of independence - typically weak, moderate, and strong.

E. Statistical Analysis GUI

The GUI was developed using Python and several libraries. PySimpleGUI was used to create an intuitive interface, while Pandas handles data. Numpy and SciPy were utilized for statiscal calculations, including the Chi-square test and Cramer's V, and PyInstaller was used to package the application into a standalone executable for ease of use.

IV. Results and Discussion

The dataset used for this project contains a wide range of variables that can provide insights into potential relationships and trends related to dementia diagnosis. Including charts and variables for every variable would be excessive and may dilute the focus of the discussion. This section highlights the most notable findings from the visualizations and statistical analysis. This includes frequency and percentage tables and simple and 100% stacked bar charts. However, complete results of the Chi-square tests and Cramer's V calculations are included to provide a comprehensive understanding of the demographic factors and its associations.

A. Data Visualization Inference

Initially to visualize the dataset we used tables to quantify frequencies and proportions of the demographic categories in the dataset. They were first tabulated independently and then further broken down with respect to 'Yes' and 'No.' This method highlighted notable patterns in categories such as education level, age, and race/ethnicity.

TABLE I. Education level and dementia diagnosis Percentages

	Dementia Diagnosis		
Education Level	No	Yes	Total
Bachelor's Degree	99.50%	0.50%	100.00%
Graduate and Professional Degree	99.40%	0.60%	100.00%
High School Level	98.40%	1.60%	100.00%
No Formal Education or Some Schooling (No Diploma)	97.30%	2.70%	100.00%
Some College or Associate Degree	99.10%	0.90%	100.00%
Total	98.90%	1.10%	100.00%

From Table I., we see an inverse relationship between education level and dementia diagnosis. Individuals with "No Formal Education or Some Schooling (No Diploma)" have the highest prevalence (0.2%), followed by those with only a "High School Level" education (0.4%). Those with a bachelor or higher have the lowest prevalence, highlighting the potential protective effect of higher education.

TABLE II. Age and dementia diagnosis Percentages

	Dementia Diagnosis		
Age Group	No	Yes	Total
Older Elderly (85+)	91.00%	9.00%	100.00%
Elderly (75-84)	96.10%	3.90%	100.00%
Senior (65-74)	98.80%	1.20%	100.00%
Older Adult (50-64)	99.50%	0.50%	100.00%
Middle-Aged Adult (35-49)	99.80%	0.20%	100.00%
Young Adult (18-34)	99.90%	0.10%	100.00%
Total	98.90%	1.10%	100.00%

From Table II., we see a strong age-related risk factor. The highest prevalence is observed with "Older Elderly (85+)" (9.0%), followed by "Elderly (75-84)" (3.9%). Younger groups showed almost no dementia cases.

TABLE III. Race and Dementia Diagnosis Percentages

	Dementia Diagnosis		
Race/Ethnicity	**No**	**Yes**	**Total Count**
Hispanic	18132	139	18271
% within Race/Ethnicity	99.20%	0.80%	100.00%
American Indian and Alaska Native and Other, Not Hispanic	973	15	988
% within Race/Ethnicity	98.50%	1.50%	100.00%
American Indian and Alaska Native, Not Hispanic	845	10	855
% within Race/Ethnicity	98.80%	1.20%	100.00%
Asian Only, Not Hispanic	7490	44	7534
% within Race/Ethnicity	99.40%	0.60%	100.00%
Black/African American Only, Not Hispanic	14074	194	14268
% within Race/Ethnicity	98.60%	1.40%	100.00%
White Only, Not Hispanic	91634	1099	92733
% within Race/Ethnicity	98.80%	1.20%	100.00%
Other single and multiple races	1502	9	1511
% within Race/Ethnicity	99.40%	0.60%	100.00%
Total Count	134650	1510	136160
% within Race/Ethnicity	98.90%	1.10%	100.00%

From Table III., we see that Black/African Americans (1.4%) show a slightly higher prevalence of dementia, compared to other races/ethnicities. "American Indian and Alaska Native and Other, Not Hispanic" (1.5%), however the sample size is much smaller.

Supplemental to tabularizing the dataset, creating stacked bar charts help identify trends in the demographic data. These stacked bar charts reinforced the insights from the tabular data, particularly for age and education level categories as seen in Fig. 1 and Fig. 2. We see a trend of increasing incidence of dementia diagnosis with increased age and lower education levels.

Figure 1. Age and Dementia Diagnosis Bar Chart

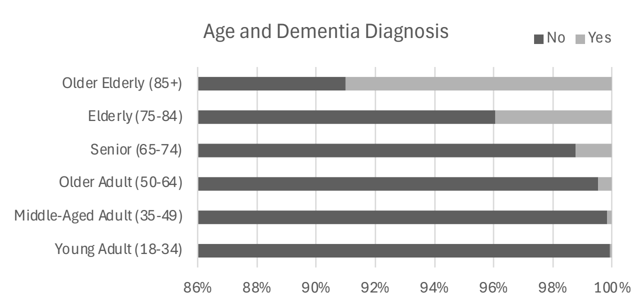
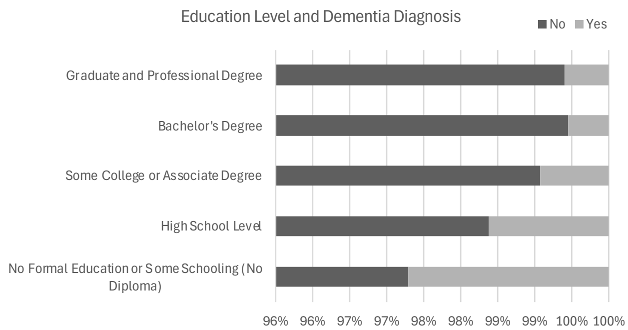


Figure 2. Education Level and Dementia Diagnosis Bar Chart



B. Statistical Analysis Inference

The Chi-square test of independence was used to assess whether there were significant associations between demographic factors and dementia diagnosis. A Chi-square value and an asymptotic value (p-value) were calculated, where those with less than 0.05 p-value have an association and the null-hypothesis is rejected. Additionally, we use the Chi-square results to calculate the Cramér's V value, which is the magnitude of association. Cramér's V provide a value between 0 and 1, where 0 is no association and 1 is perfect association.

TABLE IV. STATISTICAL TEST RESULTS

Variable Name	Chi-Square Value (X^2)	Degrees of Freedom (df)	p-value	Cramér's V Value
Urban/Rural Classification	36.174	3	<.001	0.016
U.S. Region	16.828	3	<.001	0.011
Sex	2.123	1	0.145	0.004
Race/Ethnicity	57.38	6	<.001	0.021
BMI	93.322	3	<.001	0.026
Income-to-Poverty Ratio	259.395	13	<.001	0.044
Weight	170.674	7	<.001	0.035
Height	82.175	5	<.001	0.025
Age	4425.936	5	<.001	0.18
Education Level	492.704	4	<.001	0.06

A total of 136,160 valid cases were analyzed. The results, including the Chi-square value, degrees of freedom, p-values, and Cramér's V values, are summarized in **Error! Reference source not found.**. From the p-value column we see that all variables were found to have a statistically significant association with dementia status ($p < 0.001$), except for sex ($p = 0.145$). However, of the variables showing association, the strength was generally weak indicated by Cramér's V, with most values below 0.05. The strongest associations were age groups ($X^2(5) = 4425.936, p < 0.001$, *Cramér's V* $= 0.18$) and educational level ($X^2(4) = 492.704, p < 0.001$, *Cramér's V* $= 0.06$). Urban-rural classification ($X^2(3) = 36.174, p < 0.001$, *Cramér's V* $= 0.016$) and region ($X^2(3) = 16.828, p < 0.001$, *Cramér's V* $= 0.011$) showed weak but statistically significant association. Race/ethnicity, body mass index, poverty-to-income ratio, weight groups, and height groups all show a similar pattern.

Age having the strongest association and educational level having a moderate association between dementia diagnosis aligns with existing research. Advancing age has been identified as the most significant risk factor for the condition, highlighting the important of age-based screening and care. The educational level association also supports the hypothesis which suggests that higher education levels may protect against cognitive decline, highlighting policies to promote educational access as a public health strategy. While factors such as urban-rural classification, U.S. region, race/ethnicity, BMI, weight, height, and poverty-to-income ratio show weak association, these findings may reflect a complex interplay between other factors that may tie into dementia diagnosis. Additionally, an absence in an association for sex needs to be highlight since other studies have shown an association. This maybe be due to our data, out methodology, or other factors.

These results imply public health efforts should focus on at-risk older adults and provide lifelong learning opportunities to reduce the impact of dementia. Addressing income inequality and health disparities may also help focus on the demographics that showed weak associations.

C. Data Analysis GUI

A GUI that allows researchers to run the Chi-square test of independence and Cramer's V on categorical datasets was created. This application allows researchers to analyze multiple files sequentially by selecting a dataset and the target comparison variable. Additionally, this application allows the researcher to save the results for further analysis and discussion.

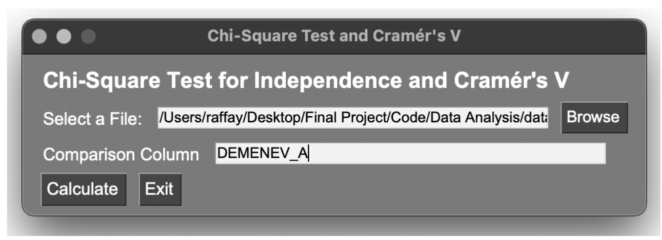


Figure 3. Main Window of GUI Application

V. CONCLUSION

This study has explored the relationship between various demographic variables and dementia diagnosis in the compiled NHIS dataset. These demographic variables include geological location, population identity, physical characteristics, and socioeconomic backgrounds. The methods used in this study included data visualization, which encompassed frequency tables and stacked bar charts, and statistical methods, which encompassed Chi-square test of independence and Cramer's V. Through these methods we found that age ($X^2(5) = 4425.936, p < 0.001$, *Cramér's V* $= 0.18$) and educational level ($X^2(4) = 492.704, p < 0.001$, *Cramér's V* $= 0.06$) were statistically significant to dementia diagnosis and had strong and moderate strength in association respectively. We also found that urban-rural classification, U.S. region, race/ethnicity, BMI, weight, height, and poverty-to-income ratio all were statistically significant, but had weak association strength. These findings highlight how demographic factors play a complex part in dementia diagnosis and provides support to already established studies and opens avenues for further research.

VI. References

[1] J. P. Koplan and D. W. Fleming, "Current and Future Public Health Challenges," *JAMA*, vol. 284, no. 13, pp. 1696–1698, Oct. 2000, doi: 10.1001/JAMA.284.13.1696.

[2] "Ten Great Public Health Achievements -- United States, 1900-1999." Accessed: Nov. 27, 2024. [Online]. Available: https://www.cdc.gov/mmwr/preview/mmwrhtml/00056796.htm

[3] S. Akwafuo, J. Urbanovsky, C. Ihinegbu, and A. R. Mikler, "Dynamic Heuristic Algorithm for Management of Public Health Emergencies in Unreliable Settings.," in *Proceedings of the 8th IEEE Internal Conf. On Healthcare Informatics*,: ACM, 2020.

[4] "Ageing and health." Accessed: Nov. 29, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[5] "The Top 10 Most Common Chronic Diseases for Older Adults." Accessed: Nov. 29, 2024. [Online]. Available: https://www.ncoa.org/article/the-top-10-most-common-chronic-conditions-in-older-adults/

[6] "What Is Dementia?" Accessed: Oct. 13, 2024. [Online]. Available: https://www.alzheimers.gov/alzheimers-dementias/what-is-dementia

[7] "Dementia - Symptoms and causes - Mayo Clinic." Accessed: Oct. 13, 2024. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/dementia/symptoms-causes/syc-20352013

[8] Alzheimer and Association, "Alzheimer's Association 2024 Alzheimer's Disease Facts and Figures".

[9] B. L. Plassman *et al.*, "Prevalence of dementia in the United States: the aging, demographics, and memory study," *Neuroepidemiology*, vol. 29, no. 1–2, pp. 125–132, Nov. 2007, doi: 10.1159/000109998.

[10] K. M. Langa *et al.*, "A Comparison of the Prevalence of Dementia in the United States in 2000 and 2012," *JAMA Intern Med*, vol. 177, no. 1, pp. 51–58, Jan. 2017, doi: 10.1001/JAMAINTERNMED.2016.6807.

[11] E. Nichols *et al.*, "Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019," *Lancet Public Health*, vol. 7, no. 2, pp. e105–e125, Feb. 2022, doi: 10.1016/S2468-2667(21)00249-8/ATTACHMENT/60E03FD1-38B2-4B40-A91D-9AFDDA22B45E/MMC1.PDF.

[12] S. Akwafuo, A. R. Mikler, and F. A. Irany, "Optimization Models for Emergency Response and Post-Disaster Delivery Logistics: A Review of Current Approaches," *International Journal of Engineering Technologies and Management Research*, no. 08, 2020, doi: 10.29121/ijetmr.v7.i8.2020.738.

[13] M. Prince, M. Guerchet, and M. Prina, "The Epidemiology and Impact of Dementia-Current State and Future Trends. WHO Thematic Briefing", Accessed: Oct. 25, 2024. [Online]. Available: https://hal.science/hal-03517019v1

[14] S. Hendriks *et al.*, "Global Prevalence of Young-Onset Dementia: A Systematic Review and Meta-analysis," *JAMA Neurol*, vol. 78, no. 9, pp. 1080–1090, Sep. 2021, doi: 10.1001/JAMANEUROL.2021.2161.

[15] F. H. Frech *et al.*, "Economic Impact of Progression from Mild Cognitive Impairment to Alzheimer Disease in the United States," *Journal of Prevention of Alzheimer's Disease*, vol. 11, no. 4, pp. 983–991, Aug. 2024, doi: 10.14283/JPAD.2024.68/TABLES/3.

[16] L. X. Tay, S. C. Ong, L. J. Tay, T. Ng, and T. Parumasivam, "Economic Burden of Alzheimer's Disease: A Systematic Review," *Value Health Reg Issues*, vol. 40, pp. 1–12, Mar. 2024, doi: 10.1016/J.VHRI.2023.09.008.

[17] S. Chen *et al.*, "The global macroeconomic burden of Alzheimer's disease and other dementias: estimates and projections for 152 countries or territories," *Lancet Glob Health*, vol. 12, no. 9, pp. e1534–e1543, Sep. 2024, doi: 10.1016/S2214-109X(24)00264-X.

[18] A. Wimo *et al.*, "The worldwide costs of dementia in 2019," *Alzheimers Dement*, vol. 19, no. 7, pp. 2865–2873, Jul. 2023, doi: 10.1002/ALZ.12901.

[19] E. Kornblith, A. Bahorik, W. J. Boscardin, F. Xia, D. E. Barnes, and K. Yaffe, "Association of Race and Ethnicity With Incidence of Dementia Among Older Adults," *JAMA*, vol. 327, no. 15, pp. 1488–1495, Apr. 2022, doi: 10.1001/JAMA.2022.3550.

[20] J. E. Balls-Berry and G. M. Babulal, "Health Disparities in Dementia," *Continuum (Minneap Minn)*, vol. 28, no. 3, p. 872, Jun. 2022, doi: 10.1212/CON.0000000000001088.

[21] E. Tsoy *et al.*, "Assessment of Racial/Ethnic Disparities in Timeliness and Comprehensiveness of Dementia Diagnosis in California," *JAMA Neurol*, vol. 78, no. 6, pp. 657–665, Jun. 2021, doi: 10.1001/JAMANEUROL.2021.0399.

[22] P. J. Lin *et al.*, "Dementia Diagnosis Disparities by Race and Ethnicity," *Med Care*, vol. 59, no. 8, p. 679, Aug. 2021, doi: 10.1097/MLR.0000000000001577.

[23] "NHIS Questionnaires, Datasets, and Documentation | National Health Interview Survey | CDC." Accessed: Dec. 01, 2024. [Online]. Available: https://www.cdc.gov/nchs/nhis/documentation/?CDC_AAref_Val=https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm

[24] M. L. McHugh, "The Chi-square test of independence," *Biochem Med (Zagreb)*, vol. 23, no. 2, p. 143, 2013, doi: 10.11613/BM.2013.018.

[25] M. Allen, "Cramér's V," *The SAGE Encyclopedia of Communication Research Methods*, Jul. 2017, doi: 10.4135/9781483381411.N107.