What is the association between Dementia and Cancer in the U.S.?

Ann Xiao
yx3493@nyu.edu

Abstract

In this study, I examined the association between cancer and dementia in the United States using 2022 data from the CDC's annual face-to-face interview survey, supplemented with data from 2019 to 2022. Contrary to previous studies suggesting an inverse relationship, my stratified analysis by gender, race, age, and cancer type consistently indicated a higher incidence of dementia in individuals with cancer. Employing causal forest analysis revealed a negligible negative association, yet the distribution of p-values suggested data imbalance. The causal forest analysis also identified age as the most influential variable in the likelihood of developing dementia, with depression, diabetes, and education level emerging as other notable contributing factors. These findings challenge existing paradigms and highlight the need for deeper investigation into the complex interplay between cancer and dementia.

1 Introduction

Dementia is caused by a variety of diseases that cause damage to brain cells. When brain cells cannot communicate normally, thinking, behavior, and feelings can be affected. Alzheimer's disease is the most common form of dementia and may contribute to 60–80% of cases[9]. Dementia affects millions of people and is more common as people grow older (about one-third of all people age 85 or older may have some form of dementia) but it is not a normal part of aging. Dementia is currently the seventh leading cause of death and one of the major causes of disability and dependency among older people globally. Currently, more than 55 million people have dementia worldwide, over 60% of whom live in low-and middle-income countries. Every year, there are nearly 10 million new cases. In 2019, dementia cost economies globally 1.3 trillion US dollars, approximately 50% of these costs are attributable to care provided by informal carers (e.g. family members and close friends)[1]. Some risk factors for dementia, such as age and genetics, cannot be changed. However, researchers continue to explore the impact of other risk factors on brain health and the prevention of dementia[8].

The interrelationship between cancer and dementia has been a subject of intrigue and investigation within the medical research community. Foundational studies from disparate geographical regions, including Britain[10], Taiwan[4], Korea[3], and Sweden[7], have underscored an inverse association between these two conditions, positing that individuals with cancer exhibit a lower risk of dementia. The Swedish research extended these observations by examining the link between dementia and specific cancer types. The Korean study also examined Alzheimer's Disease, a specific type of dementia. Such findings are seemingly counter-intuitive, given the shared risk factors and biological mechanisms that underpin both pathologies, including age-related changes, signaling pathways, and the role of specific enzymes such as Pin1.

2 Methods

This study seeks to replicate and extend these findings within the United States utilizing the National Health Interview Survey (NHIS) 2022 data from the Centers for Disease Control and Prevention

(CDC). In the second half of the study, data from 2019 to 2022 were combined to ensure a larger sample size.

2.1 Study Population

This dataset is derived from an annual, anonymized face-to-face interview survey meticulously designed to reflect the entire U.S. demographic landscape. The target population for the NHIS is the civilian noninstitutionalized population residing within the 50 states and the District of Columbia at the time of the interview. They use geographically clustered sampling techniques to select the sample of dwelling units for the NHIS. The sample is designed in such a way that each month's sample is nationally representative[5].

2.2 Covariates

In conducting this study, I sought to control for a variety of demographic and clinical factors that could potentially confound the results. Initially, I selected age, gender, and race as covariates. They were chosen because of their relevance in similar studies. Subsequently, I stratified the dataset by the types of cancer to further study the relationships. In the final stage of the study, I incorporated a comprehensive set of covariates as recommended by the World Health Organization (WHO), which includes gender, race, age, individual education level, highest education level attained within the household, and a range of health-related variables. These health variables included hypertension, prediabetes, diabetes, obesity, smoking status, and depression, as reported in the dataset. Each covariate was carefully considered to ensure a robust analysis and to identify any potential associations with the risk and prevalence of dementia among cancer patients.

2.3 Statistical Analysis

Each row in the data matrix represents an individual participant and each column corresponds to their responses to various questions. Responses were encoded as integers. For example, under the variable "sex" there is '1' indicating a 'yes' response, and '2' for 'no'; under the variable "age" there are '18' - '84' indicating the corresponding age, and '85' for '85 and over'. Other response options provided by the interviewers—such as '97' for 'refused', '98' for 'not ascertained', and '99' for 'don't know'—were considered non-contributory to the study due to their negligible frequency and indeterminate nature; hence, any records containing these responses under critical variables (cancer, dementia, age, gender, race) were excluded from the analysis.

For the first part of the study, I chose age, gender, and race as covariates. Given the constraints of sample size, it was not feasible to stratify the data concurrently by all three covariates. Therefore, two separate analyses were conducted: one stratifying by age and gender, and the other by race alone. Standardized Incidence Ratios (SIRs) were calculated to compare the observed number of dementia cases in individuals diagnosed with cancer to the expected number in a matched cancer-free cohort. These expected counts were derived from age-, sex-, and race-specific incidence rates within the control group. Additionally, to discern the relationship between dementia and specific types of cancer, I stratified the dataset by cancer types and calculated the corresponding SIRs. Cancer types represented by fewer than 5 dementia cases were omitted to maintain statistical robustness. After determining the SIRs, 95% Confidence Intervals (CIs) were calculated, assuming a Poisson distribution. This approach allowed for the estimation of the relative risk of dementia following a cancer diagnosis and facilitated comparison with those without a cancer history.

For the second part of the study, I consolidated the datasets from 2019 to 2022, thereby significantly increasing the sample size. I replicated the earlier analytical process on this larger dataset, stratifying by cancer types and recalculating the Standardized Incidence Ratios (SIRs). With the augmented dataset, I then applied a causal forest analysis, initially incorporating the covariates age, race, and gender to account for basic demographic factors. Subsequently, to capture a more detailed and comprehensive picture of potential influences on the association between cancer and dementia, the causal forest analysis was extended to include additional covariates: gender, race, age, education level, maximum education level in the household, hypertension, prediabetes, diabetes, obesity, smoking status, and depression. This approach aimed to unravel all the possible factors that could contribute to the development of dementia.

3 Results

The demographic and clinical factors among patients with cancer and the matched controls are presented in Table 1. A total of 15,052 individuals diagnosed with cancer were retrieved from the databases. Female patients (58.02%) outnumbered male patients (41.98%).

Table 1: Basic demographic and clinical characteristics among patients with cancer and cancer-free controls for the 2022 dataset and the 2019-2022 combined dataset.

	2022	
Characteristics	**Cancer Group (n=3429)**	**Cancer-Free Group (n=24174)**
Gender		
Male	1448	11127
Female	1981	13047
Age		
< 45	208	9816
$45 - 64$	868	7901
$65 - 84$	2009	5741
≥ 85	344	716
Dementia		
Yes	92	257
No	3337	23917
	2019-2022	
Characteristics	**Cancer Group (n=15052)**	**Cancer-Free Group (n=105448)**
Gender		
Male	6320	48813
Female	8732	56635
Age		
< 45	997	42839
$45 - 64$	4083	35402
$65 - 84$	8483	23997
≥ 85	1489	3210
Dementia		
Yes	357	1030
No	14695	104418

The overall risk of dementia was significantly higher among patients with cancer (SIR = 2.46), unlike the results from previous papers (SIR = 0.79)[7]. Age appeared to inversely correlate with risk; as age increased, the risk of dementia decreased for both genders. Notably, while males exhibited a consistently positive association between cancer and dementia across all age groups, females showed a reverse trend with inverse associations emerging beyond the age of 65. Across racial demographics, White, Black/African American, Asian, and Native American populations all demonstrated SIRs exceeding 1, denoting an increased risk. However, the Hispanic group was an exception, presenting a lower risk with an SIR of 0.76, as detailed in Table 2. SIR calculations for individuals under the age of 45 were excluded from the analysis due to the limited size of the sample population in that cohort.

In the 2022 dataset, SIRs for all cancer types uniformly exceeded 1, indicating a positive correlation between each type of cancer and the incidence of dementia within the U.S. population. This finding stands in contrast to the conclusions drawn from the Swedish study, which reported varied associations[7]. In order to make sure these results are not due to a small sample size, the same analytical procedures were applied to the combined dataset from 2019 to 2022. The resulting patterns remained consistent, reinforcing the initial observations (refer to Table 3 for detailed results).

The causal forest analysis, when using race, gender, and age as covariates, yielded an estimate of -0.00056 with a standard error of 0.00078. This marginal negative estimate suggests an inverse association between cancer and dementia; however, it is very close to 0 and has a very small t-value (0.71), suggesting that it has minimal statistical significance. After expanding the model to include a broader range of covariates (gender, race, age, education level, maximum education level in the

Table 2: Risk of dementia among patients with cancer stratified by gender, age, and race.

Female		
Age	**SIRs**	**95% CI**
$45-64$	1.05	(0, 2.16)
$65-84$	0.90	(0.55, 1.29)
≥ 85	0.85	(0.70, 1.01)
Male		
Age	**SIRs**	**95% CI**
$45-64$	1.91	(0.0, 3.87)
$65-84$	1.31	(0.80, 1.87)
≥ 85	1.03	(0.83, 1.25)
Race	**SIRs**	**95% CI**
Hispanics	0.76	(0.29, 1.37)
White	2.42	(1.05, 4.00)
African American	2.94	(1.43, 4.75)
Asian	2.57	(1.03, 4.47)
Native Americans	3.31	(1.87, 4.85)

Table 3: Risk of dementia among patients with specific cancer types.

Cancer Type	**SIR (2022)**	**SIR (2019-2022)**
Skin melanoma	1.55	2.59
Skin non-melanoma	1.67	1.64
Breast	3.21	2.60
Colorectal	2.38	3.55
Colon		3.21
Prostate	3.99	2.97
Bladder		3.81
Skin other	2.60	2.93
Lung	4.86	4.75
Other	1.91	2.04

household, hypertension, prediabetes, diabetes, obesity, smoking status, and depression), there is a slightly more negative estimate of -0.00075 and a standard error of 0.00074. Although the estimate shifted further into negative territory, it remains very close to zero.

The variable importance metrics derived from the causal forest model reinforce the predominance of age as the most significant factor in the development of dementia. Other notable covariates, such as depression, education level, and diabetes, also emerged as influential factors (as shown in Table 4). Additionally, the distribution of p-values, as depicted in the histogram, indicates an imbalance in the data, which is a concern for the robustness of the statistical analysis (Figure 1).

In the initial regression model applied to the dataset, the precision score for predicting cancer presence was a mere 0.07. Subsequent efforts to rectify data imbalance using SMOTE and undersampling techniques improved precision to 0.30 and recall to 0.47, yielding an ROC AUC score of 0.66 (Table 5). Despite these enhancements, the model's predictive performance remains moderately adequate. It continues to be adversely affected by the residual imbalances within the dataset and carries a risk of overfitting, which compromises its reliability for practical applications.

4 Discussion

This study of the association between various types of cancer and the incidence of dementia in the U.S. has resulted in many unexpected discoveries. It is surprising that the data indicates a two-fold elevation in the likelihood of dementia among cancer patients overall, in comparison with the

Table 4: Variable Importance from Causal Forest Model.

Variable	VI
Gender	0.22
Race	0.016
Age	0.583
Education Level	0.058
Max Edu Level in Household	0.076
Hypertension	0.035
Pre-diabetes	0.035
Diabetes	0.060
Obesity	0.015
Smoke	0.020
Depression	0.081

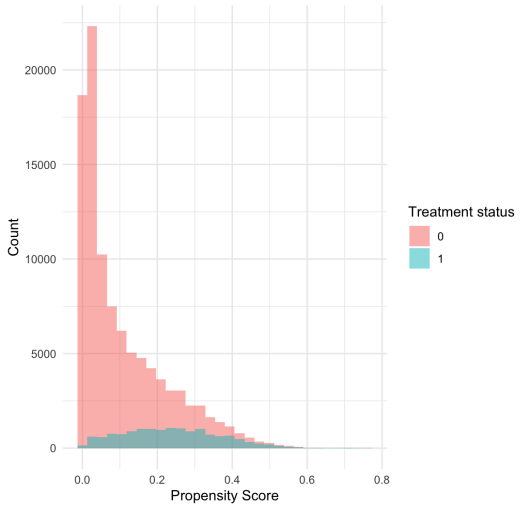

Figure 1: The Propensity Score Distribution for Causal Forest Model

Table 5: Classification Report of the Regression Model.

Before Re-sampling				
Group	**Precision**	**Recall**	**F1**	**support**
Cancer	0.07	0.00	0.00	3037
Cancer-free	0.87	1.00	0.93	21063
Weighted-Avg	0.77	0.87	0.81	24100
ROC AUC	0.50			
After Re-sampling				
Group	**Precision**	**Recall**	**F1**	**support**
Cancer	0.30	0.47	0.37	3037
Cancer-free	0.92	0.85	0.88	21063
Weighted-Avg	0.84	0.80	0.82	24100
ROC AUC	0.66			

significantly lower risks from previous papers. Interestingly, our observations resonate with another study within a German cohort, which similarly reports heightened dementia risks among cancer patients, suggesting the need for a more holistic understanding of this association on a global scale[6].

One of the most unexpected findings is the inverse association of dementia with cancer observed solely in the Hispanic demographic. This could be attributed to a myriad of factors, including genetic predispositions, lifestyle factors, or even healthcare access and utilization patterns that vary across ethnic groups. Cultural factors and social support systems, which are typically strong in Hispanic communities, might also play a role in mitigating the cognitive decline associated with dementia.

Conversely, the general inverse trend observed in women as opposed to the positive association in men raises questions about the interplay of hormonal, genetic, or environmental factors that differ by gender. Women's longer life expectancy, combined with the potential neuroprotective effects of estrogen, may contribute to this disparity, although these factors do not fully explain the observed differences and warrant further investigation.

Particularly striking is the finding that individuals with lung cancer have a quadrupled risk of developing dementia compared to those without cancer. This association might be more logically correlated to brain cancer intuitively. This observation may point to shared risk factors between lung cancer and dementia, such as smoking, which is known to affect both pulmonary and cognitive health, or perhaps the systemic effects of lung cancer and its treatments on cognitive function.

Finally, the impact of education and demography on the dataset should not be overlooked. Education level often correlates with health literacy, which can influence both the likelihood of seeking and receiving medical care and the capacity to manage complex health conditions, potentially affecting the observed association between cancer and dementia. Demographic variations may reflect differences in environmental exposures, access to healthcare, and other socioeconomic factors that can alter disease risk and outcomes.

Future studies may benefit from more balanced datasets, the inclusion of additional covariates, and longitudinal designs that can better capture causal relationships and the long-term trajectories of these diseases.

References

[1] "Dementia." World Health Organization, World Health Organization, www.who.int/news-room/fact-sheets/detail/dementia. Accessed 18 Dec. 2023.

[2] Ganguli, Mary. "Cancer and Dementia: It's Complicated." Alzheimer Disease and Associated Disorders, U.S. National Library of Medicine, 2015, www.ncbi.nlm.nih.gov/pmc/articles/PMC4437917/.

[3] Lee, Ji Eun, et al. "Association between Alzheimer's Disease and Cancer Risk in South Korea: An 11-Year Nationwide Population-Based Study." Dementia and Neurocognitive Disorders, U.S. National Library of Medicine, Dec. 2018

[4] Lin HL;Lin HC;Tseng YF;Chen SC;Hsu CY; "Inverse Association between Cancer and Dementia: A Population-Based Registry Study in Taiwan." Alzheimer Disease and Associated Disorders, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/26523711/. Accessed 10 Dec. 2023.

[5]"NHIS - about the National Health Interview Survey." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 22 Nov. 2023, www.cdc.gov/nchs/nhis/about$_n his.htm$.

[6] Roderburg, Christoph, et al. "Cancer Patients Have an Increased Incidence of Dementia: A Retrospective Cohort Study of 185,736 Outpatients in Germany." MDPI, Multidisciplinary Digital Publishing Institute, 22 Apr. 2021, www.mdpi.com/2072-6694/13/9/2027.

[7] Sun, Ming, et al. "The Association between Cancer and Dementia: A National Cohort Study in Sweden." Frontiers, Frontiers, 15 Jan. 2020, www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2020.00073/full#h6

[8] "What Is Dementia?" Alzheimer's Disease and Dementia, www.alz.org/alzheimers-dementia/what-is-dementia. Accessed 18 Dec. 2023.

[9] "What Is Dementia? Symptoms, Types, and Diagnosis." National Institute on Aging, U.S. Department of Health and Human Services, www.nia.nih.gov/health/alzheimers-and-dementia/what-dementia-symptoms-types-and-diagnosis. Accessed 18 Dec. 2023.

[10] Zhang, Dan-Dan, et al. "Investigating the Association between Cancer and Dementia Risk: A Longitudinal Cohort Study." Alzheimer's Research & Therapy, U.S. National Library of Medicine, 5 Oct. 2022, www.ncbi.nlm.nih.gov/pmc/articles/PMC9533604/.