Keyword Approximated Social and Behavioral Determinants of Health and Alzheimer's Disease Risk: A 15-year Longitudinal EHR Study

Rumeng Li[1,2], MS; Dan Berlowitz[3,4], MD; Brian Silver[5], MD; Xun Wang[2], PhD; Hong Yu[1,2,4,6], PhD

[1]Manning College of Information & Computer Sciences, University of Massachusetts Amherst, Amherst, MA, USA;

[2]Center for Healthcare Organization & Implementation Research, VA Bedford Health Care System, Bedford, MA, USA;

[3]Department of Public Health, University of Massachusetts Lowell, Lowell, MA, USA;

[4]Center of Biomedical and Health Research in Data Sciences, University of Massachusetts Lowell, Lowell, MA, USA;

[5]Department of Neurology, UMass Chan Medical School, Worcester, MA, USA

[6]Miner School of Computer & Information Sciences, University of Massachusetts Lowell, Lowell, MA, USA;

ABSTRACT

BACKGROUND: Investigating the associations between social and behavioral determinants of health (SBDH) and Alzheimer's disease (AD) is essential for the prevention or mitigation of AD risk factors related to lifestyle and social environment. We used longitudinal electronic health record (EHR) notes to examine how SBDH affect AD.

METHODS: Using longitudinal EHRs from the U.S. Department of Veterans Affairs Veterans Health Administration (VHA) (2006-2021), we conducted a case-control study of selected AD

cases diagnosed after 1/1/2016 and matched each case with nine controls including replacement by age, sex and clinical utilization. We applied natural language processing (NLP) methods to extract keywords suggesting SBDH vulnerability from EHR notes up to 15 years before diagnosis. We compared the frequency and patterns of these keywords between the two groups and analyzed their associations with AD risk. We also explored the SBDH keyword trends by demographics and SBDH categories.

Results: We found that SBDH-related keywords were more frequent in the EHRs of AD cases than in those of controls, and that their frequency changed as the disease progressed. The changes differed between cases and controls and correlated with AD progression and risk. In the 15 years leading up to the first AD diagnosis, the keywords per patient per year increased exponentially from 14.0 to 47.4 for AD cases and linearly from 11.6 to 25.0 for controls. The patterns remained consistent across all subgroups with different frequencies implying that different subgroups may face different needs and challenges in coping with AD.

Conclusions: This study demonstrates the feasibility and utility of using EHR notes and NLP methods to measure and analyze SBDH and AD. Our findings reveal the role and impact of SBDH on AD, offering insights for effective and tailored interventions for AD prevention and management at individual and population levels.

Keywords: Alzheimer's Disease, social and behavioral determinants of health, Electronic Health Records

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that affects memory and other cognitive functions.[1] AD poses a huge public health and economic challenge, as it increases the costs and mortality rates of health care and social services.[2] Since there is no cure for AD,[3,4] it is important to identify and address the risk factors that can prevent or delay its onset and progression.[5] Social and behavioral determinants of health (SBDH), which include behavioral, social, economic, environmental, and occupational factors, greatly influence a person's health and their risk of AD.[6] Studies estimate that SBDH account for as much as 50 percent of the variation in health outcomes, while clinical care impacts only 20 percent.[7] SBDH can influence AD risk and progression through multiple mechanisms, such as affecting brain development, vascular health, inflammation, oxidative stress, lifestyle behaviors, and cognitive reserve.[8–11] SBDH can also affect the access to and quality of health care and social services for people with AD and their caregivers, as well as their ability to cope with the disease and its consequences.[12–14] Therefore, addressing SBDH is essential for reducing the risk and burden of AD and improving the quality of life for people with AD and their caregivers.

Previous studies have shown that SBDH play a role in AD,[6,8,15–17] However, earlier studies on this topic have been conducted primarily on single SBDH factors such as food insecurity, socioeconomic status, neighborhood disadvantage measured by the Area Deprivation Index (ADI), education, or social isolation etc., and neglect the interactions and synergies among multiple SBDH.[18–22] Moreover, most studies are limited by small sample sizes or single-hospital settings, which may not capture the complexity and dynamics of SBDH and AD.[18] This study examines the impact of SBDH on AD using a comprehensive and longitudinal approach. We analyzed a large national EHR dataset from the US Department of Veterans Affairs Veteran Health Administration

(VHA), which covers multiple hospitals and spans 15 years. We investigated how SBDH affect AD over time and across different populations. Existing research has shown that SBDH are more frequently documented in unstructured clinical notes than structured data of EHR.[23–26] However, chart review of EHRs' unstructured free-text is not a feasible approach for screening a large population of patients, as it is time-consuming and subjective. Natural language processing (NLP) can help alleviate this problem. In this study, we proposed a cost-efficient NLP method by using a list of expert-curated keywords indicating SBDH vulnerability to identify and analyze SBDH from longitudinal EHR notes. We aimed to gain a better understanding of the long-term role and impact of SBDH on AD by addressing the following three questions. First, what are the trends of SBDH over time for AD patients and controls and whether the SBDH trends are associated with AD? Second, how do different SBDH factors (such as housing, food, etc.) influence the risk of AD? Third, how do the associations between SBDH and AD differ by age, race/ethnicity, sex or education level? Our study may help identify AD risk factors and design effective interventions for AD prevention and management, and to improve the health and well-being of individuals and populations affected by AD.

METHODS

Data Source

This study used longitudinal EHR notes from patients receiving care in the VHA from January 1, 2006 to October 1, 2021. The study was approved by the Institutional Review Board at the VHA Bedford Healthcare System. We obtained a waiver of informed consent due to minimal risk to participants.

Study Population

An incidence-based, case-control design was used in this study. Figure S1 shows the inclusion-exclusion criteria for generating the cohort. The cases are patients with first AD diagnosis between January 1, 2016 and October 1, 2021 based on ICD codes (Table S1). To improve the accuracy of AD case ascertainment, an AD case must either have been diagnosed: 1) once at a specialty dementia clinic in VHA or 2) at least twice at separate time points with at least one diagnosis generated at specialty clinics including neurology, geriatrics, GeriPACT (Geriatric Patient Aligned Care Team), mental health, psychiatry, psychology, or geriatric psychiatry with a provider type restricted to neurology, vascular neurology, psychiatry, neuropsychology and geriatric medicine. The clinic types are identified by Stop Codes (Table S2) which are used by VHA to specify the type of outpatient care and workload of a visit.[27]

For the longitudinal analysis, we excluded AD patients with fewer than five years continuous EHR notes prior to first AD diagnosis. We matched nine controls using non-AD patients with replacement by age (±1 year), sex, and clinical utilization. Cases and controls had their first visit in the same year within the study period and at least one visit in the year of the case diagnosis. We included patients younger than age 65 because early-onset AD comprises about 5% to 6% of AD cases.[28] By design, a patient could be a control participant for multiple case participants. A case participant before diagnosis could not serve as a control participant for another case participant.

There are hundreds of different note types in the EHRs. In this study, we focused on notes from clinical visits related to AD patients, such as primary care, memory clinics, neurology, psychology, psychiatry, geriatric psychiatry, geriatric medicine, cognitive care nurses, mental health clinics,

and consultation visits. Primary care notes comprised 40.6% of all notes for cases and 56.2% for controls. Figure S2 shows the note type distribution in the cohort.

SBDH-related Keywords Extraction & Frequency Pattern Analysis

We developed computationally efficient NLP methods to identify SBDH keywords from AD and controls, and to examine the SBDH frequency patterns and trends in their clinical documentation. Through an extensive literature review, domain experts HK, RG and WH curated a widely inclusive list of keywords indicating SBDH vulnerability, resulting in a final list of 522 keywords. The curated keywords list covered thirteen SBDH domains including food insecurity, economic instability (including financial or employment difficulties), housing instability, health care access barriers (including transportation or language challenges), social isolation, violence, legal problems, stigma/shame, substance abuse, negative mood & affect, sleep disturbances, pain, and disability. Table S1 in the Supplementary shows the definitions and keywords for each domain. In addition, we included modifiers such as "poor, inadequate, lack of, no, not enough, without and insecurity." We identified SBDH occurrences in the notes, and tracked their occurrences over time, by aligning them to patients' age or year before diagnosis. Clinicians document patients' SBDH status in their EHRs to assess their needs, risks, and outcomes. We used the frequency of keywords in the EHRs to measure the SBDH status of the patients and to examine how it is associated with the AD outcomes.

We calculated the frequencies of SBDH keywords by year for up to 15 years before AD diagnosis using: 1) all the notes for our selected note types and 2) only primary care notes because the primary care setting is often a patient's entry point into the healthcare system.[29] To account for the possible bias of AD cases having more physician visits and notes than controls, we adjusted the

frequency of SBDH keywords by the number of total notes. We explored the patterns of these keywords and their associations with AD, considering biological (sex, age, race/ethnicity) and sociobiological variable (education levels). We also analyzed the SBDH keyword patterns in each category, reflecting patients' SBDH needs in different domains.

RESULTS

Sample Characteristics

Our case cohort comprised 16,701 AD patients (488 [2.9%] women; mean [SD] age, 76.7 [8.6] years) and the control cohort comprised 39,097 patients (1,013 [2.6%] women; mean [SD] age, 76.6 [8.9] years). All ages were as of 1/1/2016. White patients accounted for 77.4% in the case patients and 79.9% in the control. For Black or African American, the ratios were 14.1% and 10.3% respectively. Hispanic or Latino ethnicity accounted for 11.2% in the case patients and 4.5% in the control group. Detailed characteristics of the study cohort are in Table 1. Among AD cases (16,701 patients), the majority of 16,404 patients (98.2%) had multiple ICD codes for AD with at least one diagnosis code from specialty clinics and the remaining 297 patients (1.8%) had a single dementia clinic-based diagnosis. The Charlson Comorbidity Index,[30,31] which measures the burden of comorbid conditions, was not significantly different between AD cases and controls. The mean [SD] of the index was 3.74 [3.11] for cases and 2.92 [2.91] for controls. Both groups also had some similar top comorbidities, such as hypertension, diabetes, and hyperlipidemia (Table S3).

SBDH-related Keywords Frequencies in Longitudinal EHRs

Our results show that SBDH keywords were more frequently mentioned in the EHR notes of AD case patients than control patients. On average, a case patient had 338.7 keywords extracted from an average of 75.0 notes, spanning 10.7 years of their EHR history, compared with a control who

had 193.0 keywords from 47.5 notes, spanning 9.1 years. We calculated the average number of mentions for each keyword group representing its corresponding domain of SBDH for both AD cases and controls (Figure S4(a)). Our results show that AD cases had higher SBDH keyword frequency than controls across all domains. The SBDH domains with the most frequent keywords are: pain, substance abuse and negative mood & affect, which accounted for 72.0% and 74.74% of the total keyword frequency in AD cases and controls. The next domains with the most frequent keywords were violence, disability, social isolation, and food insecurity, which compromised 20.6% and 18.9% of the total keyword frequency in AD cases and controls, respectively. Figure S4(b) displays the keywords appearance ratio for 13 SBDH domains in AD case and control patients. The ratio is the average number of keywords for case patients divided by the average number for control patients. A ratio above 1 means that case patients have higher SBDH keyword frequency than control patients in that domain, and vice versa. We observed the ratio above one in all domains and the ratio is above two in some domains: stigma/shame, health care access barriers, violence, and negative mood/affect.

SBDH-related Keyword Occurrence Patterns in Longitudinal EHRs

We found that SBDH keywords were more frequent in the EHRs of AD cases than in those of controls. In addition, the frequencies change as AD progresses and there is a difference in these changes between cases and controls that correlates with AD progression and AD risks. In the 15 years leading up to the first AD diagnosis, the number of SBDH keywords per AD patient per year increased exponentially rising from 14.0 to 47.4. During the same time period, the number of keywords in the notes of the controls increased only linearly from 11.6 to 25.0 in the same time period (Figure 1a). We replicated this finding using only primary care notes (Figure 1b), and using a normalized keyword frequency by dividing it by the notes number (Figure 1c). This pattern was consistent across different ages at diagnosis (Figure 2a). The pattern generally remains when stratifying the patient cohort by sex, and race/ethnicity (Figure 2b, Figure 3).

Figure 2b shows the SBDH keyword frequencies stratified by sex. The pattern for female patients shows fluctuation due to the limited number of female patients (case: 488 (2.9%), control: 1,013 (2.6%)). However, it shows that female controls have higher SBDH keyword frequencies than male controls, and female cases have equal or greater SBDH keyword frequencies than male cases along with AD progression.

Figure 3(a) shows the SBDH keyword frequencies stratified by ethnicity. Hispanic/Latino patients have much higher SBDH keyword frequencies than non-Hispanic patients on average. Hispanic/Latino controls have similar keyword frequencies as non-Hispanic/Latino AD cases, and both surpass non-Hispanic/Latino AD controls. Hispanic/Latino AD cases have a sharper rise in SBDH keyword frequencies from 15 years before to the year of their first AD diagnosis. At

diagnosis, Hispanic/Latino AD cases have 1.7 times more keyword frequency than non-Hispanic/Latino AD cases on average.

We compared the SBDH keyword frequencies across domains for Hispanic/Latino and non-Hispanic/Latino patients (Figure S5(a)). Figure S5(b) displays the ratio of SBDH keywords for AD cases and controls of Hispanic and non-Hispanic ethnicity in 13 domains. The ratio is the number of keywords for Hispanic patients divided by the number of keywords for non-Hispanic patients, for both AD cases and controls. A ratio above 1 means that Hispanic patients have more SBDH keyword mentions than non-Hispanic patients, and vice versa. The plot reveals that 1) Hispanic patients have higher SBDH keyword frequencies than non-Hispanic patients (ratio>1) in every domain, irrespective of their AD status. 2) The ratio exceeds 2 for violence and stigma/shame domains, indicating the difficulties for Hispanic patients in these domains. 3) Some domains are more strongly linked to AD risk among Hispanic patients than others. For instance, the ratios of cases are much higher than controls in domains such as stigma/shame, social isolation, sleep, housing instability, food insecurity, substance abuse, and disability.

Figure 3b illustrates the SBDH keyword frequencies by race. Black/African American patients have more SBDH keyword frequencies than White patients. The other races such as Native Hawaiian or Other Pacific Islander, American Indian or Alaska Native, and Asian show variations in the trends due to a relatively limited number of patients (all <1% of cases and controls), but we can infer that Native Hawaiian or Other Pacific Islander and American Indian or Alaska Native have moderate SBDH keyword frequencies and Asians have the lowest.

For all SBDH domain, AD case patients had a faster rise in SBDH keyword frequencies than control (Figure 4).

Among the three education levels, we found a small variation in SBDH keyword frequencies. Graduate-educated patients had the lowest SBDH keyword frequencies, followed by college-educated and high school or vocational-technical-educated patients.

DISCUSSION

Our study showed that AD patients had higher SBDH frequency than control patients 15 years before diagnosis, and this gap widened over time, suggesting associations between the SBDH factors and AD disease.

We found that the SBDH frequency is higher in female AD patients than male AD patients. This is consistent with existing studies that show that female veterans have lower socioeconomic status than male veterans, such as lower median household income, higher rates of divorce or singlehood, higher rates of service-connected disability, no personal income, and poverty.[33]

We observed an age difference in the documentation of SBDH keywords in EHRs. As people aged, they had fewer SBDH-related keywords in their EHR notes. Existing studies show that very old individuals tend to have a general preference for acceptance and emotion control, which is also known as a developmental coping shift in old age. They may use different coping strategies than younger adults to deal with their challenges and losses.[34] In addition, older adults may be reluctant or unable to disclose their SBDH vulnerability due to stigma, shame, privacy concerns or communication barriers.[35]

Compared to non-Hispanic/Latino patients, Hispanic/Latino patients had significantly higher frequency of SBDH keywords. Moreover, the SBDH keyword frequency increased more sharply in Hispanic AD cases, which aligns with previous studies that suggest faster AD progression among Hispanics/Latinos than other ethnoracial groups.[36] We found that Hispanic/Latino patients have more SBDH keyword frequencies in every SBDH domains than non-Hispanic/Latino patients. For the violence and stigma/shame category, the ratio is >2 signifying the challenges for Hispanic patients in those two domains. We further checked the violence category, for the top frequent keywords, we compared the frequency of three keywords: trauma, abuse, and neglect. We found that Hispanic/Latino AD case patients and controls had higher mentions of these keywords than non-Hispanic/Latino AD cases and controls. The ratios were as follows: 1) Trauma: 8.27 times for AD cases and 8.64 times for controls; 2) Abuse: 1.72 times for AD cases and 1.68 times for controls; 3) Neglect: 1.57 times for AD cases and 1.56 times for controls. Studies have shown that Posttraumatic stress disorder (PTSD) and traumatic brain injury (TBI)symptom severity are associated with accelerated cognitive decline.[37,38]Moreover, it is important to emphasize that racism itself is traumatic and could surround the victims' life course and affect their physical and mental health, behavior, cognition, interpersonal relationships, self-concept, and socioeconomic status.[39] Existing studies also reveals that Hispanic /Latinos often lack access to healthcare and have poor health outcomes due to social and structural issues such as xenophobia, language barriers, poverty, mistrust of the health care system, and limited access to high-quality health care.[40,41] These issues may be exacerbated by the experiences of violence and stigma/shame that our analysis revealed. Therefore, it is important to address these barriers and provide culturally competent and trauma-informed care for Hispanic patients.

As they neared diagnosis, AD cases had higher and faster-increasing frequency of keywords related to food insecurity. Previous work also showed that older adults living with food insecurity are more likely to have lower nutrient intakes and experience poorer health outcomes, such as cardiovascular and metabolic diseases, increased stress and depression, and increased AD risk.[6-10] Food security, unlike other risk factors such as education and exercise, may be improved by existing federal programs like the Supplemental Nutrition Assistance Program (SNAP).[63] By increasing SNAP participation among older adults, we may reduce AD disparities at the population level.

Studies found that individuals who experience high socioeconomic deprivation measured using income/wealth, unemployment rates etc. are significantly more likely to develop dementia than individuals of better socioeconomic status, even at high genetic risk.[42] Results on the economic instability category, examining financial difficulties and job insecurity, also shows the correlations between economic instability and AD.

Our results, which showed higher and faster-increasing frequency of keywords related to housing instability in AD case patients than in control patients, are consistent with existing studies that suggest a bidirectional link between homelessness and AD.[43,44]

Social isolation in midlife and late life is a major risk factor for AD, while social engagement has been studied as a protective factor that boosts cognitive reserve and beneficial behaviors.[15,43] Our study results also showed that AD case patients had higher and increasing frequency of keywords

related to social isolation as they neared diagnosis, confirming a correlation between AD and social isolation.

Community and domestic violence have been studied as associated with AD.[45,46] Our result on the violence category also confirms such association.

We observed similar patterns for legal problems-related keywords, showing higher increase in AD case patients than control patients. Existing studies mostly focus on whether AD affects a person's legal capacity[47] and on the criminal behaviors caused by the cognitive problems of patients with AD.[48,49] Few studies have investigated legal problems as a risk factor of AD. However, some studies investigated the effects of the justice system on mental health and claimed that the uncertainty about future events is a great source of stress. It generates learned helplessness and therefore worsens one's mental health.[50] Some work found that legal problems had the largest impact, as well as the most significant association with increased risk of death by suicide among the social determinants of health investigated.[24] Our study suggests an association between legal problems and AD.

Previous studies have shown that some drugs and heavy or hard liquor consumption can accelerate cognitive decline in AD patients, implying that they may worsen AD progression.[51,52] Our results on the substance abuse category also support this correlation. Reducing tobacco smoking and harmful alcohol or drug use through health promotion programs may lower the risk of cognitive impairment and AD in later life.

With the decline from a previously attained cognitive level that affects activities of daily living or social functioning, AD has become one of the main causes of disability in elderly people.[53] In the meantime, hearing loss, vision impairment and physical inactivity have been studied as potential modifiable risk factors for AD.[15,54] Our results on the disability category also suggest an association between disability and AD.

Negative mood & affect impact AD outcomes,[55,56] and a positive outlook engendered by spiritual beliefs may benefit patients in the long run. The neurodegeneration of emotional regions and pathways can trigger anxiety and depression, which may further accelerate neurodegeneration.[57] Our findings also confirm associations between negative mood & affect and AD, and suggest that alleviating negative mood symptoms could be an effective strategy for preventing AD.

Chronic pain affects 25% to 33% of older adults, and its prevalence increases with age.[58,59] Studies showed that chronic pain was associated with accelerated memory decline and increased probability of AD.[60–63] Our study confirms this association.

Sleep is crucial to consolidate memory and to remove the excess of beta-amyloid and hyperphosphorylated tau accumulated in AD patients' brains. Sleep disturbances have been studied as both clinical manifestations and risk factors of AD.[64–66] This is consistent with our study.

Access to health care is a critical factor in AD risk and management, especially for under-resourced communities. The keywords pattern divergence between AD cases and controls in this category

also suggests that AD case patients may face greater challenges in accessing and utilizing health care services that could potentially prevent or delay the onset of AD.

Stigma and shame are powerful forces that affect health, illness, and health-related behaviors. They can influence the quality and outcome of the clinical encounter. They also contribute to health disparities and disadvantage marginalized groups in society. However, stigma and shame are often overlooked and understudied in health and medicine.[67,68] Despite the sparse documentation in EHRs, we can still observe the difference in keyword patterns between AD cases and controls.

Education may influence AD development in various ways: it mirrors early-life and socioeconomic factors; it impacts work, income, and wealth; it molds occupational and adult life aspects (e.g., job complexity, work stress, environmental exposures); and it provides lifelong skills for cognition and mastery. Education also signals cognitive reserve, which may guard against cognitive impairment and dementia risk, by boosting neural network and compensation.[19] Our finding suggests that higher education levels may be associated with lower SBDH frequency and lower AD risk. However, the variation in SBDH frequency among education levels was small, indicating that other factors may also contribute to SBDH and AD. Further research is needed to explore the complex interactions between education, SBDH, and AD.

Our study suggests several policy implications for AD prevention and management. First, we found that some important SBDH factors, such as health literacy, dietary patterns, and physical activity, which are linked to better cognition and lower AD risk, are poorly recorded in EHRs. This may reflect a gap in awareness, screening, or documentation of SBDH by health care providers

and staff. We recommend enhancing the evaluation and recording of SBDH in these domains to identify and assist AD patients and those at risk, and to address their SBDH needs and challenges.

Second, we found that the SBDH keyword occurrences were associated with the odds of AD, after matching the case and control cohorts by sex, age, and clinical utilization, as well as with similar comorbidities found between those two groups. This suggests that the management of SBDH and AD should be integrated and coordinated across the health care and social service sectors, rather than isolated or fragmented, to provide comprehensive and holistic care for the patients and their caregivers. A comprehensive understanding of the impact of social determinants on health can ultimately improve health in individuals and populations as clinicians can offer more effective treatments, improved social screening, timely referrals to legal and social services, and initiation of research to understand the mechanisms by which social factors affect health. Healthcare organizations should adopt these interventions to improve the health status of older adults at the local, national, and global level.[69]

Third, we found that the frequency of SBDH keywords varies by age, sex, race/ethnicity, and SBDH domains, although they all show similar increase in patterns. This indicates that the management of SBDH and AD should be personalized and tailored to the specific needs and characteristics of patients and their caregivers to accommodate individual preferences, goals and circumstances, rather than one-size-fits-all.

Limitations

Exploring EHRs as a source of data for studying SBDH saves the effort of using screening tools to track these variables. However, our study has some limitations that need to be addressed by

future work. First, the VA population is not representative of the US population, as it is demographically skewed. Veterans also face unique challenges and disparities related to their military service such as homelessness, unemployment, disability, social isolation, and stigma, that may affect their SBDH and their AD outcomes. Second, our case-control design cannot infer causality and may have biases and confounding factors. A prospective cohort or a trial would be better, but more costly and time-consuming. Third, our study uses the EHRs as the data source, which may have limitations. The EHRs may not capture the true or complete picture of the SBDH and the AD status of the patients. For example, social engagement/support, which may have a positive impact on patients with AD, especially in terms of having someone to advocate for them, assist with activities of daily living (ADLs)/ Instrumental activities of daily living (iADLs), etc., may not be adequately captured by the EHR notes. Moreover, the EHRs for the studied period may not capture some of the early SBDH in young adulthood that could affect AD outcomes, since most AD diagnoses occur in later life. In addition,, we exclude patients with fewer than five years of continuous history for our longitudinal analysis, these patients might have higher SBDH risks and lower healthcare access and resources. Another limitation of using EHR notes is that clinicians can tag or import parts of previous notes, creating redundancy instead of originality. Fourth, we base our list of SBDH-related keywords on literature review and expert consultation, but it might not reflect the diversity and subtlety of the SBDH domains, or account for the clinical context. Fifth, due to data limitations, we could not conduct more detailed analysis considering the intricate links among health determinants. For example, we could not adjust for education levels to explore how race/ethnicity affects economic stability, among other factors etc.

Despite these limitations, this study, based on a large national EHR database from the VHA covering 1,298 health care settings across US, using one of the largest patient cohorts for AD research with 16,701 AD patients and over 15 years of longitudinal analysis, contributes to population health management by introducing a new method to measure the long-term impact of SBDH on AD outcome.

CONCLUSIONS

This study examines the relationship between SBDH and AD using longitudinal EHR notes and NLP methods. We compare the frequency of SBDH-related keywords in the EHR notes of AD cases and controls, and track the changes in keyword frequency over time as the disease progressed. We find that SBDH-related keywords are more frequent and increase more rapidly in the EHR notes of AD cases than in those of controls, regardless of demographic and SBDH characteristics of the patients. These findings suggest that SBDH plays an important role in AD development and progression. Our study demonstrates the feasibility of NLP methods on EHR notes to measure and analyze SBDH and AD; to inform the design and implementation of effective interventions and prevention for improving the health and well-being of individuals and populations affected by AD.

ACKNOWLEDGMENTS

Conflict of Interest: The authors have no conflicts of interest.

REFERENCES

1. Mucke L. Alzheimer's disease. *Nature*. 2009;461(7266):895-897. doi:10.1038/461895a

2. 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2023;19(4):1598-1695. doi:10.1002/alz.13016

3. Mehta D, Jackson R, Paul G, Shi J, Sabbagh M. Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015. *Expert Opin Investig Drugs*. 2017;26(6):735-739. doi:10.1080/13543784.2017.1323868

4. Yiannopoulou KG, Anastasiou AI, Zachariou V, Pelidou SH. Reasons for Failed Trials of Disease-Modifying Treatments for Alzheimer Disease and Their Contribution in Recent Research. *Biomedicines*. 2019;7(4):97. doi:10.3390/biomedicines7040097

5. Rao RV, Subramaniam KG, Gregory J, et al. Rationale for a Multi-Factorial Approach for the Reversal of Cognitive Decline in Alzheimer's Disease and MCI: A Review. *Int J Mol Sci*. 2023;24(2):1659. doi:10.3390/ijms24021659

6. Social Determinants of Health and Alzheimer's Disease and Related Dementias | Aging. Published September 1, 2023. Accessed December 13, 2023. https://www.cdc.gov/aging/disparities/social-determinants-alzheimers.html

7. Whitman A, Lew ND, Chappel A, Aysola V, Zuckerman R, Sommers BD. Addressing Social Determinants of Health: Examples of Successful Evidence-Based Strategies and Current Federal Efforts. Published online 2022.

8. Adkins-Jackson PB, George KM, Besser LM, et al. The structural and social determinants of Alzheimer's disease related dementias. *Alzheimer's & Dementia*. 2023;19(7):3171-3185. doi:10.1002/alz.13027

9. Roe CM, Mintun MA, D'Angelo G, Xiong C, Grant EA, Morris JC. Alzheimer Disease and Cognitive Reserve: Variation of Education Effect With Carbon 11–Labeled Pittsburgh Compound B Uptake. *Archives of Neurology*. 2008;65(11):1467-1471. doi:10.1001/archneur.65.11.1467

10. Suglia SF, Campo RA, Brown AGM, et al. Social Determinants of Cardiovascular Health: Early Life Adversity as a Contributor to Disparities in Cardiovascular Diseases. *J Pediatr*. 2020;219:267-273. doi:10.1016/j.jpeds.2019.12.063

11. Zhao YL, Qu Y, Ou YN, Zhang YR, Tan L, Yu JT. Environmental factors and risks of cognitive impairment and dementia: A systematic review and meta-analysis. *Ageing Research Reviews*. 2021;72:101504. doi:10.1016/j.arr.2021.101504

12. Peterson RL, George KM, Tran D, et al. Operationalizing Social Environments in Cognitive Aging and Dementia Research: A Scoping Review. *Int J Environ Res Public Health*. 2021;18(13):7166. doi:10.3390/ijerph18137166

13. Chen X, Lee C, Huang H. Neighborhood built environment associated with cognition and dementia risk among older adults: A systematic literature review. *Social Science & Medicine*. 2022;292:114560. doi:10.1016/j.socscimed.2021.114560

14. Besser L. Outdoor green space exposure and brain health measures related to Alzheimer's disease: a rapid review. *BMJ Open*. 2021;11(5):e043456. doi:10.1136/bmjopen-2020-043456

15. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396(10248):413-446. doi:10.1016/S0140-6736(20)30367-6

16. Li JM, Boustani MA, French DD. Social Determinants of Health in Community-Dwelling Dementia Patients Aged 65 and Over: Analysis of the 2019 National Health Interview Survey. *Gerontol Geriatr Med*. 2023;9:23337214231190244. doi:10.1177/23337214231190244

17. Majoka MA, Schimming C. Effect of Social Determinants of Health on Cognition and Risk of Alzheimer Disease and Related Dementias. *Clinical Therapeutics*. 2021;43(6):922-929. doi:10.1016/j.clinthera.2021.05.005

18. Powell WR, Buckingham WR, Larson JL, et al. Association of Neighborhood-Level Disadvantage With Alzheimer Disease Neuropathology. *JAMA Network Open*. 2020;3(6):e207559. doi:10.1001/jamanetworkopen.2020.7559

19. Cadar D, Lassale C, Davies H, Llewellyn DJ, Batty GD, Steptoe A. Individual and Area-Based Socioeconomic Factors Associated With Dementia Incidence in England: Evidence From a 12-Year Follow-up in the English Longitudinal Study of Ageing. *JAMA Psychiatry*. 2018;75(7):723. doi:10.1001/jamapsychiatry.2018.1012

20. Na M, Dou N, Ji N, et al. Food Insecurity and Cognitive Function in Middle to Older Adulthood: A Systematic Review. *Advances in Nutrition*. 2020;11(3):667-676. doi:10.1093/advances/nmz122

21. Qian H, Khadka A, Martinez SM, et al. Food Insecurity, Memory, and Dementia Among US Adults Aged 50 Years and Older. *JAMA Network Open*. 2023;6(11):e2344186. doi:10.1001/jamanetworkopen.2023.44186

22. Wong CG, Miller JB, Zhang F, et al. Evaluation of Neighborhood-Level Disadvantage and Cognition in Mexican American and Non-Hispanic White Adults 50 Years and Older in the US. *JAMA Network Open*. 2023;6(8):e2325325. doi:10.1001/jamanetworkopen.2023.25325

23. Mitra A, Ahsan H, Li W, et al. Risk Factors Associated With Nonfatal Opioid Overdose Leading to Intensive Care Unit Admission: A Cross-sectional Study. *JMIR Med Inform*. 2021;9(11):e32851. doi:10.2196/32851

24. Mitra A, Pradhan R, Melamed RD, et al. Associations Between Natural Language Processing–Enriched Social Determinants of Health and Suicide Death Among US Veterans. *JAMA Netw Open*. 2023;6(3):e233079. doi:10.1001/jamanetworkopen.2023.3079

25. Ahsan H, Ohnuki E, Mitra A, You H. MIMIC-SBDH: A Dataset for Social and Behavioral Determinants of Health. In: *Proceedings of the 6th Machine Learning for Healthcare Conference*. PMLR; 2021:391-413. Accessed November 13, 2023. https://proceedings.mlr.press/v149/ahsan21a.html

26. Oreskovic NM, Maniates J, Weilburg J, Choy G. Optimizing the Use of Electronic Health Records to Identify High-Risk Psychosocial Determinants of Health. *JMIR Med Inform*. 2017;5(3):e25. doi:10.2196/medinform.8240

27. Affairs D of V, others. VHA Handbook 1006.02, VHA site classifications and definitions.

28. Mendez MF. Early-onset Alzheimer Disease and Its Variants: *CONTINUUM: Lifelong Learning in Neurology*. 2019;25(1):34-51. doi:10.1212/CON.0000000000000687

29. Liss JL, Seleri Assunção S, Cummings J, et al. Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis. *Journal of Internal Medicine*. 2021;290(2):310-334. doi:10.1111/joim.13244

30. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8

31. CIPHER - Charlson Index (MVP Core). Accessed January 9, 2024. https://phenomics.va.ornl.gov/web/cipher/phenotype-viewer?uqid=6d7233cdee384f0d9de153fc108121bd&name=Charlson_Index__MVP_Core_

32. Röhr S, Pabst A, Baber R, et al. Social determinants and lifestyle factors for brain health: implications for risk reduction of cognitive decline and dementia. *Sci Rep*. 2022;12(1):12965. doi:10.1038/s41598-022-16771-6

33. Tamborini CR. Female Veterans: Socioeconomic Characteristics and Disability Patterns Among Social Security Beneficiaries. *Social Security Bulletin*. 2023;83(2).

34. Uittenhove K, Jopp DS, Lampraki C, Boerner K. Coping Patterns in Advanced Old Age: Findings from the Fordham Centenarian Study. *Gerontology*. 2023;69(7):888-898. doi:10.1159/000529896

35. Yorkston KM, Bourgeois MS, Baylor CR. Communication and aging. *Phys Med Rehabil Clin N Am*. 2010;21(2):309-319. doi:10.1016/j.pmr.2009.12.011

36. Novak P, Chu J, Ali MM, Chen J. Racial and Ethnic Disparities in Serious Psychological Distress among Those with Alzheimer's Disease and Related Dementias. *Am J Geriatr Psychiatry*. 2020;28(4):478-490. doi:10.1016/j.jagp.2019.08.010

37. Prieto S, Nolan KE, Moody JN, Hayes SM, Hayes JP. Posttraumatic stress symptom severity predicts cognitive decline beyond the effect of Alzheimer's disease biomarkers in Veterans. *Transl Psychiatry*. 2023;13(1):1-9. doi:10.1038/s41398-023-02354-0

38. Mielke MM, Ransom JE, Mandrekar J, Turcano P, Savica R, Brown AW. Traumatic Brain Injury and Risk of Alzheimer's Disease and Related Dementias in the Population. *J Alzheimers Dis*. 2022;88(3):1049-1059. doi:10.3233/JAD-220159

39. Cénat JM. Complex Racial Trauma: Evidence, Theory, Assessment, and Treatment. *Perspect Psychol Sci*. 2023;18(3):675-687. doi:10.1177/17456916221120428

40. Velasco-Mondragon E, Jimenez A, Palladino-Davis AG, Davis D, Escamilla-Cejudo JA. Hispanic health in the USA: a scoping review of the literature. *Public Health Reviews*. 2016;37(1):31. doi:10.1186/s40985-016-0043-2

41. Vega WA, Rodriguez MA, Gruskin E. Health Disparities in the Latino Population. *Epidemiol Rev*. 2009;31:99-112. doi:10.1093/epirev/mxp008

42. Trani JF, Zhu Y, Park S, et al. Multidimensional poverty is associated with dementia among adults in Afghanistan. *eClinicalMedicine*. 2023;58:101906. doi:10.1016/j.eclinm.2023.101906

43. Associations between Homelessness and Alzheimer's Disease and Related Dementia: A Systematic Review. doi:10.1177/07334648221109747

44. Roncarati JS, DeVone F, Halladay C, Tsai J, Jutkowitz E. Risk of dementia among veterans experiencing homelessness and housing instability. *J Am Geriatr Soc*. Published online December 6, 2023. doi:10.1111/jgs.18680

45. Leung FH, Thompson K, Weaver DF. Evaluating spousal abuse as a potential risk factor for Alzheimer's disease: rationale, needs and challenges. *Neuroepidemiology*. 2006;27(1):13-16. doi:10.1159/000093894

46. News. Studying the impact of community violence on Alzheimer's and dementia in African Americans. news. Accessed January 19, 2024. https://health.ucdavis.edu/news/headlines/studying-the-impact-of-community-violence-on-alzheimers-and-dementia-in-african-americans/2021/03

47. Arias JJ, Flicker LS. A Matter of Intent: A Social Obligation to Improve Criminal Procedures for Individuals with Dementia. *J Law Med Ethics*. 2020;48(2):318-327. doi:10.1177/1073110520935345

48. Liljegren M, Naasan G, Temlett J, et al. Criminal Behavior in Frontotemporal Dementia and Alzheimer Disease. *JAMA Neurology*. 2015;72(3):295-300. doi:10.1001/jamaneurol.2014.3781

49. Kuffel RL, Byers AL, Williams B, et al. Prevalence of Dementia and Mild Cognitive Impairment Prior to Incarceration. *J Am Geriatr Soc*. 2022;70(6):1792-1799. doi:10.1111/jgs.17724

50. Clemente M, Padilla-Racero D. The effects of the justice system on mental health. *Psychiatr Psychol Law*. 27(5):865-879. doi:10.1080/13218719.2020.1751327

51. Hulse GK, Lautenschlager NT, Tait RJ, Almeida OP. Dementia associated with alcohol and other drug use. *Int Psychogeriatr*. 2005;17 Suppl 1:S109-127. doi:10.1017/s1041610205001985

52. Heymann D, Stern Y, Cosentino S, Tatarina-Nulman O, Dorrejo JN, Gu Y. The Association Between Alcohol Use and the Progression of Alzheimer's Disease. *Curr Alzheimer Res*. 2016;13(12):1356-1362. doi:10.2174/1567205013666160603005035

53. Litke R, Garcharna LC, Jiwani S, Neugroschl J. Modifiable Risk Factors in Alzheimer Disease and Related Dementias: A Review. *Clin Ther*. 2021;43(6):953-965. doi:10.1016/j.clinthera.2021.05.006

54. Ehrlich JR, Goldstein J, Swenor BK, Whitson H, Langa KM, Veliz P. Addition of Vision Impairment to a Life-Course Model of Potentially Modifiable Dementia Risk Factors in the US. *JAMA Neurol*. 2022;79(6):623-626. doi:10.1001/jamaneurol.2022.0723

55. Appleby BS, Roy P, Valenti A, Lee HB. Diagnosis and treatment of depression in Alzheimer's disease: impact on mood and cognition. *Panminerva Med*. 2007;49(3):139-149.

56. Robert PH, Berr C, Volteau M, et al. Importance of lack of interest in patients with mild cognitive impairment. *Am J Geriatr Psychiatry*. 2008;16(9):770-776. doi:10.1097/JGP.0b013e31817e73db

57. Botto R, Callai N, Cermelli A, Causarano L, Rainero I. Anxiety and depression in Alzheimer's disease: a systematic review of pathogenetic mechanisms and relation to cognitive decline. *Neurol Sci*. 2022;43(7):4107-4124. doi:10.1007/s10072-022-06068-x

58. Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA. The epidemiology of chronic pain in the community. *Lancet*. 1999;354(9186):1248-1252. doi:10.1016/s0140-6736(99)03057-3

59. Mäntyselkä PT, Turunen JHO, Ahonen RS, Kumpusalo EA. Chronic pain and poor self-rated health. *JAMA*. 2003;290(18):2435-2442. doi:10.1001/jama.290.18.2435

60. Banik RK, Peng S, Hussain N, et al. The Relationship Between Chronic Pain and Cognitive Decline in Older Population: A Cautionary Tale from Current Literature. *Pain Medicine*. 2023;24(2):110-112. doi:10.1093/pm/pnac144

61. Chen J, Wang X, Xu Z. The Relationship Between Chronic Pain and Cognitive Impairment in the Elderly: A Review of Current Evidence. *J Pain Res*. 2023;16:2309-2319. doi:10.2147/JPR.S416253

62. Whitlock EL, Diaz-Ramirez LG, Glymour MM, Boscardin WJ, Covinsky KE, Smith AK. Association Between Persistent Pain and Memory Decline and Dementia in a Longitudinal Cohort of Elders. *JAMA Internal Medicine*. 2017;177(8):1146-1153. doi:10.1001/jamainternmed.2017.1622

63. Zhang X, Gao R, Zhang C, et al. Evidence for Cognitive Decline in Chronic Pain: A Systematic Review and Meta-Analysis. *Front Neurosci*. 2021;15:737874. doi:10.3389/fnins.2021.737874

64. Brzecka A, Leszek J, Ashraf GM, et al. Sleep Disorders Associated With Alzheimer's Disease: A Perspective. *Front Neurosci*. 2018;12:330. doi:10.3389/fnins.2018.00330

65. Sadeghmousavi S, Eskian M, Rahmani F, Rezaei N. The effect of insomnia on development of Alzheimer's disease. *Journal of Neuroinflammation*. 2020;17(1):289. doi:10.1186/s12974-020-01960-9

66. Lloret MA, Cervera-Ferri A, Nepomuceno M, Monllor P, Esteve D, Lloret A. Is Sleep Disruption a Cause or Consequence of Alzheimer's Disease? Reviewing Its Possible Role as a Biomarker. *Int J Mol Sci*. 2020;21(3):1168. doi:10.3390/ijms21031168

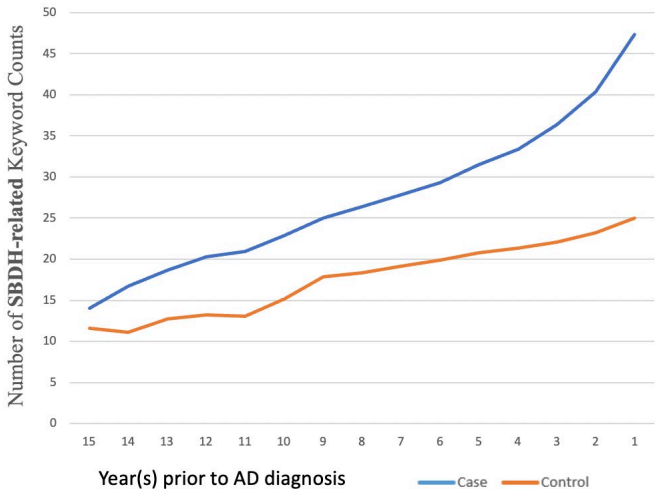
67. Dolezal L, Lyons B. Health-related shame: an affective determinant of health? *Medical Humanities*. 2017;43(4):257-263. doi:10.1136/medhum-2017-011186

68. Addison M, Lhussier M, Bambra C. Relational stigma as a social determinant of health: "I'm not what you _____see me as." *SSM - Qualitative Research in Health*. 2023;4:100295. doi:10.1016/j.ssmqr.2023.100295

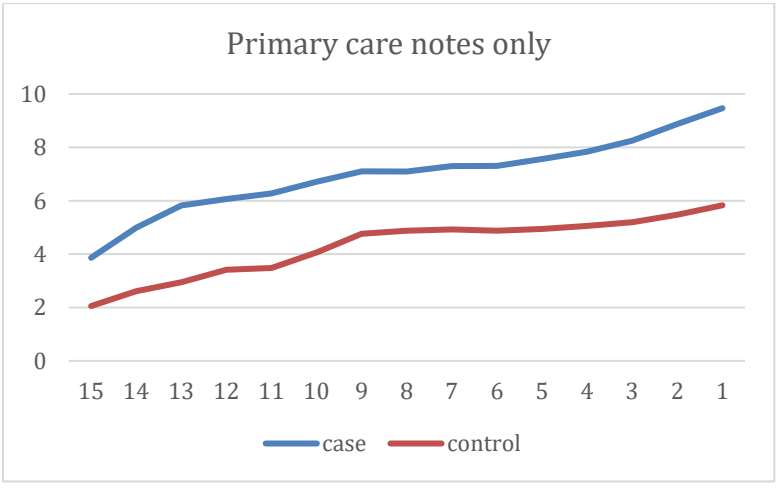
69. Iott BE, Pantell MS, Adler-Milstein J, Gottlieb LM. Physician awareness of social determinants of health documentation capability in the electronic health record. *J Am Med Inform Assoc*. 2022;29(12):2110-2116. doi:10.1093/jamia/ocac154

Figure 1. Comparison of SBDH-related keyword counts between AD case patients and matched control patients: (a). Average number of SBDH keywords per patient by year prior to AD diagnosis. (b). Average number of SBDH keywords per patient by year prior to AD diagnosis in primary care notes only. (c). Average number of SBDH keywords per note by year prior to AD diagnosis.

(a)



(b).



(c).

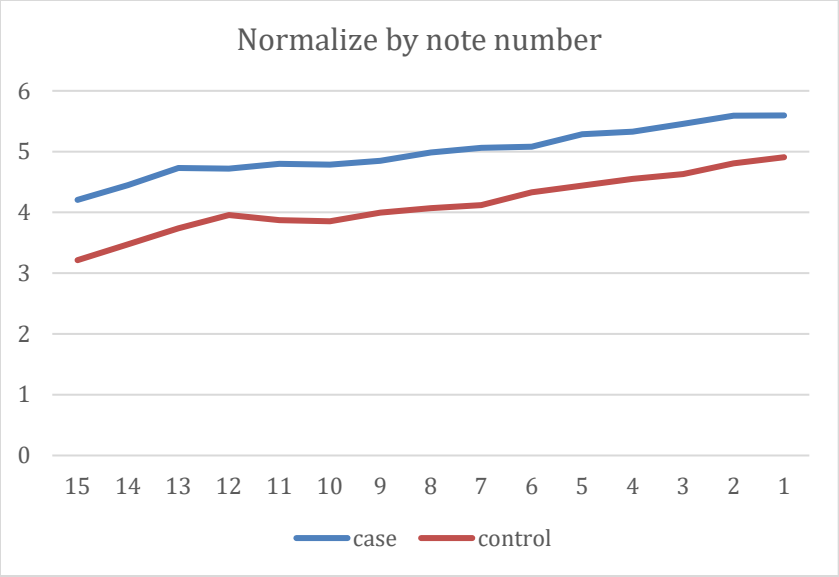
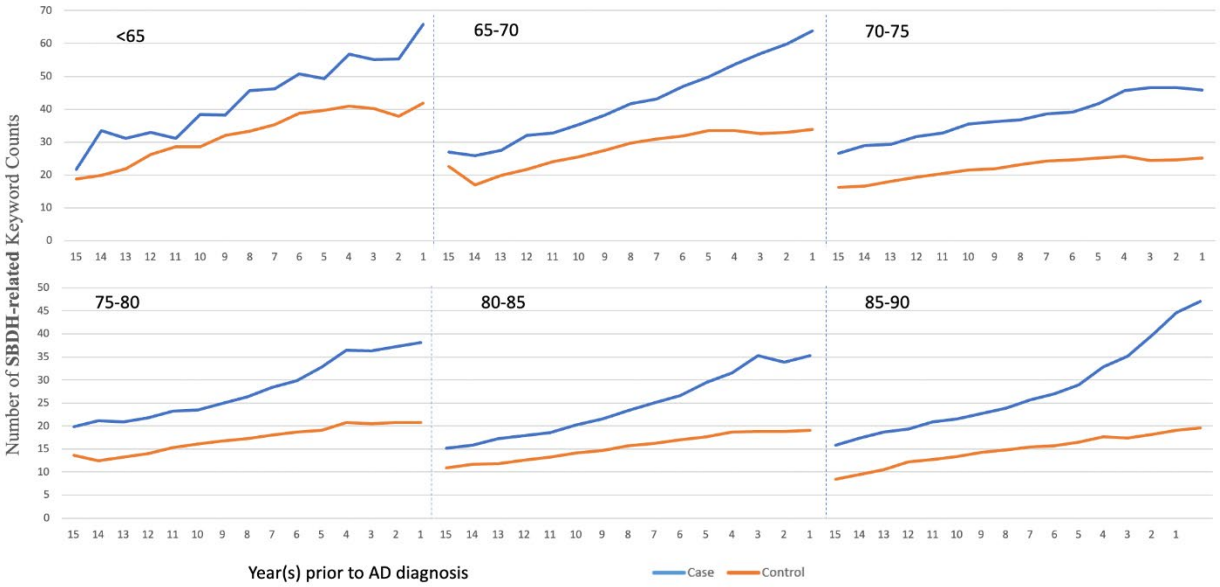

Normalize by note number
6
5
4
3
2
1
0
15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
case
control

Figure 2. Comparison of SBDH keywords by age group and sex for AD case patients and matched control patients. (a) Average number of SBDH keywords by age group at the time of AD diagnosis (<65, 65-70, 70-75, 75-80, 80-85, 85-90). (b) Average number of SBDH keywords prior to AD diagnosis by sex.

(a).



(b)

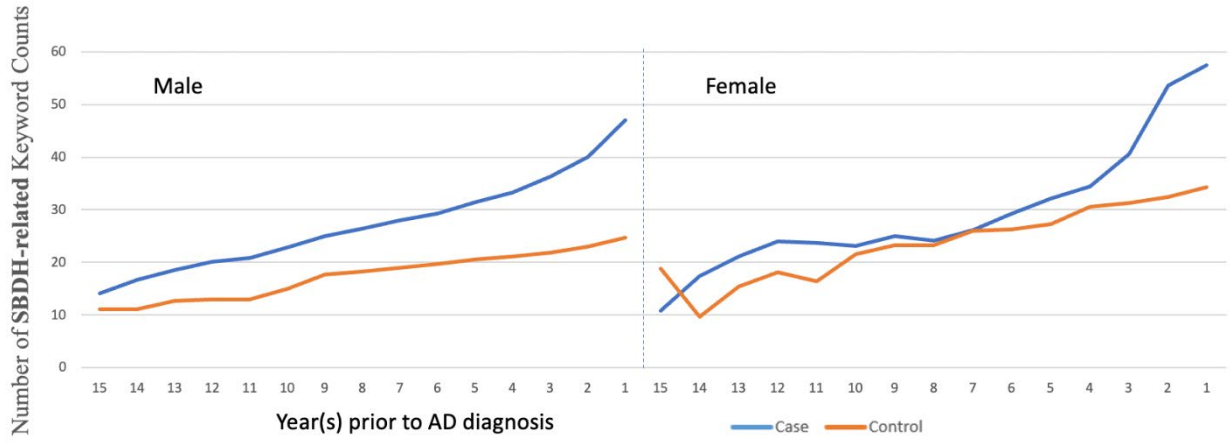


Figure 3. Comparison of the average number of SBDH keywords between AD case patients and matched control patients by ethnicity/race.

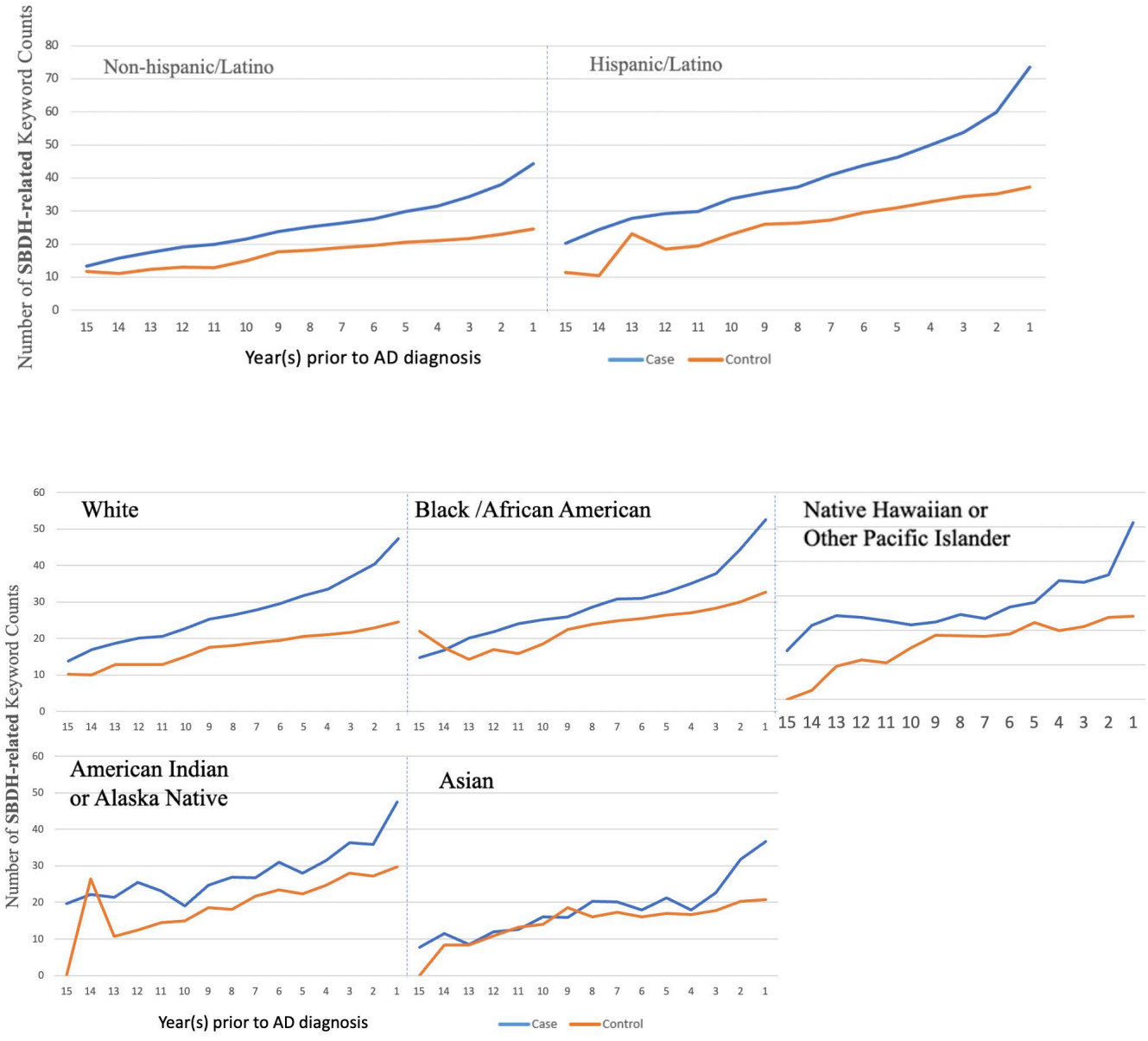
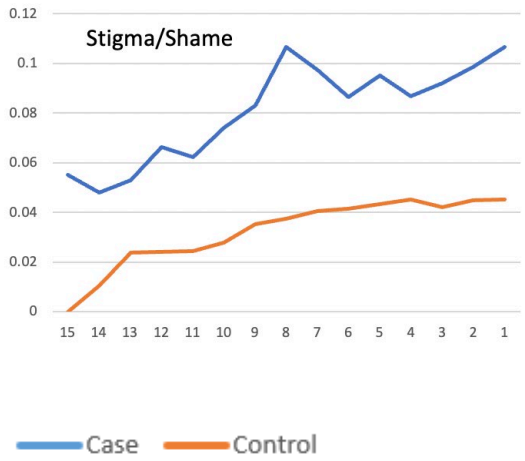


Figure 4. Comparison of the average number of SBDH keywords between AD case patients and matched control patients by SBDH domains.



X-axis: Year(s) prior to AD diagnosis

Y-axis: Number of SBDH-related Keyword Counts

Figure 5: Comparison of the average number of SBDH keywords between AD case patients and matched control patients by education levels.

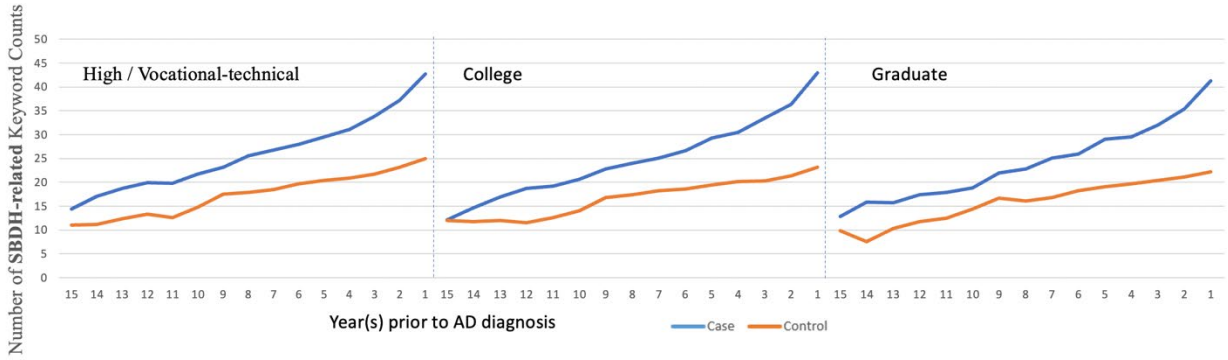


Table 1. Characteristics of the study cohort

Characteristic	**AD Case (n=16,701)**	**Control (n=39,097)**
Unique patients, No.	16,701	39,097
Age, Mean (SD), y	76.7 (8.6)	76.6 (8.9)
Sex/Gender, No. (%)		
Female	488 (2.9%)	1,013 (2.6%)
Male	16,213 (97.1%)	38,084 (97.4%)
Race/Ethnicity, No. (%)		
White	12,927 (77.4%)	31,256 (79.9%)
Black or African American	2,356 (14.1%)	4,024 (10.3%)
Asian	121 (0.7%)	260 (0.7%)
Native Hawaiian or Other Pacific Islander	168 (1.0%)	297 (0.8%)
American Indian or Alaska Native	86 (0.5%)	225 (0.6%)
Unknown	1,043 (6.2%)	3,035 (7.8%)
Hispanic or Latino	1,865 (11.2%)	1,759 (4.5%)

Characteristic	**AD Case (n=16,701)**	**Control (n=39,097)**
Patient No. with education level data	11718	29363
High school / Vocational-technical school	5395 (46.0%)	13881 (47.3%)
College	4534 (38.7%)	11175 (38.1%)
Graduate School	1789 (15.3%)	4307 (14.7%)
Charlson Comorbidity Index, Mean (SD)	3.74 (3.11)	2.92 (2.91)
	(Before AD onset)	(Across the period)
Number of notes with keywords (average)	75.0	47.5
Keyword occurrence number (average)	338.7	193.0