Original Research Article

Comparative Analysis of Dementia Risk Patterns in Community and Facility Settings: Evidence from NHATS Round 13

Gabriel Fang

Hunter College High School, 71 East 94th Street, NY 10128, United States

ABSTRACT

Dementia is a growing public health burden in the United States, yet little is known about how risk factor profiles differ across various living settings. Using Round 13 of the National Health and Aging Trends Study (NHATS), this study compared correlates of dementia among community-dwelling and facility-residing older adults. Bivariate analyses were used to establish the baseline associations; CART (Classification and Regression Trees), a nonparametric method, was used to detect complex interactions and compare risk patterns in different settings. In both community and facility settings, bivariate analyses showed physical activity to be consistently associated with lower dementia prevalence. However, socioeconomic and demographic factors, as well as comorbidity burden, were significant correlates of dementia only in community settings, whereas co-residing with a spouse or others showed stronger and more pronounced associations in facility settings. CART models were also used to visualize shared and distinct patterns of interplay among multiple risk factors. Vigorous physical activity was the first splitting variable across settings, but subsequent branches differed: community trees emphasized medical conditions, income, and types of favorite activities, whereas facility trees highlighted co-residence with a spouse or others, along with walking exercise as a further differentiating factor. Such differentiated risk profiles stress the value of setting-specific approaches to understanding dementia risk.

Keywords: CART (classification and regression trees); comorbidity; dementia; NHATS (National Health and Aging Trends Study); socioeconomic determinants

INTRODUCTION

Dementia is a progressive, incurable illness that results in severe decrements in cognition, function, and quality of life for people with dementia and their caregivers. According to the National Institutes of Health, dementia affects more than 6 million Americans and accounts for more than 100,000 deaths each year in the U.S. (1). Globally, over 55 million people are living with dementia, with the total societal cost being estimated to be $1.3 trillion as of 2019 (2). Dementia is also ranked as the seventh leading cause of death worldwide (2). Its substantial and growing global burden has been documented in large-scale prevalence analyses (3).

Dementia is a broad category encompassing multiple neurodegenerative diseases, with Alzheimer's disease being the most common subtype (4). Beyond biological processes, social determinants of health have been identified as relevant correlations of dementia in nationally representative studies (5). A recent study identified key demographic, behavioral, and health-related predictors of Alzheimer's disease using machine learning approaches. The authors examined risk factors

Corresponding author: Gabriel Fang, E-mail: gabrielfang2025@gmail.com.

Accepted February 12, 2026
https://doi.org/10.70251/HYJR2348.41687694

such as vascular comorbidities, osteoporosis, sex, and COVID-19 and their associations with Alzheimer's disease (6). However, this study aggregated the analytic sample into a single group, limiting insight into environment-specific differences in dementia risk factors. Health outcomes such as mortality and hospitalization rates among older adults differ substantially across community, residential-care, and nursing-facility settings (7). Harrison et al. (8) investigated social, functional, and medical characteristics of older adults living in home, residential-care, and nursing-facility settings to provide setting-specific insights into people with dementia. However, this study did not examine how dementia risk factors and their interactions differ across these settings.

To fill these gaps, this study examined Round 13 data from the National Health and Aging Trends Study (NHATS), a nationally representative sample of older U.S. adults, to identify which risk correlates are shared and which differ in community and facility settings (9). Bivariate analyses were used to establish baseline associations between each risk correlate and dementia and to compare the direction as well as the magnitude of these associations. Subsequently, CART (Classification and Regression Trees) modeling was applied to visualize factor interactions. The information derived from these analyses may guide customized interventions and care strategies that are responsive to the varied social and health-related backgrounds of these two residential settings.

METHODS AND MATERIALS

Data Source and Sample

Data were drawn from NHATS Round 13 (data collection period May 2023–April 2024). NHATS dataset is a nationally representative study of Medicare beneficiaries aged 65 years and older in the United States (9). NHATS is designed by the Johns Hopkins Bloomberg School of Public Health and the University of Michigan Institute for Social Research and is sponsored by the National Institute on Aging (NIA/NIH). In-person interviews are conducted by Westat, a nationally recognized survey research organization.

Variables and Measures

Dementia was defined by the NHATS variable hc13disescn9, which asks whether a doctor has ever told the sample person that they have dementia. Those responding 1 ("Yes") or 7 ("Previously reported") were considered to have dementia. The variable was coded as either having dementia or not having dementia. Of the 8,597 survey participants in Round 13 who were included in the initial sample, I excluded from the subsequent analyses 601 individuals with missing dementia status. The analytic sample was reduced to 7,996 participants, including 575 with dementia and 7,421 without dementia, at a ratio of about 12.9:1 in non-dementia vs. dementia cases.

With the NHATS variable r13dresid, participants were classified as being community-dwellers or facility-based residents (residential care or nursing home environments providing structured assistance with daily activities). In the final analytic sample of 7,996 individuals, there were 7,509 with r13dresid=1 (community), 417 with r13dresid=2 (residential care, not nursing home), and 70 participants with r13dresid=4 (nursing home). In this analysis, participants with r13dresid=2 or 4 were classified as facility-based residents.

Candidate risk factors of dementia were grouped into three conceptual domains to reflect medical, behavioral, and socioeconomic dimensions of risk:

(1) Comorbidity (Medical Burden): A comorbidity index was computed as the unweighted sum of medical conditions excluding dementia, including heart attack, heart disease, hypertension, arthritis, osteoporosis, diabetes, lung disease, stroke, and cancer. Higher scores indicate a greater medical burden.

(2) Physical Activity and Preferences: Physical activity measures included past-month vigorous activity (NHATS variable pa13vigoractv) as well as walking for exercise (pa13evrgowalk). Both serve as proxies for current physical activity engagement. Additionally, the favorite activity type (pa13dfavact) captured forms of activity that participants most preferred. This variable was not time-restricted (not just last month). It reflected broader lifestyle preferences. This domain is considered behavioral.

(3) Socioeconomic and Demographic Variables: Socioeconomic and demographic covariates included education level, marital status, sex, ethnicity, annual income ($\log_{10}$-transformed to correct right-skew), as well as living arrangements, such as living with a spouse (NHATS variable hh13livwthspo) or living with others (NHATS variable hh13dlvngarrg).

Analytic Approach and Statistical Methods

All statistical analyses were conducted using R (R Foundation for Statistical Computing, Vienna, Austria) with the rpart package (version 4.1.24) within RStudio (version 2025.05.1; Posit Software, PBC) (10–

12). Bivariate analyses were used to examine baseline associations between dementia and each candidate correlate. Fisher's exact test was used for categorical variables (e.g., physical activity, favorite activity, sex, education, marital status, ethnicity, co-residence with a spouse or others) to assess whether the odds of dementia differed across categories, given its robustness to small or uneven cell sizes. Income was treated as a continuous variable and $\log_{10}$-transformed to reduce right skewness and stabilize variance. Welch's two-sample t-test was applied to the $\log_{10}$-transformed income to accommodate unequal variances between groups. Bivariate logistic regression was used for the Comorbidity Index, modeled as a continuous variable to estimate the odds ratio (OR) for dementia per additional medical condition. Statistical significance was set at two-tailed $\alpha = 0.05$. For bivariate analyses, 95% CIs for estimates were calculated, if applicable, to measure precision. For each bivariate analysis, participants with missing values for the variable being analyzed were excluded from that specific analysis; no imputation methods were applied.

A CART model was also used to examine complex interactions among multiple risk correlations using the Gini impurity criterion. CART does not assume any functional form and uses a recursive-partitioning method that identifies informative splits. CART allows categorical as well as continuous variables to produce interpretable tree structures. Because the primary aim was descriptive characterization of dementia risk structures rather than prediction, no train and test split was performed. Trees were grown with permissive complexity settings and then subsequently pruned using the 1-standard-error (1-SE) rule. This rule selects the simplest tree whose cross-validated error is within one standard deviation of the minimum. For the community model, xerror, the cross-validated relative error reported by the *rpart* package in R, decreased sharply from 1.06 at the root node to approximately 0.70 by four splits, with minimal improvement thereafter (optimal CP ≈ 0.0065). For the facility model, xerror declined from approximately 1.10 at the root to 0.88 after three splits (optimal CP ≈ 0.0229). To reduce the impact of sampling imbalance, class weighting was used in the CART model based on the ratio of empirical dementia prevalence within each residential group. Covariates were added to the model according to variable type: income and comorbidity index as continuous variables, while other explanatory variables were categorical. In the CART models, categorical correlates with missing responses were retained as explicit "missing" (NA) categories, whereas continuous correlates with missing or invalid values were left as missing and handled internally by the rpart CART algorithm during tree construction, including the use of surrogate splits when available. No imputation was performed. Variable importance scores were calculated in order to determine which variables were most related to dementia risk.

RESULTS

Sample Characteristics

Of 8,597 Round 13 respondents, 601 were excluded for missing dementia status. The final analytic sample included 7,996 people in total (575 dementia cases and 7,421 non-dementia cases): there was approximately a 12.9:1 ratio of non-dementia cases to dementia cases. Of the 7,996 people, 7,509 (94.0%) were community-dwelling and 487 (6.0%) resided in facilities, resulting in a community-to-facility ratio of 15.42:1. The prevalence of dementia was 6.1% for the community-dwelling vs. 23.61% for the facility-dwelling population with dementia in the sample. In communities, the ratio of non-dementia to dementia cases was about 15.32. In facilities, the ratio of non-dementia to dementia cases was approximately 3.23.

Bivariate Analyses

Bivariate associations showed common as well as context-specific associations with dementia (Table 1). Both vigorous activity and walking for exercise were associated with lower dementia prevalence in both environments, with different magnitudes. In the community, there were 2.95 times higher odds of dementia (95% CI: 2.32–3.80; $p < 0.001$) for those who were not engaged in vigorous activity compared to those who were, whereas this OR was 2.77 (95% CI: 1.58–5.09; $p < 0.001$) among residents living in facilities. Walking exercise was associated with lower dementia odds and the association was marginally stronger in facilities (OR = 2.28; 95% CI: 1.46–3.57; $p < 0.01$) than in communities (OR = 1.99; 95% CI: 1.64–2.42; $p < 0.001$).

A gradient in dementia odds across multiple health, socioeconomic and demographic factors was observed at the community level. For each additional comorbidity index medical condition present, the odds of dementia increased by 40% (OR = 1.40; 95% CI: 1.31–1.49; $p < 0.001$). Less education (< high school) was associated with higher odds of dementia (OR = 2.48; 95% CI: 2.02–3.04; $p < 0.001$), as were non-active leisure preferences (OR = 2.33; 95% CI: 1.81–3.00; $p < 0.001$). There were

Table 1. Bivariate Associations Between Dementia and Candidate Correlates in Community and Facility Settings, With Cross-Setting Comparison

Factor	Community Environment	Facility Environment	Comparative Pattern Across Settings
Comorbidity index	Each additional point in the Comorbidity Index increases the odds of dementia by 40%, statistically significant. Odds Ratio (OR) = 1.40 (95% CI 1.31–1.49, p < 0.001)	Each additional point in the Comorbidity Index increases the odds of dementia by 3%, not statistically significant. Odds Ratio (OR) = 1.03 (95% CI 0.89–1.19, p = 0.68)	Only statistically significant in community
Vigorous Activity (No vs Yes)	No Vigorous Activity → 2.95 × higher odds (Fisher's exact OR 2.95, 95% CI 2.32–3.80; p < 0.001)	No Vigorous Activity → 2.77× higher odds (Fisher's exact OR 2.77, 95% CI 1.58–5.09; p = 0.0001)	Same protective effect, stronger in community
Walking Exercise (No vs Yes)	Walking protective: No walking → 1.99 × higher odds (Fisher's exact OR 1.99, 95% CI 1.64–2.42; p < 0.001)	Walking protective: No walking → 2.28× higher odds (Fisher's exact OR 2.28, 95% CI 1.46–3.57; p = 0.0002)	Same protective effect, stronger in facility.
Favorite activities (Non-active vs Active)	Non-active leisure→ 2.33 × higher odds compared to those whose favorite activity is active leisure (Fisher's exact OR 2.33, 95% CI 1.81–3.00; p < 0.001)	Non-active leisure→ 1.70 × higher odds compared to those whose favorite activity is active leisure (Fisher's exact OR 1.70, 95% CI 0.96–3.09; p = 0.0621)	Significant risk factor in community
Education (< HS vs HS+)	< HS → 2.48 × higher odds compared to HS+ (Fisher's exact OR 2.48, 95% CI 2.02–3.04; p < 0.001)	< HS → 1.13 × higher odds compared to HS+ (Fisher's exact OR 1.13, 95% CI 0.60–2.05; p = 0.6581)	Major risk in community
Marital Status (Widowed vs Married)	Widowed individuals had 1.98× higher odds of dementia compared to those who are married (Fisher's exact OR 1.98, 95% CI 1.36–2.90; p = 0.0003)	Widowed individuals had 1.81× higher odds (Fisher's exact OR 1.81, 95% CI 0.43–11.06; p = 0.5347)	Same direction, but effect only significant in community
Sex (Male vs Female)	Males →0.85 × odds vs females (Fisher's exact OR 0.85, 95% CI 0.70–1.03; p = 0.0985)	Males →0.66 × odds vs females (Fisher's exact OR 0.66, 95% CI 0.40–1.07; p = 0.0925)	Both non-significant.
Ethnicity (Hispanic vs White non-Hispanic)	Hispanic → 1.82× higher odds of having dementia compared to White (non-Hispanic) individuals (Fisher's exact OR 1.82, 95% CI 1.18–2.80; p = 0.0051).	Hispanic →2.51 times the odds of having dementia compared to White (non-Hispanic) individuals (Fisher's exact OR 2.51, 95% CI 0.46–11.85; p = 0.2271)	Same direction
Income	Mean income was lower in the dementia group ($22,000) compared to the no-dementia group ($40,200), and this difference was statistically significant (p < 0.001).	Mean income was lower in the dementia group ($32,700) compared to the no-dementia group ($40,400), but the difference was not statistically significant (p = 0.50).	Similar protective trend, significant only in community.
Living With Spouse	Not living with spouse → 3.02× higher odds (Fisher's exact OR 3.02, 95% CI 1.14–6.86; p = 0.0138).	Not living with spouse → 10.36x higher odds (Fisher's exact OR 10.36, 95% CI 3.53–32.53; p < 0.001)	Same direction, stronger in facility.
Living With Others	Not live with others → 0.64× odds (Fisher's exact OR 0.64, 95% CI 0.50–0.81; p < 0.001)	Not living with others → 2.22× higher odds (Fisher's exact OR 2.22, 95% CI 1.16–4.53; p = 0.0105)	Opposite directions, significant in both settings

Note. Odds ratios (ORs) were with 95% confidence intervals from bivariate analyses. ORs > 1 indicate higher odds of dementia for the first category listed; ORs < 1 indicate lower odds relative to the reference category. Statistical significance was p < 0.05. Conceptual domains: pink = medical burden; green = physical activity and activity preferences; yellow = socioeconomic or demographic factors; gray = not statistically significant. The "Comparative Pattern Across Settings" column summarizes cross-setting consistency in direction and significance.

higher odds among those who were widowed compared with those who were married (OR = 1.98; 95% CI: 1.36–2.90; $p < 0.01$) and among those with lower income, with a geometric mean income of approximately $22,000 among participants with dementia compared to $40,200 among those without dementia ($p < 0.001$). Hispanic ethnicity was associated with higher odds of dementia compared with White (non-Hispanic) ethnicity (OR = 1.82, 95% CI: 1.18–2.80; $p = 0.005$) in community settings.

Education and income were not statistically significant at the facility level, whereas co-residence with others or spouses was statistically significant. Living without a spouse was associated with odds of 10.36 for dementia (95% CI: 3.53–32.53; $p < 0.001$), and living alone was associated with odds of 2.22 (95% CI: 1.16–4.53; $p = 0.01$). These findings indicate a strong association between co-residence status and dementia prevalence in facility settings.

There was no significant association between dementia and sex in either setting.

Classification and Regression Trees (CART) Analysis

CART models showed that patterns of risk for dementia vary across settings (Figures 1 and 2). The prevalence of dementia in the community was 6.13%. People who were not actively engaged in vigorous physical activity, including those with missing activity data grouped with this category, had a higher dementia prevalence (8.17%). In this group, a comorbidity index of 4.5 or higher was associated with a higher prevalence of dementia (18.07%). Among participants with a comorbidity index below 4.5, income further stratified dementia prevalence. Participants with annual income less than $56,500 had a

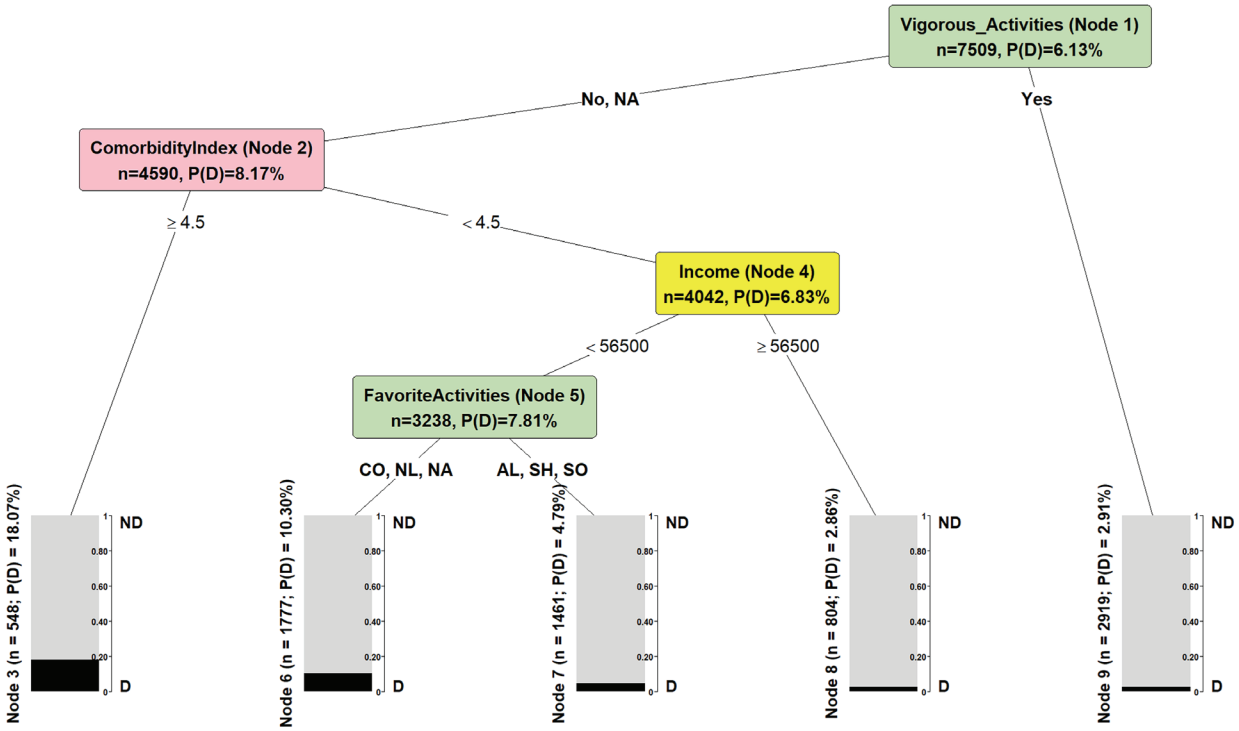

Figure 1. *Classification and Regression Tree (CART) model illustrating patterns of dementia risk among community-dwelling older adults in NHATS Round 13 (analytic sample n = 7,509; community residents only; participants with missing dementia status excluded). The tree shows splits based on vigorous physical activity, comorbidity burden, income, and favorite activity type (vigorous activity refers to self-reported vigorous physical activity; comorbidity burden represents the summed count of medical conditions excluding dementia; favorite activity indicates the participant's most preferred activity category). Each internal node reports the splitting variable, the number of individuals in that node (n), and the observed dementia prevalence [P(D)]. Terminal nodes display bar plots showing the proportion of participants with dementia (D) and without dementia (ND). Continuous variables were split at thresholds (comorbidity index ≥ 4.5; annual income < $56,500). Node colors represent conceptual domains: green = physical activity or activity preference; pink = medical burden; yellow = socioeconomic or demographic factors.*

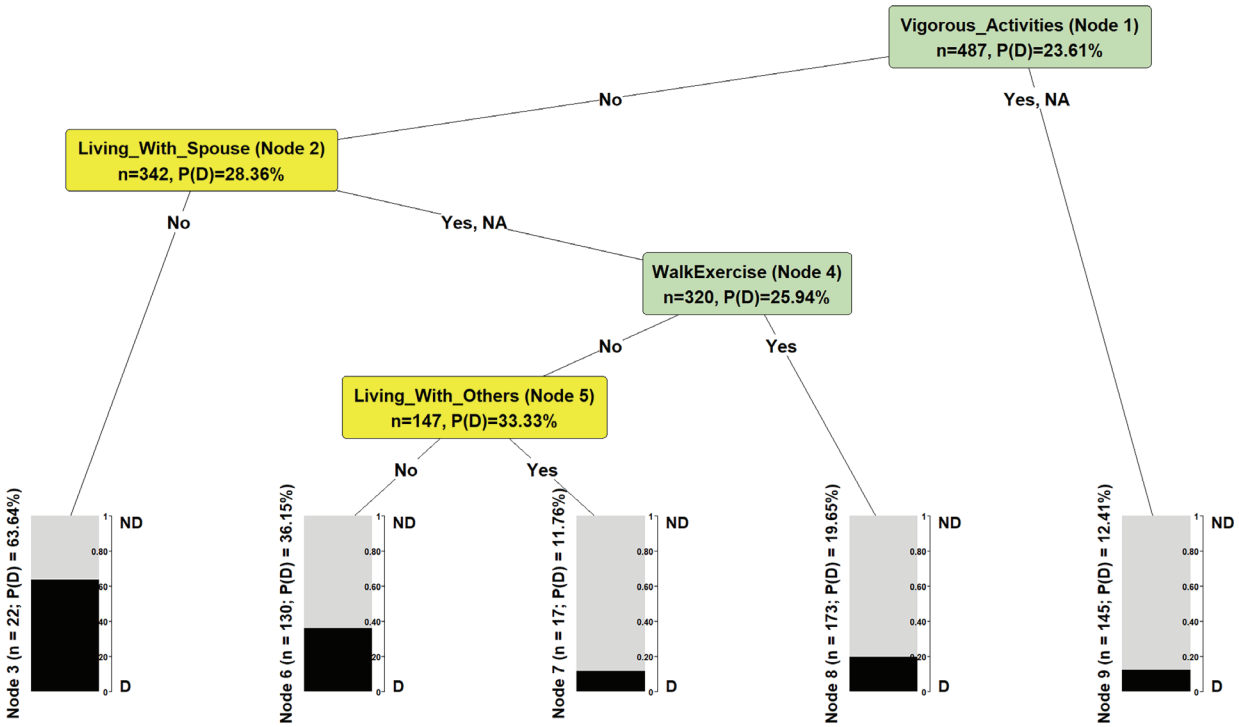

Figure 2. *Classification and Regression Tree (CART) model illustrating patterns of dementia risk among facility-dwelling older adults in NHATS Round 13 (analytic sample n = 487; facility residents only; participants with missing dementia status excluded). Facility-dwelling refers to participants residing in residential care or nursing home settings. The tree displays splits based on vigorous physical activity, co-residence with a spouse, walking exercise, and living with others (vigorous activity and walking exercise reflect self-reported physical activity engagement; co-residence variables indicate whether the participant lives with a spouse or with other individuals). Each internal node reports the splitting variable, the number of individuals in that node (n), and the observed dementia prevalence [P(D)]. Terminal nodes display bar plots showing the proportion of participants with dementia (D) and without dementia (ND). Node colors represent conceptual domains: green = physical activity; yellow = socioeconomic or demographic factors.*

dementia prevalence of 7.81%, whereas those with higher income had a prevalence of 2.86%. Among the lower-income group, favorite activity further differentiated the prevalence of dementia. Those who reported their favorite activity was socialization or active leisure had a lower prevalence of dementia (4.79%) compared with those who did not report such preferences (10.30%). The overall prevalence of dementia was lower when residents engaged in vigorous physical activity (2.91%).

In facility settings, vigorous physical activity was identified as the root node of the CART model, indicating its importance in dementia risk. Residents classified as engaging in vigorous physical activity, including those grouped with this category due to missing data, had a lower dementia prevalence of 12.41%, compared with the overall dementia prevalence of 23.61%. Among individuals who did not engage in vigorous physical activity, dementia prevalence varied by co-residence status. The highest prevalence of dementia in this subgroup was in those not living with a spouse (63.64%), indicating an association between spouse co-residence status and dementia prevalence. Among those living with a spouse, walking exercise was the next differentiating factor. In this subgroup, those who reported walking regularly had a lower prevalence of dementia of 19.65%, whereas those who reported not walking regularly had a prevalence of dementia of 33.33%. Living with others also emerged as a downstream stratifying factor in the group that did not engage in walking exercise. Overall, this facility model structure indicates that dementia prevalence is patterned by both physical activity and social connections.

Variable importance results from the CART models (Table 2) showed shared as well as environment-specific

Table 2. Comparison of Variable Importance in Community vs. Facility CART Models for Dementia Risk.

Comparison	Community Environment		Facility Environment	
Factor Importance Rank (Top 4)	**Factor**	**Relative CART Variable Importance (% of Total Impurity Reduction)**	**Factor**	**Relative CART Variable Importance (% of Total Impurity Reduction)**
1	Vigorous Activities	50.05%	Vigorous Activities	36.14%
2	Comorbidity Index	23.18%	Living With Spouse	31.26%
3	Income	13.63%	Walk Exercise	16.68%
4	Favorite Activities	12.25%	Living With Others	11.45%

factors contributing to dementia risk and quantified each factor's contribution to overall impurity reduction. Vigorous physical activity was the single most important factor in both community and facility environments. It contributed more to total impurity reduction in the community model (50.01%) than in the facility model (36.14%). The second-ranked factor varied by environment: the comorbidity index contributed 23.18% of total impurity reduction in the community model, whereas living with a spouse contributed 31.26% in the facility model. Income in the community environment (13.63%) and walking exercise in facilities (16.68%) were ranked next in the community and facility models, respectively. Favorite activities had a contribution to risk partitioning in the community model, while living with others was among the top-ranked factors in the facility model. Other variables, including education, sex, and ethnicity contributed less than 5% of total impurity reduction and did not substantially affect the structure of either tree. Variable importance indicates to what extent each factor contributed to risk partitioning in the CART models and should be interpreted together with tree structure and bivariate analyses rather than as an effect size.

DISCUSSION

In bivariate analyses, vigorous physical activity and walking exercise were strongly associated with dementia status in both settings. Vigorous physical activity showed greater salience in the community setting, whereas walking exercise played a more prominent role in facilities. In community settings, comorbidity, type of favorite activities, education, marital status, ethnicity, income, and co-residence all demonstrated substantial and statistically significant associations with dementia, together reflecting a multidimensional risk profile in the community. By comparison, medical burden, socioeconomic factors (education and income), and demographic factors such as sex were not statistically significant in facility settings. Relational factors (living with a spouse and living with others) were consistently associated with dementia across settings. Regarding CART results, engagement in vigorous physical activity was a shared and important factor in both settings. In the community environment, the downstream factors emphasized medical burden, income, and active forms of activities. In facility settings, the branches that came after the first split pointed to co-residence and walking exercise as key factors. CART results visualized these environment-specific risk patterns.

Through a combination of bivariate analyses and CART modeling, the findings address a gap in the literature by adopting a multivariable, interaction-oriented, environment-sensitive descriptive framework. These findings reflect the need for setting-specific dementia prevention strategies. Prevention at the community level should prioritize disease management along with economic and behavioral support, while prevention at the facility level should emphasize greater relational continuity as well as routine physical engagement.

CONCLUSION

Dementia risk varies by residential setting. This study provides evidence for environment-specific patterns of dementia correlates rather than a uniform risk structure across settings. Physical activity was a common correlation of lower dementia prevalence in both environments; however, dementia among community-dwelling older adults was additionally associated with medical burden and socioeconomic disadvantage, whereas dementia among facility-dwelling older adults

was more strongly associated with walking exercise and co-residence with a spouse or others. Using bivariate analyses and CART methods, this study showed that dementia risk is associated with multiple factors whose relative importance varies by setting. The results support context-specific prevention strategies, in which community interventions should prioritize medical and socioeconomic risks, while facility-based interventions should prioritize social engagement and ongoing physical activity. For future studies, longitudinal research that includes biological markers would help to test and extend the setting-specific structure of dementia risk.

REFERENCES

1. National Institutes of Health (NIH). Risk and future burden of dementia in the United States. Available from: https://www.nih.gov/news-events/nih-research-matters/risk-future-burden-dementia-united-states (accessed on 2026-02-09)
2. World Health Organization. Global Dementia Observatory (GDO). Available from: https://www.who.int/data/gho/data/themes/global-dementia-observatory-gdo (accessed on 2026-02-09)
3. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: A systematic review and meta-analysis. *Alzheimer's & Dementia.* 2013; 9 (1): 63-75. https://doi.org/10.1016/j.jalz.2012.11.007
4. Centers for Disease Control and Prevention. About dementia. Available from: https://www.cdc.gov/alzheimers-dementia/about/index.html (accessed on 2026-02-09)
5. 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. *Gerontology & Geriatric Medicine*. 2023; 9: 23337214231190244. https://doi.org/10.1177/23337214231190244
6. Ahmmad MR, Hossain E, Khan MTF, Paudel S. Using machine learning to identify risk factors for Alzheimer's disease among older adults in the United States: The role of chronic and behavioral health. *Journal of Alzheimer's Disease Reports*. 2025; 9: 25424823251377691. https://doi.org/10.1177/25424823251377691
7. Boucaud-Maitre D, Letenneur L, Dramé M, Taubé-Teguo N, *et al*. Comparison of mortality and hospitalizations of older adults living in residential care facilities versus nursing homes or the community: A systematic review. *PLOS ONE*. 2023; 18 (5): e0286527. https://doi.org/10.1371/journal.pone.0286527
8. Harrison KL, Ritchie CS, Patel K, Hunt LJ, *et al*. Care settings and clinical characteristics of older adults with moderately severe dementia. *Journal of the American Geriatrics Society*. 2019; 67 (9): 1907–1912. https://doi.org/10.1111/jgs.16054
9. Welcome Researchers | NHATS. National Health and Aging Trends Study. Available from: https://www.nhats.org/researcher (accessed on 2026-02-09)
10. R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2025. Available from: https://www.R-project.org/ (accessed on 2026-02-09).
11. Therneau TM, Atkinson B. *rpart: Recursive Partitioning and Regression Trees* (R package version 4.1.24). Available from: https://CRAN.R-project.org/package=rpart (accessed on 2026-02-09).
12. Posit Software, PBC. *RStudio: Integrated Development Environment for R* (Version 2025.05.1). Boston, MA: Posit Software, PBC; 2025. Available from: https://posit.co/download/rstudio-desktop/ (accessed on 2026-02-09).