

Original Article  
Psychiatry & Psychology



# Dementia Overdiagnosis in Younger, Higher Educated Individuals Based on MMSE Alone: Analysis Using Deep Learning Technology

Hye-Geum Kim ,<sup>1</sup> Dai-Seg Bai ,<sup>2</sup> Bon-Hoon Koo ,<sup>1</sup> Eun-Jin Cheon ,<sup>1</sup> Seokho Yun ,<sup>1</sup> So Hye Jo ,<sup>1</sup> and Byoungyoung Gu <sup>3</sup>

<sup>1</sup>Department of Psychiatry, Yeungnam University Medical Center, Yeungnam University College of Medicine, Daegu, Korea

<sup>2</sup>Department of Psychology, Yeungnam University Medical Center, Yeungnam University College of Medicine, Daegu, Korea

<sup>3</sup>KimCheon Medical Center, Gimcheon, Korea

OPEN ACCESS

Received: Jul 15, 2024

Accepted: Oct 7, 2024

Published online: Nov 8, 2024

Address for Correspondence:

Hye-Geum Kim, MD, PhD

Department of Psychiatry, Yeungnam University Medical Center, Yeungnam University College of Medicine, 170 Hyeonchung-ro, Nam-gu, Daegu 42415, Republic of Korea.  
Email: psykhg@yu.ac.kr

© 2025 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Hye-Geum Kim

<https://orcid.org/0000-0002-9677-7011>

Dai-Seg Bai

<https://orcid.org/0000-0002-6907-6201>

Bon-Hoon Koo

<https://orcid.org/0000-0001-9633-3835>

Eun-Jin Cheon

<https://orcid.org/0000-0002-6986-0768>

Seokho Yun

<https://orcid.org/0000-0002-2200-3494>

So Hye Jo

<https://orcid.org/0000-0001-9471-0713>

## ABSTRACT

**Background:** Dementia is a multifaceted disorder that affects cognitive function, necessitating accurate diagnosis for effective management and treatment. Although the Mini-Mental State Examination (MMSE) is widely used to assess cognitive impairment, its standalone efficacy is debated. This study examined the effectiveness of the MMSE alone versus in combination with other cognitive assessments in predicting dementia diagnosis, with the aim of refining the diagnostic accuracy for dementia.

**Methods:** A total of 2,863 participants with subjective cognitive complaints who underwent comprehensive neuropsychological assessments were included. We developed two random forest models: one using only the MMSE and another incorporating additional cognitive tests. These models were evaluated based on their accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC) on a 70:30 training-to-testing split.

**Results:** The MMSE-alone model predicted dementia with an accuracy of 86% and AUC of 0.872. The expanded model demonstrated increased accuracy (88%) and an AUC of 0.934. Notably, 17.46% of the cases were reclassified from dementia to non-dementia category upon including additional tests. Higher educational level and younger age were associated with these shifts.

**Conclusion:** The findings suggest that although the MMSE is a valuable screening tool, it should not be used in isolation to determine dementia severity. The addition of diverse cognitive assessments can significantly enhance diagnostic precision, particularly in younger and more educated populations. Future diagnostic protocols should integrate multifaceted cognitive evaluations to reflect the complexity of dementia accurately.

**Keywords:** Dementia; MMSE; Neuropsychology

## INTRODUCTION

Dementia is a debilitating condition that significantly affects cognitive function in millions of individuals worldwide. Accurate early diagnosis is crucial for effective management and

Byoungyoung Gu   
<https://orcid.org/0009-0004-5212-0570>

### Funding

This study was supported by a grant from the Chunma Medical Research Foundation, Korea (2023).

### Disclosure

The authors have no potential conflicts of interest to disclose.

### Author Contributions

Conceptualization: Kim HG. Data curation: Kim HG, Bai DS, Gu B. Formal analysis: Kim HG, Gu B. Funding acquisition: Kim HG. Investigation: Kim HG, Gu B. Methodology: Kim HG, Gu B. Project administration: Kim HG. Resources: Kim HG, Koo BH, Cheon EJ, Yun S, Jo S. Supervision: Kim HG, Koo BH, Cheon EJ, Yun S, Jo S. Validation: Kim HG. Visualization: Kim HG, Gu B. Writing - original draft: Kim HG. Writing - review & editing: Kim HG.

treatment of dementia, potentially slowing its progression and improving the quality of life of patients.<sup>1</sup>

The Mini-Mental State Examination (MMSE) is one of the primary tools used to assess cognitive impairment associated with dementia.<sup>2</sup> However, relying solely on MMSE scores may not capture the complexity and nuances of the disease, necessitating the exploration of additional cognitive tests that could enhance diagnostic accuracy. Research indicates that integrating multiple cognitive and functional assessments can provide a more comprehensive view of a patient's cognitive health, potentially leading to a more precise diagnosis and better tailored treatment plans.<sup>3</sup>

Jak et al.<sup>4</sup> proposed that comprehensive neuropsychological (NP) criteria provide an ideal balance of sensitivity and reliability in diagnosing cognitive impairment. Their findings showed that around one-third of individuals initially diagnosed with mild cognitive impairment (MCI) using standard criteria were actually cognitively normal (CN) when reassessed using NP criteria. Additionally, these re-diagnosed CN individuals displayed various characteristics, such as imaging results, genetic biomarkers, and pathological findings, that were more aligned with CN profiles rather than those with MCI.<sup>5</sup>

This study aimed to evaluate the efficacy of the MMSE alone versus a combination of the MMSE and other comprehensive NP tests in predicting dementia. We employed machine learning techniques to develop predictive models that distinguished between dementia and non-dementia scores. In this study, we sought to answer the following questions. Can a model based solely on MMSE scores adequately predict dementia? Furthermore, does the incorporation of additional cognitive assessments enhance the predictive accuracy of these models? The findings of this study provide insights into the comparative effectiveness of these approaches, thereby guiding clinicians in making informed decisions regarding diagnostic strategies.

## METHODS

### Data collection

All participants were outpatients or inpatients with subjective cognitive complaints who visited a local university hospital for neuropsychiatric evaluation between January 2017 and December 2023. The dataset comprises demographic information such as age and years of education, along with scores from NP assessment including MMSE, Clinical Dementia Rating (CDR), Seoul Neuropsychological Screening Battery (SNSB), and daily living capability assessments such as basic activities of daily living (ADL) and Korean Instrumental ADL (Table 1). Participants were excluded if they had incomplete or unreliable test data, which could compromise the validity of the analysis. Additionally, individuals with neuropsychiatric conditions, such as major depressive disorder or schizophrenia, which could potentially confound the cognitive assessments, were excluded.

**Table 1.** Demographic data

Demographic Variable	Overall participants	Non-dementia	Dementia
No. of participants	2,864	672	2,192
Age, yr, mean $\pm$ SD	70 $\pm$ 10	68 $\pm$ 9	72 $\pm$ 11
Sex (% male)	55%	50%	60%
Education, yr, mean $\pm$ SD	14 $\pm$ 3	15 $\pm$ 2	13 $\pm$ 4

SD = standard deviation.

SNSB is a comprehensive tool used to assess five key cognitive domains: memory, language, attention, visuospatial abilities, and executive (frontal) functions. Memory is evaluated through the Seoul Verbal Learning Test-Elderly, which measures immediate and delayed recall, recognition abilities, as well as similar tasks using the Rey Complex Figure Test (RCFT). Language function is assessed by examining spontaneous speech, comprehension, repetition, reading, writing, and naming ability through the Korean-Boston Naming Test (K-BNT). Attention is measured using the Vigilance Test, Digit Span Test, and Letter Cancellation tasks. Visuospatial abilities are evaluated by the RCFT, focusing on copying accuracy and time, and also through the Clock Drawing Test. Executive functions are tested with a series of tasks, including the Contrasting Program, Go-No-Go Test, Fist-Edge-Palm, alternating hand movements, the Luria loop, Controlled Oral Word Association Test, Korean-Color Word Stroop Test, Digit Symbol Coding, and the Korean-Trail Making Test-Elderly. The results are analyzed using z-scores adjusted for age, sex, and education level.

The dementia diagnosis for each patient was based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and CDR scores.<sup>6</sup> Patients were considered to have dementia if they met the diagnostic criteria for a major neurocognitive disorder with DSM-5 and CDR scores  $\geq 1$ .

### Data preprocessing

In the preprocessing phase, missing values for diagnosis were excluded to ensure the integrity of the analysis. The remaining missing values across other variables were imputed with the median value of each column to maintain consistency across the data. The dataset was divided into training (70%) and testing (30%) sets to facilitate robust model training and unbiased evaluation.

### Model development

Two predictive models were developed using the random forest classifier, which is a decision-tree-based ensemble machine learning algorithm known for its robustness and efficacy in handling biomedical data. The first model was based on a single variable, MMSE total score, to establish baseline performance. Subsequently, a composite model incorporating additional cognitive test scores was developed to explore the impact of integrating multiple diagnostic tests on the predictive accuracy. Feature importance was evaluated based on the Gini impurity decrease caused by each feature, providing insights into which tests were the most influential in predicting dementia (Table 2).

**Table 2.** Feature importances

Feature	Importance
MMSE_total_score	0.177208378
IADL_Score	0.127844977
B_ADL	0.047550685
Education_years	0.04272619
Age	0.041538817
Digit_span_Backward_z	0.017330207
RCFT_immediate_recall_z	0.017176673
SVLT_recognition_discriminability_index_z	0.017085217
RCFT_recognition_score_z	0.016932898
COWAT_siut_z	0.016238523

MMSE = Mini-Mental State Examination, IADL = instrumental activities of daily living, RCFT = Rey Complex Figure Test, SVLT = Seoul Verbal Learning Test, COWAT = Controlled Oral Word Association Test.

### Model evaluation

Model performance was assessed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC). The ROC curve provides a comprehensive measure of model performance at various threshold settings, with the AUC representing the probability that the model ranks a randomly chosen positive instance more highly than a randomly chosen negative instance. Higher AUC values indicate better diagnostic capabilities of the model.

### Statistical analysis

All statistical analyses were conducted using Python, specifically leveraging libraries such as scikit-learn for machine learning tasks and pandas for data manipulation. Statistical significance was set at  $P < 0.05$ . (Python 3.8.5). The methods used for data preprocessing and model development were implemented using scikit-learn and pandas, and references have been added to ensure reproducibility.<sup>7</sup>

### Ethics statement

This study was approved by the Institutional Review Board (IRB) of Yeungnam University Medical Center (IRB No. YUMC 2021-06-039), and the requirement for informed consent was waived owing to the retrospective design of the study.

## RESULTS

### Diagnostic accuracy of the MMSE alone versus combined cognitive tests

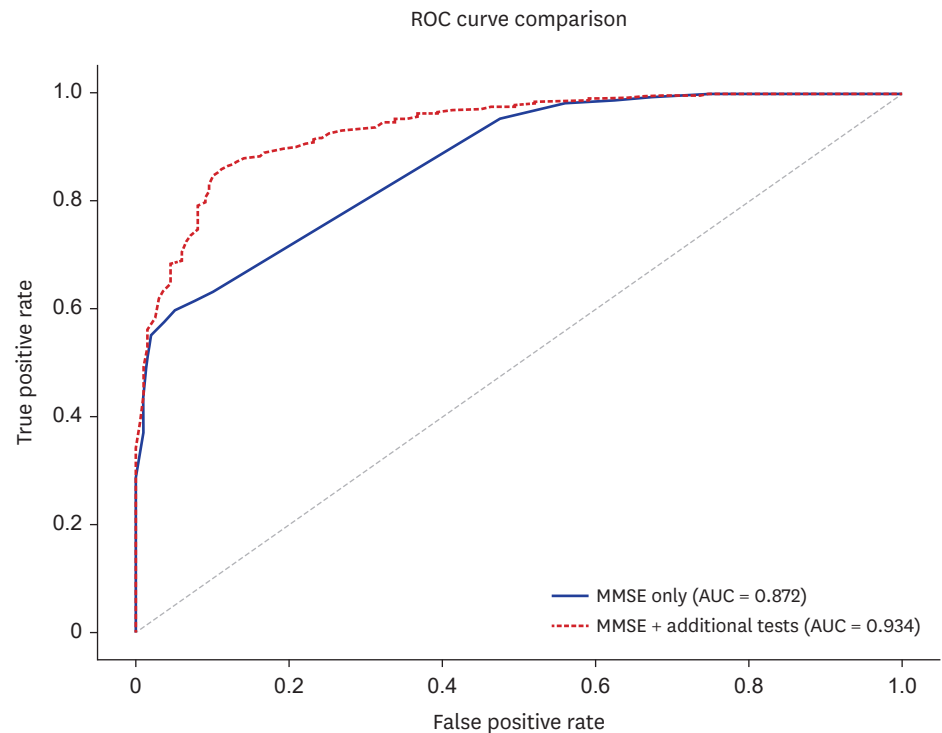
The primary objective was to compare the efficacy of the MMSE alone against a combination of the MMSE and additional cognitive tests in predicting CDR categories. The model utilizing only the MMSE showed an accuracy of 86%, with a precision of 85% for identifying dementia and a recall of 98% for the same category. This model achieved an AUC of 0.872 (Fig. 1).

In contrast, the composite model incorporating additional cognitive assessments demonstrated improved diagnostic accuracy, with an overall accuracy of 88%, increased precision for the dementia class to 90%, and a slightly reduced recall of 95%. The AUC for this model was significantly high (0.934), indicating a more precise diagnostic capability (Table 3).

### Analysis of cases with changed predictions

A critical aspect of this study was the analysis of cases in which the predictive outcome changed from dementia using the MMSE alone to non-dementia when additional tests were included. The data revealed that 150 cases, representing 17.46% of the total number of cases tested, shifted from the dementia to non-dementia category when evaluated using the comprehensive model. This shift was statistically significant, with a  $\chi^2$  test yielding a  $P = 0.01$ , underscoring the impact of incorporating a broader range of cognitive assessments into the diagnostic process.

These results highlight the potential for overdiagnosis of dementia when relying solely on the MMSE and underscore the benefits of a multifaceted diagnostic approach in reflecting true cognitive impairment more accurately.



**Fig. 1.** ROC curve and area under the curve values of the trained random forest classifier. ROC = receiver operating characteristic, AUC = area under the ROC curve.

**Table 3.** Model performance

Metric	MMSE only	MMSE + additional tests
Accuracy	86%	88%
Precision - Class 0 (non-dementia)	87%	79%
Precision - Class 1 (dementia)	85%	90%
Recall - Class 0 (non-dementia)	44%	66%
Recall - Class 1 (dementia)	98%	95%
F1-Score - Class 0 (non-dementia)	58%	72%
F1-Score - Class 1 (dementia)	91%	92%
AUC	0.872	0.934

MMSE = Mini-Mental State Examination, AUC = area under the ROC curve.

### Characteristics of the group with changed diagnostic predictions

Upon further analysis of the 150 cases in which dementia prediction changed from dementia to non-dementia, several distinctive patterns emerged. The average age of this subgroup was notably lower than that of individuals whose dementia classification remained consistent, suggesting that younger patients may be more prone to overdiagnosis when evaluated using only the MMSE. Additionally, this subgroup exhibited significantly higher scores than their initial MMSE scores on tests measuring executive function and memory recall (Table 4).

**Table 4.** Characteristics and statistical analysis of the group with changed predictions

Characteristic	Group with changed predictions	Control group	Statistical test	P value
Average age, yr	65	72	<i>t</i> -test	0.003
Education level, yr	16	12	<i>t</i> -test	0.010
Average MMSE score	24	26	<i>t</i> -test	0.050
Average IADL score	0.5	0.2	<i>t</i> -test	0.002
Gender distribution (%)	40% male	50% male	$\chi^2$	0.050

MMSE = Mini-Mental State Exam, IADL = instrumental activities of daily living.

Statistical analysis revealed that educational level significantly influenced the prediction changes. Individuals with higher educational levels were more likely to be overclassified using the MMSE alone, which was corrected when additional cognitive assessments were considered. A logistic regression model indicated that a combination of age, education, and specific cognitive test scores (such as Digit Span Backward and IADL scores) was a strong predictor of a change in diagnosis.

These findings highlight the need for a nuanced approach to dementia screening, particularly in younger and more educated populations, where the MMSE alone may not accurately reflect cognitive impairment.

## DISCUSSION

The findings of this study highlight the intricacies of dementia diagnosis and the potential limitations of relying solely on the MMSE to determine the severity of cognitive impairment. Our investigation revealed a substantive discrepancy when additional cognitive assessments were incorporated, as evidenced by 17.46% of the cases in which the diagnostic predictions shifted from the dementia category to non-dementia. The significance of this shift, statistically validated through a  $\chi^2$  test with  $P = 0.01$ , underscores the multifaceted nature of dementia and the inadequacy of a singular approach to its diagnosis.

The MMSE has long been the cornerstone of cognitive impairment screening. However, our results align with emerging evidence suggesting that the MMSE, although valuable, is not exhaustive in its diagnostic capability, particularly for patients with certain demographic characteristics. Younger, more educated individuals in our study cohort were more susceptible to overdiagnosis when evaluated using the MMSE alone. The MMSE is influenced by non-cognitive factors, such as language proficiency, literacy, and cultural background, which can lead to false positives or negatives in diverse populations. It is particularly less effective in accurately diagnosing dementia in individuals from different cultural and educational backgrounds.<sup>8</sup> This overdiagnosis is concerning as it may lead to unnecessary stress for patients and families, inappropriate allocation of resources, and potential delays in addressing the needs of those with different or more subtle forms of cognitive impairment.

Our study further demonstrated that executive function and memory recall assessments are pivotal for reclassifying patients into non-dementia categories. Many studies have reported that memory tests are the best predictors of future cognitive impairment and are useful when combined with executive function tests or language tests.<sup>9-14</sup> These cognitive domains are not adequately captured by the MMSE, which may account for the initial overestimation of dementia severity. Consequently, the inclusion of a broader battery of cognitive tests appears to correct this bias, allowing for a more nuanced and individualized assessment.

Our decision to employ a deep learning-based model, as opposed to traditional statistical methods, was driven by the need to accurately capture the complex, non-linear relationships inherent in cognitive assessment data. The superior performance of our model, particularly in terms of predictive accuracy, underscores the advantages of deep learning in this context. Deep learning models are uniquely suited to handle the large and intricate datasets used in this study, allowing for more nuanced predictions. This methodological choice reflects the



growing recognition of deep learning's potential to advance the field of cognitive diagnostics, providing clinicians with tools that are both powerful and precise.

The enhanced diagnostic model incorporating additional tests provided greater sensitivity and specificity, as reflected by a higher AUC value of 0.934. This improvement is not merely in statistical terms but also a clinically relevant enhancement that may inform the development of refined screening protocols, with implications for the timely and accurate identification of dementia.

Educational level emerged as a particularly salient factor influencing diagnostic predictions. This finding highlights the need to carefully consider sociodemographic factors during dementia screening and diagnosis. It suggests that individual patient characteristics, including education, play a central role in the interpretation of cognitive test results. The risk of misclassification due to these factors could potentially be mitigated through the use of adjusted scoring systems or the adoption of additional tests that are more sensitive to cognitive changes associated with different educational backgrounds.<sup>15</sup>

This study had several limitations. The study was conducted at a single university hospital, which limits its generalizability to broader populations. The retrospective design introduced potential biases related to data collection and accuracy. The follow-up period may have been insufficient for capturing long-term cognitive changes. Imputation of missing data may have affected the robustness of the results. Reliance on the DSM-5 and CDR may not have captured all the nuances of cognitive impairment. Despite these limitations, this study provides valuable insights into the diagnostic process of dementia, highlighting the inadequacy of relying solely on the MMSE, especially in highly educated younger populations.

In summary, our findings advocate for a more comprehensive and tailored approach for the cognitive assessment of dementia. They question the sufficiency of the MMSE as a standalone diagnostic tool, especially in younger and more educated populations. The study also highlights the need for further research to establish standardized protocols that incorporate diverse cognitive assessments. Future studies should aim to replicate and expand upon our findings in broader populations and settings, as well as explore the longitudinal impact of utilizing more extensive diagnostic assessments on patient outcomes.

## REFERENCES

1. Rosas AG, Stögmänn E, Lehrner J. Neuropsychological prediction of dementia using the neuropsychological test battery Vienna: a retrospective study. *Brain Disorders* 2022;5:100028. [CROSSREF](#)
2. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12(3):189-98. [PUBMED](#) | [CROSSREF](#)
3. Arevalo-Rodriguez I, Smailagic N, Roqué-Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, et al. Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 2021;7(7):CD010783. [PUBMED](#)
4. Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, et al. Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am J Geriatr Psychiatry* 2009;17(5):368-75. [PUBMED](#) | [CROSSREF](#)
5. Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis* 2014;42(1):275-89. [PUBMED](#) | [CROSSREF](#)

6. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: classification and criteria changes. *World Psychiatry* 2013;12(2):92-8. [PUBMED](#) | [CROSSREF](#)
7. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12(85):2825-30.
8. Devenney E, Hodges JR. The Mini-Mental State Examination: pitfalls and limitations. *Pract Neurol* 2017;17(1):79-80. [PUBMED](#) | [CROSSREF](#)
9. Rabin LA, Paré N, Saykin AJ, Brown MJ, Wishart HA, Flashman LA, et al. Differential memory test sensitivity for diagnosing amnesic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2009;16(3):357-76. [PUBMED](#) | [CROSSREF](#)
10. Eckerström C, Olsson E, Bjerke M, Malmgren H, Edman A, Wallin A, et al. A combination of neuropsychological, neuroimaging, and cerebrospinal fluid markers predicts conversion from mild cognitive impairment to dementia. *J Alzheimers Dis* 2013;36(3):421-31. [PUBMED](#) | [CROSSREF](#)
11. Galton CJ, Erzinçlioglu S, Sahakian BJ, Antoun N, Hodges JR. A comparison of the Addenbrooke's Cognitive Examination (ACE), conventional neuropsychological assessment, and simple MRI-based medial temporal lobe evaluation in the early diagnosis of Alzheimer's disease. *Cogn Behav Neurol* 2005;18(3):144-50. [PUBMED](#) | [CROSSREF](#)
12. Didic M, Felician O, Barbeau EJ, Mancini J, Latger-Florence C, Tramon E, et al. Impaired visual recognition memory predicts Alzheimer's disease in amnesic mild cognitive impairment. *Dement Geriatr Cogn Disord* 2013;35(5-6):291-9. [PUBMED](#) | [CROSSREF](#)
13. Guarch J, Marcos T, Salamero M, Gastó C, Blesa R. Mild cognitive impairment: a risk indicator of later dementia, or a preclinical phase of the disease? *Int J Geriatr Psychiatry* 2008;23(3):257-65. [PUBMED](#) | [CROSSREF](#)
14. Albert MS, Moss MB, Tanzi R, Jones K. Preclinical prediction of AD using neuropsychological tests. *J Int Neuropsychol Soc* 2001;7(5):631-9. [PUBMED](#) | [CROSSREF](#)
15. Spering CC, Hobson V, Lucas JA, Menon CV, Hall JR, O'Bryant SE. Diagnostic accuracy of the MMSE in detecting probable and possible Alzheimer's disease in ethnically diverse highly educated individuals: an analysis of the NACC database. *J Gerontol A Biol Sci Med Sci* 2012;67(8):890-6. [PUBMED](#) | [CROSSREF](#)