Alzheimer's & Dementia: Translational Research & Clinical Interventions 5 (2019) 264-271

Alzheimer's & Dementia

Featured Article

Using data science to diagnose and characterize heterogeneity of Alzheimer's disease

Ting F. A. Ang[a,b,c,1], Ning An[d,1], Huitong Ding[a,d], Sherral Devine[a,c], Sanford H. Auerbach[c,e], Joseph Massaro[c,f], Prajakta Joshi[a,c], Xue Liu[a,c], Yulin Liu[a,c], Elizabeth Mahon[a,c], Rhoda Au[a,b,c,e,*], Honghuang Lin[c,g,**]

[a]*Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA*
[b]*Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA*
[c]*The Framingham Heart Study, Framingham, MA, USA*
[d]*School of Computer and Information, Hefei University of Technology, Hefei, China*
[e]*Department of Neurology, Boston University School of Medicine, Boston, MA, USA*
[f]*Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*
[g]*Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA*

Abstract

Introduction: Despite the availability of age- and education-adjusted standardized scores for most neuropsychological tests, there is a lack of objective rules in how to interpret multiple concurrent neuropsychological test scores that characterize the heterogeneity of Alzheimer's disease.

Methods: Using neuropsychological test scores of 2091 participants from the Framingham Heart Study, we devised an automated algorithm that follows general diagnostic criteria and explores the heterogeneity of Alzheimer's disease.

Results: We developed a series of stepwise diagnosis rules that evaluate information from multiple neuropsychological tests to produce an intuitive and objective Alzheimer's disease dementia diagnosis with more than 80% accuracy.

Discussion: A data-driven stepwise diagnosis system is useful for diagnosis of Alzheimer's disease from neuropsychological tests. It demonstrated better performance than the traditional dichotomization of individuals' performance into satisfactory and unsatisfactory outcomes, making it more reflective of dementia as a spectrum disorder. This algorithm can be applied to both within clinic and outside-of-clinic settings.

Keywords: Alzheimer's disease; Dementia; Machine-learning; Decision tree; Neuropsychological assessment; Dementia screening

1. Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder, which makes up more than 60% of all dementia cases [1,2]. With a rapidly aging population, projected number of cases will triple by 2050 [3]. Cognitive decline is a key symptom of AD, and neuropsychological (NP) tests are widely used to assess varying degree of cognitive dysfunction, especially those affecting attention, memory, and executive functions [4,5]. Although cognitive impairment is a sine qua non criterion in AD diagnosis, the complexity of NP test data poses a challenge for consistent and accurate interpretation and the number of experts available to do so are limited, particularly in non-Western countries. Further, clinical AD trial studies have largely failed, partly due to the presumption of a more

The authors have declared that no conflict of interest exists.

[1]These authors contributed equally to this work.

*Corresponding author. Tel.: +508-935-3422; Fax: +1 617-358-5677.
**Corresponding authors. Tel.: +617-358-0091; Fax: +617-358-5677.
E-mail address: rhodaau@bu.edu (R.A.), hhlin@bu.edu (H.L.)

https://doi.org/10.1016/j.trci.2019.05.002

homogeneous clinical progression [6]. Therefore, variability in NP profiles based on different AD risk factors and its implication warrant further investigation.

Currently, age- and education-adjusted standardized norms are available to serve as a reference for individual NP test scores [7,8], but performance variability does not lend itself easily to a set of decision rules, nor are discrete cutoff values generalizable across all influencing factors. For example, although the Mini-Mental State Examination has well-established threshold scores [9–11], the score can vary significantly among people based on education and/or age [12]; floor and ceiling effects are additional important limitations. The Mini-Mental State Examination is also insensitive in detecting cognitive abnormalities during the earliest stages of AD [13–15], and its cutoff values rely on a single total score, making it difficult to determine the cognitive etiology of poor performance and the subtype of dementia.

Another challenge in AD diagnosis is the evaluation and interpretation of NP test results. Deciding cognitive status based on NP performance is clinician-subjective. Most conventional analyses also assume linear correlation between cognitive diagnoses and a single test, which is not reflective of dementia as a spectrum disorder. Therefore, there is clinical utility in applying new analytic approaches that can assess cognitive performance objectively across its multiple dimensions.

Machine learning techniques can readily derive information from complex data such as NP scores and uncover new knowledge to predict disease outcomes and improve the clinical decision-making process. Decision tree is one of the most widely used machine learning methods that involves breaking up a complex diagnostic process into a series of simpler rules, eventually leading to a multistage decision-making algorithm, and overcoming the knowledge bottleneck imposed by human experts [16]. It has been applied to a broad range of tasks from credit risk assessment to medical diagnosis [17,18]. Although the receiver-operating characteristic curve is often used to determine cutoff values of medical measures [19], isolation of a single cutoff value for a given NP test may compromise the overall accuracy of the prediction model. As an alternative, a decision tree can use multilevel cutoff values determined via discretization technology, which could enhance overall prediction accuracy for complex diseases such as AD. It is also important to consider the relevance of various NP tests in the diagnostic process as it has been widely accepted that certain tests are more sensitive in detecting cognitive decline than others [20,21]. Given the heterogeneity of AD, cognitive impairment may affect different cognitive domains for different subpopulations. It is, thus, important to use feature selection techniques to distinguish subsets of NP measures that are predictive of AD based on different demographic and AD risk factors.

Leveraging the rich collection of NP tests available at the Framingham Heart Study (FHS), the objective of this study is two-fold: (1) identify the most informative NP tests and (2) build a multilevel diagnostic decision tree to systematically screen for dementia.

2. Methods

2.1. Study population

The FHS is a longitudinal community-based multigenerational observational study initiated in 1948. In 1976, the FHS started cognitive screening of a subset of Original participants and subsequently extended it to all participants in all cohorts. Details of the dementia surveillance have been previously reported [22–24]. Given that sporadic AD is a disease that primarily affects individuals of advanced age, only participants from the Original, Offspring, Omni 1, and New Offspring Spouse cohorts [25], aged 70 years and older, were included in our study sample [26,27].

Thirty-two tests comprise the NP test protocol. Given that some tests were administered only to a subset of participants, the current analysis focused on 11 tests that were conducted on more than 85% of all participants. These tests included Wechsler Memory Scale Logical Memory [28]—Immediate Recall (LMi), Delayed Recall (LMd), and Recognition; Visual Reproductions[28]—Immediate Recall (VRi), Delayed Recall (VRd), and Recognition; Paired Associate Learning[28]—Immediate Recall (PASi), and the Similarities Test from the Wechsler Adult Intelligence Scale [29]. Additional tests included the Boston Naming Test (30-item Even Version; BNT30) and recall of hard scores from PAS immediate (PASi_h) and delayed conditions (PASd_h), measures of confrontational word retrieval and verbal learning, respectively [28,30].

FHS dementia diagnosis is based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [31], and AD diagnosis met criteria as specified by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association [32]. Dementia diagnosis is evaluated and verified through an adjudication panel, which includes at least one neuropsychologist and one neurologist and has been previously described [33]. Each NP assessment was assigned to one of these three outcomes: healthy control, AD, and non-Alzheimer's dementia (NAD). Refer to Supplementary Fig. 1 for sample selection flowchart.

2.2. Decision tree for dementia diagnosis

We implemented a supervised machine learning approach to recognize dementia diagnosis—both AD and NAD—from the NP assessment perspective. Our key approach is a Chi-square Automatic Interaction Detection decision tree [34], which identifies a series of diagnostics rules and arranges them in a tree-like manner in order of importance. Starting from the root node—the topmost decision step—Chi-square Automatic Interaction Detection adopts a top-down approach to select the optimal NP test that is most

Table 1
Demographics, NP test scores, and APOE genotypes of the studied population

Characteristics	Healthy control (n = 3514)	Alzheimer's disease (n = 555)	Non-Alzheimer's dementia (n = 443)
Age at NP examination			
Mean (SD)	79 (6)	85 (6)	84 (6)
Range	70–101	70–103	70–97
Male, no. (%)	1521 (43.3)	179 (32.3)	220 (49.7)
Highest level of education attained			
Valid education data, no. (%)	3510 (99.9)	549 (98.9)	442 (99.8)
High school and below, no. (%)*	1491 (42.5)	358 (65.2)	241 (54.5)
Beyond high school, no. (%)*	2019 (57.5)	191 (34.8)	201 (45.5)
APOE ε4 allele			
Valid genetic data, no. (%)†	3369 (95.9)	530 (95.5)	413 (93.2)
APOE ε4 (−), no. (%)*	2794 (82.9)	346 (65.3)	327 (79.2)
APOE ε4 (+), no. (%)*	575 (17.1)	184 (34.7)	86 (20.8)
NP test scores, mean (SD)			
LMi	11.2 (3.7)	4.8 (3.8)	7.9 (3.9)
LMd	10.2 (3.9)	3.0 (4.0)	6.5 (4.1)
LMr	9.4 (1.4)	7.1 (2.3)	8.5 (1.7)
VRi	7.1 (3.0)	3.1 (2.3)	4.0 (2.5)
VRd	6.1 (3.1)	1.6 (1.9)	2.7 (2.4)
VRr	2.6 (1.1)	1.3 (1.1)	1.7 (1.1)
PASi	12.8 (3.3)	8.4 (2.9)	9.9 (2.8)
PASd_h	2.0 (1.3)	0.5 (0.9)	1.0 (1.1)
PASi_h	4.4 (3.0)	1.1 (1.7)	2.0 (2.0)
SIM	15.5 (3.9)	9.8 (5.0)	11.6 (4.7)
BNT30	26.1 (3.4)	19.4 (5.9)	22.3 (5.4)

Abbreviations: APOE, apolipoprotein E; BNT30, Boston Naming Test (30-item Even Version); NP, neuropsychological; LMd, Logical Memory (Delayed Recall); LMi, Logical Memory (Immediate Recall); LMr, Logical Memory (Recognition); PASd_h, Hard score of Paired Associate Learning (Delayed Recall); PASi, Paired Associate Learning (Immediate Recall); PASi_h, Hard Score of Paired Associate Learning (Immediate Recall); SD, standard deviation; SIM, Similarities Test; VRd, Visual Reproductions (Delayed Recall); VRi, Visual Reproductions (Immediate Recall); VRr, Visual Reproductions (Recognition).

*Values are calculated based on a subset with valid data.

†Participants who did not consent to genetic analyses, had an APOE ε2/ε4 genotype, or with no APOE information were excluded.

important to cognitive outcomes. It designates a set of cutoff values for the chosen NP test via ChiMerge [35] and branches out to two or more lower-level (internal) nodes. This process is repeated at every internal node until the sample size in a specific node is less than 50 [36,37]. The performance of the model was evaluated by ten-fold cross validation [38].

2.3. Identification of most informative NP tests for cognitive diagnosis

The 11 NP tests were ranked, based on their strength of association with cognitive outcomes, using three feature selection techniques, namely Information Theory-based filtering [39], Correlation-based Feature Selection Adapting Greedy Search [40], and Classification and Regression Trees (CART) [41]—each representative of a class of feature selection methods (filter, wrapper, and embedded, respectively). The top five most informative NP tests were selected using majority voting. To demonstrate AD heterogeneity, a similar selection process was performed for each subpopulation, stratified by sex (male/female), education level (beyond high school/high school graduate and below), and apolipoprotein E (APOE) ε4 status [OMIM 107741]. For APOE-stratified analyses, participants who did not consent to genetic analyses or without APOE information were excluded (200 observations). Similarly, participants with missing education information were excluded from the education-stratified analyses (11 observations). Results from feature selection were further validated using k-means cluster analysis [42] and hierarchical clustering [43]. Additional decision trees were constructed using only the selected tests to avoid model overfitting and to increase generalizability of the algorithm. Refer to Supplemental materials for further details.

Written informed consent was obtained from all participants, and this study was approved by the Institutional Review Board of Boston University Medical Campus. All data collection methods used in this study were monitored by a National Heart, Lung, and Blood Institute Observational Study Monitoring Board and followed the Strengthening the Reporting of Observational Studies in Epidemiology reporting guideline.

3. Results

This study included 4512 sets of NP scores from 2091 participants (55.8% female), aged 79 ± 6 years. On average, each participant underwent 2.2 NP examinations. Among these observations, 555 were marked as AD, 443 as NAD, and the remaining were healthy controls (Table 1).

3.1. Dementia diagnosis from NP tests by decision tree

Fig. 1 shows the decision tree for dementia diagnosis. An illustration of tree generation is described in the Supplemental Results. The tree consists of five levels, with 27 internal nodes and 48 terminal nodes. Among the 11 NP tests, all but Logical Memory (Recognition) were represented. LMd was selected as the root node ($P < 1.0 \times 10^{-15}$), which branches to six internal nodes. The highest AD diagnostic accuracy yielded by this single decision step was 68.6%—the leftmost branch where LMd $\leq$1.0. With the introduction of other NP tests at subsequent nodes, the model appraises the individual overall cognitive performance, based on a set of NP scores rather than one single test score, and provides the diagnostic accuracy accordingly. For example, the leftmost path, besides LMd $\leq$1.0, is comprised of a BNT30 score from 0 to 23

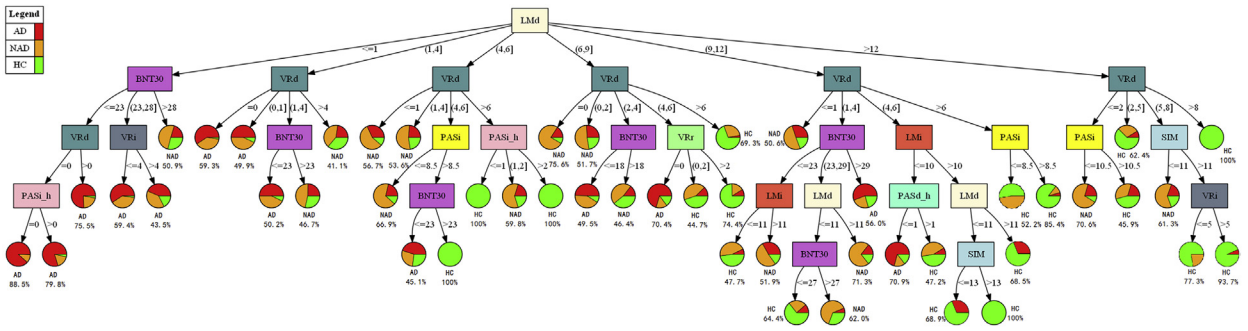

Fig. 1. Clinical cognitive screen decision tree based on all NP tests in total population. Each rectangle represents a branch node, which is a decision step where participants are divided into different subgroups based on the designated NP test score. Each pie chart represents a terminal node and is divided into color-coded slices to illustrate the probability of the three cognitive outcomes (AD, NAD, and HC). The outcome with the highest probability is indicated alongside each pie chart. Abbreviations: AD, Alzheimer's disease; BNT30, Boston Naming Test (30-item Even Version); HC, healthy control; NP, neuropsychological; LMd, Logical Memory (Delayed Recall); LMi, Logical Memory (Immediate Recall); NAD, non-Alzheimer's dementia; PASd_h, Hard score of Paired Associate Learning (Delayed Recall); PASi, Paired Associate Learning (Immediate Recall); PASi_h, Hard Score of Paired Associate Learning (Immediate Recall); SD, standard deviation; SIM, Similarities Test; VRd, Visual Reproductions (Delayed Recall); VRi, Visual Reproductions (Immediate Recall).

($P < 1.0 \times 10^{-11}$), VRd score of 0.0 ($P = .0036$), and PASi_h score of 0.0 ($P = .0023$), and this set of decision rules yielded the highest AD diagnostic accuracy sensitivity of 88.1%. This decision tree has an overall accuracy of 73.9%, with an all-cause dementia sensitivity of 85.0%.

Decision trees for different subpopulations were presented in Supplementary Figs. 2–7, with their individual performances and NP test cutoff values reported in Supplementary Table 1 and Supplementary Table 7, respectively.

3.2. Most informative NP tests for AD diagnosis

Table 2 shows the top five most informative NP tests with regard to cognitive outcomes, determined via each of the three feature selection methods. Both CART and Information Theory approaches identified the same set of five NP tests—LMd, VRd, LMi, VRi, and BNT30—for the total sample population, while the Correlation-based Feature Selection Adapting Greedy Search approach differed by picking PASi over VRi. As demonstrated in a previous study[21], LMd was consistently selected as an important feature for dementia diagnoses. While the BNT30 played a more important role for dementia diagnosis in men, PAS was preferred for women. A similar trend was observed in the stratified analyses for education and APOE ε4 status.

Fig. 2 represents the decision tree derived based on the top five most informative NP tests for total sample population. It consisted of five levels, with 19 internal nodes and 38 terminal nodes. Similar to Fig. 1, LMd was chosen as the root node. Its overall accuracy was 73.3%, with an all-cause dementia sensitivity of 84.5%. Supplementary Figs. 8–13 are decision trees created using only the optimal NP tests for different subpopulations. Their individual overall performance and NP test cutoff values are summarized in Supplementary Tables 2 and 8, respectively. Based on these results, the optimal tests had not only comparable performance with their full NP test-set counterparts but also reduced tree nodes, which would promote better ease of use for health-care workers.

4. Discussion

Cognitive domains are interconnected and may be simultaneously affected under diseased state, hence the complex and heterogeneous nature of AD [44]. Accurate dementia diagnosis requires the understanding of these relationships across all cognitive domains and the appreciation of various NP test outcomes concurrently. The current diagnostic process, however, depends heavily on the prior knowledge and experience of specialty clinicians, who often subjectively evaluate selected NP tests when making an AD diagnosis. This study comprehensively evaluated the relationships among various NP tests in a data-driven manner. None of NP tests alone was sufficient to separate participants with or without dementia. It is thus important to consider multiple NP tests for dementia diagnosis. The grading system for each NP test not only aids differential diagnoses but also transcends the traditional dichotomization—acceptable and unsatisfactory results—of individuals' neurocognitive performance, making it more reflective of dementia as a disease with a continuous spectrum of cognitive impairment. To our knowledge, this is the first study that uses a data-driven approach to leverage the multitude of NP test scores and simplify them into a set of intuitive instructions. Our approach could facilitate AD diagnosis for experienced clinicians in minimizing the subjectivity that is introduced in practitioners' decision-making process. Other health-care providers, who might lack sufficient clinical knowledge and training for AD diagnosis, could also potentially apply it.

We also evaluated the contribution of each NP test to the diagnosis of dementia. LMd was consistently identified as the most important performance indicator for AD diagnosis,

Table 2
NP tests selected by different feature selection methods for different subpopulations

	CART	CBFSGS	Information gain	Majority voting
Total	LMd, VRd, LMi, VRi, BNT30	LMd, VRd, BNT30, PASi, LMi	VRd, LMd, LMi, VRi, BNT30	LMd, VRd, LMi, VRi, BNT30
Sex				
Male	LMd, VRd, LMi, BNT30, VRi	LMd, VRd, BNT30, LMi, SIM	LMd, VRd, LMi, BNT30, VRi	LMd, VRd, LMi, BNT30, VRi
Female	LMd, VRd, PASi_h, LMi, PASi	VRd, LMd, BNT30, PASi, LMi	VRd, LMd, LMi, PASi, PASi_h	LMd, VRd, PASi_h, LMi, PASi
APOE ε4 allele*				
APOE ε4 (−)	LMd, VRd, LMi, BNT30, VRi	LMd, VRd, BNT30, SIM, LMi	VRd, LMd, LMi, VRi, BNT30	LMd, VRd, LMi, BNT30, VRi
APOE ε4 (+)	LMd, VRd, LMi, PASd_h, PASi_h	LMd, VRd, LMi, BNT30, VRi	LMd, VRd, LMi, PASi, VRi	LMd, VRd, LMi, VRi, PASi
Education				
High school and below	LMd, VRd, BNT30, LMi, VRi	LMd, VRd, BNT30, LMi, SIM	LMd, VRd, LMi, BNT30, VRi	LMd, VRd, BNT30, LMi, VRi
Beyond high school	LMd, VRd, LMi, PASd_h, PASi	VRd, LMd, PASi, BNT30, VRi	VRd, LMd, VRi, PASi, LMi	LMd, VRd, LMi, VRi, PASi

NOTE. Participants who did not consent to genetic analyses, had an APOE ε2/ε4 genotype, or with no APOE information were excluded.

Abbreviations: APOE, apolipoprotein E; BNT30, Boston Naming Test (30-item Even Version); CBFSGS, Correlation-based Feature Selection Adapting Greedy Search; LMd, Logical Memory (Delayed Recall); LMi, Logical Memory (Immediate Recall); NP, neuropsychological; PASd_h, Hard score of Paired Associate Learning (Delayed Recall); PASi, Paired Associate Learning (Immediate Recall); PASi_h, Hard Score of Paired Associate Learning (Immediate Recall); SIM, Similarities Test; VRd, Visual Reproductions (Delayed Recall); VRi, Visual Reproductions (Immediate Recall).

*APOE ε4 (−): APOE genotype ε2/ε2, ε2/ε3 or ε3/ε3; APOE ε4 (+): APOE genotype ε3/ε4 or ε4/ε4.

which aligned with the widespread use of verbal memory as a diagnostic tool. The overall performance of the reduced feature set decision tree (Fig. 2) was comparable to that of the decision tree based on all NP tests (Fig. 1): overall accuracy (73.9% vs. 73.3%) and all-cause dementia sensitivity (85.0% vs. 84.5%). This approach minimized model overfitting, and it can also potentially reduce time and effort required for clinical screening of dementia. We validated our findings using hierarchical clustering (Supplementary Fig. 14). Although both dendrograms identified three distinct clusters, the distinguishability was more pronounced for the reduced feature set dendrogram, which indicated that the use of optimal NP tests would potentially minimize data redundancy and better represent the inherent patterns within the NP data. In addition, with fewer NP tests to consider, we were able to reintroduce an additional 310 observations with valid NP scores for selected NP tests (LMd, VRd, LMi, VRi, and BNT30) and observed similar prediction accuracy (Supplementary Fig. 15).

This study effectively demonstrated the cognitive heterogeneity of AD and more importantly the need to consider the multiplicity during the diagnostic process. For example, results of the sex-stratified analysis revealed different optimal NP profiles that are most predictive for AD diagnoses in both sexes (Table 2, Supplementary Figs. 2 and 3), which is in agreement with previous findings of sex differences observed in various NP tests [45,46]. Heterogeneity of a disease is not unique to AD, as evident by various risk prediction models and diagnostic criteria having sex-specific algorithms to account for the effect modification by sex [47–49]. Despite well-established sex differences in cognitive performance, none of the current AD diagnostic criteria offer sex-specific decision rules. To meet the objectives of AD precision medicine, accurate patient stratification is crucial, and this study showed machine learning as one of the viable approaches that can help to do so. It is important to appreciate the effects of selected demographic and AD risk factors, as these not only enable more comprehensive dementia diagnosis decision-making but also have implications on patient selection in clinical trials.

Our study has several strengths. First, FHS started cognitive assessment in 1976 and has continued to monitor the participants for dementia over the next 4 decades. The long follow-up period and minimal loss to follow-up makes FHS an ideal population to examine late-onset diseases such as AD [25]. Second, dementia diagnosis of FHS participants were adjudicated by a panel of subject-matter experts, who evaluated multiple sources of information, thus minimizing outcome misclassification bias. Third, the FHS NP test battery consists of a wide array of commonly administered NP tests, which is ideal in translating the results for practical uses for clinicians and researchers. With feature selection, experts can focus on a subset of relevant NP tests to efficiently appreciate the overall data. Fourth, this data-driven approach surpasses the conventional model of dichotomizing individuals' performance into normal and impaired categories, by adopting a performance scale that is more representative of the spectrum of symptoms often exhibited by individuals with AD. Fifth, in contrast to other dimensionality reduction techniques such as those based on projection or compression, we chose to use feature selection, to avoid transforming the original values of the NP scores. With the original semantic nature of variables preserved, the discrete cutoff values allows easy interpretability, hence making it easy for assessors to follow the decision tree. Finally, the standard set of if-then diagnostic rules not only renders the implementation easy and scalable but also encourages reproducible science. As data accumulate, the accuracy of the algorithm will improve as well.

In terms of limitations, our study participants have higher levels of educational attainment compared to the general

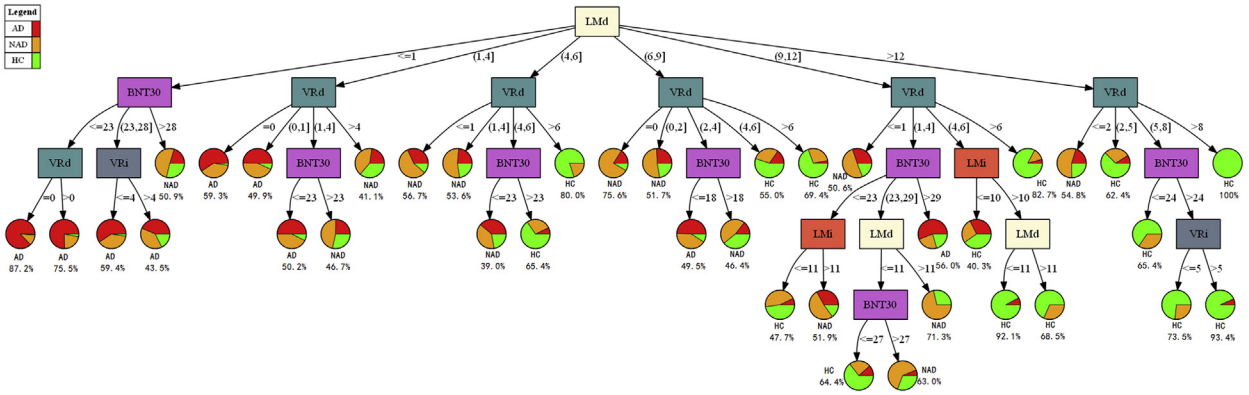

Fig. 2. Clinical cognitive screen decision tree based on optimal NP profiles (five tests) in total population. Abbreviations: AD, Alzheimer's disease; BNT30, Boston Naming Test (30-item Even Version); HC, healthy control; NP, neuropsychological; LMd, Logical Memory (Delayed Recall); LMi, Logical Memory (Immediate Recall); NAD, non-Alzheimer's dementia; PASd_h, Hard score of Paired Associate Learning (Delayed Recall); PASi, Paired Associate Learning (Immediate Recall); PASi_h, Hard Score of Paired Associate Learning (Immediate Recall); SD, standard deviation; SIM, Similarities Test; VRd, Visual Reproductions (Delayed Recall); VRi, Visual Reproductions (Immediate Recall).

public and are individuals predominantly of European descent. NP examinations were restricted to those conducted in English because of the limited number of evaluations done in Spanish. Therefore, results of this study may not be generalizable to populations of lower educational status, other races and non-English–speaking groups. In addition, the decision tree presented in this study solely uses information from NP tests. Given that FHS adjudication panel diagnosed dementia cases using multiple information sources, NP test results alone may not be adequate for definitive AD diagnosis. Hence, it should be viewed as an objective screening algorithm to identify high-risk individuals for further investigations to confirm AD diagnosis and potentially help reduce health-care costs related to overtesting. Further, only a subset of 11 tests were used and thus did not represent the full spectrum of cognitive domains assessed. It is possible that applied to a broader range of tests, a different profile of important NP features could emerge across the various AD risk factors. Similar to all clinical guidelines, these diagnosis instructions need to be periodically updated with the accumulation of additional data.

5. Conclusion

A summary of the critical achievements of our study are as follows: (1) intuitive and objective diagnostic criteria has been created as a set of if-then rules, which can be translated for actual clinical use that accounts for the complexity of AD clinical expression; (2) cutoff values of different tests have been identified with the ability to indicate a scale of severity and accurately reflect the spectrum of symptoms related to the heterogeneity of AD; and (3) the heterogeneity of AD in the context of NP tests has been verified by identifying important NP tests and predictive NP profiles for AD in subpopulations.

Future work includes development of an AD diagnosis support system based on a heterogeneous set of rules. When the individual's NP record is obtained, the system can automatically match the corresponding rule and make a diagnosis in a stepwise way that reflects a distinct AD subtype. Using concept learning methods, we can then build a general definition of AD that includes heterogeneous representation. With accumulation of additional longitudinal NP data, we will focus on the diagnosis of preclinical AD that is anticipated to have even a broader range of heterogeneity. We anticipate developing methods to diagnose conversion to AD within 5 to 10 years.

Acknowledgment

The authors want to express their thanks to the FHS study staff for their many years of hard work in the examination of subjects and acquisition of data. This work was supported by the National Heart, Lung, and Blood Institute contract (N01-HC-25195) and by grants from the National Institute on Aging AG-008122, AG-16495, and AG-062109 and from the National Institute of Neurological Disorders and Stroke, NS017950. It was also supported by Pfizer, Inc, the Boston University Digital Health Initiative, Boston University Alzheimer's Disease Center Pilot Grant, and the National Center for Advancing Translational Sciences, National Institutes of Health, through BU-CTSI Grant Number 1UL1TR001430. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institutes of Health or the U.S. Department of Health and Human Services.

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.trci.2019.05.002.

RESEARCH IN CONTEXT

1. Systematic review: Despite the availability of age- and education-adjusted standardized scores for most neuropsychological tests, there is a lack of objective rules on how to interpret multiple concurrent neuropsychological test scores that characterize the heterogeneity of Alzheimer's disease (AD). Relevant studies are cited.
2. Interpretation: Stepwise diagnosis rules that evaluate information from multiple neuropsychological tests were derived to produce an intuitive and objective AD dementia diagnosis with more than 80% accuracy. Heterogeneous AD profiles based on specific AD risk factors were also identified.
3. Future directions: Future work includes the development of an AD diagnosis support system based on a heterogeneous set of rules. Automated diagnosis rules have potential applications in both within clinic and outside-of-clinic settings. With accumulation of additional longitudinal NP data, we will focus on the diagnosis of preclinical AD that is anticipated to have an even broader range of heterogeneity.

References

[1] Burns A, Iliffe S. Alzheimer's disease. BMJ 2009;5:b158.
[2] Lam B, Masellis M, Freedman M, Stuss DT, Black SE. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. Alzheimers Res Ther 2013;5:1.
[3] Towards a dementia plan: a WHO guide. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.
[4] Elias MF, D'Agostino RB, Elias PK, Wolf PA. Neuropsychological test performance, cognitive functioning, blood pressure, and age: the Framingham Heart Study. Exp Aging Res 1995;21:369–91.
[5] Elias MF, Beiser A, Wolf PA, Au R, White RF, D'agostino RB. The preclinical phase of Alzheimer disease: a 22-year prospective study of the Framingham Cohort. Arch Neurol 2000;57:808–13.
[6] Au R, Piers RJ, Lancashire L. Back to the future: Alzheimer's disease heterogeneity revisited. Alzheimers Dement 2015;1:368.
[7] Reas ET, Laughlin GA, Bergstrom J, Kritz-Silverstein D, Barrett-Connor E, McEvoy LK. Effects of sex and education on cognitive change over a 27-year period in older adults: the Rancho Bernardo study. Am J Geriatr Psychiatry 2017;25:889–99.
[8] Mungas D, Reed BR, Farias ST, DeCarli C. Age and education effects on relationships of cognitive test scores with brain structure in demographically diverse older persons. Psychol Aging 2009; 24:116.
[9] Pangman VC, Sloan J, Guse L. An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. Appl Nurs Res 2000;13:209–13.
[10] Brayne C. The mini-mental state examination, will we be using it in 2001? Int J Geriatr Psychiatry 1998;13:285–90.
[11] Cobb JL, Wolf PA, Au R, White R, D'agostino RB. The effect of education on the incidence of dementia and Alzheimer's disease in the Framingham Study. Neurology 1995;45:1707–12.
[12] Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. JAMA 1993;269:2386–91.
[13] Borson S, Scanlan JM, Watanabe J, Tu SP, Lessig M. Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic sample. J Am Geriatr Soc 2005;53:871–4.
[14] Petersen RC, Smith GE, Ivnik RJ, Kokmen E, Tangalos EG. Memory function in very early Alzheimer's disease. Neurology 1994;44:867.
[15] Meyer JS, Xu G, Thornby J, Chowdhury M, Quach M. Longitudinal analysis of abnormal domains comprising mild cognitive impairment (MCI) during aging. J Neurol Sci 2002;201:19–25.
[16] Schmid U, Kitzelmann E. Inductive rule learning on the knowledge level. Cogn Syst Res 2011;12:237–48.
[17] Alessi L, Detken C. Identifying excessive credit growth and leverage. J Financial Stab 2018;35:215–25.
[18] Kasbekar PU, Goel P, Jadhav SP. A decision tree analysis of diabetic foot amputation risk in indian patients. Front Endocrinol 2017;8:25.
[19] Steenholdt C, Bendtzen K, Brynskov J, Thomsen OØ, Ainsworth MA. Cut-off levels and diagnostic accuracy of infliximab trough levels and anti-infliximab antibodies in Crohn's disease. Scand J Gastroenterol 2011;46:310–8.
[20] Elkana O, Eisikovits OR, Oren N, Betzale V, Giladi N, Ash EL. Sensitivity of neuropsychological tests to identify cognitive decline in highly educated elderly individuals: 12 months follow up. J Alzheimers Dis 2016;49:607–16.
[21] Johnson DK, Storandt M, Balota DA. Discourse analysis of logical memory recall in normal aging and in dementia of the Alzheimer type. Neuropsychology 2003;17:82.
[22] Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, Seshadri S. Incidence of dementia over three decades in the Framingham Heart Study. N Engl J Med 2016;374:523–32.
[23] Farmer ME, White LR, Kittner SJ, Kaplan E, Moes E, McNamara P, et al. Neuropsychological test performance in Framingham: a descriptive study. Psychol Rep 1987;60(3 Pt 2):1023–40.
[24] 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;12:189–98.
[25] Tsao CW, Vasan RS. Cohort Profile: the Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol 2015;44:1800–13.
[26] Gustafson D, Rothenberg E, Blennow K, Steen B, Skoog I. An 18-year follow-up of overweight and risk of Alzheimer disease. Arch Intern Med 2003;163:1524–8.
[27] Skoog I, Lernfelt B, Landahl S, Palmertz B, Andreasson LA, Nilsson L, et al. 15-year longitudinal study of blood pressure and dementia. The Lancet 1996;347:1141–5.
[28] Wechsler D, Stone CP. Wechsler Memory Scale (WMS). New York: The Psychological Corporation; 1948.
[29] Wechsler D. Wechsler Adult Intelligence Scale (WAIS). New York: The Psychological Corporation; 1955.
[30] Kaplan E, Goodglass H, Weintraub S, Segal O. Boston Naming Test. Philadephia: Lea & Febiger; 1983.
[31] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed.; 1994. Washington D.C.
[32] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–9.
[33] Seshadri S, Beiser A, Au R, Wolf PA, Evans DA, Wilson RS, et al. Operationalizing diagnostic criteria for Alzheimer's disease and other age-related cognitive impairment—Part 2. Alzheimers Dement 2011;7:35–52.

[34] Van Diepen M, Franses PH. Evaluating chi-squared automatic interaction detection. Inf Syst 2006;31:814–31.
[35] Kerber R. Chimerge: Discretization of numeric attributes. In: In Proceedings of the tenth national conference on Artificial intelligence; 1992. p. 123–8.
[36] Laliberte AS, Fredrickson EL, Rango A. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. Photogrammetric Eng Remote sensing 2007;73:197–207.
[37] McKee LA, Fabres J, Howard G, Peralta-Carcelen M, Carlo WA, Ambalavanan N. PaCO2 and neurodevelopment in extremely low birth weight infants. J Pediatr 2009;155:217–21.
[38] Seni G, Elder JF. Ensemble methods in data mining: improving accuracy through combining predictions. Synth Lectures Data Mining Knowledge Discov 2010;2:1–26.
[39] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.
[40] Hall MA. Correlation-based feature selection of discrete and numeric class machine learning; 2000.
[41] Breiman Leo. Classification and regression trees. Routledge; 2017.
[42] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowledge Inf Syst 2008;14:1–37.
[43] Xu D, Tian Y. A comprehensive survey of clustering algorithms. Ann Data Sci 2015;2:165–93.
[44] Dineen RA, Vilisaar J, Hlinka J, Bradshaw CM, Morgan PS, Constantinescu CS, et al. Disconnection as a mechanism for cognitive dysfunction in multiple sclerosis. Brain 2009;132:239–49.
[45] Miller DI, Halpern DF. The new science of cognitive sex differences. Trends Cogn Sci 2014;18:37–45.
[46] Zec RF, Burkett NR, Markwell SJ, Larsen DL. Normative data stratified for age, education, and gender on the Boston Naming Test. Clin Neuropsychol 2007;21:617–37.
[47] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837–47.
[48] D'agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008; 117:743–53.
[49] Dufouil C, Beiser A, McLure LA, Wolf PA, Tzourio C, Howard VJ, et al. Revised Framingham Stroke Risk Profile to Reflect Temporal Trends Clinical Perspective. Circulation 2017; 135:1145–59.