Using the ATN system as a guide for the neuropsychological assessment of Alzheimer's disease

Irene Florean , Barbara Penolazzi , Alina Menichelli , Massimiliano Pastore , Tatiana Cattaruzza , Giulia Mazzon and Paolo Manganotti

[a]Department of Life Sciences, University of Trieste, Trieste, Italy; [b]Rehabilitation Unit, Department of Medicine, Surgery and Health Sciences, Maggiore City Hospital Asugi, Trieste, Italy; [c]Department of Developmental and Social Psychology, University of Padua, Padua, Italy; [d]Clinical Unit of Neurology, Department of Medicine, Surgery and Health Sciences, Cattinara University Hospital Asugi, University of Trieste, Trieste, Italy

ABSTRACT

Introduction:: Many studies have attempted to determine whether Alzheimer's disease (AD) in-vivo biomarkers can predict neuropsychological performance since pathophysiological changes precede cognitive changes by several years. Nonetheless, neuropsychological measures can also detect cognitive deterioration in cognitively normal individuals with AD-positive biomarkers. Recent studies have investigated whether cognitive measures can be used as a proxy for biomarkers. This is a crucial issue since biomarker analysis is expensive, invasive, and not yet widespread in clinical practice. However, these studies have so far considered only one or two classes of AD biomarkers. Here, we aim at preliminarily evaluating whether and which neuropsychological measures can discriminate individuals that have been classified according to the full scheme of biomarkers known as ATN system. This scheme groups biomarkers as a function of the three main AD-related pathologic processes they measure (i.e., β-amyloidosis, tauopathy, and neurodegeneration) to provide an unbiased and descriptive definition of the Alzheimer's continuum.
Method:: Biomarkers and neuropsychological data from 78 patients (70.01 ± 9.15 years; 38 females) with suspected cognitive decline were extracted from a medical database. Participants' biomarker profiles were classified into the following ATN categories: normal AD biomarkers; Alzheimer's continuum; non-AD pathologic change. Data were analyzed using a Bayesian approach, to guarantee reliable result interpretation of data stemming from small samples.
Results:: The discrimination ability of each neuropsychological measure varied depending on the pairs of ATN categories compared. The best-discriminating predictor in the Alzheimer's continuum vs. normal biomarkers comparison was the figure naming ability. In contrast, in the Alzheimer's continuum vs. non-AD pathologic change comparison the best predictor was the wordlist forgetting rate.
Conclusions:: Although the study was exploratory in nature, the proposed methodological approach may have the potential to identify the best neuropsychological measures for estimating AD neuropathological changes, leading to a more biologically informed use of neuropsychological assessment.

KEYWORDS
Alzheimer's disease; neuropsychological assessment; CSF biomarkers; diagnosis; AT(N) scheme; ATN scheme

Introduction

Over recent years, most of the studies on associations between biomarkers for Alzheimer's disease (AD) and its cognitive manifestations have typically considered biomarkers as predictors and performance on neuropsychological measures as outcomes (Altomare et al., 2019; Bilgel et al., 2018; Soldan et al., 2016). The underlying rationale was that AD pathophysiological changes – now detectable in vivo through biomarkers – underlie cognitive changes and can precede them by several years. Nonetheless, neuropsychological assessment is still fundamental, firstly because neuropsychological changes can be found also when individuals with biomarkers positive for AD are still cognitively normal (that is, when they do not meet criteria for MCI or dementia; Altomare et al., 2019; Baker et al., 2017; Bilgel et al., 2018; Mortamais et al., 2017; Soldan et al., 2016), and secondly because biomarker examinations are currently expensive, invasive, and not always readily available in routine clinical practice. For instance, a survey reported that, in Italy, only 12% of the neurological centers use cerebrospinal fluid (CSF)

biomarkers for the diagnosis of AD (Sancesario et al., 2017). This is even more surprising if one considers that the Italian healthcare system is ranked among the best in the world (GBD 2016 Healthcare Access and Quality Collaborators, Fulman et al., 2018). Therefore, it is crucial to determine the extent to which a low-cost, non-invasive, and widely available tool such as neuropsychological assessment can be used to indicate the likely presence of pathophysiological processes underlying AD for early diagnosis and treatment.

In order to create a common language for Alzheimer's research focused on relations and interactions among pathophysiologic processes – as revealed by biomarkers – and clinical manifestations, Jack et al. (2016) proposed the amyloid/tau/neurodegeneration - ATN- system to classify biomarkers and to define the entire continuum of AD in an unbiased, descriptive way. In 2018, the system was incorporated in the new research framework by the National Institute on Aging-Alzheimer's Association (NIA-AA; Jack et al., 2018). This resulted in a radical change since this framework no longer conceives AD as a syndromic construct defined by its clinical consequences – signs and symptoms – but as a biological construct defined by its underlying pathologic processes documented through in vivo biomarker analyses. The ATN scheme classifies the main AD biomarkers into three groups: "A," measuring brain β-amyloidosis (CSF β-amyloid42 – Aβ42 – levels or amyloid-PET); "T," measuring brain tauopathy (CSF phosphorylated-tau – p-tau – levels or tau-PET); "N," measuring neurodegeneration or brain injury (CSF total-tau – t-tau – levels, 18 F-fluorodeoxyglucose-PET, or MRI). In the 2018 NIA-AA-research framework, N is bracketed because the biomarkers belonging to this group are not specific for AD. The use of positive/negative cut-points in each group of biomarkers allows for the identification of eight profiles, based on whether ("+") or not ("-") the pathological cut-point is reached for each biomarker. Therefore, these profiles can be grouped into three categories: negative (i.e., normal) AD biomarkers (A-T-N-); Alzheimer's continuum, including all profiles with positive A-biomarkers (A + T-N-; A + T-N+; A + T + N-; A + T + N+); non-AD pathologic change, including all profiles with negative A-biomarkers but positive T- and/or positive N-biomarkers (A-T-N+; A-T + N-; A-T + N+).

Although several studies have already explored the 2018 NIA-AA-research framework, only a few of them have considered the role that neuropsychological assessment could play in its context. Moscoso et al. (2019) found that episodic memory assessment can provide complementary information to the ATN biomarkers on AD by predicting its clinical and pathological progression, although they only assessed this cognitive function on a selected sample of amyloid-positive participants with mild cognitive impairment (MCI). Altomare et al. (2019) studied a sample of memory clinic patients and found that, compared to normal biomarker profiles, some Alzheimer's continuum profiles showed worse global cognition, memory, and visuospatial abilities at baseline and faster global cognitive decline over time, whereas some non-AD pathologic change profiles exhibited worse language performance. Although the neuropsychological assessment was well designed in the study, the final description of ATN profiles consisted of composite measures, which may be less than ideal for a fine-grained cognitive characterization of the patients. Using individual (rather than composite) neuropsychological measures could indeed be more effective in discriminating cognitive profiles belonging to the Alzheimer's continuum from conditions other than AD (e.g., an identical composite score in the linguistic domain may result from semantic as well as executive deficits).

Only a few studies have recently attempted to reverse the main trend in the investigations of AD biomarkers and cognitive manifestations by exploring whether neuropsychological measures can be usefully employed for predicting the status of AD biomarkers.

Mueller et al. (2020) found that an innovative memory measure, the delayed recall of proper names from a short story, significantly predicted β-amyloid status (A ±) in a sample of cognitively unimpaired adults. However, they considered only measures from short story recall as predictor variables and did not consider the "T" and "N" biomarker classes. Alves et al. (2021) found that delayed free recall of a short story was the only neuropsychological measure able to predict the amyloid status (A±) among those tested. Yet, they also investigated neuropsychological associations with only one biomarker class from the ATN system and in a large sample that however, included only patients with amnestic MCI. Interestingly, Stricker et al. (2020) found that the performance on a single wordlist delayed recall test was able to differentiate A + T+ and A + T- from A-T- participants. However, the authors categorized participants according to their amyloid and tau (AT) status, instead of using the full ATN classification scheme.

In light of the picture described above, our study had two main purposes. Our first goal was to characterize in greater detail the cognitive profile of individuals classified into ATN categories by considering a wide range of neuropsychological measures. Our second goal was to explore whether and which neuropsychological measures would allow clinicians to discriminate individuals

that have been classified according to ATN categories. Specifically, we aimed at evaluating whether the neuropsychological tests used were able to distinguish patients with Alzheimer's continuum profiles from those with normal biomarker profiles (i.e., profiles negative for all the ATN biomarker groups) and those with non-AD pathologic change profiles. Additionally, although different biological and syndromic conditions may fall within the non-AD pathologic change category, we also explored whether and which neuropsychological measures could discriminate individuals with these ATN profiles from those with normal biomarker profiles.

We analyzed data included in a medical database of patients who had undergone neuropsychological assessment and CSF analysis for suspected cognitive decline. Neuropsychological assessment takes into consideration a wide range of neuropsychological measures largely used in clinical practice. This is important since the literature has produced rich data in the AD field only for some measures (e.g., verbal episodic memory measures from wordlist presentations). We then used an analysis method that is innovative in neuropsychological research – a Bayesian approach – to preliminarily evaluate whether and which neuropsychological measures could discriminate between ATN categories. We adopted the Bayesian approach because it is particularly suitable to robustly manage and interpret data from small samples (Van de Schoot & Miocević, 2020)

It is important to note that in the wider investigation on the relationships between biological and neuropsychological measures in AD, the ultimate goal of our archival study was to describe a methodological approach potentially able to lead to a more biologically informed use of neuropsychological assessment, rather than testing formal hypotheses. However, drawing from studies that have used AD biomarkers as predictors of neuropsychological performance (e.g., Baker et al., 2017; Mortamais et al., 2017), we expect to find at least some memory (both episodic and semantic) and language measures to satisfactorily discriminate between normal and Alzheimer's continuum biomarker profiles. On the contrary, we did not make specific assumptions about the specific neuropsychological measures, which could adequately discriminate between Alzheimer's continuum and non-AD pathologic change profiles, and between these latter and normal profiles, given the very limited evidence found in the literature on these comparisons.

We believe that investigating which cognitive measures can be useful in predicting underlying pathophysiological processes is crucial not only for syndromic diagnosis, prognosis, patient monitoring, and management, but also for clinical trial enrollment and treatment selection, especially when neuroimaging or CSF biomarker profiling is not an available option.

Materials and Methods

Participants

The data analyzed in the study were extracted from an archival record of older adults admitted for outpatient treatment at the Neurologic Clinic of the "Azienda Sanitaria Universitaria Integrata di Trieste," Italy, between June 2016 and July 2019 for suspected cognitive decline. Only patients with available CSF and neuropsychological assessment data were extracted from the archive (n = 121). Patients with documented diagnoses of psychiatric syndromes, presence of more than one neuropathology (e.g., Parkinson's disease and CSF biomarkers positive for AD), and less than 5 years of formal education were excluded. The final sample included 78 people (F = 38, M = 40). Basic demographic characteristics for each ATN group and for the entire sample are reported in Table 1.

This study was approved by the local health ethics committee and was performed according to the Declaration of Helsinki. All participants released their informed consent.

Consistent with the 2018 NIA-AA research framework (Jack et al., 2018), we used CSF biomarkers to classify subjects into three binary categories: A±, T±, and N±. For each category, a given CSF biomarker value was considered pathologic (+) according to the following cut-points: Aβ42 < 550 pg/mL; p-tau > 52 pg/mL; t-tau > 375 pg/mL (Duits et al., 2014; Mulder et al., 2010). The resulting participants' biomarker profiles were classified into three groups: normal AD biomarkers (A-T-N-); Alzheimer's continuum profiles (A + T-N-; A + T + N-; A + T + N+); non-AD pathologic change profiles (A-T + N-; A-T-N+; A-T + N+). The normal biomarker group included 20 participants (age: M = 63.70, SD = 10.99; 10 F); the Alzheimer's continuum group included 37 participants

Table 1. Basic demographic characteristics for each ATN group and for the entire sample.

	N	NAD	AD	Total
N	20	21	37	78
Sex (% F)	50.00%	47.62%	54.05%	51.28%
Age (y)	63.70 (10.99)	69.52 (9.68)	73.70 (5.19)	70.01 (9.15)
Education (y)	11.75 (4.36)	12.48 (3.72)	10.89 (4.04)	11.55 (4.05)

N = normal biomarkers; NAD = non-Alzheimer's disease pathologic change; AD = Alzheimer's continuum.
For Age and Education, mean (standard deviation) values are reported.

(age: M = 73.70, SD = 5.19; 17 F); the non-AD pathologic change group included 21 participants (age: M = 69.52, SD = 9.68; 11 F).

By using current clinical criteria (Aarsland et al., 2007; Emre et al., 2007; G. M. McKhann et al., 2011; McKeith et al., 2017; T O'Brien et al., 2003; Rascovsky et al., 2011), a diagnostic team of expert neurologists, neuroradiologists, and one neuropsychologist identified different diagnostic entities in each group based on neuroimaging, CSF, and neuropsychological examinations. For cognitive staging, we used a syndromal categorical scheme defined as applicable to the entire population by the 2018 NIA-AA research framework (Jack et al., 2018). This scheme classifies the cognitive continuum into three well-known categories: cognitive unimpairment, mild cognitive impairment (MCI), and dementia. Table 2 shows diagnostic entities and cognitive staging for each biomarker group. Among participants with clinical and instrumental evidence of frontotemporal dementia included in the non-AD pathologic change group, six had the behavioral variant, and two had the linguistic variant, whereas in the Alzheimer's continuum group, 32 participants had the amnestic variant and five had the non-amnestic variant (three language variants; two visuospatial variants; G. M. McKhann et al., 2011) of the disease.

Table 2. Diagnostic entities and cognitive staging for each ATN group.

		Biomarker groups		
		N group	NAD group	AD group
		(n = 20)	(n = 21)	(n = 37)
Diagnostic entities	AD	-	-	37
	CI-uncertain etiology	4	9	-
	FTD	1	8	-
	LBD	1	-	-
	PDD	1	1	-
	SCI	10	-	-
	VCI	3	3	-
Cognitive staging (%)	Cognitive unimpairment	50.00%	-	-
	MCI	40.00%	52.38%	51.35%
	Dementia	10.00%	47.62%	48.65%

Biomarker groups: N = Normal AD biomarker group; NAD = non-AD pathologic change group; AD = Alzheimer's continuum group. Diagnostic entities: AD = Alzheimer's disease; CI-uncertain etiology = Cognitive impairment with uncertain etiology; FTD = Frontotemporal dementia; LBD = Lewy's body dementia; PDD = Parkinson's disease dementia; SCI = Subjective cognitive impairment; VCI = Vascular cognitive impairment. Cognitive staging: MCI = Mild cognitive impairment.

CSF biochemical analysis

CSF Aβ42, t-tau, and p-tau were analyzed as ATN biomarkers. The CSF analysis procedure followed the Alzheimer's Association Flow Chart for CSF biomarkers (Blennow et al., 2010). Collection, centrifugation, and storage of CSF samples were performed following a standardized protocol and using polypropylene tubes. The analysis of CSF samples was performed in the same laboratory. CSF Aβ42, t-tau, and p-tau were analyzed with INNOTEST® β-AMYLOID (1–42), INNOTEST® hTAU Ag, and INNOTEST® PHOSPHOTAU(181P), respectively (Fujirebio Europe, Ghent, Belgium).

Neuropsychological assessment

According to the consensus document of the Joint Program for Neurodegenerative Diseases Working Group (Costa et al., 2017), neuropsychological tests designed to assess cognitive domains that are critical for detecting changes related to neurodegenerative dementias were used to evaluate the participants' cognitive profile. Table 3 shows the selected tests and, for each of them, the cognitive functions primarily involved in its execution and the related measures. Test administrations followed the standard procedure for each of them, and the related measures were adjusted for age, education, and gender, according to the relevant Italian normative data (see, Table 3). Episodic memory tests based on wordlists usually assess learning by aggregating the scores of several learning trials into a single score (e.g., Mauri et al., 1997). In order to investigate in more detail the encoding process, we consider two additional scores, i.e., the percentage of words retrieved in the first and fifth learning trials.

The number of participants varied across the test measures (see, Table 4, Table 5, N column) as data were extracted from a database created for clinical purposes. However, a series of chi-squared tests (a chi-squared test for each of the cognitive measure computed) supported homogeneous severity levels of syndromal-staging (i.e., cognitive unimpairment, MCI, dementia) in the compositions of abnormal biomarkers groups (i.e., Alzheimer's continuum group, non-AD pathologic change group) for each test measure (all n.s.).

Statistical analyses

Data were analyzed using a fully Bayesian approach and interpreted descriptively, rather than inferentially, given the exploratory intent of the study (Amrhein et al., 2019). This approach can provide important insights about the neuropsychological measures that are most capable of predicting patients' ATN biomarker categories. All analyses

Table 3. Tests used for neuropsychological assessment and their related cognitive functions and measures.

Cognitive Tests	Cognitive Functions	Measures
Semantically unrelated wordlist (Mauri et al., 1997)	• Long-term verbal episodic memory	• Number of words correctly retrieved across the five learning trials (global immediate recall) [wordlist IR] • Percentage of words retrieved in the first learning trial [wordlist IR trial1] • Percentage of words retrieved in the fifth learning trial [wordlist IR trial5] • Number of words produced but not included in the list across the five learning trials (intrusions in immediate recall) [wordlist IR intrusions] • Number of words correctly retrieved 15 minutes after the fifth learning trial (delayed recall) [wordlist DR] • Number of words produced but not included in the list during the delayed recall trial (intrusions in delayed recall) [wordlist DR intrusions] • Percentage of word retrieved in the fifth learning trial but not in the delayed trial (forgetting) [wordlist forgetting] • Number of words correctly identified as included or not included in the studied list (recognition) [wordlist recognition]
Short story (Spinnler & Tognoni, 1987)	• Long-term verbal episodic memory	• Sum of the two hierarchical scores obtained on the subtasks of immediate and delayed recall[short story]
Rey Complex Figure Test (Caffarra et al., 2002)	• Visuospatial and visuo-constructive abilities • Long-term non-verbal episodic memory • Attention and executive functions	• Number of figure elements correctly copied [CRFT-copy] • Number of figure elements correctly reproduced in the delayed reproduction trial [CRFT-DR]
Oral confrontation naming task (from CAGI battery; Catricala et al., 2013)	• Visual perception, detection, and recognition • Language • Semantic memory	• Number of figures correctly named [naming]
Semantic fluency task – animal (Costa et al., 2014)	• Language • Attention and executive functions	• Total number of words produced with a category-cue in one 60s trial [verbal fluency-semantic]
Phonemic fluency task – FAS (Costa et al., 2014)	• Language • Attention and executive functions	• Total number of words produced with an initial letter-cue in three 60s trials [verbal fluency-phonemic]
Digit Span Forward test (Monaco et al., 2015)	• Short-term verbal memory	• Length of the longest digit list correctly recalled in the order presented [forward digit span]
Digit Span Backward test (Monaco et al., 2015)	• Attention and executive functions	• Length of the longest digit list correctly recalled in reverse order [backward digit span]
Forward Corsi span test (Monaco et al., 2015)	• Short-term non-verbal memory	• Length of the longest block-tapping sequence correctly reproduced [Corsi]
Modified Card Sorting Test (Caffarra et al., 2004)	• Attention and executive functions	• Number of categories achieved [MCST-category] • Number of perseverations [MCST-perseverations]
Cognitive Estimation Task-part B (Scarpina et al., 2015)	• Attention and executive functions • Semantic memory	• Scores assigned as a function of the distance between the subject's estimations and the range of the normative estimations [CET]
Stroop test-color word item task (Brugnolo et al., 2016)	• Attention and executive functions	• Number of color-words whose ink was correctly named in 30s, i.e., Stroop interference trial (Stroop-Color Word Items) [Stroop-CWI]
Trail Making Test-part B (Giovagnoli et al., 1996)	• Attention and executive functions	• Time in seconds to complete the task [TMT-B]

The first column shows the tests administered and – in brackets – the Italian normative studies used in their administration and scoring. The second column shows the cognitive functions mainly involved in the execution of each test. The third column shows the measures derived from the tests and, in square brackets, the verbal labels used to name them in the following tables and figures.

were performed with R software (R Core Team, 2019), with the following packages: readx1 (Wickham & Bryan, 2019), ggplot2 (Wickham, 2016), psych (Revelle, 2019), GGally (Schloerke et al., 2020), dataExplorer (Cui, 2020), brms (Bürkner, 2017).

Data analyses consisted of two main steps: (1) univariate and bivariate descriptive statistics, and (2) a set of categorical logit models to evaluate the relationship between each of our predictors (i.e., neuropsychological measures) and the outcome (i.e., ATN groups). For parameter estimations, the Stan statistical platform was used (Bürkner, 2017; Carpenter et al., 2017; Stan Development Team, 2018). Default priors from the brms package, which are Student's t for regression coefficients and residual variances (the latter truncated to zero), and LKJ for correlations (Lewandowski et al., 2009) were used. The estimates of each model were based on four Monte Carlo Markov Chains, with 4000 replications (2000 warm-up iterations discarded) each; therefore, the posterior distributions of the parameters had 8000 actual samples each. Since a clinical database was used, the sample size available for each predictor varied from 39 to 76. For this reason, a strategy to maximize the information using all available data for

Table 4. Univariate descriptive statistics of measures considered for each neuropsychological test.

	N	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Wordlist												
IR	63	31.21	10.24	29.60	30.56	9.01	13.50	55.30	41.80	0.60	−0.50	1.29
IR trial1 (%)	55	23.07	13.22	25.00	22.50	9.27	0.00	68.75	68.75	0.65	1.29	1.78
IR trial5 (%)	56	43.53	20.84	40.62	43.08	18.53	0.00	87.50	87.58	0.20	−0.41	2.78
IR-intrusions	63	2.03	2.89	1.00	1.41	1.48	0.00	12.00	12.00	1.97	3.64	0.36
DR	63	4.49	3.82	4.00	4.17	5.19	0.00	13.30	13.30	0.49	−0.82	0.48
DR- intrusions	63	0.52	0.88	0.00	0.35	0.00	0.00	3.00	3.00	1.48	1.01	0.11
Forgetting (%)	63	51.40	33.78	48.40	51.26	44.48	0.00	100.00	100.00	0.19	−1.35	4.26
recognition	60	25.71	4.03	26.10	25.97	4.89	15.90	32.00	16.10	−0.50	−0.69	0.52
Short story	51	6.23	5.11	6.10	6.03	8.01	0.00	16.00	16.00	0.09	−1.42	0.72
RCFT												
copy	69	27.23	6.70	29.00	27.82	5.56	9.50	36.00	26.50	−0.87	−0.32	0.81
DR	65	8.64	6.82	8.50	8.10	6.30	0.00	23.25	23.25	0.47	−0.65	0.85
Naming	66	42.11	5.15	43.54	42.68	4.13	25.27	48.00	22.73	−1.12	0.92	0.63
Verbal fluency												
semantic	60	12.30	7.26	10.21	11.58	6.24	1.00	32.90	31.90	0.82	0.06	0.94
phonemic	72	26.72	12.26	28.11	26.87	11.25	0.00	60.88	60.88	−0.06	−0.21	1.44
Digit span												
forward	76	5.48	1.06	5.32	5.44	1.19	2.99	8.74	5.75	0.41	0.36	0.12
backward	76	4.01	1.24	4.19	4.15	1.07	0.00	6.08	6.08	−1.30	2.25	0.14
Corsi span	76	4.75	1.28	4.50	4.77	1.26	1.00	8.15	7.15	−0.16	0.35	0.15
MCST												
category	39	3.64	1.83	3.00	3.67	2.97	1.00	6.00	5.00	0.04	1.54	0.29
perseverations	39	5.57	4.49	4.50	5.21	4.08	0.00	17.50	17.50	0,72	−0.25	0.72
CET	69	10.09	5.30	9.74	9.92	5.93	0.00	20.34	20.34	0.25	−0.87	0.64
Stroop – CWI	52	17.25	8.88	17.34	17.34	8.35	0.00	35.65	35.65	−0.01	−0.51	1.23
TMT – B (sec)	43	131.51	97.45	103.00	119.37	74.13	6.00	404.00	398.00	1.04	0.23	14.86

For each neuropsychological measure, the following statistics are reported: number of participants (N), mean, standard deviation (sd), 10% trimmed mean (trimmed), median absolute deviation (mad), minimum (min), and maximum (max) observed values, range of values (range), skewness (skew) and kurtosis indexes for score distribution, and mean standard error (se). Reported values refer to raw scores adjusted for age, education, and gender, according to Italian normative data. Wordlist IR = immediate recall; Wordlist IR trial1 = percentage of words retrieved in the first learning trial; Wordlist IR trial5 = percentage of words retrieved in the fifth learning trial; Wordlist IR intrusions: intrusions in immediate recall; Wordlist DR = delayed recall; Wordlist DR intrusions = intrusions in delayed recall; RCFT = Rey complex figure test; RCFT-DR = Rey complex figure test-delayed reproduction; MCST = modified card sorting test; CET = cognitive estimation task; CWI = color word items; TMT-B = trial making test-part B.

each predictor was adopted for categorical models, with comparisons being based on the indexes outlined below. Categorical models give for each participant the probability to be classified into one of the three ATN categories depending on the value of the predictor taken into consideration. For any *i* subject, the probability distribution is Prob (R*i*= k), with R*i* = expected category according to the model for the *i* subject, and k = [Alzheimer's continuum, non-AD pathologic change, normal]. Starting from those probabilities, we used three indexes. The first, Cramer's V (Cramer, 1946), compares the observed category with the most likely category expected by the model according to a 3 × 3 contingency table. The higher the value of Cramer's V within the 0–1 range, the better the quality of model classification (Cramer's V is one when all expected categories match the observed ones).

The second index, the mean expected probability of correct classification (referred to with μP), was computed by considering each participant's probability predicted for the observed category and by averaging probabilities. The higher the μP value, the more the model tends to classify participants consistently with observed categories.

These two indexes are useful to evaluate the goodness-of-fit of the models but do not provide information on the predictor's ability to discriminate between pairs of categories, which is the main goal of our study. Therefore, a third index was computed to quantify the discrimination ability of predictors. This index (referred to as η) quantifies the overlap degree between 90% credibility intervals around the estimated probabilities (Kruschke & Liddell, 2018). To make the interpretation easier, η also varies within the 0–1 range, where zero indicates the maximum and one the minimum discrimination capacity. This index can be used to discriminate between pairs of categories (i.e., Alzheimer's continuum vs. non-AD pathologic change: ηAD_NAD; Alzheimer's continuum vs. normal biomarkers: ηAD_N; non-AD pathologic change vs. normal biomarkers: ηNAD_N), and a global average (η) index can also be computed.

Data availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

Table 5. Indexes to quantify the ability of each neuropsychological measure to discriminate between pairs of ATN categories and goodness-of-fit.

	N	N_{AD}	N_{NAD}	N_N	ηAD_N	ηAD_NAD	ηNAD_N	η	CV	μP
Wordlist										
IR	63	29	15	19	**0.09**	0.37	0.20	0.22	**0.48**	0.45
IR trial1	54	27	12	15	**0.06**	0.20	0.16	0.14	**0.44**	**0.55**
IR trial5	55	27	12	16	0.13	0.33	0.23	0.23	0.31	**0.48**
IR intrusions	63	29	15	19	0.5	0.42	0.47	0.47	0.22	0.37
DR	63	29	15	19	0.15	0.37	0.25	0.26	0.41	0.42
DR intrusions	63	29	15	19	0.5	0.21	0.43	0.38	0.26	0.4
forgetting	63	29	15	19	0.23	**0.15**	0.40	0.26	0.25	0.39
recognition	60	28	15	17	0.1	0.19	**0.12**	0.14	0.35	0.42
Short story	51	24	14	13	0.29	0.4	0.54	0.41	0.26	0.39
RCFT										
copy	69	31	19	19	0.17	0.28	0.50	0.31	NA	0.36
DR	65	29	18	18	0.13	0.45	0.19	0.26	0.38	0.42
Naming	66	31	17	18	**0.04**	**0.16**	**0.05**	0.08	**0.49**	**0.49**
Verbal fluency										
semantic	60	29	15	16	0.11	0.37	0.15	0.21	0,44	0.44
phonemic	72	35	18	19	0.31	0.17	0.49	0.32	0.12	0.37
Digit span										
forward	76	37	19	20	0.28	0.33	0.75	0.45	0.20	0.38
backward	76	37	19	20	0.15	0.23	0.46	0.28	NA	0.37
Corsi	76	37	20	19	0.09	0.28	0.21	0.19	0.36	0.42
MCST										
category	39	13	8	18	0.19	0.29	**0.12**	0.20	0.40	0.44
perseverations	39	13	8	18	0.36	0.31	0.27	0.31	0.02	0.39
CET	69	33	18	18	0.2	**0.16**	0.3	0.22	0.14	0.39
Stroop – CWI	52	19	14	19	0.15	0.48	0.35	0.33	0.42	0.38
TMT – B	43	17	11	15	0.27	0.52	0.2	0.33	0.31	0.39
Null	78	37	21	20	NA	NA	NA	NA	NA	0.36

For each neuropsychological measure, the following values have been reported: total number of participants (N), number of participants for the Alzheimer's continuum group (NAD), number of participants for the non-Alzheimer's pathologic change group (NNAD), number of participants for the normal biomarker group (NN); discriminative indexes between pairs of ATN categories: Alzheimer's continuum vs. normal biomarkers (ηAD_N); Alzheimer's continuum vs. non-Alzheimer's disease pathologic change (ηAD_NAD); non-Alzheimer's disease pathologic change vs. normal biomarkers (ηNAD_N); global discriminative index (η); Cramer's V index (CV); mean expected probability of correct classification (μP). Null refers to the null model, i.e., the model without predictors. To note, the smaller the values of discriminative indexes, the higher the ability to discriminate between AT(N) categories; the higher the CV value, the higher the matching between expected and observed categories; the higher the μP value, the better the ability to predict the AT(N) category to which participants actually belong. All the values of the above indexes fall within the 0–1 range. Wordlist IR = immediate recall; Wordlist IR trial1 = percentage of words retrieved in the first learning trial; Wordlist IR trial5 = percentage of words retrieved in the fifth learning trial; Wordlist IR intrusions: intrusions in immediate recall; Wordlist DR = delayed recall; Wordlist DR intrusions = intrusions in delayed recall; RCFT = Rey complex figure test; RCFT-DR = Rey complex figure test-delayed reproduction; MCST = modified card sorting test; CET = cognitive estimation task; CWI = color word items; TMT-B = trial making test-part B.

ηAD_N, ηAD_NAD, and ηNAD_N values are in bold for the (first three) most discriminative measures between pairs of ATN categories; CV and μP values are in bold for the (first three) best-fit models.

Results

Descriptive statistics

reports the univariate descriptive statistics of the 22 neuropsychological measures from the 13 tasks taken into consideration. Raw scores were adjusted according to the respective Italian normative data for all measures with only two exceptions, i.e., the percentage of words retrieved in the first and fifth learning trials (wordlist RI trial1 and trial5), for which there is no available data in the Italian population. Therefore, the computation of categorical models based on these two measures was adjusted for age and education of participants by directly controlling their impact in the analyses. Figure 1 shows the value distributions for the 22 quantitative variables depending on ATN categories.

Correlation analyses between indexes

Pearson's correlation analyses between indexes of all the neuropsychological measures taken into consideration (Supplementary Fig. S1) provided evidence for a consistent ranking of models. Correlations indicate a substantial consistency of computed indexes, except for the discrimination index for the Alzheimer's continuum vs. non-AD pathologic change comparison (ηAD_NAD), which weakly correlates with Cramer's V index (r = .04) and with the other two discrimination indexes (r = .18 for ηAD_NAD-ηAD_N; r = .07 for ηAD_NAD-ηNAD_N). As suggested in the following sections, these low correlations are likely due to the fact that a neuropsychological measure that is good at discriminating between two ATN categories (e.g., Alzheimer's continuum vs. non-AD pathologic change comparison) was not necessarily equally good at

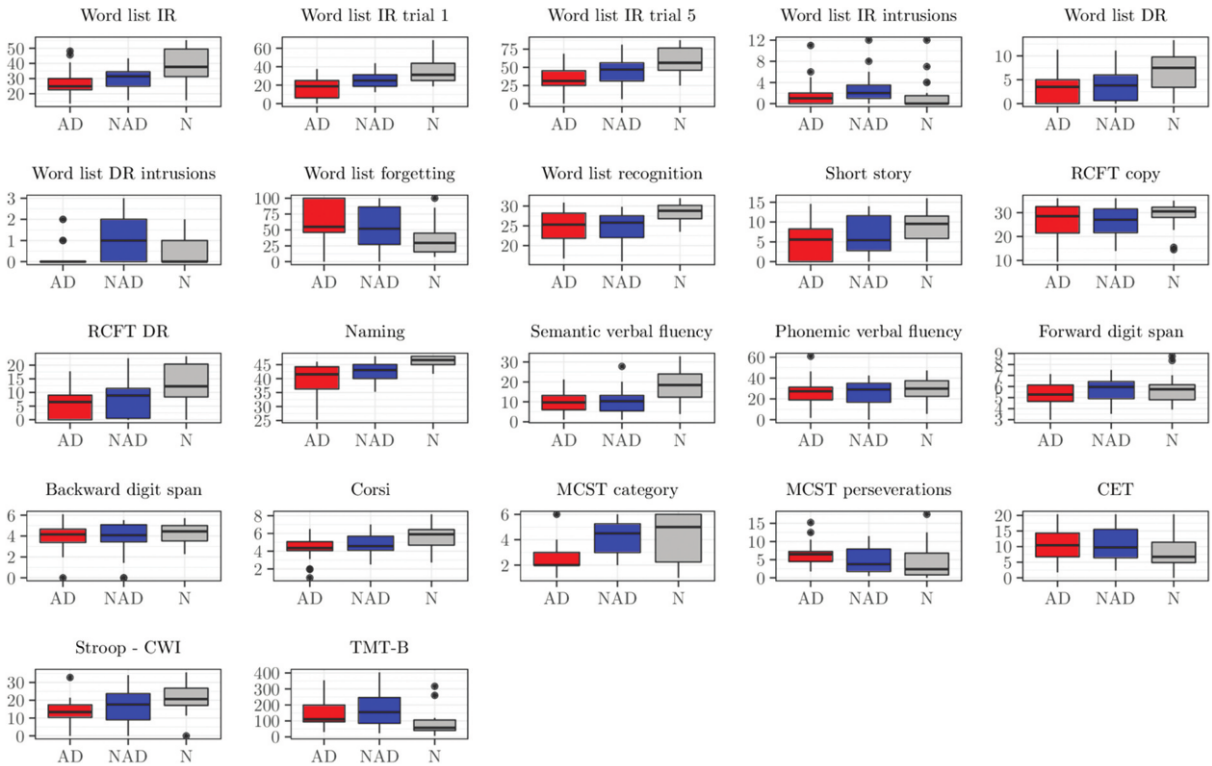

Figure 1. Score distributions on neuropsychological measures considered as a function of the three ATN categories. Scores were adjusted for age, education, and gender, according to Italian normative data. AD = Alzheimer's continuum; NAD = non-Alzheimer's disease pathologic change; N = normal biomarkers; Wordlist IR = immediate recall; Wordlist IR trial1 = percentage of words retrieved in the first learning trial; Wordlist IR trial5 = percentage of words retrieved in the fifth learning trial; Wordlist IR intrusions: intrusions in immediate recall; Wordlist DR = delayed recall; Wordlist DR intrusions = intrusions in delayed recall; RCFT = Rey complex figure test; RCFT-DR = Rey complex figure test-delayed reproduction; MCST = modified card sorting test; CET = cognitive estimation task; CWI = color word items; TMT-B = trial making test-part B.

discriminating between the other two pairs of categories (e.g., Alzheimer's continuum vs. normal biomarkers, and/or normal biomarkers vs. non-AD pathologic change). This may also explain why the indexes used to evaluate the goodness-of-fit of the models (i.e., Cramer's V and μP) showed quite modest values.

Categorical models

Table 5 reports findings on the indexes computed (see Statistical Analysis section) for all categorical models (one for each neuropsychological measure considered). Because the main goal of our study was to test the ability of neuropsychological measures to discriminate between ATN categories and the number of analyzed measures was rather broad, the presentation and discussion of results will be focused on the categorical models with the highest ability to discriminate between pairs of categories, i.e., with lower overlap degree between 90% credibility intervals around estimated probability distributions of categories in pairs. By examining our findings on categorical models (that is, on neuropsychological measures), we found that the 90% credibility intervals around the estimated probability distributions of groups were clearly differentiable for η-values < .20, whereas they overlapped progressively more for η-values > .20. This means that the predicted probability of belonging to a certain ATN group is similar to the predicted probability of belonging to the compared ATN group. For these reasons, a value < .20 for ηAD_NAD, ηAD_N, and ηNAD_N was assumed as an appropriate criterion for selecting neuropsychological measures with the highest ability to discriminate between pairs of categories

Best neuropsychological measures in discriminating between Alzheimer's continuum group and normal biomarker group

The predictor with the highest discrimination ability between Alzheimer's continuum and normal biomarker groups was the oral confrontation naming task. The value of .04 on ηAD_N (see, Table 3) indicates that the overlap degree between 90% credibility intervals around the estimated probability

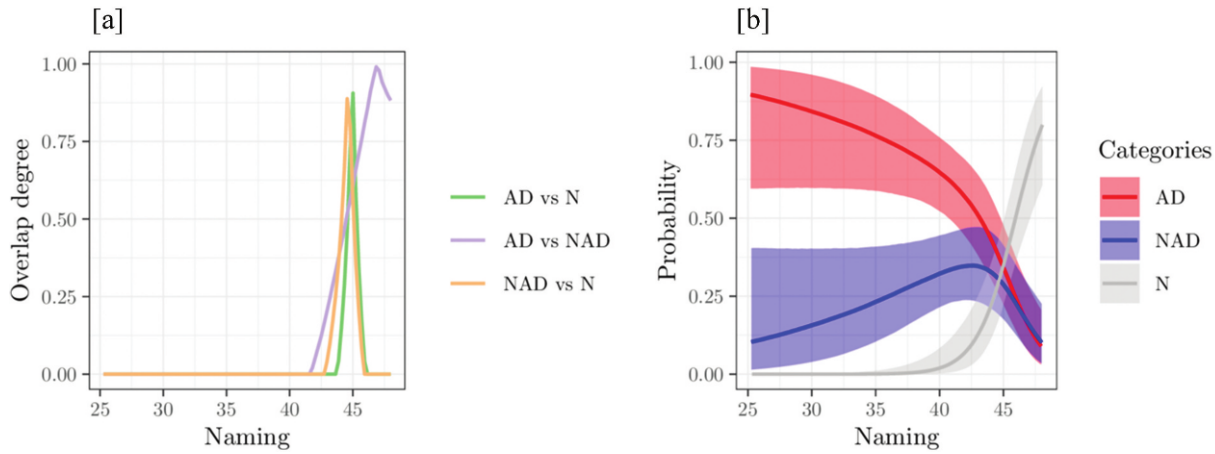

Figure 2. Ability to discriminate between biomarker groups shown by the naming task measure (number of figures correctly named). Panel A: Graphical representation of overlap degrees between 90% credibility interval around estimated probability distributions on the percentage of words retrieved in the first wordlist learning trial (wordlist IR trial1) for groups by pairs: Alzheimer's continuum – non-Alzheimer's disease pathologic change (green line); Alzheimer's continuum – normal biomarker (purple line); non-Alzheimer's disease pathologic change – normal biomarker (orange line). Scores on Wordlist IR trial1 are shown on the X-axis, whereas the overlap degree – expressed as a proportion – between 90% credibility interval around estimated probability distributions for AT(N) groups by pairs is shown on the Y-axis. Panel B: Graphical representation of 90% credibility intervals (colored bands) around estimated probability distributions (colored lines) on wordlist IR trial1 for the Alzheimer's continuum (red), non-Alzheimer's disease pathologic change (blue), and normal biomarker (gray) groups. Scores on measure are shown on the X-axis, whereas the probability of belonging to each of the three ATN groups is shown on the Y-axis. Wordlist IR trial1 = percentage of words retrieved in the first wordlist learning trial; AD = Alzheimer's continuum group; NAD = non-Alzheimer's disease pathologic change group; N = normal biomarker group.

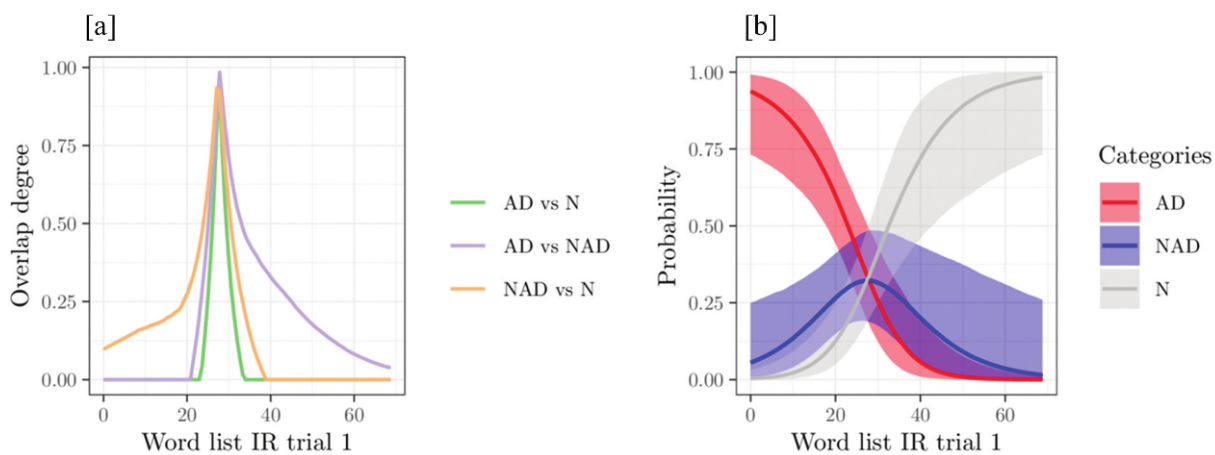

Figure 3. Ability to discriminate between biomarker groups shown by the words retrieved in the first wordlist learning trial measure (percentage). Panel A: Graphical representation of overlap degrees between 90% credibility interval around estimated probability distributions on the percentage of words retrieved in the first wordlist learning trial (wordlist IR trial1) for groups by pairs: Alzheimer's continuum – non-Alzheimer's disease pathologic change (green line); Alzheimer's continuum – normal biomarker (purple line); non-Alzheimer's disease pathologic change – normal biomarker (orange line). Scores on Wordlist IR trial1 are shown on the X-axis, whereas the overlap degree – expressed as a proportion – between 90% credibility interval around estimated probability distributions for AT(N) groups by pairs is shown on the Y-axis. Panel B: Graphical representation of 90% credibility intervals (colored bands) around estimated probability distributions (colored lines) on wordlist IR trial1 for the Alzheimer's continuum (red), non-Alzheimer's disease pathologic change (blue), and normal biomarker (gray) groups. Scores on measure are shown on the X-axis, whereas the probability of belonging to each of the three ATN groups is shown on the Y-axis. Wordlist IR trial1 = percentage of words retrieved in the first wordlist learning trial; AD = Alzheimer's continuum group; NAD = non-Alzheimer's disease pathologic change group; N = normal biomarker group.

distributions of the two groups was only 4% on this task. In particular, Figure 2 shows that 90% credibility intervals around estimated probability distributions did not overlap for scores ≤ 43 and ≥ 46 (Panel A) and that the estimated probability of belonging to the Alzheimer's continuum group was extremely high for lower scores and gradually decreased for higher scores, whereas the opposite was true for the normal biomarker group (Panel B).

The second-best predictor was the percentage of words retrieved in the first learning trial, where the overlap degree between 90% credibility intervals around the estimated probability distributions was 6%. Figure 3 shows that scores ≤ 24% were associated with a high probability of belonging to the Alzheimer's continuum group; conversely, the better the performance (scores ≥ 32%), the higher the probability of belonging to the normal biomarker group.

Similar patterns were also found for other memory measures, i.e., global immediate recall, percentage of words retrieved in the fifth learning trial, delayed recall and wordlist recognition, and Rey complex figure-delayed reproduction. For all of these measures, 90% credibility intervals around the estimated probability distributions did not overlap for low and high score bands, respectively. Thus, the lower the performance, the higher the probability of being in the Alzheimer's continuum group, whereas the better the performance, the higher the probability of belonging to the normal biomarker group (see Supplementary Fig. S2, which shows diagrams of the predictions for all the measures considered, except for those reported in the main text).

The other predictors with an ηAD_N low enough to satisfy our criterion (i.e., < .20) were Rey complex figure-copy, Corsi span and backward digit span, semantic fluency, and number of categories achieved on MCST (ηAD_N = .17, ηAD_N = .09, ηAD_N = .15, ηAD_N = .11, ηAD_N = .19, respectively; see, Table 5; for diagrams, see Supplementary Fig. S2).

Oral naming and Corsi tasks, along with the percentage of words retrieved in the first learning trial and wordlist recognition satisfied the < .20 criterion also when considering η as global average (η = .08, η = .19, η = .14 and η = .14, respectively), meaning that these measures were good discriminators in all between-category comparisons. Conversely, for all the other above-mentioned measures, the global average (η) increased, meaning that their discrimination ability was weaker in the other two between-category comparisons.

Best neuropsychological measures in discriminating between Alzheimer's continuum group and non-AD pathologic change group

The predictor with the highest discrimination ability between Alzheimer's continuum and non-AD pathologic change groups was the wordlist forgetting rate, where the overlap degree between 90% credibility intervals around the estimated probability distributions of the two groups was 15% (i.e., ηAD_NAD = .15; see, Table 5). Figure 4 shows that intervals around estimated probability distributions did not overlap for forgetting rates ≥ 50%, for which participants were more likely to have an Alzheimer's continuum rather than a non-Alzheimer's pathologic change profile. For low forgetting rates, it was instead more difficult to discriminate between the two groups; however, they both had a low probability to be found in this score range.

Cognitive Estimation Test (CET) also ranked among the best-discriminating predictors for the Alzheimer's continuum vs. non-AD pathologic change comparison (ηAD_NAD = .16; see, Table 5). In particular, the range of pathological performance (i.e., scores between ~9 and ~15) was associated with increased likelihood of belonging to the Alzheimer's continuum group as opposed to the non-AD pathologic change group, and intervals around estimated probability distributions did not overlap in this range (see Supplementary Fig. S2).

The naming task showed a good discrimination ability also for this comparison (ηAD_NAD = .16; see, Table 5). Notably, 90% credibility intervals around the estimated probability distributions did not overlap for scores ≤ 42, where the probability of belonging to the Alzheimer's continuum group was higher than that of belonging to the non-AD pathologic change group (see, Figure 2).

The ηAD_N < .20 criterion was also met by its performance on wordlist recognition and phonemic fluency tasks. Supplementary Fig. S2 shows that the score intervals have the highest discrimination ability for these predictors.

Best neuropsychological measures in discriminating between non-AD pathologic change group and normal biomarker group

The naming task was also the best-discriminating predictor for normal biomarkers vs. non-AD pathologic change comparison, with overlap degree between 90% credibility intervals around estimated probability distributions of 6%. Figure 2 shows that the two groups were differentiable, especially for high scores (≥ 46), where participants were much more likely to have normal biomarkers rather than non-AD pathologic change profiles.

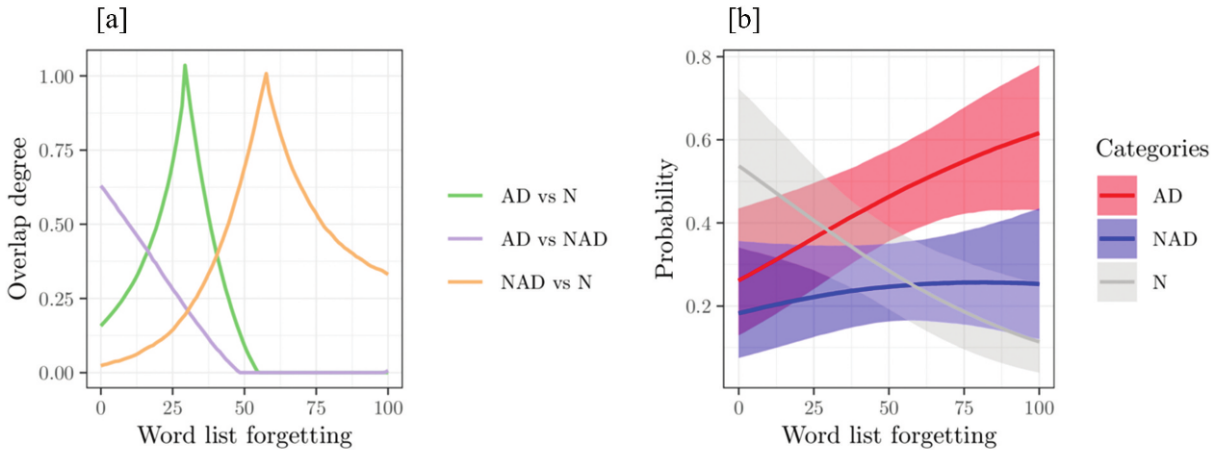

Figure 4. Ability to discriminate between biomarker groups shown by the wordlist forgetting measure (percentage). Panel A: Graphical representation of overlap degrees between 90% credibility interval around estimated probability distributions on the wordlist forgetting rate (wordlist forgetting) for groups by pairs: Alzheimer's continuum – non-Alzheimer's disease pathologic change (green line); Alzheimer's continuum – normal biomarker (purple line); non-Alzheimer's disease pathologic change – normal biomarker (orange line). Scores on wordlist forgetting – adjusted for age, education, and gender, according to Italian normative data – are shown on the X-axis, whereas the overlap degree – expressed as a proportion – between 90% credibility interval around estimated probability distributions for ATN groups by pairs is shown on the Y-axis. Panel B: Graphical representation of 90% credibility intervals (colored bands) around estimated probability distributions (colored lines) on wordlist forgetting for the Alzheimer's continuum (red), non-Alzheimer's disease pathologic change (blue), and normal biomarker (gray) groups. Scores on measure – adjusted for age, education, and gender, according to Italian normative data – are shown on the X-axis, whereas the probability of belonging to each of the three ATN groups is shown on the Y-axis.AD = Alzheimer's continuum group; NAD = non-Alzheimer's disease pathologic change group; N = normal biomarker group.

Wordlist recognition was also a good predictor in this comparison, where the non-AD pathologic change group performed worse than the normal biomarker group. Again, it was especially true for high scores ($\gtrsim$ 28) that the two categories were differentiable, with higher probability of belonging to the normal biomarker group for higher wordlist recognition performance (see Supplementary Fig. S2). A similar pattern was found for another wordlist measure, i.e., the percentage of words retrieved in the first learning trial: the higher the score, the higher the probability of belonging to the normal biomarker group and the lower that of belonging to the non-AD pathologic change group (see, Figure 3).

The other predictors with a $\eta^2_{NAD\_N}$ low enough to satisfy our criterion were the performance on Rey complex figure-delayed reproduction and semantic fluency tasks, and the number of categories achieved on MCST. Supplementary Fig. S2 shows in detail the predictions for these measures.

Discussion

Over the past few years, several studies have examined the relationship between pathophysiological processes underlying AD and its cognitive features by using AD biomarkers as predictor variables of performance on neuropsychological measures (Altomare et al., 2019; Bilgel et al., 2018; Soldan et al., 2016). This approach is justified by the fact that anomalies on AD biomarkers can be identified when individuals are still cognitively normal (Villemagne et al., 2013). Nonetheless, meta-analyses and reviews show that neuropsychological assessment can also detect cognitive deterioration in cognitively normal individuals positive for AD biomarkers (Baker et al., 2017; Mortamais et al., 2017). Furthermore, it is crucial for health policy to consider that biomarker analyses are currently expensive, invasive, and not yet widespread in routine clinical practice, unlike neuropsychological assessment. Only a few studies have recently started to investigate whether the above rationale can be reversed to improve neuropsychological assessment, namely, to explore whether cognitive measures can also be usefully employed to predict AD biomarker status. However, they have so far taken into account only one (Alves et al., 2021; Mueller et al., 2020) or, at most, two (Stricker et al., 2020) classes of AD biomarkers.

For these reasons, in the present study, we aimed to use neuropsychological measures as predictors of the three ATN categories of biomarker profiles, namely Alzheimer's continuum, non-AD pathologic change,

and normal biomarkers – which are biological diagnoses based on the full ATN biomarker scheme – in order to understand whether and which neuropsychological measures could discriminate between them. To the best of our knowledge, our study represents the first attempt in this direction. Our ultimate goal is to promote a more biologically informed use of neuropsychological assessment, which is especially useful when biomarker analysis is not readily available. We adopt a Bayesian statistical approach to identify the most informative scores for each neuropsychological measure in terms of differential diagnosis (i.e., scores implying a higher probability of belonging to a given ATN category), and therefore to obtain useful information to guide the selection of assessment measures and their interpretation in clinical practice. This statistical approach has proven to produce more robust estimations than maximum likelihood methods also in small samples (Van de Schoot & Miocević, 2020), thus increasing the reliability of the resulting interpretation.

Overall, our results indicate that the discrimination ability of each neuropsychological measure varied depending on the pairs of ATN categories compared.

With regard to the Alzheimer's continuum vs. normal biomarkers comparison, we found that the measures with the highest discrimination ability were those obtained from the naming task, wordlist memory task (global immediate recall, percentages of words retrieved in the first and fifth learning trials, delayed recall, and recognition), and Rey complex figure task (copy and delayed reproduction), with Alzheimer's continuum profiled patients scoring worse than normal biomarkers profiled participants in all of them. Although our "control" group consisted of individuals with normal ATN biomarkers but not always cognitively unimpaired (see, Table 2), these findings are consistent with previous studies showing that performance on measures similar to those we detected as more distinctive are impaired also when AD is conventionally conceived as a syndromal construct and participants that meet conventional criteria are compared with healthy controls (e.g., Costa et al., 2017). The convergence of these results suggests that these kinds of neuropsychological measures could be useful to detect Alzheimer's continuum also when the diagnosis is more biologically driven.

Although impairments in episodic memory have historically played a pivotal role in defining AD (Albert et al., 2011; Dubois et al., 2007; G. M. McKhann et al., 2011; G. McKhann et al., 1984) and, consistently, in our study various measures related to episodic memory achieved a good discrimination ability, the best-discriminating predictor in the Alzheimer's continuum vs. normal biomarkers comparison was the oral naming task. Interestingly, Bilgel et al. (2018) found that cognitively normal individuals positive for amyloidosis and neurodegeneration (A + N+) – and, hence, in the Alzheimer's continuum – showed a steeper decline over a three-year period not only in episodic memory measures similar to a subset of those we detected (i.e., wordlist immediate recall and delayed reproduction of figures), but also in naming performance when compared to cognitively normal individuals without biomarker anomalies (A-N-). Overall, these findings suggest that naming assessment, along with wordlists and reproduction of figure measures, may also be useful in distinguishing Alzheimer's continuum from normal biomarker profiles in early clinical stages of the disease.

The naming task was also among the best-discriminating predictors in the Alzheimer's continuum vs. non-AD pathologic change comparison, along with wordlist forgetting rate and recognition, phonemic fluency, and cognitive estimations. Our non-AD pathologic change group included patients with different types of neurodegenerative diseases (see, Table 2). Several studies have shown that patients with clinical AD exhibit greater impairment in naming tasks also when compared to patients with other types of dementia. For example, Brambati et al. (2006) found poorer naming performance in AD patients when compared to both normal controls and patients with other neurodegenerative dementias. The authors also found a significant correlation between naming accuracy and gray matter volumes of areas early affected by AD. Although patients with AD are known to show early semantic deficits in addition to amnestic impairments, caution in the interpretation of this finding should be recommended. Indeed, our Alzheimer's continuum group also included patients with non-amnestic variants (see Participants section) because of the biomarker-based – rather than clinical-based – categorization. This might have slightly increased the naming task ability to discriminate our Alzheimer's continuum group from the other two groups, on the one hand and diminished the potential discrimination ability of episodic memory measures on the other hand. Nonetheless, it is worth remarking that an episodic memory measure, i.e., wordlist forgetting rate, show the highest discrimination ability when considering the Alzheimer's continuum vs. non-AD pathologic change comparison. This finding is consistent with evidence that patients with clinical AD exhibit higher forgetting rates when compared to patients with different forms of dementia but comparable levels of cognitive impairment, especially for items in the very last serial position of a wordlist (i.e., recency forgetting rate; Bruno et al., 2019;

Turchetta et al., 2018). Recency forgetting rate was also found to be strongly correlated with CSF Aβ42 (i.e., the main biomarker to identify Alzheimer's continuum profiles in our study) while being uncorrelated with other CSF biomarkers (i.e., CSF p-tau and t-tau, the main biomarkers to identify also non-AD pathologic change profiles in our study; Bruno et al., 2019). Although further studies are needed to confirm the discrimination ability of forgetting rate, one possibility is that its assessment with measures similar to or even more specific than the one used here (i.e., recency forgetting rate) might be an effective neuropsychological marker for discerning Alzheimer's continuum profiles from those with different anomalies in ATN biomarkers.

Cognitive Estimation Test (CET) also ranked among the best measures for discriminating between the two groups of anomalies in ATN biomarkers. We found that partially or fully pathological performance in this task was associated with increased likelihood of having Alzheimer's continuum profiles rather than non-AD pathologic change profiles. CET was originally developed to reveal estimation ability problems in patients with frontal lobe dysfunctions (Shallice & Evans, 1978). Given that a substantial number of patients in our non-AD pathologic change group suffer from conditions typically associated with damage to the frontal lobes and/or their connections (see, Table 2), this finding seems counterintuitive. However, Mendez et al. (1998) found that patients with clinical AD produced significantly more extreme estimates than patients with frontotemporal dementia. Moreover, several studies have failed to provide evidence that CET is specific to frontal lobe dysfunctions (Margraf et al., 2009; Spencer & Johnson-Greene, 2009; Taylor & O'Carroll, 1995). This is due to the fact that the CET requires not only the recruitment of "frontal" functions such as planning, abstract reasoning, and working memory, but also the retrieval of knowledge from semantic memory to be successfully performed (Wagner et al., 2011). Accordingly, Della Sala et al. (2004) found that performance on CET was correlated with semantic knowledge in both patients with clinical AD and healthy adults, and Brand et al. (2003) reported that Alzheimer's patients were impaired on both general knowledge and estimation abilities also when compared to patients with Korsakoff's syndrome. In summary, several studies suggest that semantic knowledge is the best predictor of CET performance. Therefore, our findings might be consistent with evidence that deficits in semantic abilities are an early and major cognitive feature in AD (Daum et al., 1996; Giffard et al., 2002) even during preclinical stages when identified by biomarkers (Baker et al., 2017; Mortamais et al., 2017; Mueller et al., 2020). This interpretation, although in need of further investigations, is also supported by our findings on the naming task.

Unlike the Alzheimer's continuum category, the non-AD pathologic change category is aimed at clustering profiles with anomalies unrelated to the Alzheimer's spectrum on ATN biomarkers rather than biomarker profiles sharing the same etiopathogenetic processes. Although various conditions may fall under this category, some of the administered measures (i.e., naming, semantic fluency, wordlist recognition, and percentage of words retrieved in the first learning trial, delayed reproduction of Rey complex figure, and number of categories achieved in the Modified Card Sorting Test) nevertheless showed a good performance in discriminating this category from the normal biomarker category. Again, the best predictor in this comparison was the naming task. Altomare et al. (2019) also found that patients in the non-AD pathologic change category (A-T-N+ and A-T + N+) performed lower on language when compared to individuals with normal biomarkers. However, the aforementioned patients also showed poor language performance when compared to A + T-N- profiles (i.e., when compared to profiles within the Alzheimer's continuum), thus suggesting that neurodegeneration alone may lead to language decline. This seems to be partially in line with our findings that non-AD pathologic change profiles were differentiable from normal biomarker profiles on naming and semantic fluency tasks because of their poorer performance. On the other hand, this result is also partially in contrast with our findings that Alzheimer's continuum profiles (always presenting with amyloidosis but not necessarily with neurodegeneration) were differentiable from profiles of the other two ATN categories on naming tasks because of their worse performance. Although both studies point to some language impairments in non-AD patients, their findings are not strictly comparable. Indeed, Altomare et al. (2019) computed a single composite measure for language, by combining scores on phonemic fluency, semantic fluency, and naming tasks, whereas in the present study, we considered these measures separately. We use this approach because it is controversial whether verbal fluency tasks are to be regarded preferably as language or executive function measures (e.g., Aita et al., 2019; Whiteside et al., 2016). Furthermore, Altomare et al. (2019) categorized participants into eight ATN biomarker profiles, whereas in the present study we categorized them into three higher-level ATN categories. Further research is therefore needed to determine whether

and which measures can reliably discriminate between non-AD profiles and both normal biomarker and Alzheimer's continuum profiles.

From a clinical perspective, the present study is based on a heterogeneous sample of patients with suspected cognitive decline, rather than patients meeting more specific criteria, and, since this mirrors more closely what happens in real clinical practice, a good ecological validity is guaranteed. However, because our research has an exploratory intent, our preliminary findings should be considered taking some limitations into account.

Our data were first taken from a medical archive originally designed for clinical, rather than research, purposes. As a consequence, the sample size varied across neuropsychological measures because not all tests were performed by all participants. However, we did our best to minimize the impact of this limitation by computing one prediction model for each neuropsychological measure taken into consideration (to maximize the use of all available data), and by controlling that each prediction model was homogeneous for abnormal biomarker groups' cognitive-staging. The availability of larger samples of individuals fully assessed with a comprehensive neuropsychological battery, in the near future will allow us to establish whether a set of measures exists that, collectively, achieves a discrimination ability between ATN categories higher than that achieved by each single measure.

Secondly, patients with diagnosed psychiatric disorders were excluded from the study. However, given the archival nature of our dataset, we did not have data on specific standardized psychiatric measures for the entire sample. For psychiatric symptoms to be controlled more strictly, future studies should include their formal assessment, especially since some psychiatric syndromes such as anxiety and depression are more common in AD than in the general population of older adults also in early stages.

Third, as our sample was relatively modest in size, we considered the three higher-level ATN categories rather than the eight ATN biomarker profiles. Although it has been argued that the ATN profiles included in the three larger ATN categories tend to overlap for clinical features (Altomare et al., 2019), we cannot rule out the possibility that results could be partially different if participants were classified into more fine-grained groups. For example, some data suggest that profiles with tauopathy in isolation do not seem to differ in baseline and longitudinal cognitive performance from normal profiles (Altomare et al., 2019; Pascoal et al., 2017).

Another potential limitation of our study is that patients with non-amnestic variants in our Alzheimer's continuum group may have decreased the discrimination ability of wordlist measures and/or increased that of naming performance, as previous studies have reported impairments or steeper cognitive decline more frequently in the former than in the latter for individuals with amyloidosis (i.e., with Alzheimer's continuum profiles; Baker et al., 2017; Mortamais et al., 2017). Again, our non-AD pathologic change group included a substantial number of patients with frontotemporal dementia (eight out of 21) compared to the normal biomarker category (one out of 20). Previous studies have reported different proportions (Carandini et al., 2019; Cousins et al., 2020), and it is not yet clear whether individuals with frontotemporal dementia are more likely to show normal or non-AD pathologic change profiles. However, these limitations are somewhat inherent in the ATN classification system, which does not always match more clinically driven categorizations. Further research should clarify how frequently clinically relevant conditions fall within each of the two non-Alzheimer's continuum ATN categories.

Unlike previous studies employing neuropsychological measures aimed to predict categorizations based on biomarker status, we explored a heterogeneous sample of patients with suspected cognitive decline who are presented for diagnostic examination rather than a selected sample of cognitively unimpaired (Mueller et al. 2020; Stricker et al., 2020) or aMCI individuals (Alves et al., 2021). As a consequence, our findings may not be extended to selectively impaired or unimpaired people. On the other hand, we look at this heterogeneity more as a strength than as a limitation, since our sample reflects more closely the actual circumstances of clinical practice and, consequently, our findings may be more salient to clinicians.

Finally, because the study is based on archival data from a relatively small sample, it does not provide neuropsychological cutoff values that can be applied in clinical practice. Nonetheless, the study's general aim was to provide initial indications of the neuropsychological measures that should be administered to increase the likelihood of a biological diagnosis of AD.

Conclusion

Within the context of neuropsychological research, we proposed a novel methodological approach potentially able to lead to a more biologically informed use of neuropsychological assessment. We employed a Bayesian approach to explore whether performance on neuropsychological measures can be successfully

used to predict categorization of individuals into the three ATN categories (i.e., their biological diagnoses), and some investigated measures seem to be promising. On the one hand, our findings are relatively consistent with previous studies that examined cognitive features of individuals with clinical (or mixed) rather than biological diagnosis of AD. Indeed, we identified similar cognitive impairments on comparable neuropsychological measures also in patients with Alzheimer's continuum profiles. On the other side, given that the discrimination ability of each neuropsychological measure varied depending on the biological diagnosis being compared, our results also confirm once again the importance of considering multiple measures to better detail and characterize individuals' cognitive profile (Lezak, 1995), rather than relying on composite scores (which often provide an oversimplified picture of the status of specific components and processes within a single cognitive domain).

The proposed methodological approach has the potential to identify neuropsychological measures that individually or collectively, can reliably estimate the presence of specific types of neuropathological abnormalities in larger samples as documented in vivo by biomarkers, and ultimately to provide crucial diagnostic information even when data on biomarkers are lacking. This could increase the diagnostic accuracy of AD, especially when biomarker analyses are not available in routine clinical practice, thus facilitating wider access to more targeted treatments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Barbara Penolazzi http://orcid.org/0000-0001-8717-6233

References

Aarsland, D., Emre, M., Lees, A., Poewe, W., & Ballard, C. (2007). Practice parameter: Evaluation and treatment of depression, psychosis, and dementia in Parkinson disease (an evidence-based review): Report of the quality standards subcommittee of the American academy of neurology. *Neurology*, *68*(1), 80–81. https://doi.org/10.1212/01.wnl.0000252956.34585.a6

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology: Adult*, *26*(5), 441–451. https://doi.org/10.1080/23279095.2018.1439830

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Altomare, D., de Wilde, A., Ossenkoppele, R., Pelkmans, W., Bouwman, F., Groot, C., van Maurik, I., Zwan, M., Yaqub, M., Barkhof, F., van Berckel, B. N., Teunissen, C. E., Frisoni, G. B., Scheltens, P., & van der Flier, W. M. (2019). Applying the ATN scheme in a memory clinic population: The ABIDE project. *Neurology*, *93*(17), e1635–e1646. https://doi.org/10.1212/WNL.0000000000008361

Alves, L., Cardoso, S., Silva, D., Mendes, T., Marôco, J., Nogueira, J., Lima, M., Tábuas-Pereira, M., Baldeiras, I., Santana, I., Mendonça, A., & Guerreiro, M. (2021). Neuropsychological profile of amyloid-positive versus amyloid-negative amnestic Mild Cognitive Impairment. *Journal of Neuropsychology*, *15*(S1), 41–52. https://doi.org/10.1111/jnp.12218

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, *73*(sup1), 262–270. https://doi.org/10.1080/00031305.2018.1543137

Baker, J. E., Lim, Y. Y., Pietrzak, R. H., Hassenstab, J., Snyder, P. J., Masters, C. L., & Maruff, P. (2017). Cognitive impairment and decline in cognitively normal older adults with high amyloid-β: A meta-analysis. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *6*(1), 108–121. https://doi.org/10.1016/j.dadm.2016.09.002

Bilgel, M., An, Y., Helphrey, J., Elkins, W., Gomez, G., Wong, D. F., … Resnick, S. M. (2018). Effects of amyloid pathology and neurodegeneration on cognitive change in cognitively normal adults. *Brain*, *141*(8), 2475–2485. https://doi.org/10.1093/brain/awy150

Blennow, K., Hampel, H., Weiner, M., & Zetterberg, H. (2010). Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology*, *6*(3), 131–144. https://doi.org/10.1038/nrneurol.2010.4

Brambati, S. M., Myers, D., Wilson, A., Rankin, K. P., Allison, S. C., Rosen, H. J., Miller, B. L., & Gorno-Tempini, M. L. (2006). The anatomy of category-specific object naming in neurodegenerative diseases. *Journal of Cognitive Neuroscience*, *18*(10), 1644–1653. https://doi.org/10.1162/jocn.2006.18.10.1644

Brand, M., Kalbe, E., Fujiwara, E., Huber, M., & Markowitsch, H. J. (2003). Cognitive estimation in patients with probable Alzheimer's disease and alcoholic Korsakoff patients. *Neuropsychologia*, *41*(5), 575–584. https://doi.org/10.1016/s0028-3932(02)00183-5

Brugnolo, A., De Carli, F., Accardo, J., Amore, M., Bosia, L. E., Bruzzaniti, C., Cappa, S. F., Cocito, L., Colazzo, G., Ferrara, M., Ghio, L., Magi, E., Mancardi, G. L., Nobili, F., Pardini, M., Rissotto, R.,

Serrati, C., & Girtler, N. (2016). An updated Italian normative dataset for the Stroop color word test (SCWT). *Neurological Sciences*, *37*(3), 365–372. https://doi.org/10.1007/s10072-015-2428-2

Bruno, D., Gleason, C. E., Koscik, R. L., Pomara, N., Zetterberg, H., Blennow, K., & Johnson, S. C. (2019). The recency ratio is related to CSF amyloid beta 1-42 levels in MCI-AD. *International Journal of Geriatric Psychiatry*, *34*(3), 415–419. https://doi.org/10.1002/gps.5029

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., & Venneri, A. (2002). Rey-Osterrieth complex figure: Normative values in an Italian population sample. *Neurological Sciences*, *22*(6), 443–447. https://doi.org/10.1007/s100720200003

Caffarra, P., Vezzadini, G., Dieci, F., Zonato, F., & Venneri, A. (2004). Modified card sorting test: Normative data. *Journal of Clinical and Experimental Neuropsychology*, *26*(2), 246–250. https://doi.org/10.1076/jcen.26.2.246.28087

Carandini, T., Arighi, A., Sacchi, L., Fumagalli, G. G., Pietroboni, A. M., Ghezzi, L., … Galimberti, D. (2019). Testing the 2018 NIA-AA research framework in a retrospective large cohort of patients with cognitive impairment: From biological biomarkers to clinical syndromes. *Alzheimer's Research & Therapy*, *11*(1), 1–11. https://doi.org/10.1186/s13195-019-0543-7

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Catricala, E., Della Rosa, P. A., Ginex, V., Mussetti, Z., Plebani, V., & Cappa, S. F. (2013). An Italian battery for the assessment of semantic memory disorders. *Neurological Sciences*, *34*(6), 985–993. https://doi.org/10.1007/s10072-012-1181-z

Costa, A., Bagoj, E., Monaco, M., Zabberoni, S., De Rosa, S., Papantonio, A. M., Mundi, C., Caltagirone, C., & Carlesimo, G. A. (2014). Standardization and normative data obtained in the Italian population for a new verbal fluency instrument, the phonemic/semantic alternate fluency test. *Neurological Sciences*, *35*(3), 365–372. https://doi.org/10.1007/s10072-013-1520-8

Costa, A., Bak, T., Caffarra, P., Caltagirone, C., Ceccaldi, M., Collette, F., … Cappa, S. F. (2017). The need for harmonisation and innovation of neuropsychological assessment in neurodegenerative dementias in Europe: Consensus document of the Joint Program for Neurodegenerative Diseases Working Group. *Alzheimer's Research & Therapy*, *9*(1), 1–15. https://doi.org/10.1186/s13195-017-0254-x

Cousins, K. A., Irwin, D. J., Wolk, D. A., Lee, E. B., Shaw, L. M., Trojanowski, J. Q., Da Re, F., Gibbons, G. S., Grossman, M., & Phillips, J. S. (2020). ATN status in amnestic and non-amnestic Alzheimer's disease and frontotemporal lobar degeneration. *Brain*, *143*(7), 2295–2311. https://doi.org/10.1093/brain/awaa165

Cramer, H. (1946). *Mathematical methods of statistics*. Princeton Univ. Press.

Cui, B. (2020). *DataExplorer: Automate data exploration and treatment*. R package version 0.8.1. https://CRAN.R-project.org/package=DataExplorer

Daum, I., Riesch, G., Sartori, G., & Birbaumer, N. (1996). Semantic memory impairment in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *18*(5), 648–665. https://doi.org/10.1080/01688639608408289

Della Sala, S., MacPherson, S. E., Phillips, L. H., Sacco, L., & Spinnler, H. (2004). The role of semantic knowledge on the cognitive estimation task. *Journal of Neurology*, *251*(2), 156–164. https://doi.org/10.1007/s00415-004-0292-8

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, *6*(8), 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3

Duits, F. H., Teunissen, C. E., Bouwman, F. H., Visser, P. J., Mattsson, N., Zetterberg, H., van der Flier, W. M., Minthon, L., Andreasen, N., Marcusson, J., Wallin, A., Rikkert, M. O., Tsolaki, M., Parnetti, L., Herukka, S.-K., Hampel, H., De Leon, M. J., Schröder, J., Aarsland, D., … Blennow, K. (2014). The cerebrospinal fluid "Alzheimer profile": Easily said, but what does it mean? *Alzheimer's & Dementia*, *10*(6), 713–723. https://doi.org/10.1016/j.jalz.2013.12.023

Emre, M., Aarsland, D., Brown, R., Burn, D. J., Duyckaerts, C., Mizuno, Y., Broe, G. A., Cummings, J., Dickson, D. W., Gauthier, S., Goldman, J., Goetz, C., Korczyn, A., Lees, A., Levy, R., Litvan, I., McKeith, I., Olanow, W., Poewe, W., … Dubois, B. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *22*(12), 1689–1707. https://doi.org/10.1002/mds.21507

Fullman, N., Yearwood, J., Abay, S. M., Abbafati, C., Abd-Allah, F., Abdela, J., … Chang, H. Y. (2018). Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: A systematic analysis from the Global Burden of Disease Study 2016. *The Lancet*, *391*(10136), 2236–2271. https://doi.org/10.1016/S0140-6736(18)30994-2

Giffard, B., Desgranges, B., Nore-Mary, F., Lalevée, C., Beaunieux, H., de la Sayette, V., … Eustache, F. (2002). The dynamic time course of semantic memory impairment in Alzheimer's disease: Clues from hyperpriming and hypopriming effects. *Brain*, *125*(9), 2044–2057. https://doi.org/10.1093/brain/awf209

Giovagnoli, A. R., Del Pesce, M., Mascheroni, S., Simoncelli, M., Laiacona, M., & Capitani, E. (1996). Trail making test: Normative values from 287 normal adult controls. *The Italian Journal of Neurological Sciences*, *17*(4), 305–309. https://doi.org/10.1007/BF01997792

Jack, C. R., Jr, Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., Hampel, H., Jagust, W. J., Johnson, K. A., Knopman, D. S., Petersen, R. C., Scheltens, P., Sperling, R. A., & Dubois, B. (2016). A/T/N An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539–547. https://doi.org/10.1212/WNL.0000000000002923

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). Oxford University Press.

Margraf, N., Bachmann, T., Schwandner, W., Gottschalk, S., & Seidel, G. (2009). Bedside screening for executive dysfunction in patients with subcortical ischemic vascular disease. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, *24*(9), 1002–1009. https://doi.org/10.1002/gps.2212

Mauri, M., Carlesino, G. A., Graceffa, A. M. S., Loasses, A., Lorusso, S., Sinforiani, E., … Caltagirone, C. (1997). Standardizzazione di due nuovi test di memoria: Apprendimento di liste di parole correlate e non correlate semanticamente. *Archivio Di Psicologia Neurologia E Psichiatria*, *58*, 621–645.

McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J. P., Weintraub, D., … Kosaka, K. (2017). Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology*, *89*(1), 88–100. https://doi.org/10.1212/WNL.0000000000004058

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939-939. https://doi.org/10.1212/wnl.34.7.939

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr, Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). 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. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Mendez, M. F., Doss, R. C., & Cherrier, M. M. (1998). Use of the cognitive estimations test to discriminate frontotemporal dementia from Alzheimer's disease. *Journal of Geriatric Psychiatry and Neurology*, *11*(1), 2–6. https://doi.org/10.1177/089198879801100102

Monaco, M., Costa, A., Caltagirone, C., & Carlesimo, G. A. (2015). "Forward and backward span for verbal and visuo-spatial data: Standardization and normative data from an Italian adult population". *Erratum. Neurological Sciences*, 36(2), 345–347. https://doi.org/10.1007/s10072-014-2019-7

Mortamais, M., Ash, J. A., Harrison, J., Kaye, J., Kramer, J., Randolph, C., Pose, C., Albala, B., Ropacki, M., Ritchie, C. W., & Ritchie, K. (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, *13*(4), 468–492. https://doi.org/10.1016/j.jalz.2016.06.2365

Moscoso, A., Silva-Rodriguez, J., Aldrey, J. M., Cortes, J., Fernandez-Ferreiro, A., Gomez-Lado, N., Ruibal, Á., & Aguiar, P., & Alzheimer's Disease Neuroimaging Initiative. (2019). Staging the cognitive continuum in prodromal Alzheimer's disease with episodic memory. *Neurobiology of Aging*, *84*, 1–8. https://doi.org/10.1016/j.neurobiolaging.2019.07.014

Mueller, K. D., Koscik, R. L., Du, L., Bruno, D., Jonaitis, E. M., Koscik, A. Z., Christian, B. T., Betthauser, T. J., Chin, N. A., Hermann, B. P., & Johnson, S. C. (2020). Proper names from story recall are associated with beta-amyloid in cognitively unimpaired adults at risk for Alzheimer's disease. *Cortex*, *131*, 137–150. https://doi.org/10.1016/j.cortex.2020.07.008

Mulder, C., Verwey, N. A., van der Flier, W. M., Bouwman, F. H., Kok, A., van Elk, E. J., Scheltens, P, and Blankenstein, M. A. (2010). Amyloid-β(1–42), Total Tau, and Phosphorylated Tau as Cerebrospinal Fluid Biomarkers for the Diagnosis of Alzheimer Disease. *Clinical Chemistry*, *56*(2), 248–253. 10.1373/clinchem.2009.130518.

Pascoal, T. A., Mathotaarachchi, S., Shin, M., Benedet, A. L., Mohades, S., Wang, S., Beaudry, T., Kang, M. S., Soucy, J.-P., Labbe, A., Gauthier, S., & Rosa-Neto, P., & Alzheimer's Disease Neuroimaging Initiative. (2017). Synergistic interaction between amyloid and tau predicts the progression to dementia. *Alzheimer's & Dementia*, *13*(6), 644–653. doi:10.1016/j.jalz.2016.11.005.

R Core Team. (2019). *R: A language and environment for statistical computing [Computer software manual]*. https://www.R-project.org/

Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., van Swieten, J. C., Seelaar, H., Dopper, E. G. P., Onyike, C. U., Hillis, A. E., Josephs, K. A., Boeve, B. F., Kertesz, A., Seeley, W. W., Rankin, K. P., Johnson, J. K., Gorno-Tempini, M.-L., Rosen, H., … Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, *134*(9), 2456–2477. https://doi.org/10.1093/brain/awr179

Revelle, W. (2019). *psych: Procedures for psychological, psychometric, and personality research [Computer software manual]*. (R package version 1.9.12). https://CRAN.R-project.org/package=psych

Sancesario, G. M., Toniolo, S., Chiasserini, D., Di Santo, S. G., Zegeer, J., Bernardi, G., Musicco, M., Caltagirone, C., Parnetti, L., & Bernardini, S. (2017). The clinical use of cerebrospinal fluid biomarkers for Alzheimer's Disease diagnosis: The Italian selfie. *Journal of Alzheimer's Disease*, *55*(4), 1659–1666. https://doi.org/10.3233/JAD-160975

Scarpina, F., D'Aniello, G. E., Mauro, A., Castelnuovo, G., & MacPherson, S. E. (2015). How many segments are there in an Orange: Normative data for the new Cognitive Estimation Task in an Italian population. *Neurological Sciences*, *36*(10), 1889–1895. https://doi.org/10.1007/s10072-015-2276-0

Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E., … Larmarange, J. (2020). *Ggally: Extension to 'ggplot2' [Computer software manual]*. (R package version 1.5.0). https://CRAN.R-project.org/package=GGally

Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex*, *14*(2), 294–303. https://doi.org/10.1016/s0010-9452(78)80055-0

Soldan, A., Pettigrew, C., Cai, Q., Wang, M. C., Moghekar, A. R., O'Brien, R. J., Selnes, O. A., & Albert, M. S., & BIOCARD Research Team. (2016). Hypothetical preclinical Alzheimer disease groups and longitudinal cognitive change. *JAMA Neurology*, *73*(6), 698–705. doi:10.1001/jamaneurol.2016.0194.

Spencer, R. J., & Johnson-Greene, D. (2009). The Cognitive Estimation Test (CET): Psychometric limitations in neurorehabilitation populations. *Journal of Clinical and Experimental Neuropsychology*, *31*(3), 373–377. https://doi.org/10.1080/13803390802206398

Spinnler, H., & Tognoni, G. (1987). *Standardizzazione e taratura italiana di test neuropsicologici: Gruppo italiano per lo studio neuropsicologico dell'invecchiamento*. Masson Italia Periodici.

Stan Development Team. (2018). *Stan modeling language user's guide and reference manual*. (Version 2.18.0). http://mc-stan.org/

Stricker, N. H., Lundt, E. S., Albertson, S. M., Machulda, M. M., Pudumjee, S. B., Kremers, W. K., … Mielke, M. M. (2020). Diagnostic and prognostic accuracy of the Cogstate Brief Battery and Auditory Verbal Learning Test in preclinical Alzheimer's disease and incident Mild Cognitive Impairment: Implications for defining subtle objective cognitive impairment. *Journal of Alzheimer's Disease*, *76*(1), 261–274. https://doi.org/10.3233/JAD-200087

T O'Brien, J., Erkinjuntti, T., Reisberg, B., Roman, G., Sawada, T., Pantoni, L., Bowler, J. V., Ballard, C., DeCarli, C., Gorelick, P. B., Rockwood, K., Burns, A., Gauthier, S., & DeKosky, S. T. (2003). Vascular cognitive impairment. *The Lancet Neurology*, *2*(2), 89–98. https://doi.org/10.1016/s1474-4422(03)00305-3

Taylor, R., & O'Carroll, R. (1995). Cognitive estimation in neurological disorders. *British Journal of Clinical Psychology*, *34*(2), 223–228. https://doi.org/10.1111/j.2044-8260.1995.tb01456.x

Turchetta, C. S., Perri, R., Fadda, L., Caruso, G., De Simone, M. S., Caltagirone, C., & Carlesimo, G. A. (2018). Forgetting rate on the recency portion of a word list differentiates mild to moderate Alzheimer's disease from other forms of dementi. *Journal of Alzheimer's Disease*, *66*(2), 461–470. https://doi.org/10.3233/JAD-180690

Van de Schoot, R., & Miocević, M. (2020). *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 284). Taylor & Francis.

Villemagne, V. L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K. A., Salvado, O., Szoeke, C., Macaulay, S. L., Martins, R., Maruff, P., Ames, D., Rowe, C. C., & Masters, C. L. (2013). Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study. *The Lancet Neurology*, *12*(4), 357–367. https://doi.org/10.1016/S1474-4422(13)70044-9

Wagner, G. P., MacPherson, S. E., Parente, M. A. M., & Trentini, C. M. (2011). Cognitive estimation abilities in healthy and clinical populations: The use of the Cognitive Estimation Test. *Neurological Sciences*, *32*(2), 203–210. https://doi.org/10.1007/s10072-010-0449-4

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., & Bryan, J. (2019). *readxl: Read excel files [Computer software manual]*. (R package version 1.3.1). https://CRAN.R-project.org/package=readxl