FEATURED ARTICLE

Harmonizing neuropsychological assessment for mild neurocognitive disorders in Europe

Marina Boccardi[1,2] | **Andreas U. Monsch**[3] | **Clarissa Ferrari**[4] | **Daniele Altomare**[2,5] | **Manfred Berres**[6] | **Isabelle Bos**[7] | **Andreas Buchmann**[8] | **Chiara Cerami**[9,10] | **Mira Didic**[11,12] | **Cristina Festari**[13] | **Valentina Nicolosi**[13] | **Leonardo Sacco**[14] | **Liesbeth Aerts**[15] | **Emiliano Albanese**[16] | **Jean-Marie Annoni**[17] | **Nicola Ballhausen**[18] | **Christian Chicherio**[5] | **Jean-François Démonet**[19] | **Virginie Descloux**[17] | **Suzie Diener**[20] | **Daniel Ferreira**[21] | **Jean Georges**[22] | **Anton Gietl**[8] | **Nicola Girtler**[23,35] | **Ingo Kilimann**[1] | **Stefan Klöppel**[25] | **Nicole Kustyniuk**[25] | **Patrizia Mecocci**[26] | **Nathalie Mella**[27] | **Martina Pigliautile**[26] | **Katrin Seeher**[15] | **Steven D. Shirk**[28] | **Alessio Toraldo**[29] | **Andrea Brioschi-Guevara**[19] | **Kwun C. G. Chan**[30] | **Paul K. Crane**[31] | **Alessandra Dodich**[32,33] | **Alice Grazia**[1] | **Nicole A Kochan**[15] | **Fabricio Ferreira de Oliveira**[34] | **Flavio Nobili**[24,35] | **Walter Kukull**[30] | **Oliver Peters**[36] | **Inez Ramakers**[7] | **Perminder S. Sachdev**[15] | **Stefan Teipel**[1] | **Pieter Jelle Visser**[7] | **Michael Wagner**[37,38] | **Sandra Weintraub**[39] | **Eric Westman**[21] | **Lutz Froelich**[40] | **Henry Brodaty**[15] | **Bruno Dubois**[41] | **Stefano F. Cappa**[9,10] | **David Salmon**[42] | **Bengt Winblad**[43] | **Giovanni B. Frisoni**[2,5] | **Matthias Kliegel**[44] | on behalf of the Consortium for the Harmonization of Neuropsychological Assessment for Neurocognitive Disorders (https://nextcloud.dzne.de/index.php/s/EwXjLab9caQTbQe)

[1] DZNE - Deutsches Zentrum für Neurodegenerative Erkrankungen, Rostock-Greifswald site, Rostock, Germany

[2] LANVIE - Laboratory of Neuroimaging of Aging, University of Geneva, Geneva, Switzerland

[3] Memory Clinic, University Department of Geriatric Medicine FELIX PLATTER, Faculty of Psychology, University of Basel, Basel, Switzerland

[4] Unit of Statistics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[5] Memory Center, Geneva University Hospitals, Geneva, Switzerland

[6] Department of Mathematics and Technology, University of Applied Sciences Koblenz, Koblenz, Germany

[7] Department of Psychiatry and Neuropsychology, School of Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht University, Maastricht, The Netherlands

[8] Institute for Regenerative MedicineUniversity of Zurich, Schlieren, Switzerland

[9] Institute for Advanced Studies (IUSS-Pavia), Pavia, Italy, Pavia, Italy

[10] IRCCS Mondino Foundation, Pavia, Italy

[11] APHM, Timone, Service de Neurologie et Neuropsychologie, Hôpital Timone Adultes, Marseille, France

[12] Aix-Marseille Université, Inserm, INS, UMR_S 1106, 13005, Marseille, France

[13] Laboratory of Alzheimer's Neuroimaging and Epidemiology, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[14] Clinic of Neurology, Neurocenter of Southern Switzerland, EOC, Lugano, Switzerland

[15] Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, Australia

[16] Department of Psychiatry, University of Geneva, Geneva, Switzerland

[17] Department of Neuroscience and Movement Sciences, University of Geneva and Fribourg Hospital, Geneva, Switzerland

[18] Department of Developmental Psychology, Tilburg University, Tilburg, The Netherlands

[19] Leenaards Memory Centre-CHUV, Clinical Neuroscience Department, Cité Hospitalière CHUV, Lausanne, Switzerland

[20] Department of Neurology, Kantonsspital St. Gallen, St. Gallen, Switzerland

[21] Division of Clinical Geriatrics, Center for Alzheimer Research, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[22] Alzheimer Europe, Luxembourg, Luxembourg

[23] Clinical Psychology and Psychotherapy, IRCCS Ospedale Policlinico San Martino, Genova, Italy

[24] Neurology Clinic, IRCCS Ospedale Policlinico San Martino, Genova, Italy

[25] Hospital of Old Age Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland

[26] Department of Medicine and Surgery, Institute of Gerontology and GeriatricsUniversity of Perugia, Perugia, Italy

[27] Cognitive Aging Lab, University of Geneva, Geneva, Switzerland

[28] VISN 1 New England MIRECC and VISN 1 New England GRECC, Bedford VA Healthcare System, Bedford, Department of Psychiatry and Population and Quantitative Health Sciences, University of Massachusetts Medical School, Massachusetts, USA

[29] Department of Brain and Behavioural Sciences, University of Pavia, Pavia, ItalyMilan Center for Neuroscience, Milan, Italy

[30] National Alzheimer's Coordination Center (NACC), Department of Epidemiology, University of Washington, Seattle, Washington, USA

[31] Department of Medicine, University of Washington, Seattle, Washington, USA

[32] Neuroimaging and Innovative Molecular Tracers Laboratory, and Division of Nuclear Medicine, Diagnostic Departement, University of GenevaUniversity Hospitals of Geneva, Geneva, Switzerland

[33] Centre for Mind/Brain Sciences, University of Trento, Rovereto, Italy

[34] Department of Neurology and Neurosurgery, Federal University of São Paulo (UNIFESP), São Paulo, Brazil

[35] Dept of Neuroscience (DINOGMI), University of Genoa, Genoa, Italy

[36] Department of Psychiatry and Psychotherapy, Campus Benjamin Franklin, CharitéUniversitätsmedizin Berlin, Berlin, GermanyZNE, German Center for Neurodegenerative Diseases, Berlin, Germany

[37] DZNE, German Center for Neurodegenerative Diseases, Bonn, Germany

[38] Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany

[39] Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern Feinberg School of Medicine, Chicago, Illinois

[40] University of Heidelberg, Heidelberg, Central Institute of Mental Health, Medical Faculty Mannheim, Mannheim, Germany

[41] Hôpital Pitié-Salpêtrière, AP-HP, Alzheimer Research Institute (IM2A) and Institut du cerveau et la moelle (ICM) Sorbonne Université, Paris, France

[42] Department of Neurosciences, University of California San Diego School of Medicine, San Diego, USA

[43] Dept NVS, Center for Alzheimer Research, Division of Neurogeriatrics, Karolinska Institutet, Stockholm, Sweden

[44] Cognitive Aging Lab, Department of Psychology, University of Geneva, Geneva, Switzerland

Correspondence
Marina Boccardi, Deutsches Zentrum für Neurodegenerative Erkrankungen e.V. (DZNE) in der Helmholtz-Gemeinschaft - Standort Rostock/ Greifswald, Gehlsheimer Straße 20, D- 18147 Rostock, Germany.
Email: marina.boccardi@dzne.de

Abstract

Introduction: Harmonized neuropsychological assessment for neurocognitive disorders, an international priority for valid and reliable diagnostic procedures, has been achieved only in specific countries or research contexts.

Methods: To harmonize the assessment of mild cognitive impairment in Europe, a workshop (Geneva, May 2018) convened stakeholders, methodologists, academic, and non-academic clinicians and experts from European, US, and Australian harmonization initiatives.

Results: With formal presentations and thematic working-groups we defined a standard battery consistent with the U.S. Uniform DataSet, version 3, and homogeneous methodology to obtain consistent normative data across tests and languages. Adaptations consist of including two tests specific to typical Alzheimer's disease and behavioral variant frontotemporal dementia. The methodology for harmonized normative data includes consensus definition of cognitively normal controls, classification of confounding factors (age, sex, and education), and calculation of minimum sample sizes.

Discussion: This expert consensus allows harmonizing the diagnosis of neurocognitive disorders across European countries and possibly beyond.

KEYWORDS
Alzheimer's disease, cognitive assessment, diagnosis, mild cognitive impairment, mild neurocognitive disorders, standard neuropsychological assessment

1 | INTRODUCTION

A key early step in the clinical diagnostic process for persons with cognitive complaints who are referred to memory clinics consists of ascertaining the presence of objective cognitive impairment[1,2] by examining performance on a set of neuropsychological tests. As shown in Supplemental Table S1, different tests are more sensitive to different disorders. Heterogeneous batteries could therefore result in different diagnoses for patients. Reliable clinical actions require that different diagnostic procedures operationalize the definition of the target clinical disorder consistently. One way to accomplish this is by consistent test selection. Such standard procedures would enable the use of biomarkers and treatment in line with their demonstrated informative or therapeutic value, and uniformly across centers. In the same way, valid research procedures require consistent operationalization of the same clinical construct across research settings. It is therefore desirable that the selection of patients who are eligible for the full diagnostic procedure or for research studies be based on a standard common neuropsychological assessment, operationalizing the target condition consistently.[3,4] Post hoc computations permit alignment of scores from heterogeneous batteries and pooling of data from different centers for research aims,[5,6] thereby facilitating analyses of large multi-site data sets. Such computations, however, cannot amend the upstream inclusion of heterogeneous patients.

Many efforts have been made to tackle this problem, providing resources to support harmonization (Supplemental Table S2). For example, many US research centers have been able to standardize neuropsychological assessments on a large scale.[7] Similarly, German-speaking countries widely adopted the Consortium to Establish a Registry for Alzheimer's Disease–Neuropsychological Assessment Battery (CERAD-NAB) for the diagnosis of patients with dementia (Supplemental Table S2), and recently, a Chinese effort defined a standard battery for clinical use.[8] This work aims toward standardizing neuropsychological assessments for detecting mild cognitive impairment (MCI) consistently in people attending European memory clinics with cognitive complaints. We leveraged previous initiatives and incorporated the complementary expertise of academic and non-academic clinicians to provide standard procedures that will reduce costs and effort in clinics and benefit research activities.

2 | METHODS

This initiative follows the Strategic Biomarker Roadmap, a methodological framework specific to biomarker validation, adapted from oncology to the field of dementia. This framework outlined the appropriate sequence of validation steps for diagnostic biomarkers, and the priority of standardizing neuropsychological assessment as a prerequisite for their proper validation.[4,9] Because many studies of clinical validity and utility are performed on patients from memory clinics, harmonizing neuropsychological assessment for the clinical settings would

RESEARCH IN CONTEXT

1. **Systematic review**: With inclusive strings, we identified literature, resources, projects, and participants for a workshop to harmonize neuropsychological assessment for European clinics, as existing initiatives were either limited to individual countries or to research settings (Table S2).

2. **Interpretation**: Our consensus Clinician's Uniform Dataset (cUDS), similar to UDS-3, and our methodology for generating harmonized norms would (a) improve detection of Alzheimer's disease (AD) and of non-AD or atypical-AD syndromes in mild cognitive impairment (MCI); (b) reduce costs; (c) benefit patients, health care systems, and clinical research within a consistent framework; (d) align clinical and research procedures; and (e) achieve modern, reliable, and cost-effective standard of care for neurocognitive disorders.

3. **Future directions** will consist of exploring hurdles and needs to implement the cUDS in academic and non-academic memory clinics, creating and validating local versions for European languages, and creating tools to support adoption.

have double benefit, improving research as well as clinical procedures at once.

A workshop (Geneva, May 9-11, 2018) was hosted by the European Alzheimer's Disease Consortium (EADC), the Geneva Memory Clinic, the Centre Interfacultaire de Gérontologie et d'Études de Vulnérabilité (CIGEV), and Swiss Memory Clinics. Participants were European dementia experts—physicians, neurologists, (neuro)psychologists, psychiatrists, geriatricians—from non-academic and academic memory clinics, researchers from previous pertinent harmonization initiatives worldwide, methodologists, and stakeholders (https://cigev.unige.ch/files/5015/3788/2053/hnade.pdf) (see Supplemental Box S1 for institutions and their representatives and Supplemental Box S2 for individual participants and affiliations). At the workshop, presenters described the methods, results, issues, and resources from previous harmonization initiatives, and current development of tests specific to dementing neurodegenerative disorders in plenary sessions. After the plenaries, participants were assigned to specific thematic subgroups based on their expertise and leveraging published evidence, to work in parallel and find solutions to harmonize the aspects specific to their competence (neuropsychology experts: test selection; statistics experts: modeling for the generation of normative values; digital working group: potential and issues on digital-assisted testing).

All of the consensual decisions reported in Results derive from the following procedure (Figure 1).
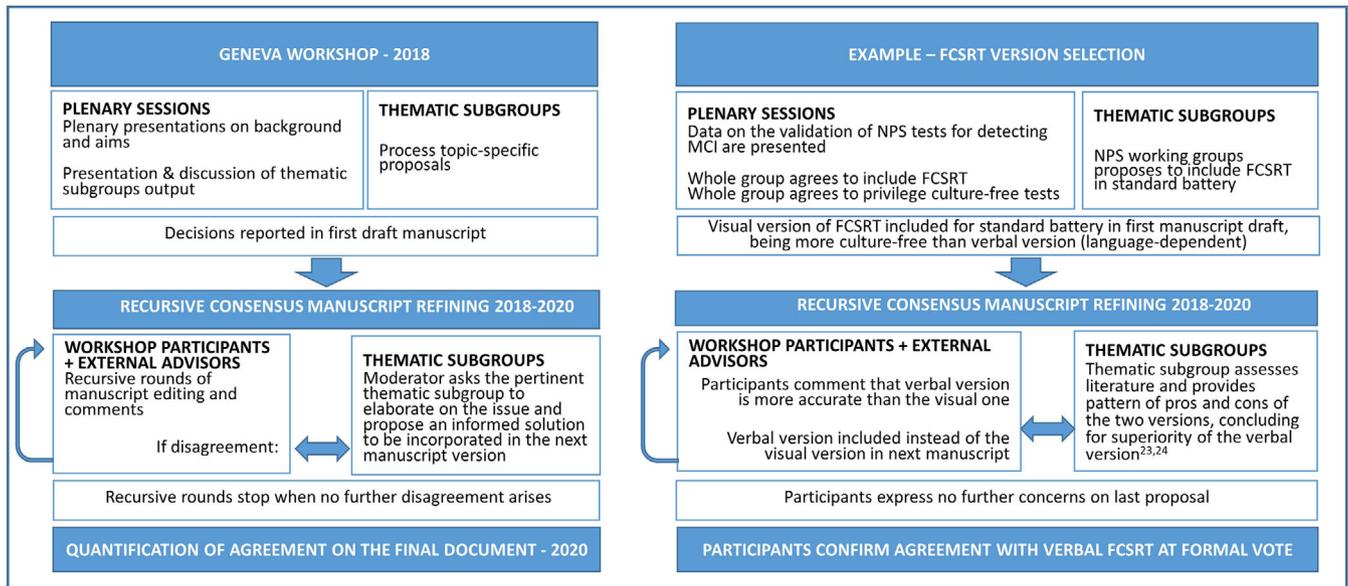
FIGURE 1 Consensus procedure used for defining the cUDS and the methods to generate harmonized normative values

2.1 | Thematic subgroups

Each thematic subgroup was led by one expert (two in the case of the statistical working group) to help discussions through a semi-structured approach. They elaborated on methodology (methods to define normative values; statistical approaches/modeling), issues related to neuropsychological tests (test selection; hypothesis-driven test generation), and digitally assisted testing. At the workshop, the subgroups developed proposals for defining the normal population, confounding factors and minimum sample size required to produce consistent normative values, test selection for standard assessment, and perspectives for future consideration. Decisions were proposed, discussed, refined, and ratified in subsequent plenary discussions with informal consensus procedures. For some aspects, the subgroups were tasked to further process their topics after the workshop. Subgroups completed the processing of their tasks in the following months through online meetings and provided written sections for the paper and the supplemental material. All participants were entitled to object and contribute. Divergence was settled based on published data, processed by the pertinent thematic working group after the workshop. When no further objections emerged for solutions to previous objections, we considered that the majority agreed on the latest solution.

2.2 | Consensus refining

After the workshop, the moderator (first author) incorporated the achieved consensus decisions into a first draft paper. Additional contributors were invited to provide information and knowledge based on their expertise in the field, and their comments were accommodated into the manuscript. Workshop participants and the additional contributors could access the manuscript at all phases, and proposed comments were accessible to all. Whenever objections and disagreement requiring specific expertise arose, the moderator consulted the thematic subgroups again for qualified processing. This was done through both in-person or remote meetings. Both expertise and published evidence were used to support decisions.

2.3 | Quantification of final consensus

At the end of this procedure, participants could express their (a) full agreement, (b) partial agreement (ie, agree as a first step, in view of next improvement), or (c) disagreement, and could propose reasons and further comments through a formal voting system. Those who took part in the workshop in person and the additional contributors based in Europe were invited to this final vote ($N = 47$). Because all of the proposed points came from a lengthy (although informal) consensus procedure, we set the threshold for agreement at 90%. Both the options "Full" and "Partial agreement in view of future improvement" were considered as agreement for the current decision, "Partial agreement in view of future improvement" meaning that the current solution is a required interim step toward harmonization. Answers of partial or lacking agreement required explanation of reasons. These were used to further improve the final manuscript, or processed as far as possible, fed back to all participants blinding the identity of individual responders, and stored to support next developments.

2.4 | Definition of sample size

The statistical subgroup started with a general-purpose estimation method to pinpoint the number of subjects required to compute normative values with correction parameters based on the consensual

classification of confounding factors.[10] A simulation procedure was then used to compute the minimum sample size per language that would allow for (a) stable computations across different scenarios (eg, different distribution properties of test scores, adjustments for confounding factors) and (b) use of complex computational procedures (eg, Item Response Theory, Structural Equation Models) to allow development of flexible composite measures.[5] Additional computational details are provided in Supplemental Section 1.

3 | RESULTS

Consensus was achieved and formalized (1) on a proposal for a Uniform Dataset analogous to that produced by the U.S. National Alzheimer's Coordinating Center (NACC) (Table 1)[7] for the context of use of diagnosis in memory clinics, and on enriching it with tests specific to the pathophysiology of typical Alzheimer's disease (AD) and behavioral variant frontotemporal dementia (bvFTD); (2) on a standard definition of normal controls (Box 1), aimed to produce consistent normative values; and (3) on a standard methodology to produce harmonized normative data across tests and languages (Box 1; Table 1). The main next steps will require refining the harmonization as emerged from the discussions and to proceed toward implementation (Box 2). Forty-two of the 47 invited participants and contributors sent their final votes and comments at the final questionnaire. Formal consensus with the decisions expressed in the final manuscript was 100% for all points. Partial agreement that could not be fully accommodated in this manuscript amounted to three voters (7%) for the adoption of UDS-3 and the inclusion of FCSRT and the Story-Based Empathy Task (SET), two (5%) for the definition of normal controls, one (2%) for the inclusion and exclusion criteria of controls, none for the methodology to provide standard normative values, and two (5%) for the next harmonization steps.

3.1 | Context of use

Workshop participants agreed on the need and timeliness of a harmonization initiative for neuropsychological assessment for memory clinics and its concomitant utility for clinical research, and on the need to encourage a collaborative and representative participation of different clinical contexts and countries. Although aimed at European clinical settings, the intrinsic European cultural and linguistic heterogeneity, the high rate of migration, and the need to leverage existing resources and to align the clinical and clinical-research contexts motivated participants to consider this effort within a broader European and non-European context. This effort has been aimed primarily at clinical use, but its applicability to clinical research is straightforward.

The proposed primary objective of the anticipated standard assessment is to reliably identify MCI and progression of cognitive decline in persons referred to memory clinics, or other specialized centers, for cognitive complaints. The assessment is not designed to detect subtle deficits in the preclinical disease phase, or to grade severity of impairment at the dementia stage. Moreover, it is not designed for population screening, case finding, or finer cognitive profiling aiming at other purposes (eg, to formulate etiopathological diagnoses based on cognitive profile, or to tailor neuropsychological rehabilitation).

3.2 | Uniform data set

Workshop participants proposed and consented to adopt version 3 of the U.S. NACC Uniform Dataset (UDS-3) neuropsychological test battery, to date the most widely implemented standard battery for diagnosing MCI and measuring progression of cognitive decline in early dementia. The main reasons for *partial agreement in view of future improvement* in the formal voting for final agreement included the length of the proposed battery, possibly excessive for some clinics, as well as requests for even more extensive and thorough assessment; a limited added value over current harmonization in countries already using standard batteries (ie, The Netherlands and German-speaking countries); and possible issues on administration, scoring, and norms.

The following integrations were required to increase sensitivity to MCI.

3.2.1 | Adaptations

The UDS-3 largely consists of new, copyright-free versions of common neuropsychological tests that are sensitive to MCI and early dementia cognitive decline and overcome test repetition effects[7,11] (Table 1). The tests were specifically developed, adapted, and normed for the elderly U.S. population. The first step for the European harmonization is to adopt European "local analogues", that is, traditional tests with local normative values analogous to those in the UDS-3, like the WAIS digit span instead of the UDS-3 number span, or the Boston Naming Test if local norms for MINT (Multilingual Naming Test) are lacking (Table 1). Subsequent steps require adapting tests across languages and acquiring local normative data (Box 2). Another possible adaptation relates to test order. If tests are already used in local batteries, with different order due to different composition of such batteries, such order discrepancies are considered compatible with this harmonization effort at this stage.

3.2.2 | Integration

Workshop participants also agreed on including tests specific to episodic memory and emotional cognition impairment (bold in Table 1) to provide better coverage of these cognitive domains in line with Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) guidelines, and to enhance the battery's sensitivity to the typical AD, atypical AD, and to bvFTD. Specifically, they proposed replacing the Craft Story Memory test (the UDS-3 replacement of the logical memory test) with the verbal version of the Free and Cued Selective Reminding Task (FCSRT),[12] and adding the Story-based Empathy Task (SET)[13] to assess social competence. The Free and Cued Selective Reminding Task (FCSRT)[12,14] is a cued word list task providing controlled learning and retrieval conditions that enable one to

TABLE 1 Uniform Dataset, based on the current US standard for research (UDS-3(1)) and the Geneva Workshop 2018 adaptation to clinical use (cUDS), mapped onto the cognitive domains and sub-functions as listed and recommended by DSM-5(2,3) for the diagnosis of neurocognitive disorders

DSM-5-addressed domains	UDS-3			cUDS			
	Test	Order	Time	Test	Order	Time	Notes
General cognitive assessment	MoCA	1	10′	MoCA	1	10′	Often administered at a first separate visit(5,6)
Attention							
sustained							
selective							
divided							
processing speed	Trail making A	7	1′	Trail making A	5	1′	
Executive function							
planning							
decision making							
working memory	Number span backward	5	1′	*Digit span backward*	3	1′	*From WAIS*
feedback use							
flexibility	Trail making B	8	3′	Trail making B	6	3′	
Perceptual-motor							
visual perception							
constructional	Benson figure copy	3	5′	Benson figure copy	9	5′	or Rey Complex figure, if lacking norms
perceptual-motor							
praxis							
gnosis							
Social cognition							
emotion recognition							
theory of mind					**SET**	7	10′	**SET: Story-based Empathy task (7);** https://forms.gle/muDpJLkqH6X8h9z99
Learning-Memory							
immediate recall	Number span forward	4	1′	*Digit span forward*	2	1′	*From WAIS*
	Craft story 21–immediate	2	5′	**FCSRT – immediate**	**4**	**10′**	**Verbal version. Normative values available in many languages** (8-15)
short-term memory	Craft story 21–delayed	9	**10′**	**FCSRT – delayed**	8	**3′**	FCSRT
	Benson figure recall	10	**5′**	Benson figure recall	13	5′	
long-term memory	*			*			Category fluency and MINT have a long-term memory component
implicit memory							

(Continues)

TABLE 1 (Continued)

DSM-5-addressed domains	UDS-3			cUDS			Notes
	Test	Order	Time	Test	Order	Time	
Language							
production	MINT	11	5-10'	MINT	12	5-10'	or Boston naming test, if lacking norms
	Category fluency (animals, vegetables)	6	3'	Category fluency	10	3'	
				Letter fluency (F, L)			
	Letter fluency (F, L)	12	3'		11	3'	
comprehension							
Total time			52-57'			60-65'	

Tests in italics denote "local analogues," that is, traditional local versions of the UDS-3 tests. Bold (besides titles) denotes tests added or replaced to UDS-3 tests. "Order" denotes the presentation order in UDS-3 as from https://www.alz.washington.edu/NONMEMBER/UDS/DOCS/VER3/UDS3_npsych_worksheets_C2.pdf and the administration order of the cUDS tests as described in the Results section. Time denotes expected duration of administration estimated for patients with MCI (with CDR test score between 0.5-1) and including instructions. MINT = Multi-lingual naming test; FCSRT = Free and cued selective reminding test, verbal version(4); MoCA Montreal Cognitive Assessment.

*Category fluency and MINT also have a long-term memory component.

References.

1. Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, et al. Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). Alzheimer Assoc Disord. marzo 2018;32(1):10-7.

2. Ganguli M, Blacker D, Blazer DG, Grant I, Jeste DV, Paulsen JS, et al. Classification of neurocognitive disorders in DSM-5: a work in progress. Am J Geriatr Psychiatry. marzo 2011;19(3):205-10.

3. APA. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington V :American PP, curatore. 2013.

4. Grober E, Buschke H. Genuine memory deficits in dementia. Dev Neuropsychol. 1987;3(1):13-36.

5. Costa A, Bak T, Caffarra P, Caltagirone C, Ceccaldi M, Collette F, et al. 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. Alzheimers Res Ther. 17 aprile 2017;9(1):27.

6. Boccardi M, Nicolosi V, Festari C, Bianchetti A, Cappa S, Chiasserini D, et al. Italian consensus recommendations for the biomarker-based etiological diagnosis in MCI patients. Eur J Neurol. 2019;Submitted(EJoN-19-0241).

7. Dodich A, Cerami C, Canessa N, Crespi C, Iannaccone S, Marcone A, et al. A novel task assessing intention and emotion attribution: Italian standardization and normative data of the Story-based Empathy Task. Neurol Sci Off J Ital Neurol Soc Ital Soc Clin Neurophysiol. ottobre 2015;36(10):1907-12.

8. Grau-Guinea L, Perez-Enriquez C, Garcia-Escobar G, Arrondo-Elizaran C, Pereira-Cutino B, Florido-Santiago M, et al. Development, equivalence study, and normative data of version B of the Spanish-language Free and Cued Selective Reminding Test. Neurologia. 8 maggio 2018;.

9. Dion M, Potvin O, Belleville S, Ferland G, Renaud M, Bherer L, et al. Normative data for the Rappel libre/Rappel indice a 16 items (16-item Free and Cued Recall) in the elderly Quebec-French population. Clin Neuropsychol. 2015;28 Suppl 1:S1-19.

10. Frasson P, Ghiretti R, Catricala E, Pomati S, Marcone A, Parisi L, et al. Free and Cued Selective Reminding Test: an Italian normative study. Neurol Sci. dicembre 2011;32(6):1057-62.

11. Girtler N, De Carli F, Amore M, Arnaldi D, Bosia LE, Bruzzaniti C, et al. A normative study of the Italian printed word version of the free and cued selective reminding test. Neurol Sci. luglio 2015;36(7):1127-34.

12. Grober E, Lipton RB, Katz M, Sliwinski M. Demographic influences on free and cued selective reminding performance in older persons. J Clin Exp Neuropsychol. aprile 1998;20(2):221-6.

13. Ivnik RJ, Smith GE, Lucas JA, Tangalos EG, Kokmen E, Petersen RC. Free and cued selective reminding test: MOANS norms. J Clin Exp Neuropsychol. ottobre 1997;19(5):676-91.

14. Mokri H, Avila-Funes JA, Meillon C, Gutierrez Robledo LM, Amieva H. Normative data for the Mini-Mental State Examination, the Free and Cued Selective Reminding Test and the Isaacs Set Test for an older adult Mexican population: the Coyoacan cohort study. Clin Neuropsychol. 2013;27(6):1004-18.

15. Vogel A, Mortensen EL, Gade A, Waldemar G. The Category Cued Recall test in very mild Alzheimer's disease: discriminative validity and correlation with semantic memory functions. Eur J Neurol. gennaio 2007;14(1):102-8. .

distinguish between impaired encoding (eg, due to attention disorders as in depression), storage (eg, due to hippocampal damage as in typical AD), and recall strategies (eg, due to frontal lobe dysfunction).[3] Impaired task performance in the FCSRT correlates with in vivo AD biomarkers,[15] even at the prodromal stage.[16,17] Low total recall performance despite retrieval facilitation with cueing discriminates MCI patients subsequently converting to AD dementia with 88.6% specificity,[18] showing better diagnostic and prognostic performance than the Logical Memory task.[16,17,19–22] The FCSRT has been validated in many EU countries, with normative values available in different languages (see Table 1). The verbal version of the FCSRT has lower ceiling effects and greater dispersion of test scores,[23,24] and is therefore preferred over the visual version for a standard assessment. The SET was specifically developed to assess social cognition in patients with dementia. It requires subjects to select the possible endings of stories told with cartoons, and assesses emotion and intention attribution separately, as well as causal inference as the control condition. Verbal interaction is used to ascertain proper comprehension of instructions; however, correct performance consists of selecting the correct story ending among the available pictures. Similar to the mini Socio-Emotional Assessment (mini-SEA),[25] performance on the SET[13] correlates with structural and functional imaging evidence of frontal lobe

BOX 1   Consensual definition of Normative Population to obtain consistent normative values across countries and tests. All of the reported criteria must be met to recruit proper harmonized control samples

Selection of normal (not super-normal) subjects

- Bias-free recruitment modalities (eg, to guarantee representativeness for the whole country population, recruitment should avoid clustered data within just one site, and data should be gathered pairwise—for age, sex, education—within each recruitment site)
- Avoid convenience samples unless compliant with the inclusion/exclusion criteria
- Avoid voluntary exclusion of otherwise healthy individuals positive to biomarkers for brain amyloidosis, tau, or other neurodegeneration that define risk or preclinical stage for neurocognitive disorders
- Avoid voluntary exclusion of subjects with subjective cognitive decline (SCD) from samples explicitly recruited as normal controls
- Do not seek demonstration of stable cognitive health with longitudinal neuropsychological and neurological assessment ("robust norms")

Inclusion Criteria

- Age: if feasible, 40 years or older
- Self-identified as "cognitively normal"
- Denies a worrying cognitive decline
- Judged to be cognitively normal by a family member (or other knowledgeable informant); cut-off of 3.3 on the short form of the Informant Questionnaire on Cognitive Decline (IQCoDe)(1) or equivalent
- MoCA score greater than or equal to 23(2,3) (MMSE greater than or equal to 27 if MoCA cannot be used). Please note: values to be corrected by age and education

Exclusion Criteria

Clinical findings
- Sensory or motor deficits interfering with test administration
- Continuous moderate-to-intense pain
- Current psychiatric diagnosis (including major depression). Geriatric Depression Scale (15-items) score of 6 or greater (4)

Medical history
- Head injury with loss of consciousness for more than 5 minutes
- General anesthesia within the last 3 months
- Prior recurrent psychiatric disorder requiring hospitalization
- Use of psychoactive drugs, alcohol abuse
- Significant cerebrovascular disease (eg, TIA, stroke, general atherosclerosis)
- Severe active medical condition (cancer, organ failure, unstable heart condition) that may interfere with test administration

Convenience samples

- May be used if compliant with the above features
- May be used *ad interim* when proper samples are unavailable
- Specific research samples (only SCD, or composed of subjects all having specific risk factors) are not appropriate

References.
1. Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. Cochrane Database Syst Rev. 10 aprile 2014;(4):Cd010079.
2. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. Alzheimers Dement. settembre 2013;9(5):529-37.
3. van Steenoven I, Aarsland D, Hurtig H, Chen-Plotkin A, Duda JE, Rick J, et al. Conversion between mini-mental state examination, montreal cognitive assessment, and dementia rating scale-2 scores in Parkinson's disease. Mov Disord. dicembre 2014;29(14):1809-15.
4. Friedman B, Heisel MJ, Delavan RL. Psychometric properties of the 15-item geriatric depression scale in functionally impaired, cognitively intact, community-dwelling elderly primary care patients. J Am Geriatr Soc. settembre 2005;53(9):1570-6.

impairment[26]. Although further studies on comparative diagnostic performance across social cognition tests and European cultural contexts are warranted, we propose the SET because it was developed specifically for bvFTD patients, and it places a minimum load on non-target cognitive functions such as language and working memory.

3.2.3 | Administration procedure

The total administration time for the cUDS is estimated to be 60-65 minutes for patients with MCI. The delay interval required between immediate and delayed trials of the FCSRT[12,14] is filled with administration of the non-verbal Trail Making A and B and SET[13], allowing the delay to be free from interference from other verbal tasks[27] (Table 1).

3.3 | Harmonized methodology to produce normative values

Workshop participants highlighted the priority of defining normative values based on standard methodology to guarantee consistent assessment of performance across tests and languages. Therefore, they

BOX 2 Roadmap of required actions towards a harmonized cognitive assessment

Immediate actions for implementation	Next actions for implementation	Medium term development of harmonized assessment
If possible, use cUDS as from Table 1	Extend representativeness of the consortium	Adapt tests across European cultures and languages
If not possible, adopt local analogues (differences in administration order are admitted, when the cUDS tests are already used in local batteries guaranteeing appropriate administration relative to delay, interference, etc.)	Perform survey to explore feasibility, hurdles, facilitators and needs for implementing cUDS in academic and non-academic memory clinics	Develop alternate test versions for repeated testing
Define a standard format for data entry and for the clinical report	Potentiate reciprocal connection of research and clinical centers	Validate cUDS:
Coordinate next steps consistently across countries to get aligned as much as possible (e.g., exact version of tests)	Offer services to clinicians, to:	• for most widespread languages first
	• connect and receive feedback	• for the 27 EU languages (include language variant sub-samples in main languages)
	• support compliance to harmonization	• based on the defined harmonized methods
	• Define copyright-free cUDS tests in the same way as done for UDS-3 (1)	• both paper-pencil and tablet version if available
	• Provide local norms with the harmonized methodology proposed here (labs of neuropsychology)	• Define backwards compatibility to shift from currently used batteries to the standard
	• Bridge with pertinent stakeholders (e.g., health refunders, regulators) for consistent implementation	• Disseminate the information about cUDS implementation capillary through clinical and professional networks and Scientific Societies
	• Identify tests most needed to complete appropriately cUDS (uncovered domains; actual administration time & tasks for interference/delay; etc.)	• Adapt tests for digitally-assisted assessment (tablet)
		• Converge informatics experts and entrepreneurs to overcome issues on digitally-assisted assessment
		• Possibly include more, or more sensitive, tests, thanks to digital advancements
		• Keep developing hypothesis-driven and culture-free tests, and validate them based on the consensually defined methodology
		• Consider the use of robust controls to compare, and possibly improve, normative values and test sensitivity in the future
		• Select newly developed tests based on diagnostic performance
		• Fine-tune cUDS and implementation based on consensus with all stakeholders

1. Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, et al. Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). Alzheimer Assoc Disord. marzo 2018;32(1):10-7. .

developed a standard definition of the normal population, defined how to code confounders of normative data consistently, and proposed the minimum necessary requirements for the normative sample size.

3.3.1 | Normal population

The normal population that will provide harmonized normative data for the proposed context of use must consist of individuals without cognitive decline, functional impairment due to cognitive deterioration, or major clinical conditions that could interfere with cognition (Box 1). Main reasons for *partial agreement in view of future improvement* in the formal voting for final agreement relate to the potential appropriateness of robust controls. At present, recruitment of normal controls from the general population should not exclude cognitively unimpaired individuals who may have positive biomarkers or specific risk factors (eg, APOE ε4 allele) for AD, nor should it target the "super-normal" individuals (also described as "robust controls") with longitudinal evidence of preserved cognition, although this will plausibly be a required future development (Box 2). Convenience samples of cognitively normal individuals may be employed only if they were recruited to serve as normal controls and if compliant with the criteria reported in Box 1. Convenience cognitively normal samples should not consist of individuals

recruited as target experimental cases (eg, people with subjective cognitive decline or otherwise at risk for neurocognitive disorders). One reason for *partial agreement in view of future improvement* in the formal voting for final agreement on the individual exclusion/inclusion criteria in Box 1 consisted in disagreement on excluding people who experienced post-traumatic loss of consciousness without significant memory loss.

3.3.2 | Demographic factors affecting neuropsychological performance

Standard classification was proposed for age, sex, and education, the main variables that often affect neuropsychological test performance and should be accounted for in normative data. Panelists agreed that normative data and test cut-off values should be stratified across 6 age decades, from the 40 to the upper 90+ year age categories, with balanced sex and education level in each age category. To code education consistently across countries, we adopted the three-level coding system from the International Standard Classification of Education (ISCED) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) https://ec.europa.eu/education/international-standard-classification-of-education-isced_en). This

system includes (1) compulsory (primary and secondary) education (age range: min = 5, max = 17), (2) upper-secondary education (age range for EU countries: min = 12, max = 19), and (3) post-secondary education (age range: min = 18, max $\geq$22) (Supplemental Table S3).

3.3.3 | Sample size

The above classification scheme defines 36 cells that encompass six age classes by three educational levels by two sexes (Supplemental Table S4). Our sample size computation estimated that 10 subjects per cell would suffice to perform general linear models (see Supplemental Section-1 for computational details). Data from this carefully constructed sample can be used to normalize standard scores or modern psychometric scores derived from Item Response Theory or Structural Equation Modeling. An estimated sample size of $N = 330$ subjects per language[28-30] provides a fair trade-off between feasibility and reliability, whereas taking into account the stratifications for age, education, and sex (although cells covering rare populations, eg, age 90+ with high education, may be hard to fill) (Supplemental Table S4).

3.3.4 | Test validation

The cUDS neuropsychological tests should ideally be copyright-free versions, analogous to the U.S. UDS-3[7], which are adapted to the various European target languages. They should be validated in all European languages, taking care to include linguistic minorities in validation studies.

Achieving a standard assessment of this kind requires additional research steps (see roadmap outlined in Box 2). A reason for *partial agreement in view of future improvement* in the formal voting for final agreement on such roadmap consisted in the difficulty of harmonizing and producing culture-free social cognition among countries. Another main hurdle lies in the fact that some countries already use CERAD or other batteries as local standards. However, using the cUDS, as a common set of UDS-3 or analogous tests with locally appropriate normative data was seen as a first practical step towards harmonization.

4 | DISCUSSION

This work defines the first steps toward a standard assessment of people with cognitive complaints attending European memory clinics or participating in clinical research on MCI in Europe. This assessment includes a standard neuropsychological test battery with a harmonized methodology to produce normative data and cut-off values for impairment. Besides leveraging on previous harmonization efforts, the bi-directional collaboration with clinicians from specialized non-academic centers is a new and important step necessary to fit the needs and constraints of both clinical practice and research. With evolving clinical criteria and the availability of biomarkers with specific diagnostic value, it is now even more important to assess patients consistently

with a precise definition of the target disorder, with the demonstrated informative value of diagnostic biomarkers, and across centers. With standard assessment, patients could get second opinions or receive follow-up examinations in different centers without the need to repeat existing baseline assessments. Costs, as well as practice effects, would be reduced; benefits for data pooling and comparability of studies in clinical research are straightforward. The reliability of diagnostic procedures for MCI may improve and approach the quality standards of other clinical conditions.

The proposed standard assessment is not designed to ascertain the pathophysiology underlying cognitive impairment, but to detect the presence and possibly progression of objective cognitive decline that may be due to a neurodegenerative condition. Further exploration, increasingly performed through biomarkers, is required to formulate an etiopathological diagnosis or to identify specific clinical needs.

The cUDS is based on the U.S. UDS-3.[7] Extensive work was undertaken to develop the UDS-3 for the cognitive assessment of individuals with MCI, before it was adopted as a standard cognitive assessment in all federally funded U.S. AD research centers. Although defined to improve CERAD-NAB performance,[31] to date there is no systematic evidence on the ability of UDS-3 in detecting MCI. The cUDS is expected to outperform the CERAD-NAB in the detection of MCI. Although overlapping with CERAD-NAB and its "-plus" version (Trail Making, Figures copy/recall; Boston Naming Test and Verbal Fluency tasks), the Montreal Cognitive Assessment (MoCA), included in the cUDS, was shown to clearly outperform the CERAD's Mini-Mental State Examination.[32] The CERAD word list was demonstrated to be less sensitive than the California Verbal Learning test,[33] and its diagnostic and prognostic performance for AD-MCI was demonstrated to increase by adding the FCSRT[34] (32). In addition, the SET allows early detection of impairment also in patients affected by frontotemporal lobar degeneration, who may perform well on typical cognitive tests. Overall, cUDS overlaps with CERAD to a considerable extent but is devised to be more sensitive to mild, atypical, and non-AD conditions. Although CERAD-plus has specific normative values for US-English and German populations (Supplemental Table S2), UDS-3 local analogues are frequently used in European memory clinics with local norms for most countries. Alignment with the NACC UDS-3 sets this European proposal up for improved sensitivity to MCI due to different neurodegenerative causes and for a possibly wider international consensus.

In addition to CERAD and UDS-3, the Neuropsychological Normative Project (CN-NORM)[8] (Supplemental Table S2) has selected and recommended tests for harmonized assessment of MCI for Chinese memory clinics. Besides some overlap with the cUDS (eg, trail making, digit span and fluency tests, and the memory binding test[35] that, like the FCSRT, uses a controlled learning paradigm minimizing the use of individual strategies), the CN-NORM battery covers cognitive domains more extensively than the cUDS. In particular, it allows a more thorough assessment of attentional, perceptual-motor, and social functions. This main difference between the cUDS and the CN-NORM consists of the fact that the cUDS aims to provide objective evidence of impairment without the aim to identify pathophysiology, because it is devised for a biomarker-based diagnostic procedure. This differs from

the CN-NORM initiative, that is not expressly restricted to biomarker-based procedures and consistently provides more thorough cognitive assessment requiring additional measures. A short battery, however, has greater potential to be adopted in EU countries. Nevertheless, future developments aimed to improve the assessment of cognitive and social functions not assessed by the current cUDS may consider the Chinese standard and seek further consistency across Western and Eastern countries.

Previous initiatives selecting and recommending tests for standard assessment relate to research on preclinical AD (European Prevention of Alzheimer's Disease [EPAD] and Alzheimer's disease cognitive composite [PACC], Supplemental Table S2),[36,37] a different context of use relative to the cUDS. The brevity of PACC makes it an interesting option for detecting MCI, and decreased scores were associated with MCI in research cohorts.[37,38] Its potential in detecting MCI patients in memory clinics, not yet explored, is expected to be lower than cUDS. First, cUDS also includes a test for MCI due to non- or atypical AD. Moreover, the mentioned limitations of logical memory test[16,17,19–22] and of the Mini-Mental State Examination (MMSE)[32] are consistent with the demonstrated improved performance of PACC after removing MMSE and including semantic and executive assessment.[39,40] A PACC-like composite can be derived from the cUDS; however, future developments of initiatives aimed at assessing individuals at preclinical (PACC, EPAD) and clinical (cUDS, UDS-3, CN-NORM) disease stage may try to seek consistency across each other (eg, FCSRT is already included both in PACC and cUDS). Bridging these different contexts of use may provide continuity of assessment, thereby sparing money and effort, if preclinical assessment should be adopted for future population screening in the future.

The methodology we propose to produce normative data and cut-off values for impairment is not new in terms of the need for correction itself but is new as it tries to align the validation of different tests and of tests in different languages to a common methodological standard. The age range of controls providing normative values (40+) allows the detection of early onset cases; the standard classification of education, although based on only three levels, guarantees a reliable comparison across countries characterized by very different educational systems, which can hardly be captured in a harmonized framework. Although such compromises are required to achieve a minimum and feasible harmonization, neuropsychological research groups are encouraged to provide additional, finer normative values for research aims, for other contexts of use, or to better account for less represented groups (eg, elderly with very low educational attainments). Further stratification for variables like residence in urban versus rural areas may also be included if possible. Other possibly confounding variables, like ethnicity, may have a less consistent effect in Europe than in the United States, as migration waves are currently variable and variably handled. On the other hand, norms based on "robust" control samples (ie, subjects whose normal cognition is documented over longitudinal evaluations), or on individuals free from pathology or risk factors for neurocognitive disorders, may also be of interest for research aims, and may be required in future clinical applications, but are not pertinent to the proposed standard assessment aimed at detecting

mild impairment from a clinical rather than pathophysiological point of view.

Further fine-tuning of the cUDS is warranted to meet practical constraints to implementation, to assess other dimensions (eg, motivational level, malingering), or to provide more balanced testing of the different cognitive domains, similar to the CN-NORM battery. Newly developed hypothesis-driven tests may increase sensitivity (Supplemental Section 2). Moreover, consistency of administration should also be achieved across centers, raters, and time through specific methodological procedures defining standard administration, data entry, score computation, and ascertaining compliance and reliability over time, similar to other diagnostic procedures. Score conversion tools and demonstration of backward compatibility with local batteries are also required, to shift to the new standard, as is the development of copyright-free tests analogous to UDS-3, of local normative data (Box 2), and of digital infrastructure to assist the harmonized assessment (Supplemental Section 3).

Within this initiative, we have tried to take advantage of the knowledge and experience gained in different research and clinical contexts. Although not entirely new in nature, this is among the first efforts trying to converge such knowledge into a comprehensive harmonized procedure for the reliable assessment of patients with MCI possibly serving both clinical and research aims. This approach was maximized by the support and participation of several relevant consortia (Supplemental Box S1), most importantly the European Alzheimer's Disease Consortium. The Alzheimer's Association International Society to Advance Alzheimer's Research and Treatment (ISTAART) Professional Interest Area on Cognition has initiated a Workgroup on Harmonization of Assessment that will also help bring the relevant groups together to review, evaluate, and make recommendations on convergence. It should be noted, however, that representation in this process was limited by logistics and feasibility, penalizing clinicians not engaged in research at this step. Another major limitation that we did not address consists of the heterogeneous definition of MCI. Different definitions have been applied, but ideally a consensus should be achieved upstream of a test battery selection. Nevertheless, a standard selection of tests is a valuable starting point for subsequent more thorough harmonization. The opportunity to benefit from digital technology (Supplemental Section 3) to assist the standardized assessment exists but requires additional consensual development and implementation while respecting clinical and scientific principles underlying standard assessment for neurocognitive disorders. Future developments may allow for more accurate and reliable assessment, especially relative to confounders that cannot be stratified and corrected for in a satisfactory way based on pen-and-pencil tests and traditionally collected normative values. Finally, the number and complexity of the variables processed in this consensus was so high that formal traditional consensus procedures (eg, Delphi panel) could not be applied from the beginning to each of them for reasons of feasibility. Future developments should try to use formal methods. Despite these limitations, this first step toward harmonization in memory clinics can help develop more modern and efficient clinical procedures for neurocognitive disorders through the consistent definition of normal controls and

methodological procedures for the production of norms, the harmonization across tests and languages, and the attempt to seek for extra-European convergence. It will optimize costs and reliability, aligning diagnostic procedures across centers, with the demonstrated informative value of diagnostic tools, and with the therapeutic value of available treatment expected to slow down progression and improve the quality of life of patients and caregivers[41] (34).

CONFLICT OF INTEREST

I. Bos received grants from Innovative Medicine Initiative Joint Undertaking. H. Brodaty received personal fees from Nutrivia Australia. S. Cappa received personal fees from Biogen. W. Kukull received grants from National Institutes of Health/National Institute on Aging (NIH/NIA). F. Oliveira received personal fees from Gerson Lehrman Group and Atheneum Partners.

O. Peters received personal fees from Biogen; grants and personal fees from Roche; and grants from Novartis, Janssen, Pharmatrophix, and Eisai.

P. Sachdev received grants from NHMRC Australia and personal fees from Biogen Pharmaceuticals.

D. Salmon received consulting fees from Biogen, Inc. and Aptinyx, Inc.

Teipel received personal fees from Roche Pharma AG, MSD Sharp & Dohme GmbH, and Biogen.

P.J. Visser received grants from Innovative Medicine Initiative, ZONmW, and Biogen.

B. Winblad received advisory board fees from AlzeCure, Alzinova, and Axon Neuroscience.

The other authors report no conflict of interest.

ORCID

Marina Boccardi (iD) https://orcid.org/0000-0001-6744-8246

REFERENCES

1. APA. Diagnostic and Statistical Manual of Mental Disorders. 5th. 2013.
2. Albert MS, DeKosky ST, Dickson D, et al. 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. *Alzheimers Dement.* 2011;7(3):270-279.
3. Cerami C, Dubois B, Boccardi M, Monsch AU, Demonet JF, Cappa SF. Clinical validity of delayed recall tests as a gateway biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. *Neurobiol Aging.* 2017;52:153-166.
4. Frisoni GB, Boccardi M, Barkhof F, et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol.* 2017;16(8):661-676.
5. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav.* 2012;6(4):517-527.
6. Mukherjee S, Mez J, Trittschuh EH, et al. Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Mol Psychiatry.* 2018.
7. Weintraub S, Besser L, Dodge HH, et al. Version 3 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (UDS). Alzheimer disease and associated disorders. 2018;32(1):10-17. https://doi.org/10.1097/wad.0000000000000223
8. Wang H, Fan Z, Shi C, et al. Consensus statement on the neurocognitive outcomes for early detection of mild cognitive impairment and Alzheimer dementia from the Chinese Neuropsychological Normative (CN-NORM) Project. *J Glob Health.* 2019;9(2):20320.
9. Boccardi M, Gallo V, Yasui Y, et al. The biomarker-based diagnosis of Alzheimer's disease. 2-lessons from oncology. *Neurobiol Aging.* 2017;52:141-152.
10. Jennen-Steinmetz C, Wellek S. A new approach to sample size calculation for reference interval studies. *Stat Med.* 2005;24(20):3199-3212.
11. Shirk SD, Mitchell MB, Shaughnessy LW, et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. Alzheimer's. *Research Therapy.* 2011;3(6):32.
12. Grober E, Buschke H. Genuine memory deficits in dementia. *Dev Neuropsychol.* 1987;3(1):13-36.
13. Dodich A, Cerami C, Cappa SF, et al. Combined Socio-Behavioral Evaluation Improves the Differential Diagnosis Between the Behavioral Variant of Frontotemporal Dementia and Alzheimer's Disease: in Search of Neuropsychological Markers. *J Alzheimers Dis.* 2018;61(2):761-772.
14. Buschke H. Cued recall in amnesia. *J Clin Neuropsychol.* 1984;6(4):433-440.
15. Rami L, Sole-Padulles C, Fortea J, et al. Applying the new research diagnostic criteria: mRI findings and neuropsychological correlations of prodromal AD. *Int J Geriatr Psychiatry.* 2012;27(2):127-134.
16. Derby CA, Burns LC, Wang C, et al. Screening for predementia AD: time-dependent operating characteristics of episodic memory tests. *Neurology.* 2013;80(14):1307-1314.

17. Wagner M, Wolf S, Reischies FM, et al. Biomarker validation of a cued recall memory deficit in prodromal Alzheimer disease. *Neurology*. 2012;78(6):379-386.

18. Sarazin M, Chauvire V, Gerardin E, et al. The amnestic syndrome of hippocampal type in Alzheimer's disease: an MRI study. *J Alzheimers Dis*. 2010;22(1):285-294.

19. Sarazin M, Berr C, Rotrou Jde, et al. Amnestic syndrome of the medial temporal type identifies prodromal AD: a longitudinal study. *Neurology*. 2007;69(19):1859-1867.

20. Grande G, Vanacore N, Vetrano DL, et al. Free and cued selective reminding test predicts progression to Alzheimer's disease in people with mild cognitive impairment. *Neurol Sci*. 2018;39(11):1867-1875.

21. Small BJ, Rosnick CB, Fratiglioni L, Bäckman L. Apolipoprotein E and cognitive performance: a meta-analysis. *Psychol Aging*. 2004;19(4):592-600.

22. Teichmann M, Epelbaum S, Samri D, et al. Free and Cued Selective Reminding Test - accuracy for the differential diagnosis of Alzheimer's and neurodegenerative diseases: a large-scale biomarker-characterized monocenter cohort study (ClinAD). *Alzheimers Dement*. 2017;13(8):913-923.

23. Zimmerman ME, Katz MJ, Wang C, et al. Comparison of "Word" vs. "Picture" Version of the Free and Cued Selective Reminding Test (FCSRT) in Older Adults. *Alzheimer's & Dementia (Amsterdam, Netherlands)*. 2015;1(1):94-100.

24. Delgado C, Munoz-Neira C, Soto A, et al. Comparison of the Psychometric Properties of the "Word" and "Picture" Versions of the Free and Cued Selective Reminding Test in a Spanish-Speaking Cohort of Patients with Mild Alzheimer's Disease and Cognitively Healthy Controls. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*. 2016;31(2):165-175.

25. Bertoux M, Funkiewiez A, O'Callaghan C, Dubois B, Hornberger M. Sensitivity and specificity of ventromedial prefrontal cortex tests in behavioral variant frontotemporal dementia. *Alzheimers Dement*. 2013;9(5 Suppl):S84-94.

26. Caminiti SP, Canessa N, Cerami C, et al. Affective mentalizing and brain activity at rest in the behavioral variant frontotemporal dementia. *NeuroImage Clinical*. 2015;9:484-497.

27. Strauss E, Sherman E, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford University Press; 2006.

28. Wolf EJ, Kelly M, Harrington KM, Clark SL, Miller MW. Sample Size Requirements for Structural Equation Models: an Evaluation of Power, Bias, and Solution Propriety. *Educ Psychol Meas*. 2013;76(6):913-934.

29. Muthén LK, Muthén BO. How To Use A Monte Carlo Study To Decide On Sample Size and Determine Power. *Structural Equation Modeling: A Multidisciplinary Journal*. 2002;9(4):599-620.9(4):599-620.

30. Myers ND, Ahn S, Jin Y. Sample Size and Power Estimates for a Confirmatory Factor Analytic Model in Exercise and Sport: a Monte Carlo Approach. *Res Q Exerc Sport*. 2011;82(3):412-423.

31. Morris JC, Weintraub S, Chui HC, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis Assoc Disord*. 2006;20(4):210-216.

32. RE Nieuwenhuis-Mark. The death knoll for the MMSE: has it outlived its purpose?. *J Geriatr Psychiatry Neurol*. 2010;23(3):151-157.

33. Beck IR, Gagneux-Zurbriggen A, Berres M, Taylor KI, Monsch AU. Comparison of verbal episodic memory measures: consortium to establish a registry for Alzheimer's disease–Neuropsychological Assessment Battery (CERAD-NAB) versus California Verbal Learning Test (CVLT). *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*. 2012;27(5):510-519.

34. Sala I, Illán-Gala I, Alcolea D, et al. Diagnostic and Prognostic Value of the Combination of Two Measures of Verbal Memory in Mild Cognitive Impairment due to Alzheimer's Disease. *Journal of Alzheimer's disease: JAD*. 2017;58(3):909-918.

35. Buschke H, Mowrey WB, Ramratan WS, et al. Memory Binding Test Distinguishes Amnestic Mild Cognitive Impairment and Dementia from Cognitively Normal Elderly. *Arch Clin Neuropsychol*. 2017;32(8):1037-1038.

36. Ritchie K, Ropacki M, Albala B, et al. Recommended cognitive outcomes in preclinical Alzheimer's disease: consensus statement from the European Prevention of Alzheimer's Dementia project. *Alzheimers Dement*. 2017;13(2):186-195.

37. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurol*. 2014;71(8):961-970.

38. Burnham SC, Coloma PM, Li Q-X, et al. Application of the NIA-AA Research Framework: towards a Biological Definition of Alzheimer's Disease Using Cerebrospinal Fluid Biomarkers in the AIBL Study. *J Prev Alzheimers Dis*. 2019;6(4):248-255.

39. Lim YY, Snyder PJ, Pietrzak RH, et al. Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: introducing the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults composite score. *Alzheimers Dement (Amst)*. 2016;2:19-26.

40. Papp KV, Rentz DM, Orlovsky I, Sperling RA, Mormino EC. Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: the PACC5. *Alzheimer's & dementia (New York, N Y)*. 2017;3(4):668-677.

41. Bahar-Fuchs A, Martyr A, Goh AM, Sabates J, Clare L. Cognitive training for people with mild to moderate dementia. *Cochrane Database Syst Rev*. 2019;3:CD013069.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

APPENDIX: COLLABORATORS

Ingo Frommann,[1,2] Sandra Roeske,[1] Steffen Wolfsgruber,[1] L'ubomira Anderkova,[3] Felbecker Ansgar,[4] Andrea Chincarini,[5] Maria Cotelli,[6] Eva Ntanasi,[7] Daniel Damian,[8] Janine Diehl-Schmid,[9] Valentina Garibotto,[10] Angélique Gruters,[11] Manuela Guerreiro,[12] Maximilian Haas,[13] Göran Hagman,[14] Ilona Hallikainen,[15] Inga Mehrani,[16] Tuomo Hanninen,[17] Oskar Hansson,[18] Jakub Hort,[19] Adrian Ivanoiu,[20] Frank Jessen,[21] Luka Kulic,[22] Gabriela Latour,[23] Alberto Lleó,[24] Rosa Manenti,[6] Shima Mehrabian,[25] Sharon Naismith,[26] Pierre-Jean Ousset,[27] Hans-Albert Pihan,[28] Tideman Pontus,[29] Geraint Price,[30] Margarita Raycheva,[28] Irena Rektorova,[31] Stefania Rossi,[32] Isabel Sala Matavera,[24] Eric Salmon,[33] Isabel Santana,[34] Egemen Savaskan,[35] Nikolaos Scarmeas,[36] Ann-Katrin Schild,[21] Tanja Richter-Schmidinger,[37] Hilkka Soininen,[15] Franziska Stalder,[28] Katya Stoyanova,[25] Latchezar Traykov,[25] Paul Unschuld,[38] Sergi Valero,[39] Bruno Vellas,[40] Asmus Vogel,[41] Martin Vyhnalek,[19] Gunhild Waldemar,[41] Görser Yener,[42] Deniz Yerlikaya.[43]

[1] DZNE, German Center for Neurodegenerative Diseases, Bonn, Germany

[2] Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany

[3] Brain and Mind Research Programme, CEITEC Masaryk University, Brno, Czech Republic; University Hospital, Masaryk University, Brno, Czech Republic

[4] Klinik für Neurologie, Kantonsspital St. Gallen, St. Gallen, Switzerland

[5] Univ of Genoa, Clin Neurophysiology, Dept of Endocrinological and Med Sciences, Genoa, Italy

[6] IRCCS S.Giovanni di Dio-Fatebenefratelli, Brescia, Italy

[7] 1st Department of Neurology, Aiginition Hospital, Medical School, National and Kapodistrian University of Athens, Greece

[8] Centre Leenaards de la Mémoire, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

[9] Department of Psychiatry, Technical University of Munich, Munich, Germany

[10] CeRiN, Centre for Mind/Brain Sciences, University of Trento, Rovereto, Italy

[11] Department of Psychiatry and Neuropsychology, School of Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht University, Maastricht, The Netherlands.

[12] Inst of Molecular Medicine, Lab of Neurosciences, Lisbon, Portugal

[13] Laboratory of Cognitive Ageing, University of Geneva, Switzerland

[14] Karolinska Institutet, Dept NVS, Center for Alzheimer Research, Division of Neurogeriatrics, Stockholm, Sweden

[15] Institute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland;

[16] Center for Healthy Brain Ageing, University of New South Wales, Sydney, Australia

[17] Neurocenter, neurology. Kuopio University Hospital, Kuopio Finland.

[18] Department of Clinical Sciences Malmö, Lund University, Malmoe, Sweden

[19] Memory Clinic, Department of Neurology, Charles University, 2nd Faculty of Medicine and Motol University Hospital, Czech Republic

[20] Neurology Department, Cliniques Universitaires Saint-Luc, Brussels, Belgium

[21] Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, Cologne, Germany

[22] Zurich Neuroscience Center, Zürich, Switzerland

[23] Psychiatrische Universitätsklinik Zürich (PUK)

[24] Memory Unit, Department of Neurology, Institut d'Investigacions Biomèdiques Sant Pau, Hospital de la Santa Creu i Sant Pau of Barcelona, Spain

[25] Clinic of Neurology, UH "Alexandrovska", Medical University-Sofia, Sofia, Bulgaria

[26] School of Psychology, Charles Perkins Centre and the Brain & Mind Centre, University of Sydney, Sydney, Australia

[27] Centre Hospitalier Universitaire (CHU) de Toulouse

[28] Spitalzentrum Biel, Biel, Switzerland

[29] Karolinska Institutet, Center for Alzheimer Research

[30] Ageing Epidemiology Research Unit, School of Public Health, Imperial College London, UK.

[31] Movement Disorders Centre, First Department of Neurology, Faculty of Medicine, St. Anne's University Hospital, Masaryk University, Brno, Czech Republic

[32] Neurocenter of Southern Switzerland, Lugano, Switzerland

[33] Psychology and Neuroscience of Cognition Research Unit, Faculty of Psychology and Educational Sciences, University of Liège, Liège, Belgium

[34] Neurology Service, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal; Faculty of Medicine, University of Coimbra, Coimbra, Portugal.

[35] Department of Geriatric Psychiatry, University Hospital of Psychiatry, Zurich, Switzerland

[36] 1st Department of Neurology, Aiginition Hospital, Medical School, National and Kapodistrian University of Athens, Greece; Department of Neurology, Columbia University, New York, NY, USA

[37] University of Erlangen, Dept of Psychiatry and Psychotherapy, Erlangen, Germany

[38] Schweizerische Gesellschaft für Alterspsychiatrie und Psychotherapie, Zurich, Switzerland

[39] Research Center and Memory Clinic, Fundació ACE, Institut Català de Neurociències Aplicades, Universitat Internacional de Catalunya, Barcelona, Spain

[40] Chu La Grave-Casselardit, Service de Médecine Interna et Gérontologie Clinique, Toulouse, France

[41] Danish Dementia Research Centre, Department of Neurology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

[42] Department of Neurology, Dokuz Eylül University Medical School, Izmir, Turkey

[43] Department of Neurosciences, Institute of Health Science, Dokuz Eylül University, Izmir, Turkey.