Journal of Alzheimer's Disease 88 (2022) 1663–1678
DOI 10.3233/JAD-215548
IOS Press

A Comparison of Operational Definitions for Mild Cognitive Impairment

Alexandra Polcher[a,b], Steffen Wolfsgruber[b], Oliver Peters[c,d], Lutz Frölich[e], Jens Wiltfang[f,g,h], Johannes Kornhuber[i], Michael Hüll[j], Eckart Rüther[f], Piotr Lewczuk[i,k], Wolfgang Maier[a,b], Frank Jessen[l,m] and Michael Wagner[a,b,*]

[a]*Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany*
[b]*German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*
[c]*Department of Psychiatry and Psychotherapy, Charité, Campus Benjamin Franklin, Berlin, Germany*
[d]*German Center for Neurodegenerative Diseases (DZNE), Berlin, Germany*
[e]*Department of Geriatric Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany*
[f]*Department of Psychiatry and Psychotherapy, University Medical Center Göttingen (UMG), University of Göttingen, Göttingen, Germany*
[g]*German Center for Neurodegenerative Diseases (DZNE), Göttingen, Germany*
[h]*iBiMED, Medical Science Department, University of Aveiro, Aveiro, Portugal*
[i]*Department of Psychiatry and Psychotherapy, University Hospital Erlangen, and Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany*
[j]*Center for Geriatric Medicine and Gerontology, University of Freiburg, Freiburg, Germany*
[k]*Department of Neurodegeneration Diagnostics, Medical University of Biasłystok, and Department of Biochemical Diagnostics, University Hospital of Bialystok, Bialystok, Poland*
[l]*German Center for Neurodegenerative Diseases (DZNE), Cologne, Germany*
[m]*Department of Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany*

Accepted 8 June 2022
Pre-press 7 July 2022

Abstract.

Background: Consideration of many tests from different cognitive domains in defining mild cognitive impairment (MCI) is clinical routine, but guidelines for a neuropsychological operationalization of MCI are lacking.

Objective: Among different operational MCI criteria, to identify those which are best in predicting either conversion to dementia, or a biomarker profile indicative for Alzheimer's disease (AD).

Methods: Memory clinic patients without dementia ($N = 558$; mean age = 66; up to 3 years of follow-up; $n = 360$ with baseline CSF biomarkers) were included in an observational study using most liberal criteria of cognitive impairment. Four operational definitions of MCI were retrospectively applied: 1) amnestic MCI (CERAD word list delayed recall), 2) CERAD total score, 3) comprehensive criteria and 4) base rate corrected CERAD. We compared their accuracy in predicting incident all-cause dementia or AD dementia within three years, or a concurrent CSF $A\beta_{42}$/tau-ratio indicative of AD.

Results: The four definitions overlapped considerably, classified 35–58% of the original sample as impaired and were associated with markedly increased PPVs regarding incident all-cause dementia (39–46% versus 26% of the original sample), AD dementia and AD biomarker positivity. The base rate corrected MCI definition had the highest prognostic accuracy.

*Correspondence to: Prof. Dr. Michael Wagner, Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Venusberg-Campus 1/80, D-53127 Bonn, Germany. Tel.: +49 228 287 36377; E-mail: Michael.Wagner@ukbonn.de.

ISSN 1387-2877

Conclusion: The operational criteria examined seem suitable to specify MCI in memory clinic settings, as they identify subjects at high risk of clinical progression. Depending on the neuropsychological battery in use, one or several of these criteria could help to calibrate the clinical judgment of test results, reduce false-positive decisions, and define risk-enriched groups for clinical trials.

Keywords: Alzheimer's disease, biomarker, cognition, conversion, dementia, diagnosis, DSM-5 mild NCD, mild cognitive impairment, prognosis

INTRODUCTION

Dementia refers to a clinical syndrome which is characterized by a variety of cognitive difficulties that interfere with individuals daily functioning. Alzheimer's disease (AD) dementia is the most common cause of dementia, accounting for 60–80% of all dementia cases. The pathophysiological hallmarks of AD are extracellular amyloid-β (Aβ) accumulation and intracellular neurofibrillary changes in the brain. Biomarkers like reduced levels of $A\beta_{42}$ and increased levels of tau or phosphorylated tau in the cerebrospinal fluid (CSF) reflect this pathology.

Mild cognitive impairment (MCI), which refers to a transitional state between normal aging and dementia [1], has evolved from a fruitful research concept into a new category of clinical diagnosis in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5, [2]) termed mild neurocognitive disorder (mild NCD). MCI and mild NCD are defined as a decline in cognitive performance compared to a previous level in one or more cognitive domains that does not interfere with daily functioning. Although diagnostic guidelines exist for research purposes (National Institute on Aging and the Alzheimer's Association (NIA-AA): MCI, [3]) or clinical practice (DSM-5: mild NCD), they do not contain clear operational definitions of impairment. The diagnostic features section of the DSM-5 mild NCD chapter refers to a performance typically ranging from 1-2 standard deviations (SD) below the normative group, while the NIA-MCI guidelines [3] mention a deficit of 1–1.5 SD. Thus, these criteria allow to diagnose and study individuals with very mild impairments, including non-memory impairments, but require further decisions regarding test selection. The standardized neuropsychological testing of several cognitive domains with multiple tests together with a liberal threshold can result in many false-positive diagnostic decisions [4–6]. In addition, differences in the operationalization of MCI/mild NCD limits comparability, and may underlie widely varying rates of progression from MCI to dementia in different studies [7]. Thus, more specific operational criteria would be a useful complement of the generic DSM-5 NCD/MCI definition [8].

Several operational criteria have been proposed, and we compare four of them in the present study:

(1) Amnestic MCI: This is usually defined by poor performance on a single test, e.g., the Alzheimer's Disease Neuroimaging Initiative (ADNI) defined amnestic MCI as performing below 1.5 SD (demographically adjusted) in the Delayed Recall part of the Wechsler Memory Scale-Revised (WMS-R) Logical Memory Test Story A. Increasing evidence indicates that this approach to identifying memory impairment based on a single measure is vulnerable to false-positive decisions [9–12], as it provides a less reliable estimation of cognitive functioning in comparison to multiple measures [11, 13]. In fact, neuropsychologists usually apply either fixed or flexible test batteries, which result in several standardized scores.

(2) Overall cognitive performance: lower than normal overall cognitive performance as assessed with brief cognitive scales or with more detailed neuropsychological test batteries. For example, the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological test battery is a widely employed fixed test battery covering memory, language, visuoconstruction and, in an extended version, executive and attentional tests (CERAD-Plus). The CERAD total score [14] can be used to operationally define MCI, as it aggregates several test scores from different cognitive domains into a single scaled score and allows for a straightforward classification.

The CERAD total score accurately discriminates healthy elderly subjects from patients with MCI and dementia [15–17], predicts incident AD dementia [18] and tracks progression

from mild to more severe stages of AD [15, 19].

Two more elaborate algorithmic approaches have been proposed that are based on the pattern or the number of deviant test scores in a given test battery.

(3) Comprehensive criteria: Comprehensive criteria proposed by Jak and colleagues [20] "were developed in consideration of the fact that the interpretive value of an isolated impaired score is often limited" (p. 7). The criteria are intended to balance sensitivity and specificity by considering two measures for each of three cognitive domains (which allows for a more reliable estimation of impairment) and by quantifying impairment with 1 SD [20]. The strength of these criteria is that they are generic and do not depend on a specific fixed battery. There is a growing body of literature that recognizes the strength of these comprehensive criteria in terms of diagnostic and predictive accuracy. Compared to conventional approaches (as operationalized, e.g., in ADNI by considering a single measure), comprehensively defined cognitive impairment has been proven to be superior regarding progression to dementia, temporal stability [11, 21–23], identification of individuals with AD risk factors (such as *APOE* ε4) and positive biomarkers (cerebrospinal fluid (CSF) hyperphosphorylated tau, Aβ) [11, 24], and vulnerability to false-positive diagnoses [9–12, 25]. The use of comprehensive criteria has therefore been recommended for the discrimination of subjective cognitive decline from MCI for research purposes [26]. Studies have already applied this approach [26–28].

(4) Base rate correction: A base rate correction approach for the identification of cognitive impairment has been proposed [5] which statistically adjusts for the fact that even healthy elderly adults will have some (a "base rate" of) deviant test scores when many tests are given (the terms multivariate base rate correction [29] or Number of Impaired Tests [30, 31] have also been used for this MCI definition). Mistridis et al. [5] calculated the base rates of deviant scores for the German CERAD test battery using data from a normative study. These are used to gauge results of patients tested with this battery. Within this approach, cognitive impairment is assumed when < 10% of healthy older adults obtain a certain number of scores below a given cut-off. For example, having one deviant score (e.g., below the 16th percentile/deficit of 1 SD) out of 10 CERAD subtest scores is a common event (71%) in healthy normative subjects (i.e., the base rate of this event is high). However, less than ten percent of a healthy comparison group will have 5 or more scores below the 16th percentile. Applying this criterion to demographically matched groups of subjects who either converted to AD dementia ($n = 26$) or remained healthy ($n = 26$), Mistridis et al. found that none of the nonconverters but 23% of the converters initially had 5 or more scores below the 16th percentile, giving some initial validation to the predictive validity of this MCI definition. A recent study [30] applied the base rate approach to reclassify MCI participants in ADNI, using base rates for the number of impaired tests calculated from ADNI controls, and found this MCI definition to be superior to other operational definitions in terms of predictive accuracy. However, as ADNI required a verbal memory deficit > 1.5 *SD* upon inclusion, it is not clear whether the base rate criteria will also do well in less selective samples.

In sum, while the issue of operational definitions for MCI has received increasing attention recently [11, 21, 23, 24], more systematic comparisons of alternative operational definitions of MCI are needed. The predictive accuracy for progression to dementia, and the concurrent relation to AD biomarkers are criteria by which these MCI definitions can be judged, as they reflect clinical utility and physiological validity.

Here, we applied and compared four operational definitions of MCI regarding clinical progression to dementia and cross-sectional AD biomarkers in a large multicenter memory clinic cohort of the German Dementia Competence Network (DCN) [32]. DCN participants were initially included by a purposefully liberal impairment definition that required a deficit of at least 1 SD in any of 12 test scores concerning several cognitive domains to include the mildest form of cognitive decline and to examine the diagnostic and prognostic performance of biomarkers in a mixed memory clinic sample. This liberal inclusion criterion allowed for the application and comparison of more stringent *post-hoc* operational definitions, which vary regarding the number of tests and domains considered.

We aimed to examine these four criteria head-to-head in a single dataset, in order to identify those best predicting either conversion to dementia, or a biomarker profile indicative for AD. We also aimed to study which of the examined MCI criteria would add most to an AD biomarker profile regarding the joint prediction of conversion to dementia. Finally, we aimed to establish the agreement of the criteria with each other and their stability over time.

We expected that the application of each operational definition would result in a reduced proportion of cognitively unimpaired individuals compared to the initial liberal DCN inclusion criteria. Furthermore, we expected the probability for progression to dementia would be higher in operationally reclassified individuals than in those classified initially. Finally, based on the studies of Bondi et al. [11] and Oltra-Cucarella et al. [30], we also expected that the comprehensive criteria and the base rate approach would predict incident dementia particularly well.

MATERIALS AND METHODS

Subjects

We analyzed data from memory clinic patients without dementia from the observational diagnostic and prognostic study (DAP) of the German Dementia Competence Network [32]. Patients over the age of 50 referred because of memory complaints were included in the DAP study when they had a Clinical Dementia Rating (CDR) of 0.5, at least one mild cognitive impairment (1 SD below demographically adjusted norms) on an extensive test battery (see below), an informant available and no cognitive impairment due to causes other than neurodegenerative or vascular disease. Further exclusion criteria were substance abuse or dependence, insufficient German language skills, multimorbidity, comorbid condition with excess mortality, circumstances that made regular attendance at follow-up visits uncertain [32]. Individuals were followed-up yearly (2003–2007) with a clinical and neuropsychological assessment.

We included 558 individuals (42% female, mean age = 66.02 years, $SD = 8.07$ years, mean level of education = 12.47 years, $SD = 2.82$ years) who had complete neuropsychological data (CERAD-Plus [40]) at baseline and at least one year of follow-up.

The DAP study was conducted in accordance with the principles of the Helsinki Declaration, and ethical approval was obtained by the Ethics Review Board of the Erlangen medical faculty and the Ethics Committees at each center involved. Subjects gave written informed consent.

Neuropsychological measures and definition of MCI

Original definition of MCI and dementia in the DAP study

The definition of MCI at baseline and follow-up was based on a clinical and neuropsychological examination and the following criteria: 1) decline in cognitive performance, indicated by a deficit of 1 SD below a normative comparison group on at least one of the following neuropsychological tests: the WMS-R Logical Memory immediate and delayed recall, CERAD word list (verbal learning and memory; 33, 34), CERAD figures recall (nonverbal learning and memory), CERAD animal fluency (word fluency; [35]), CERAD 15-item short form of the Boston Naming Test [36], CERAD figures copy or clock-drawing test (visuoconstruction; [34]), CERAD Trail Making Test A (TMT-A, cognitive speed) or Trail Making Test B (TMT-B, executive function), 2) Complaints of cognitive problems in daily life, 3) preserved independence in daily activities, whereas marginal deficits in the performance of complex everyday activities were tolerated (Bayer Activities of Daily Living Scale (B-ADL), 4) a global CDR score of 0.5.

A diagnosis of dementia (either upon inclusion or during follow-up) required impairments in demographically adjusted test scores on at least two cognitive domains, plus persisting impairments in functional activities (B-ADL > 6) [32].

Operational reclassifications of MCI

After each of the operational criteria were applied as described below, DAP MCI participants were reclassified into operational MCI and operational non-MCI participants. This classification was based on their test scores on the German CERAD-Plus test battery. The CERAD test battery is a translation of the original CERAD test battery developed in the US, which has demonstrated good objectivity, reliability (retest-reliability) and validity (e.g., [37, 38]). Many studies using the German CERAD version have also found good reliability [39] and validity [18, 40]. Age-, gender-, and education-adjusted

Table 1
Test scores of the CERAD-Plus test battery required to operationalize the different operational definitions

	Amnestic MCI	CERAD total score	Comprehensive criteria	Base rate correction
Operationalization	Impaired score (> 1 *SD*) below age-corrected normative mean in **one test** of verbal episodic memory	Impaired score (> 1 *SD*) below age-corrected normative mean on a **global score** that is based on six scores of three cognitive domains	Impaired score (> 1 *SD*) below age-corrected normative mean) on **two measures** in the **same cognitive domain**; or one impaired score in **each of the three cognitive domains** OR Functional impairment (IADL)	Impaired scores (> 1 *SD*) below age-corrected normative mean in at least **5 out of 10 test scores** out of four cognitive domains
Language				
Verbal Fluency		**x**	**x**	**x**
Boston Naming Test		**x**	**x**	**x**
Verbal episodic memory				
Word List Learning		**x**		**x**
Word List Delayed Recall	**x**	**x**	**x**	**x**
Word List Recognition		**x**	**x**	**x**
Word List Savings				**x**
Word List Intrusion Errors				**x**
Visual episodic memory				
Figures Delayed Recall				**x**
Figures Recall Savings				**x**
Constructional praxis				
Figures Copy		**x**		**x**
Attention / Executive function				
TMT-A			**x**	
TMT-B			**x**	

TMT-A and -B, Trail Making Test A and B; x, test scores required for the operationalization of the respective criteria; Word list recognition, number of true positives minus number of false positives.

values for the German CERAD-Plus test battery are available at https://www.memoryclinic.ch [41]. Standardization was based on 1,100 healthy individuals (age: 49–92 years; education: 7–20 years) for the CERAD test battery and 604 healthy individuals (age: 55–88 years; education: 7–20 years) for the additional elements of the CERAD-Plus test battery (TMT-A and -B). Table 1 shows the test scores of the CERAD-Plus test battery employed for the different MCI operational definitions, which are described as follows.

1. *Amnestic MCI* [42]. Typically, a 1.5 *SD* memory test deficit is used to operationalize aMCI, but a 1.0 *SD* deficit has also been applied for research purposes to capture early MCI [43]. Here, subjects performing 1.0 *SD* (or 1.5 *SD*, respectively) below age-, sex-, and education-adjusted normative values in the word list delayed recall of the CERAD-Plus test battery were reclassified as aMCI. For the three other MCI definitions, which rely on several tests instead of just one, we focused on a uniform 1 *SD* deficit, as proposed for the comprehensive criteria [11].
2. *CERAD total score* [14, 17]. We calculated a global score by summing six raw scores of the CERAD test battery, shown in Table 1, with a maximum score of 100 points [14]. The total score was adjusted for influences by age, sex, and education with a multiple regression-based formula. The 1 SD cognitive impairment threshold for this adjusted score is < = 88 points, based on the German CERAD normative sample [17].
3. *Comprehensive criteria* [11, 20]. Cognitive impairment was assumed when cognitive deficits quantified by 1 SD impairment threshold were present in 1) two tests in at least one cognitive domain (memory, executive function or language) or 2) one test of each cognitive domain. Deterioration in daily functioning, in the absence of cognitive impairment, constitutes a third, alternative criterion to define MCI. In our study, criteria 1 and 2 were operationalized

using six scores of the CERAD-Plus test battery (see Table 1). As independent daily functioning was required for study participation (B-ADL < 4), criterion 3 of the comprehensive criteria was never fulfilled in the present sample.

4. *Base rate correction.* We considered base rate percentages adjusted for influences by age, sex, and education in the German CERAD test battery of a normative group [5].

   We here focused on base rates provided for the 16th percentile (< = 1 SD) of test performance. According to Mistridis et al. [5] cognitive impairment (MCI) beyond the normal scatter of performance is assumed when 5 or more scores (out of 10 CERAD scores) are below 1 SD. This number of low scores corresponds to the bottom 10% of the normative sample (i.e., 90% have a maximum of 4 scores at or below 1 SD).

Outcome criteria: Dementia, AD dementia, and the CSF AD signature

Our primary outcome was progression to all-cause dementia, as liberal case selection in the DAP study yielded a heterogeneous sample with different underlying pathologies. As additional outcomes, we considered conversion to incident AD dementia and a biomarker profile indicative of AD, because most of the operational criteria have been developed and validated in the context of AD and give more weight to memory impairment as the most common and earliest cognitive deficit in patients with AD. Because of this, incident dementia of other types ($n = 48$) here was treated as nonconverted for the analysis of conversion from MCI to incident AD dementia. Probable AD was defined according the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and Alzheimer's Disease and Related Disorders Association (ADRDA) criteria [44].

A CSF signature indicative of AD was defined by an abnormal $A\beta_{42}$/tau ratio, as this ratio was found to be a reliable predictor for clinical progression from MCI to AD dementia [45]. In this study, we used Hulstaert et al.'s [46] formula ($A\beta_{42}$ / [240 + 1.18 × tau] < 1) to define an abnormal $A\beta_{42}$/tau ratio.

In addition, we examined the stability of MCI, i.e., the proportion of MCI subjects at baseline not reverting back to normal after one year in a subsequent examination using the same operational criteria. For this analysis we applied the same normative data of the respective neuropsychological subtests of the CERAD-Plus test battery at baseline and follow- up because retest-normative data of the test battery was not available.

Statistical analysis

Statistical analyses were carried out using IBM SPSS statistics 24 (Armonk, NY, USA) and MedCalc Software (Ostend, Belgium). Cohen's κ was calculated to analyze the pairwise diagnostic agreement among the criteria (0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement; 0.81–1 = almost perfect agreement, [47]). Concordance (%) was determined by the proportion of corresponding evaluations among all subjects ($\sum$ of agreement/N).

The diagnostic/predictive accuracy of the criteria was quantified by calculations of the sensitivity and specificity, positive and negative predictive values (PPVs and NPVs) and Youden's index (%) with regard to the described outcomes. Furthermore, we give information about the area under the curve (AUC) and corresponding confidence intervals given the predetermined criterion [48]. In addition, the risk of clinical progression for each MCI definition was determined by Cox proportional hazard regression analyses, with time to incident dementia (all cause and AD) as the outcome. We included age, education and gender as covariates in these analyses. Furthermore, we used logistic regression to analyze the association between group (MCI/non MCI) and the CSF-AD biomarker profile [46] for each operational definition applied. Age, education, and gender were again considered as covariates.

In addition, the risk of clinical progression for each MCI definition in combination with biomarker information and demographic information (age, years of education, gender) was determined by stepwise Cox proportional hazard regression analyses (Tables 6 and 7). In addition, we report AUCs derived from stepwise logistic regressions with the same combined predictors.

Model assumptions for the Cox-regression models and logistic regression models were checked via graphical inspections and appropriate residual methods (e.g., Schoenfeld and Martingale residual inspection for testing proportional hazard assumption and ruling out non-linearity associations between predictors and outcome, respectively). We found no evidence for violation of assumptions in any of the models. Thus, no further steps needed to be taken in the modeling process.

RESULTS

Baseline characteristics

The baseline characteristics of the study sample are shown in Table 2. As expected, due to the liberal inclusion of individuals in the DAP cohort, deficits indicated by average z-scores were mild and were mainly present in the episodic memory domain of the CERAD-Plus test battery.

Table 2
Demographics and neuropsychological performance of MCI subjects at baseline ($N = 558$)

	M or n	*SD*	*min*	*max*
Age, y	66.02	8.07	50	89
Sex (female/male)	237/321	n.a.	n.a.	n.a.
Education, y	12.47	2.82	5	19
APOE4 (yes/no) ($n = 470$)	181/289	n.a.	n.a.	n.a.
MADRS ($n = 539$)	7.77	6.28	0	36
MMSE (max. 30 points)	27.29	2.09	20	30
CERAD-Plus test battery				
Verbal Fluency	–0.79	1.11	–4.83	1.91
Boston Naming Test	–0.27	1.27	–4.89	1.93
Word List Immediate Recall	–1.22	1.33	–6.54	2.68
Word List Delayed Recall	–1.07	1.14	–4.41	2.11
Word List Recognition	–0.76	1.40	–5.06	1.19
Word List Savings	–0.66	2.76	–5.87	6.61
Word List Intrusions	–0.35	1.15	–3.44	0.98
Figures Copy	–0.23	1.26	–4.15	1.81
Figures Delayed Recall	–1.27	1.55	–5.46	2.21
Figures Recall Savings	–1.00	1.25	–3.86	2.50
Trail Making Test-A	–0.65	1.39	–4.44	4.21
Trail Making Test-B	–0.64	1.23	–3.21	4.77

M, mean; SD, standard deviation; min, minimum; max, maximum; MMSE, Mini-Mental State Examination; MADRS, Montgomery and Asberg Depression Rating Scale; n.a., not applicable.

Proportion of MCI and non-MCI

The prevalence rates of groups are displayed in Table 4 for each criterion. As expected, the operational criteria examined herein classified fewer (35%–58%) individuals of the original sample as having MCI. The 1.5 SD aMCI criterion was the most restrictive (34.6%), followed by the base rate correction (45.5%), the 1 SD aMCI criterion (49.6%), the comprehensive criteria (54.8%) and the CERAD total score (57.9%). Pairwise diagnostic agreement among the criteria revealed considerable overlap (70%–85%, Table 3).

Prediction of progression to incident all-cause dementia

During three years of follow-up (mean time to conversion = 25 months, SD = 10.3 months) 145 patients (26% of the sample with follow-up data) converted to dementia, with 97 patients receiving a diagnosis of AD dementia. Types of non-AD dementia ($n = 48$) were frontotemporal dementia ($n = 24$), dementia due to corticobasal degeneration ($n = 1$), Huntington's disease dementia ($n = 1$), Lewy body disease dementia ($n = 4$), vascular dementia ($n = 5$), Parkinson's

Table 3
Pairwise diagnostic agreement among the operational criteria

	aMCI (1 SD)	aMCI (1.5 SD)	CERAD total score (1 SD)	Comprehensive criteria (1 SD)
aMCI (1 *SD*)	–	–	–	–
aMCI (1.5 *SD*)	84.95% $\kappa = 698$ SE = 0.029 $p < 0.001$	–	–	–
CERAD total score (1 *SD*)	73.84% $\kappa = 0.477$ SE = 0.037 $p < 0.001$	69.53% $\kappa = 0.419$ SE = 0.033 $p < 0.001$	–	–
Comprehensive criteria (1 *SD*)	76.52% $\kappa = 0.531$ SE = 0.036 $p < 0.001$	69.71% $\kappa = 0.412$ SE = 0.035 $p < 0.001$	79.93% $\kappa = 0.592$ SE = 0.034 $p < 0.001$	–
Base rate correction (1 *SD*)	83.69% $\kappa = 0.674$ SE = 0.031 $p < 0.001$	79.39% $\kappa = 0.576$ SE = 0.034 $p < 0.001$	78.32% $\kappa = 0.572$ SE = 0.033 $p < 0.001$	75.09% $\kappa = 0.507$ SE = 0.036 $p < 0.001$

aMCI, amnestic MCI; %, Concordance ($\sum$ of agreement/N); κ, Cohen's κ (0.21–0.40 = fair agreement; 0.41–0.60 = moderate agreement, 0.61–0.80 = substantial agreement; 0.81–1 = almost perfect agreement, [47]); SE, standard deviation.

Table 4
Predictive accuracy of the respective operational neuropsychological criteria with regard to incident all-cause dementia and AD dementia

	MCI at baseline	PPV		NPV		Sensitivity		Specificity		J	AUC		Hazard ratio	
Operational definitions	%	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	value	95% CI	value	95% CI
(1) Predictive accuracy regarding incident all-cause dementia within three years														
Liberal DAP-MCI criteria (1 *SD*)	100	26.0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Amnestic MCI criteria (1 *SD*)	49.6	39.7	[36.2, 43.3]	87.5	[83.9, 90.5]	75.9	[68.1, 82.6]	59.6	[54.7, 64.3]	35.5	0.677[a]	[0.64, 0.72]	4.0	[2.7, 6.0]
Amnestic MCI criteria (1.5 *SD*)	34.6	**45.6**	**[40.4, 50.9]**	84.4	[81.4, 87.0]	60.7	[52.2, 68.7]	**74.1**	**[70.1, 78.7]**	35.3	0.676	[0.64, 0.72]	3.5	[2.4, 4.9]
CERAD total score (1 *SD*)	**57.9**	38.7	[35.9, 41.6]	**91.5**	**[87.6, 94.2]**	**86.2**	**[79.5, 91.4]**	52.1	[47.1, 57.0]	38.3	0.691	[0.65, 0.73]	5.3	[3.3, 8.6]
Comprehensive criteria (1 *SD*)	54.8	38.9	[35.8, 42.0]	89.7	[85.9, 92.6]	82.1	[74.8, 87.9]	54.7	[49.8, 59.6]	36.8	0.684	[0.64, 0.72]	4.8	[3.1, 7.4]
Base rate correction (1 *SD*)	45.5	44.1	[40.2, 48.1]	89.1	[85.8, 91.8]	77.2	[69.6, 83.8]	65.6	[60.8, 70.1]	**42.8**	**0.714[a]**	**[0.68, 0.75]**	**5.4**	**[3.6, 8.0]**
(2) Predictive accuracy regarding incident AD dementia within three years														
Liberal DAP-MCI criteria (1 *SD*)	100	17.4	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Amnestic MCI criteria (1 *SD*)	49.6	28.5	[25.7, 31.5]	93.6	[90.5, 95.7]	81.4	[72.3, 88.6]	57.1	[52.4, 61.6]	38.5	0.692[a]	[0.65, 0.73]	5.4	[3.2, 9,2]
Amnestic MCI criteria (1.5 *SD*)	34.1	**33.2**	[**28.8, 37.8**]	91.0	[88.4, 93.0]	66.0	[55.7, 75.3]	**72.1**	[**67.7, 76.1**]	38.0	0.690	[0.65, 0.73]	4.1	[2.6, 6.7]
CERAD total score (1 *SD*)	**57.9**	26.0	[23.8, 28.3]	**94.5**	**[91.1, 96.6]**	**86.6**	**[78.2, 92.7]**	48.2	[43.5, 52.8]	34.8	0.674[b]	[0.63, 0.71]	5.2	[2.9, 9.4]
Comprehensive criteria (1 *SD*)	54.8	24.8	[22.3, 27.5]	91.7	[88.2, 94.2]	78.4	[68.8, 86.1]	50.1	[45.5, 54.8]	28.5	0.642[a,c]	[0.60, 0.68]	3.8	[2.3, 6.2]
Base rate correction (1 *SD*)	45.5	31.5	[28.4, 34.8]	94.4	[91.6, 96.3]	82.5	[73.4, 89.4]	62.3	[57.7, 66.7]	**44.8**	**0.724[b,c]**	**[0.69, 0.76]**	**7.1**	**[4.2, 12.1]**

PPV, positive predictive value; NPV, negative predictive value; CL, confidence interval; J, Youden's index; AUC, area under the curve; n.a., not applicable, as follow- up information about subjects not fulfilling the liberal MCI inclusion criteria is not available; Letters (a, b, c) behind AUCs mark significant differences between those AUCs with the same letters; Highest values are highlighted in bold; Sample size: $N = 558$.

disease dementia ($n = 1$), or dementia due to unspecified causes ($n = 12$).

The results related to predictive accuracy of the operational MCI definitions regarding incident all-cause dementia (PPV, NPV, sensitivity, specificity, Youden's index, AUC) and risk of progression (quantified by hazard ratios; HRs) are displayed in Table 4.

The PPV of all operational MCI criteria ranged between 38.7% and 45.6%, and was substantially higher than the PPV regarding all cause dementia for the liberal DAP definition (26%). The rate of conversion to all-cause dementia was highest for the 1.5 SD aMCI criterion and the base rate correction, which were at the same time most restrictive. The NPV ranged between 84.4% and 91.5%, with the highest value for the CERAD total score and the comprehensive criteria. Sensitivity varied between 60.7% and 86.2%, with the CERAD total score and the comprehensive criteria being most sensitive. Specificity ranged between 52.1% and 74.1%, with the aMCI criterion (1.5 SD) and the base rate correction being most specific. Youden's index was highest for the base rate correction (Table 4). It is worth noting at this point that the absolute values of Youden's index are necessarily low because of the low rate of true-negative cases in our preselected memory clinic MCI sample, which reduces specificity because subjects without any cognitive deficit, who were unlikely to convert, were not in the sample. The highest AUC was found for the base rate correction (AUC = 0.714) which is significantly different from the AUC of the aMCI criterion (1 *SD;* Table 4). In line with the increased PPV values, the results of the Cox regression analyses showed a three- to fivefold higher risk of progression to incident all-cause dementia in the reclassified MCI groups than in their corresponding non-MCI groups, depending on the criteria applied. MCI reclassified with the CERAD total score and the base rate correction was associated with the highest risk of conversion within the following years (see HRs in Table 4).

Prediction of progression to incident AD dementia

The results related to predictive accuracy to incident AD dementia in the diagnostic groups and risk of progression are displayed in the lower part of Table 4.

Considering three years of follow-up, 17.4% of 558 DAP-MCI patients converted to AD dementia. A gain, reclassification resulted in a substantial increase in the PPV in comparison to the initial MCI sample PPV

for all operational definitions (24.8%–33.2%). The rate of conversion was again highest for the 1.5 SD aMCI criterion and for the base rate correction.

The pattern of sensitivity, specificity and Youden's index for the prediction of incident AD dementia by the different MCI criteria was similar to the one found for incident all cause dementia (Table 4).

The highest AUC was found for the base rate correction (AUC = 0.724), significantly superior to the AUC of the CERAD total score and the comprehensive criteria (Table 4).

While conversions also occurred in subjects not being re-classified operationally as MCI, the risk of conversion to incident AD dementia was four to eight times higher in the reclassified MCI groups than in the corresponding non-MCI groups (Table 4). The highest risk of conversion was found for those MCI cases identified with the base rate correction and the CERAD total score.

Cross-sectional prediction of a CSF-AD profile

In a subsample with available CSF data ($n = 360$, CSF-AD-positive: $n = 167$, CSF-AD-negative: $n = 193$), reclassification of the MCI individuals resulted in an increased probability of CSF-AD biomarker positivity (PPV: 51.1%–59.6%) compared to the initial DAP-MCI criteria (46.4%; Table 4). The base rate correction and the 1 SD aMCI criterion were associated with the highest probability of biomarker positivity. Furthermore, the aMCI criterion (1 SD) showed the best balance of sensitivity and specificity (see Youden's index in Table 5).

Operationally defined MCI groups were 1.6 to 3.6 times more likely to be biomarker positive than the corresponding non-MCI groups (see odds ratios in Table 5).

Differences in AUCs among the criteria are reported in Table 5.

Prediction of progression to incident dementia: Combining AD biomarkers with different MCI criteria

In the subsample that had information on both CSF and conversion to dementia ($n = 215$), combining AD biomarkers with MCI criteria improved the prediction of conversion to all-cause dementia and AD dementia (see differences in χ^2 on Tables 6 and 7). The prediction was most improved in combinations with multidomain MCI criteria (the base rate corrected MCI, the CERAD total score or the comprehensive

Table 5
Predictive accuracy of the respective operational neuropsychological criteria with regard to a concurrent CSF biomarker profile indicative for AD [46]

	MCI at baseline	PPV		NPV		Sensitivity		Specificity		J	AUC			Odds ratio	
Operational definitions	%	%	95% CI	%	95% CI	%	95% CI	%	95% CI	%	value	95% CI	value	95% CI
Liberal DAP-MCI criteria (1 *SD*)	100	46.4	**n.a.**	**n.a.**	**n.a.**	**n.a.**	**n.a.**	n.a.	n.a.	**n.a.**	**n.a.**	**n.a.**	**n.a.**	**n.a.**
Amnestic MCI criteria (1 *SD*)	56.4	**59.6**	**[55.0, 64.1]**	**70.7**	**[64.7, 76.0]**	**72.5**	**[65.0, 79.1]**	57.5	[50.2, 64.6]	**30.0**	**0.650**[a,b]	**[0.60, 0.70]**	**3.6**	**[2.2, 5.9]**
Amnestic MCI criteria (1.5 *SD*)	39.3	57.1	[48.8, 65.1]	63.9	[58.3, 69.1]	50.5	[40.1, 61.0]	**69.8**	**[60.7, 77.8]**	20.3	0.619[c]	[0.57, 0.67]	2.7	[1.7, 4.4]
CERAD total score (1 *SD*)	**65.3**	51.1	[47.3, 54.8]	62.4	[55.2, 69.1]	71.9	[64.4, 78.5]	40.4	[33.4, 47.7]	12.3	0.561[a,c,d]	[0.51, 0.61]	1.6	[0.9, 2.5]
Comprehensive criteria (1 *SD*)	62.2	53.1	[49.1, 57.1]	64.7	[58.0, 70.9]	71.3	[63.8, 78.0]	45.6	[38.4, 52.9]	16.9	0.584[b,d,e]	[0.53, 0.64]	2.2	[1.3, 3.6]
Base rate correction (1 *SD*)	51.9	59.4	[54.3, 64.2]	67.6	[62.1, 72.7]	66.5	[58.8, 73.6]	60.6	[53.4, 67.6]	27.1	0.635[d,e]	[0.58, 0.69]	3.1	[1.9, 5.0]

PPV, positive predictive value; NPV, negative predictive value; CL, confidence interval; J, Youden's index; AUC, area under the curve; n.a., not applicable, as follow-up information about subjects not fulfilling the liberal MCI inclusion criteria is not available; Letters (a, b, c, d, e) behind AUCs mark significant differences between those AUCs with the same letters; Highest values are highlighted in bold; Sample size: $n = 360$.

Table 6
Prediction of conversion to all-cause dementia: Results of stepwise cox regression and logistic regressions analyses ($n = 215$; non converters: $n = 147$; converters: $n = 68$)

Predictors of conversion to all-cause dementia	Model fit: χ^2			Hazard ratio		AUC
	difference in χ^2	df	p	value	95.0% CI	value
Model 1: Demographic information only (gender, age, years of education)						
gender				1.1	[.68, 1.8]	
age				1.0	[1.0, 1.1]	
years of education				0.96	[0.87, 1.0]	
Model 2: Model 1 plus CSF AD biomarker risk profile	16.4	1	<0.001	3.0	[1.7, 5.2]	
Model 3: Model 2 plus operational MCI criteria						
Model 2 plus Amnestic MCI criteria (1 *SD*)	8.1	1	<0.05	2.3	[1.3, 4.1]	0.717
Model 2 plus Amnestic MCI criteria (1.5 *SD*)	7.6	1	<0.05	2.1	[1.2, 3.5]	0.721
Model 2 plus CERAD total score (1 *SD*)	16.2	1	<0.001	3.4	[1.7, 6.6]	0.742
Model 2 plus Comprehensive criteria (1 *SD*)	19.2	1	<0.001	3.8	[1.9, 7.3]	0.745
Model 2 plus Base rate correction (1 *SD*)	22.8	1	<0.001	3.8	[2.1, 6.7]	0.747

CSF, cerebrospinal fluid; SD, standard deviation; MCI, mild cognitive impairment; AUC, area under the curve.

Table 7
Prediction of conversion to AD dementia: Results of stepwise cox regression and logistic regressions analyses ($n = 215$; non converters: $n = 171$; converters: $n = 44$)

Predictors of conversion to all-cause dementia	Model fit: χ^2			Hazard ratio		AUC
	difference in χ^2	df	p	value	95.0% CI	value
Model 1: Demographic information only (gender, age, years of education)						
gender				1.5	[0.84, 2.8]	
age				1.1	[1.0, 1.1]	
years of education				0.95	[0.85, 1.1]	
Model 2: Model 1 plus CSF AD biomarker risk profile	38.3	1	<0.001	13.0	[4.5, 37.5]	
Model 3: Model 2 plus operational MCI criteria						
Model 2 plus Amnestic MCI criteria (1 *SD*)	4.8	1	<0.05	2.4	[1.1, 5.4]	0.819
Model 2 plus Amnestic MCI criteria (1.5 *SD*)	5.0	1	<0.05	2.1	[1.1, 4.1]	0.825
Model 2 plus CERAD total score (1 *SD*)	5.1	1	<0.05	2.4	[1.1, 5.3]	0.825
Model 2 plus Comprehensive criteria (1 *SD*)	6.4	1	<0.05	2.6	[1.2, 5.8]	0.825
Model 2 plus Base rate correction (1 *SD*)	13.2	1	<0.001	3.8	[1.7, 8.4]	0.842

CSF, cerebrospinal fluid; SD, standard deviation; MCI, mild cognitive impairment; AUC, area under the curve.

criteria). Again, the base rate corrected MCI came out as the best among equals (it added most to the CSF-based prediction).

The combined prediction with AD biomarkers and MCI criteria was substantial (AUC up to 0.842) and was better than prediction based on MCI criteria alone (AUCs of 0.677 to 0.724).

Temporal stability (proportion of MCI cases not reverting to normal)

Out of 558 cases, we included 493 individuals with complete neuropsychological data (=all CERAD subtests necessary for definition of the operational criteria) at the first follow-up in an analysis of temporal stability. Only for this analysis, we excluded 65 cases due to missing values at the first follow-up that precluded an operationalization according to one or more of the criteria. The number of cases included in the analysis differed slightly among the criteria due to differences in the proportion of impaired individuals at baseline. The proportion of individuals still fulfilling the respective operational criteria for MCI longitudinally also includes individuals who progressed to dementia at the first follow-up or later.

Stability was lowest for the aMCI criterion (1 SD and 1.5 SD: 70% each). Temporal stability was highest for the CERAD total score (79%), followed by the comprehensive criteria and the base rate approach (75% each).

DISCUSSION

In a large and well-characterized multi-center memory clinic sample, we compared different operational definitions of MCI that might be useful

specifications of the generic DSM-5 definition of mild NCD, or of the NIA-AA definitions of MCI. As the original sample included patients with memory complaints already when they had at least one deviant (1 SD) test score among 12 scores considered, it is not surprising the application of each of the operational MCI definitions resulted in a reduced proportion of cognitively impaired cases, with only 34.6% to 58% reclassified as MCI. At the same time, all operational criteria showed substantially enhanced rates of conversion to all-cause and AD dementia (PPV = 25–46%) as well as increased probability of CSF-AD biomarker positivity (PPV = 51–59%) in comparison to the initial criteria.

Most operational criteria, except the 1.5 SD aMCI criterion, used 1 SD deficit as threshold for cognitive impairment. In line with previous findings [11] the substantial PPVs (and HRs) suggest that even a liberal 1-SD deficit threshold can be employed to indicate a clinically meaningful impairment in memory clinic patients when using one of these operational definitions of MCI. Of course, other relevant variables (e.g., current depression or other medical conditions) will need to be considered as contributing to performance, particularly when liberal cutoffs are used. It is beyond the scope of the present paper to examine which cutoffs would be optimal for each of the criteria. However, we note that the overall predictive accuracy of 1.0 and 1.5 SD aMCI definitions, respectively, was quite similar.

Our results show that meeting any of the operationally defined MCI/mild NCD criteria indicates a rather high probability for conversion to dementia or AD dementia within the next three years, and of a cross-sectional biomarker profile indicative of AD. Individuals not fulfilling the operational criteria were at a much lower risk, as reflected by the HRs. However, the NPV was approximately 90% regarding progression to dementia for all criteria in this reanalysis, which implies that there are still true prodromal cases with only subtle cognitive impairments among those not fulfilling the operational criteria. This is in line with the idea of a pre-MCI stage of subjective cognitive decline, where memory concerns are present but cognitive impairment is either absent or below the detection threshold of current tests [49].

Progression rates were roughly similar across the examined 1 SD multidomain criteria and the traditional 1.5 SD aMCI Peterson criterion (about 40–45% progressed to dementia). This is higher than the PPV based on the liberal 1 SD inclusion criterion (26% progression to dementia). The difference between the original (liberal) and the *post hoc* operational criteria confirms that differences in MCI criteria indeed can explain substantial variance in progression rates from MCI to dementia reported for different studies, in addition to sample composition (e.g., age, comorbidities, referral pathways). However, in similarly composed samples, any of the more restrictive operational criteria examined herein should identify a group of subjects at a similar high risk for clinical progression.

Regarding the differences between the operational definitions, the overall accuracy (quantified by the Youden's index and the AUCs) in predicting the risk of progression (quantified by HRs) to all-cause dementia or AD dementia was highest for the base rate correction, with generally small (but sometimes statistically significant) differences in comparison to the AUCs other criteria. With regard to prediction of an AD biomarker profile, the base rate correction came out second, close to, and not inferior to, the 1 SD aMCI criterion, and the AUC of both criteria differed significantly from the AUCs for the CERAD total score and the comprehensive criteria. These findings validate the predictive value of the base rate approach in a much larger sample than that examined before by Mistridis et al. [5]. In a recent reanalysis of ADNI data, Oltra-Cucarella et al. [30] also found that the base rate approach to the classification of MCI performed somewhat better in predicting incident dementia than did other MCI criteria, in terms of a high PPV and a balanced sensitivity and specificity.

The base rate correction, and to a similar degree also the CERAD total score, predicted all cause dementia as well as AD dementia, despite being heavily weighted for the AD-sensitive memory measures. This may be a consequence of the memory concerns being an initial selection criterion, and the fact that a majority (67%) of the incident all-cause dementia cases in our memory clinic sample received a diagnosis of AD-dementia, but it may also suggest that these MCI definitions are clinically useful in general, predicting progression even in non-AD patients who attend a memory clinic. Furthermore, as both the CERAD total score and the base rate approach can aggregate several low scores from several different tests, they are giving information about "probably true" cognitive impairment.

In contrast to previous findings [11], the comprehensive criteria were not associated with a higher predictive accuracy than was the aMCI criterion in our sample. Rather, the aMCI criterion seems to

be slightly better than the comprehensive criteria in predicting a positive AD biomarker profile (quantified by Youden's index). This may be a result of the fact that we used the same tests for all operational MCI definitions (including aMCI), rather than contrasting the ADNI aMCI definition (based on the Wechsler Logical Memory Test) with the comprehensive criteria (using the Rey Auditory Verbal Learning Test), as in [11] or in [30]. Further comparisons of different operational definitions in other samples, based on identical neuropsychological measures, will be useful to judge the merits of different definitions.

Cognitive assessment is increasingly used in combination with biomarkers for individual diagnosis and prognosis. When we asked which MCI criterion would add the most information to a CSF AD biomarker profile in terms of dementia prediction, it were again the multidomain criteria (base rate MCI, CERAD total score, and comprehensive criteria) which performed best, which is not surprising given the close association of amnestic MCI with AD. Again, the base rate corrected MCI came out as the best among equals (it added most to the CSF-based prediction). In absolute terms, the combined prediction of dementia by the $A\beta_{42}$/tau ratio and the MCI criteria achieved AUCs between 0.72 to 0.84, as compared to AUCs of 0.68 to 0.72 without considering biomarker information (Tables 6 and 7). It needs to be stressed again that these numbers come from our "liberal" MCI sample, where an unknown number of memory clinic attendants was not included initially. In unselected memory clinic samples, the combined prediction would likely be higher.

The rates of reversion to normal cognition in our patient sample ranged from 21–30%, somewhat higher than expected for memory clinic samples, where meta-analyses found average rates of 14% [50]. Thomas et al. [51] found a one-year reversion rate of 15.8%, when applying comprehensive neuropsychological criteria (Jak/Bondi criteria) for analyzing the degree of reverting MCI individuals in ADNI. Various aspects seem to influence the degree of reversion back to normal as summarized by Thomas et al. [51], such as for example younger age, better neuropsychological and functional performance, amnestic and multiple domain MCI, absence of *APOE4* or absence of an biomarker profile indicative of AD. We note that our sample was only mildly impaired on average, and that the same norms were applied for baseline and follow-up testing. Test repetition effects might improve the scores of subjects at second testing, sometimes below the MCI threshold. The further analysis of these aspects in our sample is beyond the scope of this manuscript. Importantly, all criteria were similarly affected by possible test-repetition effects.

In general, the return to a cognitive performance within the normal range once the criteria for MCI have been met supports the idea that MCI is a heterogeneous condition with different courses or trajectories [50]. Furthermore, even reverting individuals, rather than being only false positives, have been found to remain at increased risk of future cognitive decline and dementia compared to those who have never met the criteria for MCI before [52, 53], which emphasizes the prognostic value of an MCI classification [53].

Implications

The probability of conversion to dementia within three years, in subjects fulfilling any of the analyzed operational definitions of mild NCD is substantial, about 40%. Differences between the three operational definitions using multiple tests were not pronounced and varied with the outcome studied. Furthermore, pairwise concordance is substantial (averaging 76% across all pairwise comparisons). Thus, each of these operational criteria may be used to specify the general neuropsychological mild NCD criteria in clinical routine, e.g., for indicating the need for further diagnostic procedures in those passing the threshold.

The risk of progression derived on the group level for each operational definition can be used to gauge the individual risk of progression, taking further risk factors for progression into account. In addition, fulfilling any of the criteria might also call for close longitudinal monitoring and treatment efforts targeting other risk factors for progression to dementia, such as hypertension or obesity.

Furthermore, each of these criteria could be used as an adjunctive measure, on the one hand, to calibrate differing local procedures to determine MCI/mild NCD and, on the other hand, to verify the judgement based on clinical assessment in memory clinics.

All operational criteria markedly improved the prediction of a cross-sectional AD biomarker profile. Using any of the criteria examined, the probability of having an abnormal CSF $A\beta_{42}$/tau ratio is over 50%. Using these criteria as inclusion criteria will help to select subjects for clinical AD trials.

Finally, these criteria might be used to align MCI definitions across studies with different inclusion criteria for *post hoc* integrated data analyses.

Considerations regarding the use of these operational definitions

In general, the test battery used in this study is well suited to depict all proposed criteria, as not only do the number of tests and thresholds matter [4, 6], but the characteristics of the subtest, such as the sensitivity to impairment and the reliability of the measurement, are also important.

As none of the studied operational criteria approaches was clearly superior to the others, pragmatic considerations may determine which one to use.

Where the CERAD test battery is integrated into the neuropsychological test routine, like in many European memory clinics, both the CERAD total score and the base rate approach can be recommended to derive a diagnosis of MCI, or to provide some "calibration" for the clinical judgment of the overall test pattern.

For the sake of consistency, in this study we determined 1 SD as cutoff for cognitive impairment for most operational criteria. However, within the base rate correction approach [5] base rates are provided for six thresholds of single test scores ($\leq$2.32 *SD*, $\leq$–1.96 *SD*, $\leq$–1.48 *SD*, $\leq$–1.28 *SD*, $\leq$–1.0 *SD*, $\leq$–0.67 *SD*). It is unknown, which of these thresholds might be superior in predicting conversion to all-cause—or AD dementia or a biomarker profile indicative for AD in our study and should be used accordingly in memory clinics. Thus, in an additional analysis (Supplementary Table 1), we compared these six thresholds provided by Mistridis et al. and found that the thresholds of 1 SD and 1.28 SD gave almost identical results and were superior to other thresholds in predicting the outcomes mentioned above. This suggests that many mild deficits may carry more diagnostic/prognostic information than fewer pronounced deficits, even when the base rates are properly adjusted. It also suggests that the base rate correction approach, which came out favorably in the present comparisons, works best with one of these two thresholds.

In general, the application of the base rate approach for defining cognitive impairment is not limited to the CERAD test battery, but it requires the existence of base rates calculated in normative controls. A limiting factor is that this information is rarely available. It is desirable that more test batteries provide this information. Base rates for the NACC 3.0 UDS test battery, based on NACC controls, have recently been published by Kiselica at al. [29].

Our findings regarding the base rate correction approach complement those of other working groups [29–31]. For example, Oltra-Cucarella et al. [30] found higher rates of conversion (PPVs) to dementia for individuals classified as MCI by a base rate approach with the number of impaired tests considered as for individuals classified by other operational definitions such as the Petersen criteria and Jak/Bondi criteria in ADNI.

The aMCI criterion, focusing on episodic memory only, continue to be a simple and valid method when risk enrichment strategies are pursued in clinical studies. However, these criterion does not capture the multidomain assessment approach inherent in the DSM-5 mild NCD definition. Another disadvantage with the amnestic MCI approach is that executive deficits may be among the first subtle signs of cognitive decline [54] and that other MCI subtypes cannot be identified with this criterion. Because of its close association with AD biomarkers, aMCI is well suited to preselect subjects for research studies on AD. However, in clinical settings, aMCI adds less information to an AD biomarker profile than do the other MCI criteria regarding a combined risk prediction.

In many settings where several tests reflecting numerous cognitive domains are routinely used, the comprehensive criteria might be advantageous. They conceptually match the generic concept of mild NCD as being agnostic to a specific etiology because only the comprehensive criteria give equal emphasis to the domains of memory, language, and executive function. Other operational definitions explored here were geared toward typical AD deficits and give more weight to episodic memory function (we have previously discussed that this weighting may be rational given the prevalence of prodromal AD cases in a memory clinic setting). Second, the application of the comprehensive criteria is independent of a specific test score, a specific test battery, or the existence of normative base rates. This flexibility is advantageous for use in different memory clinics with differing local assessments and in the challenging situation of harmonizing multiple study cohorts for joint scientific analysis. The comprehensive criteria, however, do not consider visual cognition and attention separately (although many executive tests, including the TMTs, capture these domains), and would need adaptation if more than three cognitive domains are considered.

Despite each of the operational definitions examined herein has some limitations, all of them have the major advantage of being explicit, in contrast to the

poorly specified clinical judgement of a more or less extensive neuropsychological test battery.

The operational MCI definitions studied here were not exhaustive and represent a selection of criteria currently in use. Thus, further comparative analyses including other criteria (e.g., based on serial assessment as suggested by NIA-AA [55]) will be interesting in future studies.

Limitations and strengths of the present study

Although the three years of follow-up time in our study represent a clinically relevant timeframe for a memory clinic sample, longer periods of follow-up would have been desirable for our study question, especially given that subtle cognitive performance deficits at baseline elevate uncertainty regarding clinical progression. It should be noted that, due to the preselection of memory clinic patients with at least a 1 SD cognitive deficit on one test, specificity, NPV and Youden's index for all operational criteria are systematically underestimated in our sample because subjects without any cognitive deficit, who are also unlikely to convert (true-negative cases), were systematically excluded (and not followed up). The same preselection also will have excluded cases in an early preclinical stage of a neurodegenerative disease (false-negative cases). These cases, however, would probably not convert to dementia within three years. Thus, the exclusion of such false negatives unlikely affected the reported values.

As our study sample was recruited from a memory clinic, it represents a specific group of individuals, which thus limits generalizability. Our sample consists of individuals with mainly episodic memory deficits, which is to be expected given that episodic memory deficits are a clinical hallmark of AD as the most common cause of dementia, accounting for approximately 70% of all dementia cases. Our results are also limited by the test battery applied in this study and used for reclassification. A more extensive test battery would have allowed for more distinct MCI operationalizations, with a lower degree of pairwise overlap. However, the CERAD-plus test battery, as a frequently used battery in clinical practice, is suited to depict all proposed criteria sufficiently and even with the unavoidable overlap between our *post hoc* MCI definitions, we were able to identify some suggestive differences between them.

To conclude, our study provides comprehensive empirical data about the prognostic accuracy of four neuropsychological MCI/mild NCD criteria. The operational criteria examined seem suitable to specify mild NCD in memory clinic settings, as they identify subjects at high risk of future clinical progression. Depending on the neuropsychological battery in use, one or several of these criteria might be useful in calibrating the clinical judgment of test results, reducing false-positive decisions, and defining risk-enriched clinical groups for clinical trials. The base rate correction approach for the definition of MCI seems to have particular merits in terms of predictive validity and might be considered as a good standard to define MCI wherever normative base rates are available.

ACKNOWLEDGMENTS

Data acquisition has been supported by a grant from the German Federal Ministry of Education and Research (BMBF): Kompetenznetz Demenzen (01GI0420). Additional funding was provided by the DZNE, German Center for Neurodegenerative Diseases, Bonn, Germany, and by the Department of Neurodegenerative Diseases and Geriatric Psychiatry, Bonn, Germany.

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-5548r2).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-215548.

REFERENCES

[1] Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, Leon M de, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B (2006) Mild cognitive impairment. *Lancet* **367**, 1262-1270.

[2] American Psychiatric Association (2013) *Diagnostic and statistical manual of mental disorders: DSM-5,* 5. ed., American Psychiatric Publishing, Washington, DC.

[3] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (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. *Alzheimers Dement* **7**, 270-279.

[4] Brooks BL, Iverson GL, White T (2007) Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *J Int Neuropsychol Soc* **13**, 490-500.

[5] Mistridis P, Egli SC, Iverson GL, Berres M, Willmes K, Welsh-Bohmer KA, Monsch AU (2015) Considering the

base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the Consortium to Establish a Registry for Alzheimer's Disease-Neuropsychological Assessment Battery (CERAD-NAB). *Eur Arch Psychiatry Clin Neurosci* **265**, 407-417.

[6] Schretlen DJ, Testa SM, Winicki JM, Pearlson GD, Gordon B (2008) Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *J Int Neuropsychol Soc* **14**, 436-445.

[7] Mitchell AJ, Shiri-Feshki M (2009) Rate of progression of mild cognitive impairment to dementia–meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* **119**, 252-265.

[8] Sachs-Ericsson N, Blazer DG (2015) The new DSM-5 diagnosis of mild neurocognitive disorder and its relation to research in mild cognitive impairment. *Aging Ment Health* **19**, 2-12.

[9] Edmonds EC, Delano-Wood L, Clark LR, Jak AJ, Nation DA, McDonald CR, Libon DJ, Au R, Galasko D, Salmon DP, Bondi MW (2015) Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimers Dement* **11**, 415-424.

[10] Eppig JS, Edmonds EC, Campbell L, Sanderson-Cimino M, Delano-Wood L, Bondi MW (2017) Statistically derived subtypes and associations with cerebrospinal fluid and genetic biomarkers in mild cognitive impairment: A latent profile analysis. *J Int Neuropsychol Soc* **23**, 564-576.

[11] Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, Nation DA, Libon DJ, Au R, Galasko D, Salmon DP (2014) Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis* **42**, 275-289.

[12] Clark LR, Delano-Wood L, Libon DJ, McDonald CR, Nation DA, Bangen KJ, Jak AJ, Au R, Salmon DP, Bondi MW (2013) Are empirically-derived subtypes of mild cognitive impairment consistent with conventional subtypes? *J Int Neuropsychol Soc* **19**, 635-645.

[13] Bondi MW, Smith GE (2014) Mild cognitive impairment: A concept and diagnostic entity in need of input from neuropsychology. *J Int Neuropsychol Soc* **20**, 129-134.

[14] Chandler MJ, Lacritz LH, Hynan LS, Barnard HD, Allen G, Deschner M, Weiner MF, Cullum CM (2005) A total score for the CERAD neuropsychological battery. *Neurology* **65**, 102-106.

[15] Seo EH, Lee DY, Lee JH, Choo IH, Kim JW, Kim SG, Park SY, Shin JH, Do YJ, Yoon JC, Jhoo JH, Kim KW, Woo JI (2010) Total scores of the CERAD neuropsychological assessment battery: Validation for mild cognitive impairment and dementia patients with diverse etiologies. *Am J Geriatr Psychiatry* **18**, 801-809.

[16] Paajanen T, Hänninen T, Tunnard C, Mecocci P, Sobow T, Tsolaki M, Vellas B, Lovestone S, Soininen H, Addneuromed Consortium (2010) CERAD neuropsychological battery total score in multinational mild cognitive impairment and control populations: The AddNeuroMed study. *J Alzheimers Dis* **22**, 1089-1097.

[17] Ehrensperger MM, Berres M, Taylor KI, Monsch AU (2010) Early detection of Alzheimer's disease with a total score of the German CERAD. *J Int Neuropsychol Soc* **16**, 910-920.

[18] Wolfsgruber S, Jessen F, Wiese B, Stein J, Bickel H, Mösch E, Weyerer S, Werle J, Pentzek M, Fuchs A, Köhler M, Bachmann C, Riedel-Heller SG, Scherer M, Maier W, Wagner M (2014) The CERAD Neuropsychological Assessment Battery Total Score detects and predicts Alzheimer disease dementia with high diagnostic accuracy. *Am J Geriatr Psychiatry* **22**, 1017-1028.

[19] Rossetti HC, Munro Cullum C, Hynan LS, Lacritz L (2010) The CERAD Neuropsychological Battery Total Score and the progression of Alzheimer's disease. *Alzheimer Dis Assoc Disord* **24**, 138-142.

[20] Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, Delis DC (2009) Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am J Geriatr Psychiatry* **17**, 368-375.

[21] Loewenstein DA, Acevedo A, Small BJ, Agron J, Crocco E, Duara R (2009) Stability of different subtypes of mild cognitive impairment among the elderly over a 2- to 3-year follow-up period. *Dement Geriatr Cogn Disord* **27**, 418-423.

[22] Wong CG, Thomas KR, Edmonds EC, Weigand AJ, Bangen KJ, Eppig JS, Jak AJ, Devine SA, Delano-Wood L, Libon DJ, Edland SD, Au R, Bondi MW (2018) Neuropsychological criteria for mild cognitive impairment in the Framingham Heart Study's Old-Old. *Dement Geriatr Cogn Disord* **46**, 253-265.

[23] Jak AJ, Preis SR, Beiser AS, Seshadri S, Wolf PA, Bondi MW, Au R (2016) Neuropsychological criteria for mild cognitive impairment and dementia risk in the Framingham Heart Study. *J Int Neuropsychol Soc* **22**, 937-943.

[24] Jak AJ, Urban S, McCauley A, Bangen KJ, Delano-Wood L, Corey-Bloom J, Bondi MW (2009) Profile of hippocampal volumes and stroke risk varies by neuropsychological definition of mild cognitive impairment. *J Int Neuropsychol Soc* **15**, 890-897.

[25] Edmonds EC, Delano-Wood L, Jak AJ, Galasko DR, Salmon DP, Bondi MW (2016) "Missed" mild cognitive impairment: High false-negative error rate based on conventional diagnostic criteria. *J Alzheimers Dis* **52**, 685-691.

[26] Molinuevo JL, Rabin LA, Amariglio R, Buckley R, Dubois B, Ellis KA, Ewers M, Hampel H, Kloppel S, Rami L, Reisberg B, Saykin AJ, Sikkes S, Smart CM, Snitz BE, Sperling R, van der Flier WM, Wagner M, Jessen F (2017) Implementation of subjective cognitive decline criteria in research studies. *Alzheimers Dement* **13**, 296-311.

[27] Wolfsgruber S, Polcher A, Koppara A, Kleineidam L, Frölich L, Peters O, Hüll M, Rüther E, Wiltfang J, Maier W, Kornhuber J, Lewczuk P, Jessen F, Wagner M (2017) Cerebrospinal fluid biomarkers and clinical progression in patients with subjective cognitive decline and mild cognitive impairment. *J Alzheimers Dis* **58**, 939-950.

[28] Eliassen CF, Reinvang I, Selnes P, Fladby T, Hessen E (2017) Convergent results from neuropsychology and from neuroimaging in patients with mild cognitive impairment. *Dement Geriatr Cogn Disord* **43**, 144-154.

[29] Kiselica AM, Webber TA, Benge JF (2020) Using multivariate base rates of low scores to understand early cognitive declines on the uniform data set 3.0 Neuropsychological Battery. *Neuropsychology* **34**, 629-640.

[30] Oltra-Cucarella J, Sánchez-SanSegundo M, Lipnicki DM, Sachdev PS, Crawford JD, Pérez-Vicente JA, Cabello-Rodríguez L, Ferrer-Cascales R (2018) Using base rate of low scores to identify progression from amnestic mild cognitive impairment to Alzheimer's disease. *J Am Geriatr Soc* **66**, 1360-1366.

[31] Oltra-Cucarella J, Sánchez-SanSegundo M, Rubio-Aparicio M, Arango-Lasprilla JC, Ferrer-Cascales R (2021) The association between the number of neuropsychological

measures and the base rate of low scores. *Assessment* **28**, 955-963.

[32] Kornhuber J, Schmidtke K, Frolich L, Perneczky R, Wolf S, Hampel H, Jessen F, Heuser I, Peters O, Weih M, Jahn H, Luckhaus C, Hüll M, Gertz H-J, Schröder J, Pantel J, Rienhoff O, Seuchter SA, Rüther E, Henn F, Maier W, Wiltfang J (2009) Early and differential diagnosis of dementia and mild cognitive impairment: Design and cohort baseline characteristics of the German Dementia Competence Network. *Dement Geriatr Cogn Disord* **27**, 404-417.

[33] Atkinson RC, Shiffrin RM (1971) The control of short-term memory. *Sci Am* **225**, 82-91.

[34] Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *Am J Psychiatry* **141**, 1356-1364.

[35] Isaacs B, Kennie AT (1973) The Set Test as an aid to the detection of dementia in old people. *Br J Psychiatry* **123**, 467-470.

[36] Kaplan E, Goodglass H, Weintraub S (2001) *Boston naming test*, Pro-ed.

[37] Morris JC, Edland S, Clark C, Galasko D, Koss E, Mohs R, van Belle G, Fillenbaum G, Heyman A (1993) The consortium to establish a registry for Alzheimer's disease (CERAD). Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. *Neurology* **43**, 2457-2465.

[38] Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C (1989) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159-1165.

[39] Stein J, Luppa M, Luck T, Maier W, Wagner M, Daerr M, van den Bussche H, Zimmermann T, Köhler M, Bickel H, Mösch E, Weyerer S, Kaufeler T, Pentzek M, Wiese B, Wollny A, König H-H, Riedel-Heller SG (2012) The assessment of changes in cognitive functioning: Age-, education-, and gender-specific reliable change indices for older adults tested on the CERAD-NP Battery: Results of the German Study on Ageing, Cognition, and Dementia in Primary Care Patients (AgeCoDe). *Am J Geriatr Psychiatry* **20**, 84-97.

[40] Schmid NS, Ehrensperger MM, Berres M, Beck IR, Monsch AU (2014) The extension of the German CERAD Neuropsychological Assessment Battery with tests assessing subcortical, executive and frontal functions improves accuracy in dementia diagnosis. *Dement Geriatr Cogn Dis Extra* **4**, 322-334.

[41] Berres M, Monsch AU, Bernasconi F, Thalmann B, Stahelin HB (2000) Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. *Stud Health Technol Inform* **77**, 195-199.

[42] Petersen RC, Morris JC (2005) Mild cognitive impairment as a clinical entity and treatment target. *Arch Neurol* **62**, 1160-1163; discussion 1167.

[43] Aisen PS, Petersen RC, Donohue MC, Gamst A, Raman R, Thomas RG, Walter S, Trojanowski JQ, Shaw LM, Beckett LA, Jack CR, Jagust W, Toga AW, Saykin AJ, Morris JC, Green RC, Weiner MW (2010) Clinical core of the Alzheimer's disease neuroimaging initiative: Progress and plans. *Alzheimers Dement* **6**, 239-246.

[44] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (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**, 939-944.

[45] Duits FH, Teunissen CE, Bouwman FH, Visser P-J, Mattsson N, Zetterberg H, Blennow K, Hansson O, Minthon L, Andreasen N, Marcusson J, Wallin A, Rikkert MO, Tsolaki M, Parnetti L, Herukka S-K, Hampel H, Leon MJ de, Schröder J, Aarsland D, Blankenstein MA, Scheltens P, van der Flier, Wiesje M (2014) The cerebrospinal fluid "Alzheimer profile": Easily said, but what does it mean? *Alzheimers Dement* **10**, 713-723.e2.

[46] Hulstaert F, Blennow K, Ivanoiu A, Schoonderwaldt HC, Riemenschneider M, Deyn PP de, Bancher C, Cras P, Wiltfang J, Mehta PD, Iqbal K, Pottel H, Vanmechelen E, Vanderstichele H (1999) Improved discrimination of AD patients using beta-amyloid(1-42) and tau levels in CSF. *Neurology* **52**, 1555-1562.

[47] Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, 159-174.

[48] DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837-845.

[49] Jessen F, Amariglio RE, Buckley RF, van der Flier WM, Han Y, Molinuevo JL, Rabin L, Rentz DM, Rodriguez-Gomez O, Saykin AJ, Sikkes SAM, Smart CM, Wolfsgruber S, Wagner M (2020) The characterisation of subjective cognitive decline. *Lancet Neurol* **19**, 271-278.

[50] Malek-Ahmadi M (2016) Reversion From mild cognitive impairment to normal cognition: A meta-analysis. *Alzheimer Dis Assoc Disord* **30**, 324-330.

[51] Thomas KR, Edmonds EC, Eppig JS, Wong CG, Weigand AJ, Bangen KJ, Jak AJ, Delano-Wood L, Galasko DR, Salmon DP, Edland SD, Bondi MW (2019) MCI-to-normal reversion using neuropsychological criteria in the Alzheimer's Disease Neuroimaging Initiative. *Alzheimers Dement* **15**, 1322-1332.

[52] Koepsell TD, Monsell SE (2012) Reversion from mild cognitive impairment to normal or near-normal cognition: Risk factors and prognosis. *Neurology* **79**, 1591-1598.

[53] Roberts RO, Knopman DS, Mielke MM, Cha RH, Pankratz VS, Christianson TJH, Geda YE, Boeve BF, Ivnik RJ, Tangalos EG, Rocca WA, Petersen RC (2014) Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology* **82**, 317-325.

[54] Kiselica AM, Kaser AN, Benge JF (2021) An initial empirical operationalization of the earliest stages of the Alzheimer's continuum. *Alzheimer Dis Assoc Disord* **35**, 62-67.

[55] Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535-562.