Abstract

Objective: To assess the role of brief neuropsychological assessments in prediction and identification of Alzheimer's disease (AD) pathology and progression to AD dementia.

Method: 255 adults (40-81 years) with self-reported cognitive decline underwent baseline and two-year follow-up clinical assessment, including a brief neuropsychological screening and lumbar puncture. Five different mild cognitive impairment (MCI) algorithms were applied on baseline cognitive test results: one conventional, three amnestic (lenient, stringent, multi-domain), and one comprehensive criterion. We compared predictive and diagnostic accuracy of these MCI criteria by performing logistic regression analyses and calculating diagnostic accuracy measures for two-year outcomes of: 1) clinical diagnosis of AD dementia, and 2) cerebrospinal fluid biomarkers in the Alzheimer's continuum. **Results:** The lenient amnestic MCI criterion showed the largest effect size for predicting progression to AD dementia (OR=13.762, 99% CI=1.969-96.194, p=.001), and AD biomarkers OR=4.855, 99% CI=1.974-11.924, p<.001) after two years. This criterion was sensitive for progression to dementia (sensitivity=92.0%, specificity=54.8%, LR+=2.03, LR-=0.15) and showed the highest overall diagnostic accuracy for AD biomarkers (sensitivity =72.7%, specificity=59.1%, LR+=1.78, LR-=0.46). The multi-domain amnestic MCI criterion produced the highest specificity for dementia (sensitivity=76.0%, specificity=73.0%, LR+=2.82, LR-=0.33) and AD biomarkers (sensitivity=46.8% specificity=70.9% LR+=1.61, LR-=0.75). **Conclusions:** Defining MCI using a brief neuropsychological battery provided limited accuracy for progression to AD dementia and CSF Aβ. The lenient amnestic MCI criterion identified the highest number of individuals who progressed to clinical AD or showed biomarker pathology, but this approach included a substantial number of false positives.

Keywords: Alzheimer's disease, mild cognitive impairment, brief neuropsychological assessment, cerebrospinal fluid biomarkers

Key points

Question: How can results from brief cognitive testing best be used to identify individuals with higher risk of developing Alzheimer's disease and -dementia within two years?

Findings: Test results from brief cognitive testing show variable ability to discriminate between healthy individuals and those with signs of disease.

Importance: Brief and low-cost cognitive assessment methods for early Alzheimer's disease can be widely employed and provide useful clinical information, but may not be sufficiently accurate to be used in isolation.

Next Steps: Future research on this topic could involve direct comparison of brief cognitive testing with larger test batteries.

Predictive and diagnostic utility of brief neuropsychological assessment in detecting Alzheimer's pathology and progression to dementia

Mild cognitive impairment (MCI) due to Alzheimer's disease (AD) is often conceptualized as a transitional stage between normal cognition and dementia (Albert et al., 2011). However, MCI is a heterogeneous condition that may be due to other conditions or etiologies than AD, and is not always followed by dementia (Matthews et al., 2008; Petersen, 2016).

Core clinical criteria for MCI due to AD require objective impairment in at least one cognitive domain (memory, executive function, attention, language or visuospatial ability), typically operationalized as one or more neuropsychological test scores falling more than 1-1.5 standard deviations (SD) below age- and education matched norms (Albert et al., 2011). Episodic memory impairments are typical, but not necessary for the diagnosis. This prevailing cognitive MCI criterion, presented by expert groups from the National Institute of Aging and the Alzheimer's disease Association (NIA-AA), was developed to be applicable in all clinical settings and to allow for clinical judgment in assessments. Consequently, diagnostic approaches can vary within this broad criterion, for instance with regard to the selection of neuropsychological tests and testing modalities, the application of different cut-off scores, and whether single or repeated assessments are used (Petersen, 2016). Variable diagnostic strategies for MCI due to AD can lead to different estimates of dementia conversion (Gothlin, Eckerstrom, Rolstad, Wallin, & Nordlund, 2017; Jak et al., 2009; Matthews et al., 2008), indicating that not all MCI diagnoses are associated with similar risk of dementia. This limits the usefulness of MCI as a diagnostic category.

Established neuroimaging and cerebrospinal fluid (CSF) AD biomarkers (signs of Aβ deposition, tau pathology and neurodegeneration) aid in determining the underlying etiology

of cognitive impairment (Jack et al., 2018). However, these methods are often costly, invasive and not always available in standard clinical settings. At the same time, the number of people with dementia worldwide is expected to double every 20 years till 2050, with a majority living in low to middle income countries (Prince et al., 2013). This supports the use of effective assessment methods that place low demands on resources. Brief neuropsychological test batteries are often preferred in memory clinics, as they are more accessible and less time consuming than comprehensive neuropsychological batteries (Hessen et al., 2019).

Memory impairment is a key feature of MCI due to AD, and meta-analytic evidence shows that impaired performance on neuropsychological memory measures in MCI predicts AD progression (Belleville et al., 2017). In a recent cross-sectional study, Hessen and colleagues (2019) compared brief neuropsychological screening criteria for MCI and their association with cerebrospinal fluid AD biomarkers. It was found that a criterion targeting amnestic MCI was more strongly associated with the NIA-AA stage 2 (lowered CSF Aβ42 concentrations and elevated CSF tau concentrations; Sperling et al., 2011), than criteria allowing for impairment in any cognitive domain. Other reports indicate that amnestic MCI subtypes have higher AD dementia conversion rates than non-amnestic MCI, particularly when memory problems are not isolated, but occur in conjunction with impairment in other cognitive domains (Gainotti, Quaranta, Vita, & Marra, 2014; Gothlin et al., 2017; Hessen et al., 2014).

As outlined above, the current clinical cognitive criterion for MCI due to AD (Albert et al., 2011) may define a patient group with heterogeneous cognitive symptom profiles and levels of impairment, probably with different risks of progression to dementia. Using a diagnostic criterion that separates incipient AD from healthy aging is important, as an MCI diagnosis can have implications for clinical practice, research, and patients' well-being. Although comprehensive and multimodal methods of assessment often are preferable, simple

and low-cost alternatives might benefit patients and clinicians with limited access to such resources.

In the current study, the main objective was to evaluate the utility of a brief neuropsychological battery, comparing different MCI criteria based on their ability to predict and correctly classify, at two years follow-up: 1) AD dementia diagnosis and 2) AD CSF biomarker category. We hypothesized that the inclusion of an amnestic component in the MCI definition would give the highest predictive and diagnostic accuracy.

Methods

Participants

Participants were between 40 and 81 years, had self-reported cognitive decline, and spoke a native language of Norwegian, Swedish or Danish. They took part in the Dementia Disease Initiation (DDI), a Norwegian multicenter longitudinal study targeting early biological and cognitive markers of dementia (Figure 1). The study was approved by the regional medical research ethics committee, and all participants provided written informed consent. Individuals with brain trauma or disorder (including stroke, dementia, severe psychiatric disease, or developmental disorder) were excluded. For detailed information on recruitment and eligibility criteria of participants, see Fladby et al. (2017).

Participants for the current study were recruited between January 2013 and January 2019, from memory clinics in collaborating hospitals and through self-referral after seeing media advertisements. This sample included individuals with cognitive complaints (symptom group participants) who had completed baseline and follow-up assessments. Participants who received a non-AD dementia diagnosis during the study period were excluded from analyses.

Procedures

All procedures were performed at baseline and at follow-up assessment after two years. Clinical assessments followed a standardized Case Report Form (CRF). The CRF includes medical history reports, neurological examination, the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), the 15 item Geriatric Depression Scale (GDS-15; Sheikh & Yesavage, 1986) and a brief neuropsychological screening.

The brief neuropsychological battery included tests of verbal learning (CERAD word list; Fillenbaum et al., 2008), delayed verbal recall (CERAD word list; Fillenbaum et al., 2008), visuoperceptual ability (VOSP silhouettes; Warrington & James, 1991), psychomotor speed (Trail Making Test A; Reitan & Wolfson, 1985), cognitive flexibility/divided attention (Trail Making Test B; Reitan & Wolfson, 1985), and phonemic word fluency (Controlled Oral Word Association Test; Benton & Hamsher, 1978). The administration time of the battery was approximately 30 minutes. Age-, gender-, and education-adjusted norms were used to convert raw scores into standardized T-scores (Heaton, Miller, Taylor, & Grant, 2004; Kirsebom et al., 2019). Norms for VOSP silhouettes were developed from a sample of healthy control participants from the DDI and the Gothenburg MCI study (Wallin et al., 2016; Supplementary material). These norms correct only for age, since education and gender did not significantly influence test scores.

Lumbar puncture was carried out according to a standardized protocol and cerebrospinal fluid concentrations of total tau (T-tau), phosphorylated tau (P-tau181) and β amyloid ($A\beta_{42}$) were analyzed using ELISA (Innotest β-Amyloid (1-42), Innotest Phospho-Tau (181P) and Innotest h-Tau Ag, Fujirebio, Ghent Belgium). Methods have been described in detail in Fladby et al. (2017).

Diagnostic and Classification Criteria

All participants were categorized using five different MCI definitions in succession. Since one aim of the current study was a longitudinal evaluation of findings from Hessen et al. (2019), we adopted the algorithms used for the conventional, comprehensive and lenient amnestic MCI criteria from this report. We added two new criteria to evaluate effects of a more stringent amnestic definition, and one of combined problems in both amnestic and other domains.

MCI criteria.

1. A version of the conventional NIA-AA MCI criterion: at least one test score in any cognitive domain similar to or below T-score 35.
2. A lenient amnestic MCI criterion: at least one memory test score below T-score 40.
3. A stringent amnestic MCI criterion: at least one memory test score below T-score 35.
4. Multi-domain amnestic MCI criterion: at least one memory test score below T-score 40, and at least one test score in any cognitive domain below T-score 40.
5. A criterion based on the comprehensive approach described by Jak et al. (2009) and Bondi et al. (2014): at least two test scores in any cognitive domain below T-score 40.

We used the following cut-off values for determining abnormality in CSF biomarkers: CSF $A\beta_{1\text{-}42}$ <708 pg/ml (based on findings from Kalheim, Fladby, Coello, Bjornerud, & Selnes, 2018), p-tau ≥80 pg/ml, and t-tau >300 pg/ml for age <50 years, >450 pg/ml for age 50–69 years, and >500 pg/ml for age ≥70 years (modified from Sjogren et al., 2001). For classification of biomarkers, we used general biomarker categories derived from the AT(N) system, as proposed by NIA-AA in a recently published research framework for AD (Jack et al., 2018). Based on binary (normal/abnormal) classification of the three CSF biomarkers, individuals were assigned to one of three categories: 1) normal AD biomarkers; 2) biomarkers

in the Alzheimer's continuum (abnormal $A\beta_{1-42}$, with or without p-tau and t-tau abnormality); and 3) suspected non-AD pathological changes ($A\beta_{1-42}$ within normal boundaries, but abnormal p-tau and/or t-tau).

The diagnosis of dementia entails verified cognitive and/or behavioral impairment compared to previous levels, as well as significant functional decline. Diagnostic evaluation is based on the Clinical Dementia Rating Scale (CDR; Hughes, Berg, Danziger, Coben, & Martin, 1982), using information collected in the clinical interview with participant and informant. If a patient was diagnosed with dementia, etiological diagnosis of probable or possible AD dementia was determined using McKhann et al. (2011) criteria. Core clinical criteria for probable AD dementia include gradual onset of symptoms and a clear-cut cognitive worsening with either amnestic or nonamnestic (language, visuospatial or executive) presentations. A diagnosis of possible AD dementia is made when the core clinical criteria are met, except for that a) cognitive symptoms have presented in an atypical course, or b) there is evidence of mixed etiology behind the cognitive symptoms. Diagnoses were determined by a medical doctor. To ensure inter-rater reliability across centers, the DDI project holds bi-annual meetings discussing diagnostic criteria and procedures.

Statistics

Statistical analysis was performed using The Statistical Package for Social Sciences (SPSS) version 25. Descriptive statistics were calculated for baseline demographic, clinical and cognitive characteristics. We used binary logistic regression analyses to investigate the relationship between independent variables and the dependent variable of AD dementia diagnosis after two years. The relatively few dementia cases at follow-up (n=25) limited the number of independent variables we could include in the analyses. We examined each MCI criterion individually, including age as a covariate. To investigate associations between

baseline MCI classification and CSF biomarker categories at follow-up, we used multinomial logistic regression in a similar binary logistic regression procedure. We used the Bonferroni correction to reduce the family-wise error rate related to multiple comparisons. A family of multiple tests was defined as the logistic regression analyses performed for each AD-related outcome (dementia diagnosis and Aβ positive biomarker status), resulting in a correction of five contrasts for each family. When applying a conventional significance level of .05, the corrected alpha level was .01 for all logistic regression analyses. Corresponding 99% confidence intervals are presented for these analyses.

The Odds Ratios from the logistic regressions give an indication of global diagnostic accuracy. Detailed diagnostic accuracy measures include sensitivity (percentage of correct classification of true positives/dementia progressors), specificity (percentage of correct classification of true negatives/nonprogressors), positive likelihood ratio (LR+; likelihood that disease is present when given an MCI diagnosis), and negative likelihood ratio (LR-; likelihood that disease is absent when not given an MCI diagnosis). These measures, with 95% confidence intervals, were calculated for both outcomes using the MedCalc diagnostic test evaluation calculator (MedCalc Software, Ostend, Belgium). Groups with normal AD biomarkers and non-AD pathologic change were collapsed for diagnostic accuracy calculations, as this procedure requires binary outcome classification. This Aβ negative group (n=110) was compared with the abnormal Aβ group (n=77), to evaluate the diagnostic accuracy of MCI classifications in the whole sample.

Results

Longitudinal data were available for 264 symptom group participants (SCD and MCI) in the DDI cohort (Figure 1). Participants with missing data from more than one neuropsychological screening test at baseline (n=8), or with a non-AD dementia diagnosis at

follow-up (n=1; dementia with Lewy bodies; diagnosed according to McKeith et al. (2005) criteria) were excluded from analyses. The analyzed sample comprised 255 participants at baseline. At follow-up, 39 of these participants were missing data from more than one cognitive test, and 85 were missing CSF data. Individuals with pathological baseline Aβ who were missing follow-up CSF data were included in longitudinal analyses, since Aβ accumulates and we assumed these individuals would therefore not have changed biomarker category after two years. Table 1 presents baseline demographic, clinical and cognitive characteristics. The distribution of cognitive classifications at baseline and follow-up is presented in Figure 2.

(Insert Figure 1, Table 1, Figure 2 here)

Outcome: Diagnosis of Dementia due to Alzheimer's Disease

During a two-year follow-up period, 25 of the participants in the study sample received an AD dementia diagnosis. As depicted in Table 2, binary logistic regression analyses indicate that all MCI criteria were associated with higher odds of AD dementia after two years, while controlling for age. The amnestic MCI criterion with a lenient cutoff (T<40) showed the largest effect size, indicating individuals classified with MCI according to this criterion had increased odds of progressing to AD dementia compared to individuals not given an MCI diagnosis. However, this parameter estimate was associated with considerable uncertainty, with odds ratios ranging from approximately two to ninety-six times also being compatible with our data. The conventional criterion showed an effect size of almost similar magnitude and uncertainty. The comprehensive and stringent amnestic criteria produced the smallest effect sizes.

For diagnostic accuracy estimates, the conventional criterion and the lenient amnestic criterion produced the highest numbers for correct classification of dementia progressors, with

sensitivity values over 90%. Specificity values were closer to 50% for these two criteria. The multi-domain amnestic criterion achieved the highest specificity value as well as the most even balance between relatively high sensitivity and specificity, with values around 75%. The positive likelihood ratio was also highest for this criterion, indicating a slight increase in the likelihood of disease given a multi-domain amnestic MCI at baseline. All negative likelihood ratios indicated reduced probability for disease in participants not diagnosed with MCI, most markedly in the lenient amnestic and conventional criteria.

(Insert Table 2 here)

Outcome: Biomarker Category on the Alzheimer's Continuum

Follow-up CSF data were available for 168 participants, and 19 participants with pathological baseline Aβ were included in the total analyzed sample (n=187). In multinomial logistic regression analyses with AT(N) biomarker categories as the dependent variable, the likelihoods of having a biomarker profile on the Alzheimer's continuum (n=77) or in the spectrum of non-AD pathologic change (n=19), were estimated with reference to the likelihood of having a normal AD biomarker profile (n=91). All analyses included age as a covariate.

Our results (Table 3) show that the lenient amnestic MCI criterion had the largest effect size for predicting CSF biomarkers on the AD continuum. AD-continuum biomarker classification at follow-up could best be predicted from baseline classifications of amnestic MCI and from robust approaches using at least two test scores. The results were less certain for the conventional MCI criterion. The point estimate indicates increased odds for AD markers after two years, but it did not meet the corrected alpha level. Likelihood ratio tests also indicate that the main effect of this predictor may not contribute significantly to the

model (χ^2 (2, *N=187*)=4.691, p=.096). The estimates for non-AD pathology were of smaller size and not statistically significant.

(Insert Table 3 here)

Diagnostic accuracy for the outcome of AD biomarker category (Table 4) was lower than for dementia progression, and associated with comparable uncertainty around the estimates. The lenient amnestic criterion showed the highest overall accuracy. This criterion was most successful in correctly classifying the individuals with AD-continuum biomarkers, while the multi-domain MCI criterion best identified individuals who were not in the AD continuum. Overall, receiving an MCI diagnosis according to any of the criteria produced only small changes in probability of having biomarkers on the AD continuum, as most likelihood ratios were close to one. Although differences between criteria were small, meeting the lenient amnestic MCI criterion gave the largest increase in likelihood of having abnormal AD biomarkers, while not fulfilling this criterion represented the largest decrease in likelihood of pathological AD biomarkers.

(Insert Table 4 here)

Discussion

Results from the current study suggest modest utility of using a brief neuropsychological battery in a clinical sample to predict AD diagnosis, and low utility for predicting biomarker pathology.

For the outcome of AD dementia after two years, all MCI definitions were associated with higher likelihood of progression. The lenient amnestic and conventional MCI criteria correctly classified over 90% of individuals who would later progress to dementia,

consequently increasing confidence in the notion that an individual who does not fulfill one of these MCI diagnoses is less likely to get a dementia diagnosis after two years. However, both criteria showed specificity around 50%, including a large number of nonprogressors in the MCI group. The multi-domain amnestic MCI criterion showed sensitivity and specificity values close to 75%. This was the most accurate criterion for classifying individuals who would not progress, thereby increasing confidence in that receiving this MCI diagnosis indicates a larger chance of progression. Considering the sum of true positives and negatives, this criterion correctly classified the most individuals.

For the outcome of biomarker categories after two years, results indicated poor diagnostic accuracy across MCI definitions. The MCI criteria had low predictive utility for biomarkers on the AD continuum (i.e. abnormal Aβ levels, isolated or in combination with abnormal p-tau and/or t-tau), and no utility for suspected non-AD pathologic changes (normal Aβ, but abnormal p-tau and/or t-tau). A baseline classification of lenient amnestic MCI gave the highest increase in odds for having biomarkers on the AD continuum and identified most Aβ positive individuals. This is in accordance with previous cross-sectional findings (Hessen et al., 2019), indicating that applying an amnestic MCI criterion on a brief neuropsychological protocol increases chances of identifying individuals with positive AD biomarkers longitudinally. The conventional MCI criterion showed sensitivity close to the level of the lenient amnestic criterion, around 70%, but otherwise poor classification and prediction of amyloid positivity.

Our results indicate that patients who enter the clinic with relatively mild amnestic impairment have an increased likelihood of progressing to dementia over time, and of displaying pathological CSF biomarkers in the AD continuum. However, the majority of individuals with such cognitive profiles will not meet either of these outcomes in a two-year perspective. Results from a brief neuropsychological battery could give an indication of risk,

but limited diagnostic accuracy suggests this may not be used as a stand-alone diagnostic method. In our findings, uncertainty around the effect size estimates, especially in prediction of dementia diagnosis, complicates evaluation of the actual differences in risk magnitude between MCI criteria.

While the role of verbal episodic memory is well-documented in early AD, several investigations also show independent predictive value of tests in other cognitive domains such as executive functions, language and visuoperceptual abilities (Belleville et al., 2017). For the outcome of dementia diagnosis, the conventional MCI criterion showed comparable diagnostic accuracy to the lenient amnestic MCI criterion in our cohort. The conventional criterion also showed higher sensitivity than the amnestic criterion with the same cutoff value (T-score ≤ 35), indicating that it identified individuals who initially present with nonamnestic cognitive impairments and later progress to AD dementia. Such a nonspecific algorithm can also capture atypical variants of AD, where for instance aphasia or visuospatial deficits are the primary problem (Alladi et al., 2007). Together, these results support the utility of this commonly applied MCI criterion in identifying atypically presenting individuals who progress to dementia.

The multi-domain amnestic MCI criterion increased the specificity and positive likelihood ratio for AD dementia in the current sample. Several studies report that combining episodic memory impairments with impairment in one or more other cognitive domains gives additional prognostic value, in particular when using combinations with executive function or language measures (Albert, Moss, Tanzi, & Jones, 2001; Chen et al., 2000; J. Mitchell, Arnold, Dawson, Nestor, & Hodges, 2009; Tabert et al., 2006). One proposed explanation of such findings is that impairments beyond memory often reflect a more advanced stage of disease than isolated amnestic deficits (Brambati et al., 2009; Gainotti et al., 2014). Increased disease severity in research participants with both memory and non-memory problems could

lead to a higher conversion rate for this group (Brambati et al., 2009). This could also explain why the multi-domain amnestic criterion showed low sensitivity for AD biomarker category, while the lenient amnestic criterion was most sensitive and predictive of this outcome. Both episodic memory impairment and abnormal Aβ may occur in earlier stages of the disease, and the multi-domain algorithm could be too restrictive to capture these more subtle changes.

Basing diagnosis or clinical decisions on isolated neuropsychological test scores is not optimal. When healthy adults take multiple neuropsychological tests, it is normal to obtain one or more scores below a given cut-off value that marks abnormal performance (Binder, Iverson, & Brooks, 2009).

The likelihood of impaired test results increases with the number of tests in the battery, and when fewer test scores and less stringent cut-off values are used to define impairment (Binder et al., 2009; Ingraham & Aiken, 1996). Using a statistical model to demonstrate these psychometric principles, Ingraham and Aiken (1996) found that the probability of obtaining at least one score below 1 SD from a battery of six tests was approximately 70% in the general population. Requiring the same cutoff level for at least two tests gave an estimated probability of abnormality above 20%. Although this model is not directly applicable to our clinical sample and test battery, it illustrates that an MCI group defined by one test score is likely to include many false positive cases who may not have underlying brain pathology. In the current study, the lenient amnestic MCI criterion produced the largest effect sizes in prediction analyses, likely because the high inclusiveness (a relatively small memory impairment needed for diagnosis) captured most progressors in the sample. As evident from the specificity value, however, this approach identified a large proportion of individuals who did not progress to dementia. Amnestic criteria applying stricter cutoffs or requiring more than one impaired score improved the correct classification of nonprogressors, but reduced the number of true positives. This demonstrates a classical

challenge of balancing sensitivity and specificity. Since none of the criteria provide excellent overall accuracy, criterion preference might depend on which of these features are deemed most important, and which type of error is considered most imperative to avoid.

However, conclusions from a "one-impaired-test-approach" can be more valid when the targeted cognitive symptoms are in concordance with the suspected condition (Binder et al., 2009), for instance episodic memory impairment in AD. In addition, using robust criteria that require more than one impaired test score usually increases stability and validity of MCI diagnosis (Bondi et al., 2014; Jak et al., 2009; Loewenstein et al., 2009). The relatively high and balanced diagnostic accuracy estimates for AD dementia in the multi-domain amnestic MCI group could reflect such effects in our cohort, as it encompasses both AD-typical memory impairment and a requirement of at least two reduced test scores.

The recent NIA-AA research framework proposes a shift toward a biological definition of AD, placing less emphasis on clinical symptoms of the disease and more on biomarker signatures (Jack et al., 2018). This is based on a body of evidence suggesting Aβ can be used as the defining feature of AD, and combinations with pathologic tau and neurodegeneration markers represent later phases in disease development. As relatively few patients converted to dementia during the follow-up period in our study, using biomarker categories allowed us to test the MCI algorithms on a larger subgroup of participants with suspected AD pathology. Our study shows that in logistic regression, the association between the conventional MCI criterion and AD biomarker pathology was smaller and less reliable than for other MCI criteria. These findings are similar to a previous study (Hessen et al., 2019). The conventional criterion did, however, predict dementia diagnosis in our cohort, albeit including many false positives. This discrepancy could indicate that this criterion's high inclusivity contributed to prediction of and sensitivity to the relatively few cases of dementia, while poor discrimination along underlying biomarker patterns suggests it may not be specific

to AD pathology. Of note, the diagnostic accuracy of CSF Aβ in MCI samples is variable (Ritchie et al., 2014), and pathological AD biomarkers occur in cognitively normal individuals (Jansen et al., 2015). This implies that the presence of pathological biomarkers may not always correspond to current and future clinical symptoms such as problems with memory and activities of daily living, which arguably are the most meaningful outcomes for patients and their families.

Some study limitations should be addressed. Although cognitive and biomarker data were not used to make clinical dementia diagnoses, clinicians who collected data and diagnosed participants were not blinded to this data. The diagnosis of dementia is based on the patient's functional abilities in the present, and is therefore independent of previous assessments. However, it is possible this information could have influenced clinicians' overall impression of patient function. Moreover, cognitive changes can manifest many years before progression to AD dementia. As we followed patients for two years, our sample probably includes individuals who did not progress during this period, but may do so on a later time point. With a mean age of 62.9 years, we also have a relatively young cohort compared to other MCI studies (e.g. Petersen et al., 2010). In meta-analysis, the prevalence of all-cause dementia in Western Europe has been estimated to approximately 6.9 % for adults ages 60 years and older, with age-stratified prevalences ranging from 1.6% for ages 60-64, to 12.9% in adults 80-84 years (Prince et al., 2013). Thus, we would expect a low occurrence of cases with AD dementia in the age groups most heavily represented in our sample. The short follow-up and relatively young sample probably also contributed to that quite few individuals (n=25) progressed to AD dementia during the observation. While the annual conversion rate from MCI to AD has been estimated in meta-analysis to approximately 8% (95% CI=6.3-10%) in clinical settings (A. J. Mitchell & Shiri-Feshki, 2009), the percentage of two-year progression to dementia was 9.8% in our clinical sample. Last, our approach to error control

was conservative, which increases the likelihood of type II errors (wrongly assuming an MCI definition is not useful for predicting AD in logistic regression). Recognizing that with a more liberal approach, the conventional criterion could be a statistically significant predictor of Aβ positivity, we interpret results with caution. However, the relative effect size of this criterion compared to the other MCI criteria supports the interpretation that it appears to be a less powerful predictor for this outcome with reference to having normal AD biomarkers.

One implication of the current study is that a brief neuropsychological battery may not have adequate diagnostic accuracy to be employed as a stand-alone method of assessment for identifying MCI due to AD. This points to the need for further research on time-efficient and low-cost cognitive assessment methods with higher diagnostic accuracy. Another implication could be that a definition with an amnestic component might be preferable over the prevailing core clinical MCI criterion. This is based on findings of the conventional and lenient amnestic criteria showing comparable association with dementia diagnosis but differential association with AD biomarkers, and of higher specificity in the multi-domain amnestic criterion. In line with current recommendations within the NIA-AA guidelines (Albert et al., 2011), emphasizing episodic memory impairments when diagnosing MCI due to AD may increase diagnostic accuracy.

In conclusion, using a brief neuropsychological battery for defining MCI due to AD can be informative of disease development in a two-year perspective, but may not be adequate to use as independent diagnostic assessment. This method of assessment demonstrated a moderate ability to discriminate between stable and progressing groups when it comes to AD dementia, and low accuracy for separating groups with normal or pathological CSF Aβ.

References

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

Albert, M. S., Moss, M. B., Tanzi, R., & Jones, K. (2001). Preclinical prediction of AD using neuropsychological tests. *J Int Neuropsychol Soc, 7*(5), 631-639. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11459114

Alladi, S., Xuereb, J., Bak, T., Nestor, P., Knibb, J., Patterson, K., & Hodges, J. R. (2007). Focal cortical presentations of Alzheimer's disease. *Brain, 130*(10), 2636-2645. doi:10.1093/brain/awm213

Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., Croteau, J., & Consortium for the Early Identification of Alzheimer's, d.-Q. (2017). Neuropsychological Measures that Predict Progression from Mild Cognitive Impairment to Alzheimer's type dementia in Older Adults: a Systematic Review and Meta-Analysis. *Neuropsychol Rev, 27*(4), 328-353. doi:10.1007/s11065-017-9361-5

Benton, A., & Hamsher, K. (1978). Multilingual aphasia examination manual. *Iowa City: University of Iowa*.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. *Arch Clin Neuropsychol, 24*(1), 31-46. doi:10.1093/arclin/acn001

Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., . . . Salmon, D. P. (2014). Neuropsychological criteria for mild cognitive impairment

improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis, 42*(1), 275-289. doi:10.3233/JAD-140276

Brambati, S. M., Belleville, S., Kergoat, M. J., Chayer, C., Gauthier, S., & Joubert, S. (2009). Single- and multiple-domain amnestic mild cognitive impairment: two sides of the same coin? *Dement Geriatr Cogn Disord, 28*(6), 541-549. doi:10.1159/000255240

Chen, P., Ratcliff, G., Belle, S., Cauley, J., DeKosky, S., & Ganguli, M. (2000). Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology, 55*(12), 1847-1853. Retrieved from https://n.neurology.org/content/55/12/1847.1.long

Fillenbaum, G. G., van Belle, G., Morris, J. C., Mohs, R. C., Mirra, S. S., Davis, P. C., . . . Heyman, A. (2008). Consortium to Establish a Registry for Alzheimer's Disease (CERAD): the first twenty years. *Alzheimers & Dementia, 4*(2), 96-109. doi:10.1016/j.jalz.2007.08.005

Fladby, T., Palhaugen, L., Selnes, P., Waterloo, K., Brathen, G., Hessen, E., . . . Aarsland, D. (2017). Detecting At-Risk Alzheimer's Disease Cases. *J Alzheimers Dis, 60*(1), 97-105. doi:10.3233/JAD-170231

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res, 12*(3), 189-198. doi:10.1016/0022-3956(75)90026-6

Gainotti, G., Quaranta, D., Vita, M. G., & Marra, C. (2014). Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease. *J Alzheimers Dis, 38*(3), 481-495. doi:10.3233/JAD-130881

Gothlin, M., Eckerstrom, M., Rolstad, S., Wallin, A., & Nordlund, A. (2017). Prognostic Accuracy of Mild Cognitive Impairment Subtypes at Different Cut-Off Levels. *Dement Geriatr Cogn Disord, 43*(5-6), 330-341. doi:10.1159/000477341

Heaton, R., Miller, S., Taylor, M. J., & Grant, I. (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. *Lutz, FL: Psychological Assessment Resources*.

Hessen, E., Kirsebom, B. E., Eriksson, C. M., Eliassen, C. F., Nakling, A. E., Brathen, G., . . . Fladby, T. (2019). In Brief Neuropsychological Assessment, Amnestic Mild Cognitive Impairment (MCI) Is associated with Cerebrospinal Fluid Biomarkers for Cognitive Decline in Contrast to the Prevailing NIA-AA MCI Criterion. *J Alzheimers Dis, 67*(2), 715-723. doi:10.3233/JAD-180964

Hessen, E., Reinvang, I., Eliassen, C. F., Nordlund, A., Gjerstad, L., Fladby, T., & Wallin, A. (2014). The Combination of Dysexecutive and Amnestic Deficits Strongly Predicts Conversion to Dementia in Young Mild Cognitive Impairment Patients: A Report from the Gothenburg-Oslo MCI Study. *Dement Geriatr Cogn Dis Extra, 4*(1), 76-85. doi:10.1159/000360282

Hughes, C. P., Berg, L., Danziger, W., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British journal of psychiatry, 140*(6), 566-572.

Ingraham, L. J., & Aiken, C. B. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology, 10*(1), 120.

Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., . . . Contributors. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers & Dementia, 14*(4), 535-562. doi:10.1016/j.jalz.2018.02.018

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of five neuropsychological approaches to defining

mild cognitive impairment. *Am J Geriatr Psychiatry, 17*(5), 368-375. doi:10.1097/JGP.0b013e31819431d5

Jansen, W. J., Ossenkoppele, R., Knol, D. L., Tijms, B. M., Scheltens, P., Verhey, F. R. J., . . . Group, a. t. A. B. S. (2015). Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia: A Meta-analysisPrevalence of Cerebral Amyloid Pathology in Persons Without DementiaPrevalence of Cerebral Amyloid Pathology in Persons Without Dementia. *JAMA, 313*(19), 1924-1938. doi:10.1001/jama.2015.4668

Kalheim, L. F., Fladby, T., Coello, C., Bjornerud, A., & Selnes, P. (2018). [18F]-Flutemetamol Uptake in Cortex and White Matter: Comparison with Cerebrospinal Fluid Biomarkers and [18F]-Fludeoxyglucose. *J Alzheimers Dis, 62*(4), 1595-1607. doi:10.3233/JAD-170582

Kirsebom, B. E., Espenes, R., Hessen, E., Waterloo, K., Johnsen, S. H., Gundersen, E., . . . Fladby, T. (2019). Demographically adjusted CERAD wordlist test norms in a Norwegian sample from 40 to 80 years. *Clin Neuropsychol*, 1-13. doi:10.1080/13854046.2019.1574902

Loewenstein, D. A., Acevedo, A., Small, B. J., 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*(5), 418-423. doi:10.1159/000211803

Matthews, F. E., Stephan, B. C., McKeith, I. G., Bond, J., Brayne, C., Medical Research Council Cognitive, F., & Ageing, S. (2008). Two-year progression from mild cognitive impairment to dementia: to what extent do different definitions agree? *J Am Geriatr Soc, 56*(8), 1424-1433. doi:10.1111/j.1532-5415.2008.01820.x

McKeith, I. G., Dickson, D. W., Lowe, J., Emre, M., O'Brien, J. T., Feldman, H., . . . Consortium on, D. L. B. (2005). Diagnosis and management of dementia with Lewy

bodies: third report of the DLB Consortium. *Neurology, 65*(12), 1863-1872. doi:10.1212/01.wnl.0000187889.17253.b1

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers & Dementia, 7*(3), 263-269. doi:10.1016/j.jalz.2011.03.005

Mitchell, A. J., & 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*(4), 252-265. doi:10.1111/j.1600-0447.2008.01326.x

Mitchell, J., Arnold, R., Dawson, K., Nestor, P. J., & Hodges, J. R. (2009). Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. *J Neurol, 256*(9), 1500-1509. doi:10.1007/s00415-009-5152-0

Oosterhuis, H. E., van der Ark, L. A., & Sijtsma, K. (2016). Sample Size Requirements for Traditional and Regression-Based Norms. *Assessment, 23*(2), 191-202. doi:10.1177/1073191115580638

Petersen, R. C. (2016). Mild Cognitive Impairment. *Continuum (Minneap Minn), 22*(2 Dementia), 404-418. doi:10.1212/CON.0000000000000313

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., . . . Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology, 74*(3), 201-209. doi:10.1212/WNL.0b013e3181cb3e25

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., & Ferri, C. P. (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers & Dementia, 9*(1), 63-75 e62. doi:10.1016/j.jalz.2012.11.007

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4): Reitan Neuropsychology.

Ritchie, C., Smailagic, N., Noel-Storr, A. H., Takwoingi, Y., Flicker, L., Mason, S. E., & McShane, R. (2014). Plasma and cerebrospinal fluid amyloid beta for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev*(6), CD008782. doi:10.1002/14651858.CD008782.pub4

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*.

Sjogren, M., Vanderstichele, H., Agren, H., Zachrisson, O., Edsbagge, M., Wikkelso, C., . . . Blennow, K. (2001). Tau and Abeta42 in cerebrospinal fluid from healthy adults 21-93 years of age: establishment of reference values. *Clin Chem, 47*(10), 1776-1781.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., . . . Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers & Dementia, 7*(3), 280-292. doi:10.1016/j.jalz.2011.03.003

Tabert, M. H., Manly, J. J., Liu, X., Pelton, G. H., Rosenblum, S., Jacobs, M., . . . Devanand, D. P. (2006). Neuropsychological prediction of conversion to Alzheimer disease in patients with mild cognitive impairment. *Arch Gen Psychiatry, 63*(8), 916-924. doi:10.1001/archpsyc.63.8.916

Testa, S. M., Winicki, J. M., Pearlson, G. D., Gordon, B., & Schretlen, D. J. (2009). Accounting for estimated IQ in neuropsychological test performance with regression-

based techniques. *J Int Neuropsychol Soc, 15*(6), 1012-1022. doi:10.1017/S1355617709990713

Wallin, A., Nordlund, A., Jonsson, M., Lind, K., Edman, A., Gothlin, M., . . . Eckerstrom, C. (2016). The Gothenburg MCI study: Design and distribution of Alzheimer's disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism, 36*(1), 114-131. doi:10.1038/jcbfm.2015.147

Warrington, E. K., & James, M. (1991). The visual object and space perception battery.

Table 1.

Baseline demographic, clinical and cognitive characteristics of patients with cognitive complaints

Variables	**Mean (SD) *Range***	**N**
Age	62.85 (9.72) *40-81*	255
Female, n (%)	137 (53.73)	255
Years of education	13.67 (3.25) *7-22*	255
MMSE	28.53 (1.75) *21-30*	255
Geriatric Depression Scale	2.75 (2.54) *0-13*	245
CDR Global score, Median, *Range*	0.5, *0-1*	229
CDR Sum of boxes, Median, *Range*	0.5, *0-4*	230
CERAD word list learning, raw score	19.24 (4.55) *2-28*	254
CERAD word list learning, T-score	44.27 (11.54) *23-71*	254
CERAD word list recall, raw score	5.88 (2.86) *0-25*	251
CERAD word list recall, T-score	43.32 (12.13) *13-71*	251
COWAT, raw score	39.62 (12.05) *14-76*	253
COWAT, T-score	49.03 (9.87) *25-77*	253
TMT-A, raw score	39.54 (16.58) *13-142*	255
TMT-A, T-score	45.94 (9.67) *14-77*	255
TMT-B, raw score	109.19 (64.13) *25-429*	254
TMT-B, T-score	45.63 (11.33) *7-78*	254
VOSP silhouettes, raw score	21.68 (4.58) *10-30*	232
VOSP silhouettes, T-score	48.74 (11.62) *20.76-73.72*	232
$A\beta_{42}$, pg/ml	931.43 (296.52) *278-1630*	233
Pathological $A\beta_{42}$, n (%)	68 (29.18)	233
P-tau, pg/ml	61.96 (31.49) *16-185*	233
Pathological p-tau, n (%)	48 (20.60)	233
T-tau, pg/ml	397.50 (256.05) *75-1370*	233
Pathological t-tau, n (%)	58 (24.89)	233

Notes. SD= standard deviation, MMSE=Mini Mental State Examination, CDR= Clinical Dementia Rating Scale, CERAD= Consortium to Establish a Registry for Alzheimer's Disease, COWAT= Controlled Oral Word Association Test, TMT= Trail Making Test, VOSP=Visual Object and Space Perception battery, Aβ42 =β amyloid, P-tau= phosphorylated tau, T-tau= total tau

Table 2.

Binary logistic regression models with dementia diagnosis at follow-up as dependent variable, and diagnostic accuracy measures for each MCI criterion

Variable	**OR (99 % CI)**	**P-value**	**TP/FP/FN/TN**	**SENS (95 % CI)**	**SPEC (95 % CI)**	**LR+ (95 % CI)**	**LR- (95 % CI)**
Conventional criterion T≤35	11.974 (1.715-83.627)	.001	23/110/2/120	92.00% (73.97%-99.02%)	52.17% (45.51%-58.78%)	1.92 (1.61-2.30)	0.15 (0.04-0.58)
Age	1.087 (1.014-1.165)	.002					
Lenient amnestic criterion T<40	13.762 (1.969-96.194)	.001	23/104/2/126	92.00% (73.97%-99.02%)	54.78% (48.11%-61.33%)	2.03 (1.69-2.44)	0.15 (0.04-0.55)
Age	1.089 (1.016-1.167)	.002					
Stringent amnestic criterion T≤35	7.587 (2.061-27.933)	<.001	19/76/6/154	76.00% (54.87%-90.64%)	66.96% (60.47%-73.00%)	2.30 (1.73-3.06)	0.36 (0.18-0.72)
Age	1.105 (1.026-1.190)	.001					
Multi-domain amnestic criterion T<40	9.049 (2.461-33.265)	<.001	19/62/6/168	76.00% (54.87%-90.64%)	73.04% (66.82%-78.66%)	2.82 (2.08-3.83)	0.33 (0.16-0.66)
Age	1.097 (1.019-1.181)	.001					
Comprehensive criterion T<40	7.408 (1.699-32.299)	<.001	21/99/4/131	84.00% (63.92%-95.46%)	56.96% (50.29%-63.45%	1.95 (1.56-2.45)	0.28 (0.11-0.69)
Age	1.097 (1.021-1.179)	.001					

Notes. OR= odds ratio, CI=confidence interval, TP=true positive, FP=false positive, TN=true negative, FN=false negative, SENS=sensitivity, SPEC=specificity, LR+=positive likelihood ratio, LR-=negative likelihood ratio. Bonferroni corrected alpha level= 0.01.

Table 3.

Multinomial logistic regression analyses with CSF biomarker category at follow-up as the dependent variable, including age and MCI criteria as independent variables

Outcome	**Variable**	**OR (99% CI)**	**P-value**
Alzheimer's continuum	Conventional criterion T≤35	2.013 (.890-4.549)	.027
	Age	1.073 (1.026-1.123)	<.001
	Lenient amnestic criterion T<40	4.855 (1.974-11.924)	<.001
	Age	1.080 (1.028-1.135)	<.001
	Stringent amnestic criterion T≤35	3.149 (1.362-7.285)	<.001
	Age	1.083 (1.034-1.135)	<.001
	Multi-domain amnestic criterion t<40	2.548 (1.079-6.018)	.005
	Age	1.076 (1.026-1.127)	<.001
	Comprehensive criterion T<40	2.764 (1.193-6.404)	.002
	Age	1.076 (1.026-1.128)	<.001
Non-AD pathologic change	Conventional criterion T≤35	1.045 (.306-3.574)	.926
	Age	1.005 (.944-1.071)	.830
	Lenient amnestic criterion T<40	1.829 (.498-6.722)	.232
	Age	1.010 (.944-1.080)	.705
	Stringent amnestic criterion T≤35	1.532 (.431-5.443)	.386
	Age	1.009 (.947-1.076)	.711
	Multi-domain amnestic criterion T<40	1.560 (.416-5.850)	.386
	Age	1.007 (.944-1.074)	.780
	Comprehensive criterion T<40	1.069 (.297-3.851)	.894
	Age	1.005 (.942-1.073)	.832

Notes. Reference category is Normal CSF biomarker profile. OR= odds ratio, CI=confidence interval. Bonferroni corrected alpha level= 0.01

Table 4.

Diagnostic accuracy measures for AD biomarker category

Variable	**TP/FP/FN/TN**	**SENS (95% CI)**	**SPEC (95% CI)**	**LR+ (95% CI)**	**LR- (95% CI)**
Conventional criterion T≤35	52/57/25/53	67.53% (55.90%-77.77%)	48.18% (38.55%-57.91%)	1.30 (1.03-1.65)	0.67 (0.46-0.98)
Lenient amnestic criterion T<40	56/45/21/65	72.73% (61.38%-82.26%)	59.09% (49.31%-68.37%)	1.78 (1.37-2.31)	0.46 (0.31-0.69)
Stringent amnestic criterion T<=35	42/38/35/72	54.55% (42.79%-65.94%)	65.45% (55.79%-74.26%)	1.58 (1.14-2.19)	0.69 (0.52-0.92)
Multi-domain amnestic criterion T<40	36/32/41/78	46.75% (35.29%-58.48%)	70.91% (61.48%-79.18%)	1.61 (1.10- 2.34)	0.75 (0.59- 0.96)
Comprehensive criterion T<40	49/45/28/65	63.64% (51.88%-74.30%)	59.09% (49.31%-68.37%)	1.56 (1.17-2.06)	0.62 (0.44-0.86)

Notes. CI=confidence interval, TP=true positive, FP=false positive, TN=true negative, FN=false negative, SENS=sensitivity, SPEC=specificity, LR+=positive likelihood ratio, LR-=negative likelihood ratio.

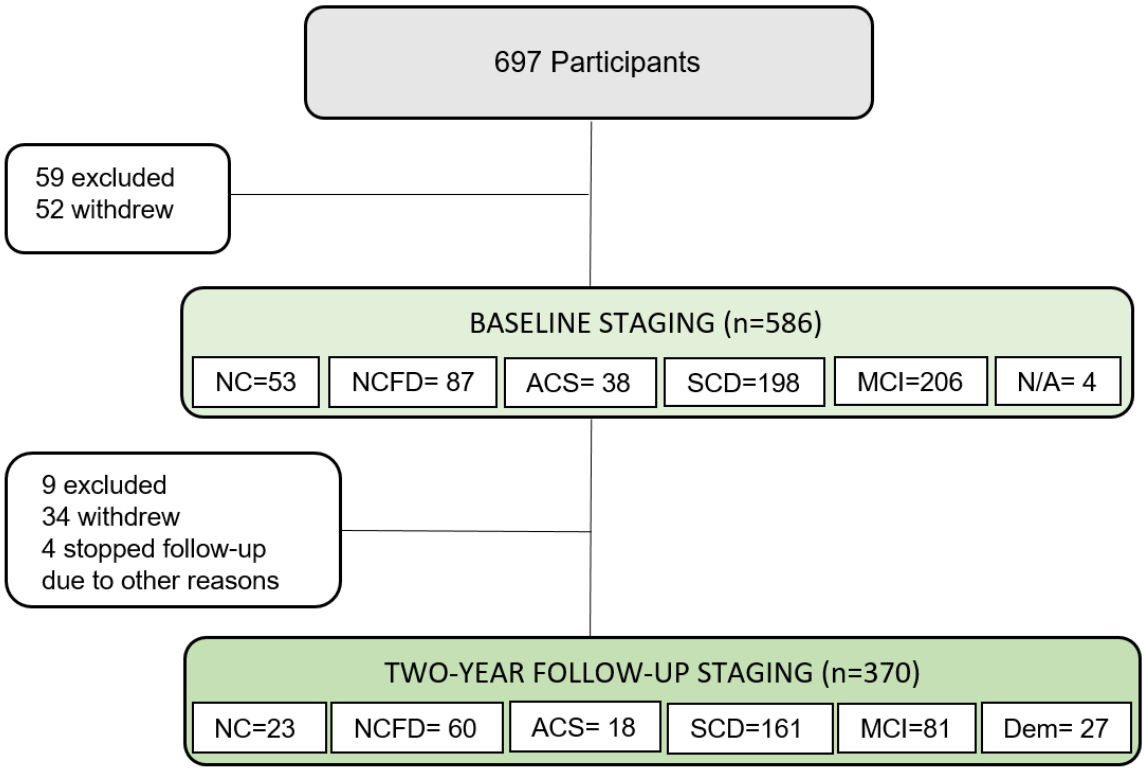

Figure 1. Participant flow chart for the DDI study. Out of the 697 subjects considered for inclusion, 586 participants were staged at baseline, and 370 at follow up. Staging categories were normal control (NC), normal control with first degree relative with dementia (NCFD), control with abnormal cognitive screening results (ACS), subjective cognitive decline (SCD), mild cognitive impairment (MCI) or dementia (Dem).

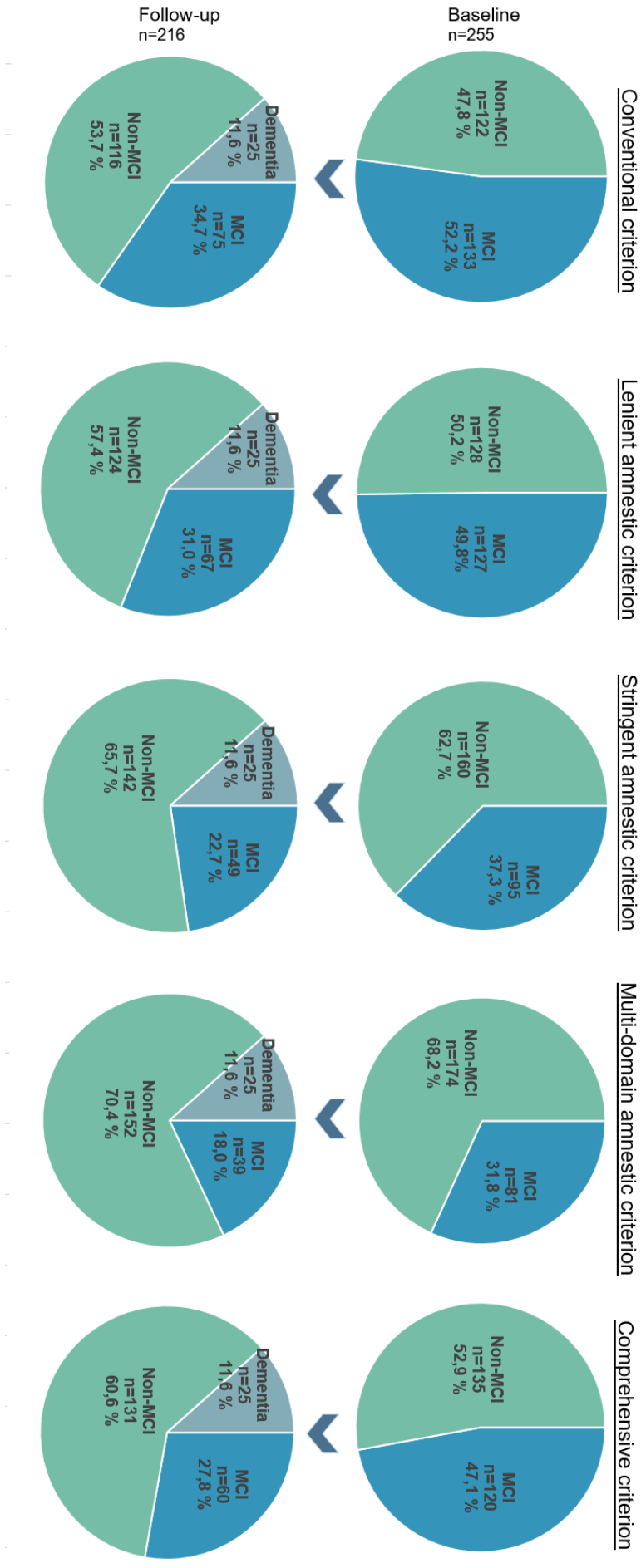

Figure 2. Longitudinal distribution of cognitive classifications for each of the MCI criteria