A Neuropsychological Perspective on Defining Cognitive Impairment in the Clinical Study of Alzheimer’s Disease: Towards a More Continuous Approach

Roos J. Jutten\textsuperscript{a,1}, Louisa Thompson\textsuperscript{b,c,1,*}, Sietske A.M. Sikkes\textsuperscript{d}, Paul Maruff\textsuperscript{e}, José Luis Molinuevo\textsuperscript{f,g}, Henrik Zetterberg\textsuperscript{h,i,j,k}, Jessica Alber\textsuperscript{l,m}, David Faust\textsuperscript{n}, Serge Gauthier\textsuperscript{o}, Michael Gold\textsuperscript{p}, John Harrison\textsuperscript{d,q,r}, Athene K.W. Lee\textsuperscript{b,c} and Peter J. Snyder\textsuperscript{d}

\textsuperscript{a}Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
\textsuperscript{b}Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, Providence, RI, USA
\textsuperscript{c}Memory and Aging Program, Butler Hospital, Providence, RI, USA
\textsuperscript{d}Alzheimer Center Amsterdam, Department of Neurology, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands
\textsuperscript{e}The Florey Institute of Neuroscience and Mental Health, Melbourne, VIC, Australia
\textsuperscript{f}Barcelona\textsuperscript{beta} Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain
\textsuperscript{g}Alzheimer’s Disease and Other Cognitive Disorders Unit, Hospital Clinic, Barcelona, Spain
\textsuperscript{h}Department of Psychiatry and Neurochemistry, Institute of Neuroscience & Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden
\textsuperscript{i}Department of Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden
\textsuperscript{j}Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, UK
\textsuperscript{k}UK Dementia Research Institute at UCL, London, UK
\textsuperscript{l}Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, The University of Rhode Island, Kingston, RI, USA
\textsuperscript{m}George and Anne Ryan Institute for Neuroscience, University of Rhode Island, Kingston, RI, USA
\textsuperscript{n}Department of Psychology, University of Rhode Island, Kingston, RI, USA
\textsuperscript{o}McGill Centre for Studies in Aging, Montréal, Canada
\textsuperscript{p}AbbVie Inc., North Chicago, IL, USA
\textsuperscript{q}Metis Cognition Ltd, Kilmington Common, UK
\textsuperscript{r}Institute of Psychiatry, Psychology & Neuroscience, King’s College London, UK

Handling Associate Editor: Katherine Gifford

Accepted 20 December 2021
Pre-press 31 January 2022

\textsuperscript{1}These authors contributed equally to this work.

\textsuperscript{*}Correspondence to: Louisa Thompson, PhD, Memory & Aging Program, Butler Hospital, 345 Blackstone Blvd., Providence, RI 02906, USA. E-mail: louisa_thompson@brown.edu.
Abstract. The global fight against Alzheimer’s disease (AD) poses unique challenges for the field of neuropsychology. Along with the increased focus on early detection of AD pathophysiology, characterizing the earliest clinical stage of the disease has become a priority. We believe this is an important time for neuropsychology to consider how our approach to the characterization of cognitive impairment can be improved to detect subtle cognitive changes during early-stage AD. The present article aims to provide a critical examination of how we define and measure cognitive status in the context of aging and AD. First, we discuss pitfalls of current methods for defining cognitive impairment within the context of research shifting to earlier (pre)symptomatic disease stages. Next, we introduce a shift towards a more continuous approach for identifying early markers of cognitive decline and characterizing progression and discuss how this may be facilitated by novel assessment approaches. Finally, we summarize potential implications and challenges of characterizing cognitive status using a continuous approach.

Keywords: Cognitive assessment, early diagnosis, neuropsychology, preclinical Alzheimer’s disease, progression

BACKGROUND

Alzheimer’s disease (AD) is currently understood as a neurodegenerative disease starting with a preclinical phase lasting 15 years or more wherein neuropathological change occurs but clinical symptoms remain subtle or absent [1–5]. Despite this presumed long presymptomatic continuum characterized by subtle and gradual cognitive change, all clinical diagnostic systems for AD imply that cognitive function should be classified as either normal or impaired as if there were an abrupt change at one point [6, 7]. While this classification is practical for clinical settings, it also leads to a lack of precision and can limit the design of research and clinical trials that focus on preventing AD-related cognitive decline.

Defined in its broadest terms, the field of neuropsychology is concerned with understanding the convergence of behavior, emotion, cognition, and brain function. In clinical neuropsychology practice, decisions about cognitive impairment are typically guided by the comparison of performance on one or more tests to cut-off scores for abnormality derived from relevant normative data [8, 9]. Measures of premorbid intellectual functioning aid in estimating the extent of cognitive change, especially when testing is only done at a single point in time. The extent to which any detected cognitive impairment reflects AD is then determined by the profile or pattern of cognitive performance as well as the presence or absence of supportive neuroimaging or other biomarker data [10]. The practice of single time point assessment has developed out of necessity and practicality for diagnostic purposes, and numerous studies have consistently and convincingly shown the diagnostic value of a cross-sectional neuropsychological evaluation in distinguishing normal cognition from mild cognitive impairment (MCI) and dementia [11].

This clinical neuropsychological approach to assessment has influenced the role of cognitive testing in AD research and clinical trials, including screening procedures targeting specific groups for recruitment and the stratification of patient cohorts. In clinical research, characterizing individuals as either cognitively healthy or impaired based on a single, cross-sectional screening assessment allows for a straightforward design. However, this approach certainly leads to the loss of a large amount of information, as the transitions between baseline cognitive function, age-related amnestic changes, and disease-related decline can be subtle and difficult to measure with precision. Therefore, it is important to examine the pitfalls and potential alternatives to this approach as we seek to better characterize the preclinical phase of AD and consider ways to update the role of clinical neuropsychology in the future of AD research.

Theories and definitions of normal cognitive aging, such as those established by Salthouse [12] and Lidenberger & von Oertzen [13], are shifting as a result of improved understanding of amyloid-β (Aβ) deposition and other neuropathological changes that can identify individuals without dementia who have neurodegenerative disease, and therefore may not have normal (i.e., non-pathological) cognitive aging. For example, in a recent longitudinal analysis of cognitively normal older adults, Harrington and colleagues [14] found that estimates of age-related decline in cognition were accentuated when cognitively unimpaired individuals with elevated Aβ pathology were considered in the “normal aging” group. Although the clinical criteria for MCI and dementia due to AD provide useful societal and medical benchmarks, clinicians and researchers are still grappling with how to address and define the preclinical stage of AD. The 2018 National Institute on Aging–Alzheimer’s Association (NIA-AA)
Alzheimer’s Diagnostic Framework provides a biomarker-based system for AD staging, while placing the clinical syndrome on a separate continuum. Advantages and limitations of the NIA-AA Framework have been discussed in depth elsewhere [15], but some key challenges with this framework include the inherent discordance between results obtained from biomarker testing and a neuropsychological evaluation (e.g., cognitively unimpaired with positive AD biomarkers), which can be confusing for both providers and patients [16]. Additionally, within-subject discrepancies between the clinical and biomarker categorization may affect the outcomes of research and clinical trials involving cognitively unimpaired individuals as a control group, given evidence that approximately 30% of such individuals likely have elevated cerebral Aβ [17].

Thus, we believe the time is right for neuropsychologists to examine how our characterization of the cognitive features of AD can evolve to better describe subtle cognitive decline in preclinical AD, and to potentially better reflect the more continuous nature of the clinical AD syndrome when possible.

CONCEPTUAL ISSUES AND CHALLENGES IN DEFINING ‘COGNITIVE HEALTH’ IN AGING

Beginning with a broad view of practices around assessing cognitive health, one of the first apparent challenges is the fact that definitions and interpretations of cognitive decline vary widely. For example, cognitive decline during older age is interpreted differently across cultures, with some being more likely to attribute memory loss and other AD-related symptoms to natural aspects of aging as opposed to pathological processes [18–20]. Clinical disciplines may also vary in their approaches to defining cognitive health. Neurologists may assess cognitive health with a combination of neurological examination, brief cognitive screening measures, and neuroimaging, whereas neuropsychologists primarily rely on multi-domain cognitive assessments, as well as informant and self-reports of cognitive and functional changes. Consequently, the discipline of the clinician assessing the patient will influence how cognitive health is defined.

In addition to these broader challenges, measuring AD-specific cognitive decline and distinguishing it from potential age-related cognitive changes remains particularly difficult because the clinical profile of incipient AD can be quite heterogeneous [21, 22]. Psychometric limitations of commonly employed tests (e.g., floor- and ceiling effects) as well as interpretive factors may impede a formal computational pattern analysis of cognitive test data, as the complexity of potential patterns across these variables increases. For example, marked variation in the level of test performance within and across individuals and the mathematics of comparing an expanding number of test scores, which can exponentially diminish reliability of test patterns as a whole [23–25].

Altogether, these challenges underline that there is no universal consensus on the actual definition of cognitive health, nor any single validated approach to assess this complex construct. We will now review two main complexities inherent in conventional neuropsychological approaches to determining cognitive impairment, that is, the use of cross-sectional normative data to define cognitive status, followed by the inclusion of subjective complaints in defining cognitive health.

The use of normative data

Neuropsychological test results are often interpreted based on cross-sectional normative data with adjustment for age, sex, and sometimes education. The reliance on historical normative data obtained from cross-sectional samples carries limitations. First, it has been shown that cognitive test performance is cohort dependent. That is, a healthy 50-year-old person performs differently on a test today relative to how a healthy 50-year-old person performed on the same test twenty years ago, even after adjusting for the higher educational level of more recent cohorts [26]. As a result, the ‘average performance’ on a given test may drift over time and, consequently, normative data (norms) may become rapidly outdated. In addition, even after correction for years of education, it remains difficult to capture cognitive decline among highly educated individuals who likely possess substantial cognitive reserve [27]. There also exists a major and well-established problem when applying normative data to both highly educated and less educated individuals, as both ends of the continuum are generally represented poorly in normative samples. Indeed, norms are based typically on relatively homogeneous samples, and may not generalize to other populations or birth cohorts that differ across a number of other key variables (e.g., socioeconomic status, ethnicity, cultural heritage, language fluency, and varying quality and models...
of education). With the rapid growth of the aging population worldwide, additional issues are amplified, such as known ethno-racial disparities in the prevalence, diagnosis, and treatment of AD [28, 29]. Despite these significant gaps in care, we nonetheless continue to have limited reference data for clinically underserved and underrepresented racial and ethnic populations [30] as well as the oldest-old segment of our population. Though recent efforts have made progress in addressing these gaps [31–33], broader issues in cultural competency among clinicians and researchers and monocultural assumptions and practices also hinder the adoption of more equitable assessment approaches [34, 35].

Another challenge is that shifts in test performance across editions of tests may be partly or largely artifactual and relate to changes in inclusionary and exclusionary criteria that alter the overall composition of normative samples. For example, more exclusionary criteria were added across the successive versions of the Wechsler Adult Intelligence Scale, thus creating more select samples of higher functioning individuals on average, which in large part explains the need to recalibrate norms and the presumed tendency of test scores to ‘improve’ in the general population over time (i.e., the Flynn effect [36]). Another problem is that drawing normative samples from older populations may unintentionally result in the inclusion of individuals with preclinical stage AD or other neurodegenerative disease, leading to a misrepresentation of the ‘normal’ population. This, in turn, results in poor sensitivity for identifying early AD-related cognitive changes when using the commonly employed cut-offs of 1.5 or 1.0 standard deviations below the mean [37, 38].

Thus, it is a complex and expensive undertaking to collect a sufficient amount of data to provide reliable norms for common cognitive tests for each specific sub-population and across ages, and more importantly, doing this would not necessarily guarantee improved sensitivity to early-stage AD. Moreover, substantial proportions of the normative data derived from such a massive project would likely become outdated in less than a decade as newer generations of neuropsychological instruments are developed and as individual differences that influence cognitive performance become better understood.

The role of subjective cognitive complaints

The process of differentiating between normal cognitive health, MCI, and AD dementia relies heavily on self-reports and partner/caregiver-reports of cognitive change [39]. Subjective cognitive decline (SCD), generally defined as the perception of worsening cognitive performance in the absence of neuropsychologically detectable cognitive deficits, may be both a normal part of cognitive aging and an important feature of early pathological changes in cognitive function [40–42]. In the clinical staging scheme proposed in the NIA-AA Framework [43], subjective complaints are described as a possible clinical feature of preclinical AD Stage 2, reflecting that, in the context of research focused on the earliest stages of AD, the utility and role of subjective reporting is still unclear. This is in part because it has been difficult to craft sensitive and broadly applicable clinical criteria in order to assess SCD [44, 45]. Currently, there are a variety of measures and criteria designed to capture the phenomenon, making it challenging to compare data across studies and to investigate whether SCD is an early, reliable marker of declining cognitive health due to a neurodegenerative disease [45].

As the case with cognitive testing, assessing SCD at only a single time point is likely to be less reliable than repeated assessments over time. SCD often worsens with stress and commonly co-occurs with anxiety and depression [41, 46]. While anxiety and depression have been associated with increased risk for AD, they are by no means specific to AD and are known to have independent deleterious effects on memory and other cognitive functions for individuals throughout the lifespan [47, 48]. Research has recently sought to parse these associations through longitudinal monitoring of depression, cognitive performance, and SCD in cognitively normal adults, with early results suggesting that changes in depression symptoms may mediate associations between SCD and declines in objective memory performance between subjects over time [49].

Comparing SCD reports with clinical outcomes, collateral reports (i.e., from a caregiver or study partner), and changes in cognitive functioning and AD biomarkers, may be the most productive approach to deciphering the predictive utility of SCD across the clinical AD spectrum, particularly when examined in longitudinal studies [51–54]. Recent progress in this direction has shown that SCD distinguishes...
between A\(\beta\)(+) and A\(\beta\)(−) older adults beyond the predictive utility of apolipoprotein E (APOE) genotype across several large community-based cohorts of older adults [55]. In a longitudinal study of cognitively normal older adults, SCD at baseline predicted more rapid cognitive decline on neuropsychological assessments among A\(\beta\)(+) individuals [56]. Furthermore, collateral reports provide a useful point of comparison for interpreting SCD. Patient- and self-reported subjective cognitive changes on the Cognitive Function Index (CFI) [57] have been shown to differentially predict longitudinal objective cognitive decline at different stages of AD, with CFI self-reports showing the greatest accuracy early in the disease course [57]. Overall, findings suggest that SCD is an important component of cognitive change on the AD continuum, which should be factored into AD research frameworks and monitored in longitudinal clinical trials.

The field of neuropsychology has made significant strides in addressing the above challenges, such as developing more inclusive normative datasets [58], and forming working groups that establish standards and unified approaches to constructs such as SCD in the context of the AD continuum [43, 44]. However, reliance on cross-sectional cut-off scores to determine cognitive impairment has remained essentially unchanged over the past 50 to 75 years. As the larger field of AD research is growing rapidly and has become intensely interdisciplinary, neuropsychology must also develop more flexible, progressive approaches to defining cognitive dysfunction using advances in methodology, technology, and analytics.

**DEFINING COGNITIVE HEALTH USING PROGRESSION MARKERS**

To tackle the aforementioned challenges, we propose a shift toward a more continuous approach to defining cognitive performance, by using multiple repeated assessments across two or more narrow and/or long-time windows. Using sensitive and reliable tools, change scores resulting from these repeated assessments would enable the identification of ‘progression markers’, that is, cognitive decline determined by within-person change scores. The use of progression markers to define cognitive status has several advantages when compared to single time-point testing. Multiple repeated assessments may reduce error from various sources associated with single time point assessment, thereby providing a more reliable method to evaluate cognitive health [59]. This point was well illustrated in a study by Darby and colleagues, in which only a minority of participants were consistently diagnosed as having MCI on the basis of assessments at multiple time points within one day [60]. Additionally, since within-person measurement will rely on an individual’s change scores (using either raw scores when they possess satisfactory psychometric properties, such as a good approximation of interval measurement, or standardized scores), the availability of normative data is no longer a strict prerequisite to benchmark one’s performance. Hence, within-person measurement is expected to be more cross-culturally applicable, though development and validation with diverse samples would still be needed to avoid potential cultural bias in test items and formats.

Furthermore, a comparison of patterns of within-person decline across persons may be a promising method to detect accelerated cognitive change due to pathological processes that are distinct from ongoing ‘age-related’ changes [61]. The greater sensitivity of this approach has also been suggested by studies on preclinical AD showing that in older adults who do not meet criteria for MCI, abnormally high A\(\beta\) is associated with within-subject cognitive decline over time, but not with cognitive impairment at baseline [14]. Determining progression markers may therefore be an especially useful application to the measurement of subjective or subtle changes to evaluate individuals who report decline compared to their previous level of functioning but whose ‘objective’ performance still falls within the normal range according to available reference data. In these cases, multiple repeated assessments may allow for the detection of subtle changes associated with the earliest stages of neurodegenerative disease.

Utilizing a multiple time point assessment approach; however, also raises several challenges, such as the fact that it is time-consuming to do with current testing paradigms. Furthermore, it relies greatly on participant adherence and assumes that the within-person variance is equal across age, education, and cross-cultural groups, which might not always be the case. In the next section, we aim to discuss several approaches to address these challenges, including the implementation of new neuropsychological tools and strategies for repeated neuropsychological assessment in the clinical study of aging and AD.
NOVEL APPROACHES TO CHARACTERIZE COGNITIVE PROGRESSION MARKERS

Digital cognitive testing

One promising direction for repeated neuropsychological assessment involves the ongoing shift toward digital assessment tools, which can be used on their own or in combination with traditional neuropsychological assessments [62]. Testing software has the potential to reduce administration and scoring errors, automatically tailor tasks to an individual’s level of ability (i.e., computerized adaptive testing), capture nuanced performance information (e.g., response latencies and sequencing), rapidly compute and compare scores, and generate meta-data. Digital assessment tools can also be used to easily capture and characterize speech and language using advanced analytics, such as machine learning [63, 64]. All of these features are particularly attractive for applying digital tools for the longitudinal assessment of subtle, early cognitive changes associated with preclinical AD. Numerous digital cognitive test batteries, such as the NIH Toolbox and Cambridge Neuropsychological Test Automated Battery (CANTAB) [65], Cogstate Brief Battery [66], TabCAT [67], as well as standalone tests, such as the DCTclock™ [68] have been developed for face-to-face administration in clinical and research settings [69]. Unsupervised online neuropsychological testing has been used in several AD clinical trials [66]. More recently, the Online Repeated Cognitive Assessment [70] tool has been found to be sensitive to detect cognitive changes during the preclinical stage of AD, using metrics such as learning curves across multiple days of assessment [71]. Mobile versions of existing cognitive tests, as well as novel tasks designed specifically for mobile use, have also been developed in recent years [72–75], including from academic research programs such as the Center For Healthy Aging at Penn State [76], the Harvard Aging Brain Study [77], the Dominantly Inherited Alzheimer Network (DIAN) Observational Study [78], and Oxford University [79]. Remote app-based cognitive testing has received increased attention recently in light of the COVID-19 pandemic and is rapidly becoming more feasible [75, 80]. App-based testing in brief repeated sessions has demonstrated similar or better reliability and validity compared to standard in-clinic assessments [76, 81]. About two-thirds of baby boomers report using smartphones, and rates of smartphone use are similar among Black, Latino/Hispanic, and White Americans (~80%), indicating that app-based testing could be useful for reaching a broad demographic [82, 83].

Collecting cognitive data with smartphones comes with several unique challenges (e.g., variable device specifications, privacy, popup interruptions), as does remote cognitive testing in any format (e.g., environmental variability, use of unallowed supports or assistance from others). However, these approaches also have the potential to reduce costs and patient burden by avoiding the need for lengthy in clinic testing, and thus allowing repeated assessments and detection of within-person change within a short time frame. Adherence is major challenge with remote cognitive assessment that must be considered at the onset of any clinical research study deploying these methods. A recent survival analysis of adherence across eight different remote digital assessment studies (total combined N=1,000) found that greater participant retention time was most strongly associated with: a) referral to the study by a clinician, b) compensation for participation, and c) having the clinical condition of interest in the study, and d) older age [84]. An additional approach to improving adherence may be the provision of feedback for participants (either in real-time or post-assessment) to improve engagement and commitment with testing [79]; however, this method carries challenges to assessment validity and requires further research.

New “burst” testing approaches (i.e., multiple brief assessments completed over a period of several days) provide multiple data points that can be averaged to generate more reliable indicators of cognitive performance and may be better tolerated and more robust to variable adherence in data collection relative to more lengthy assessments ([79, 85]). Conducting brief assessments via mobile app also has the potential to improve the ecological validity of cognitive tests by allowing patients to complete them in their typical environments across different time points both within a day and across days [76]. While empirical support for mobile app-based assessment tools is still limited, recent initiatives, such as the NIA Mobile Toolbox, aim to bring open source and easily accessible mobile assessment tools to wider scientific and clinical audiences in the next few years, which may help accelerate validation studies.

Novel approaches to analyze conventional test data

Another solution for improving the sensitivity of assessment may involve new approaches to analyzing
longitudinal cognitive data obtained using conventional measurements. Asken and colleagues [86] have recently proposed a Discrepancy-based Evidence for Loss of Thinking Abilities (DELTA) score as a new method for characterizing cognitive decline on a continuous spectrum. Using ADNI data, they derived regression-based normative reference scores using age, sex, years of education, and word-reading ability from cognitively normal participants. DELTA scores were calculated to reflect the degree of discrepancy between predicted and observed scores, and hence the strength of evidence of cognitive decline. This approach had a positive predictive value greater than 0.9 for AD biomarker classification cross-sectionally, and more recently, was shown to improve prediction of cognitive and functional decline above and beyond biomarker variables longitudinally [87]. These findings suggest that DELTA scoring could be a promising method for capturing cognitive function on a continuum; however, this work has yet to be replicated with more representative longitudinal samples of diverse older adults and in other clinical research studies with different cognitive tools and intervals between assessments.

One other statistical strategy to increase the responsiveness of existing neuropsychological instruments involves the use of item response theory (IRT) analysis. IRT links responses for a specific set of items to an underlying construct resulting in a latent trait score, assuming that items contribute differently to this latent trait score [88]. That is to say, the IRT model considers that some items may be more difficult to ‘endorse’ than others and thus carry more weight in reflecting an individual’s level of cognitive function. In addition, IRT-based items can be designed to provide high degree of separation within a narrow range of baseline capacity (sigmoid curves). Compared to classic scoring methods, such as creating a simple sum score, an IRT score maximizes the sensitivity of responses and results in greater accuracy in the assessment of individual change over time [89–91]. For example, studies have shown that IRT might improve the precision of widely-applied cognitive tests such as the Alzheimer’s Disease Assessment Scale - Cognitive subscale [92, 93]. It should be noted that calibrating robust and generalizable IRT models requires large sample sizes that cover the entire spectrum of the latent trait (i.e., cognitive function) and provide an adequate representation of the target population (i.e., ranging from older adults with normal cognition to individuals with dementia). Ideally, calibrated IRT parameters are then validated in an independent sample. This undertaking seems worthwhile since head-to-head comparison studies have consistently shown that IRT scoring outperforms classic test scoring techniques, especially when investigating an individual’s change or ‘growth’ over time or the effectiveness of clinical interventions [94, 95]. An additional advantage is that, once calibrated, IRT parameters can be used to ‘link’ scores across different test versions and anchor measurements of change to determine the clinical meaningfulness, as well as minimal important change, of the latent trait scores.

The study of practice effects (PE), also referred to as learning or retest effects, provides another opportunity to analyze repeated neuropsychological measurements. Practice effects are improvements in cognitive test performance due to repeated evaluation with the same or similar test materials [96]. It has been shown that subjects with late-life cognitive disorders show reduced practice effects as compared to their healthy peers [97]. Furthermore, diminished PE may predict future decline, a future diagnosis of MCI, and increased presence of cerebral pathology. Hassenstab and colleagues [98] showed that reduced PE on episodic memory tests were detectable in subjects with preclinical AD, and that the magnitude of these practice effects was inversely related to risk of progression to AD. All together, these findings suggest quantifying the magnitude of PE may be a valuable progression marker especially in early stages of AD [99]. However, at this point there is not much evidence that supports the use of PE on an individual level, and this warrants further investigation. Moreover, significant challenges and limitations remain in determining how we quantify PE (e.g., via simple within-subject change scores, regression-based methods, or reliable change index), and whether there is an optimal time interval to study PE (e.g., over months [100], days [101], or even within a single day [60]). A further question is how to handle variation in their magnitude related to the number of test administrations, as studies have consistently shown that PE seem to be greatest over initial exposures followed by a pattern of diminished gain over the course of repeated assessments [97].

DISCUSSION

We described several promising methods and approaches to complement conventional neuropsychological assessments and characterize cognitive
progression markers based on repeated assessments. These progression markers could result in a scale reflecting the continuous nature of cognitive aging more reliably as compared to static, single time-point testing that groups people as cognitively healthy or impaired. The use of cognitive progression markers may be particularly relevant for individuals in the preclinical stage of AD, an early, presumably transitional phase of cognitive decline without objectifiable cognitive impairment, or “Stage 2” as described in the clinical staging scheme of the NIA-AA Framework [43]. These Stage 2 individuals are considered an important and clinically relevant population, and they have become the main target population of secondary AD prevention trials [4]. However, identifying individuals in this ‘transitional stage’ remains challenging using current neuropsychological paradigms that rely on cross-sectional, single time-point testing [102]. Longitudinal assessment to monitor for cognitive progression markers could aid the identification of Stage 2 individuals, especially in combination with disease-specific biomarker information. This approach has the potential to both advance AD clinical trial screening procedures leading to more successful enrollment of Stage 2 participants, and aid in the detection of individuals with early cognitive decline in the memory clinic.

Harmonization of cognitive progression markers with biomarker models, such as the NIA-AA Framework, may improve our ability to identify high-risk individuals for secondary prevention trials, especially if progression markers could be combined with accurate AD blood biomarkers for screening [103]. This approach may aid in detecting subtle cognitive decline and assist with clinical trials by better predicting cognitive trajectories in the disease course. For example, if we know that a given biomarker pattern (e.g., tau deposition) evolves in a predictable way and predicts the onset of more rapid decline in a cognitive domain (e.g., declarative memory), then we can plan clinical trials where delays in the progression of the pathology can potentially be tied to delayed worsening of the affected cognitive domain, and we can avoid looking for signals where no treatment effect is likely to take place.

When harmonizing cognitive progression markers with a biomarker-based model of disease progression such as the NIA-AA Framework, it is tempting to consider biology as the reference standard. However, it is important to realize that the clinical value of most AD biomarkers remains inconclusive particularly in the preclinical stages, as not all individuals with AD neuropathology will go on to develop clinical symptoms [104]. More work is needed to understand what leads and what lags in terms of both neurobiology and cognitive function. Therefore, as a best practice we should avoid relying solely on single biomarker-based outcomes for validation of cognitive measures and characterization of early AD cognitive change, and instead focus on the prediction of the emergence and progression of clinical symptoms.

Ultimately, combining the predictive power of cognitive and biomarker data as complementary approaches may be particularly useful for differentiating healthy aging from preclinical AD [105] and those at risk for progression to dementia [106]. Yet many newer algorithms in development do not include cognitive data [107]. This again points to the need for innovative and more sensitive cognitive tools that are easy to administer and repeat as part of a larger, multi-modal assessment approach. If the development of novel neuropsychological methods for detecting cognitive progression markers could go hand-in-hand with the analytical and clinical validation of novel blood tests for AD pathology, longitudinal studies of the interplay between cognitive function and AD-related brain changes would become more feasible.

**Challenges and limitations**

While cross-sectional neuropsychological assessment will always be needed in certain situations, there is much to be gained by characterizing cognitive decline as points along a continuum and moving away from the oft relied on dichotomous “lumping” approach to cognitive aging assessment. However, adopting an approach that involves characterization of change on a continuum creates both methodological and practical challenges, as well as limitations that are not easily addressed. Included among these methodological challenges is establishing the reliability and longitudinal validity of measurement instruments that are used to identify progression markers. These measurement properties are not self-evident, and thus novel test paradigms require thorough validation efforts, including examination of their generalizability and sensitivity to change over time in the target population. Moreover, in the context of examining progression markers, determining what constitutes
an ‘abnormal’ within-person change, and what the appropriate length of time should be between tests, are additional methodological challenges that need to be addressed. A number of promising research studies investigating the feasibility, reliability, and validity of novel repeated assessment approaches exist and can set a precedent for future work [76, 100, 101, 108].

Another challenge relates to the use of progression markers on a group level versus on a case-by-case or individual level. Further, it raises the question of what the ideal time frame for longitudinal assessments should be to in order to reliably stage individuals on the spectrum from cognitively normal to impaired. From the patient perspective, not being able to receive immediate diagnostic feedback following an initial evaluation of memory problems could lead to heightened feelings of uncertainty and anxiety. One potential solution to this problem could involve the remote administration of repeated assessments via online or smart phone-based testing prior to conventional in-clinic cognitive evaluation, in order to avoid delays in data interpretation. From a clinical trial perspective, repeated assessment could necessitate a longer screening period, a concept that has already been applied in the use of a “trial-ready-cohort” or “run-in-study” in AD research, with the expectation of faster clinical trial enrollment in the long run. Even with brief assessments over short time-intervals, repeated measures testing is more time-, labor- and cost-intensive than single-time testing, and this is especially problematic when participants are lost to follow-up. Adherence is a major issue in implementing mobile phone-based assessments, and involvement of the target population in developing these digital tests is crucial. Relying on multiple assessments may therefore be less suitable for individuals who are already on the more impaired end of the cognitive aging spectrum, as in those cases a single cognitive screening visit will likely be sufficient to establish a diagnosis or screen for participation in a clinical trial. In fact, conventional neuropsychological tests have proven to show high sensitivity and specificity for establishing a dementia diagnosis [109].

**Implications**

We believe that the future success of many efforts in neuropsychology will depend on the ability to detect subtle changes as opposed to identifying impairments after it is too late to do much. That, in turn, will require better and revised measurement strategies, but multiple factors create challenges in this pursuit. For one, it is often difficult to pinpoint when the disease process starts, or at least the period before neurocognitive deficit is first manifested. Also, pre-disease measurement benefits may not be realized if baseline assessment is not done effectively [110, 111]. Finally, variation or error in test performance could obscure the appearance of disease specific effects, especially if one’s goal is early detection, given the 3 to 4 standard deviation range common across an individual’s highest and lowest test scores in a battery of tests [112–114].

It would be very desirable if cognitive screening tools were brief, widely available, affordable and reimbursed and if cognitive screening were part of routine healthcare, akin to measuring cholesterol levels, with an emphasis on population-based approaches. Brevity might be pursued through various strategies, such as adaptive testing approaches, which can greatly improve efficiency, and focusing in on functions that are most susceptible to change across the condition(s) of interest. As for cost and accessibility, digital technology could go a long way towards solving many of these issues.

Some neuropsychologists may experience understandable concern that brevity is incompatible with a complete neuropsychological assessment, could deprive professionals of fair economic opportunity or ownership, or may lead non-neuropsychologists to take on tasks for which they are not necessarily qualified. However, expansion of population-based, longitudinal cognitive screening should not preclude a full neuropsychological assessment when indicated. There will always be benefits to thorough testing when a clinical diagnosis is necessary, when there are complex phenotypes (e.g., posterior variant AD) to characterize, and when recommendations for support and treatment are needed. Improved cognitive screening in different contexts (e.g., primary care settings) could refine the referral pipeline for neuropsychologists and increase access for patients who are most need of a full evaluation. Additionally, a move toward using novel methods (e.g., digital technology) for developing and adopting high quality repeated assessment approaches more broadly in neuropsychological clinical practice, when appropriate for detecting change, could provide incredible clinical and research opportunities for the field and greatly enhance our contributions to human health and well-being.
CONCLUSION

This is a pivotal time for neuropsychologists to reexamine and adjust how we assess and characterize the cognitive features of AD, in order to improve harmonization with a biomarker-based disease framework and detection of subtle cognitive changes during the earliest stage of AD. We propose a multipronged, flexible approach to allow for the assessment of cognitive change on the AD continuum, particularly as a means to improve future AD prevention clinical trial design, which incorporates the development and use of ‘cognitive progression markers’ based on repeated assessments. We addressed several promising approaches and methods to define these progression markers, any or all of which may better capture the continuous nature of cognitive aging and cognitive decline in the context of AD (Box 1). Based on methodological and practical issues, we argue that multiple assessments are currently of particular relevance for defining ‘cognitive health’ and subtle cognitive impairments in the early clinical stages of AD, particularly in a research setting. Neuropsychological paradigms that reliably assess clinically meaningful cognitive progression need to be further validated in order to apply them in clinical practice.

ACKNOWLEDGMENTS

HZ is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#681712), Swedish State Support for Clinical Research (#ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), and the UK Dementia Research Institute at UCL.

AL and JA are partially supported by Institutional Development Award Number U54GM115677 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds Advance Clinical and Translational Research (Advance-CTR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors’ disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-5098r1).

REFERENCES


[67] Possin KL, Moskowitz T, Erhoff SJ, Rogers KM, Johnson ET, Steele NZR, Higgins JJ, Stiver J, Alioti AG,


