Assessment of cognition in mild cognitive impairment: A comparative study

Peter J. Snyder[a,b,*], **Colleen E. Jackson**[c], **Ronald C. Petersen**[d], **Ara S. Khachaturian**[e], **Jeffrey Kaye**[f], **Marilyn S. Albert**[g], and **Sandra Weintraub**[h]

[a]Office of the Vice President for Research, Lifespan Hospital System, Providence, RI, USA

[b]Department of Neurology, Alpert Medical School of Brown University, Providence, RI, USA

[c]Department of Psychology, University of Connecticut, Storrs, CT, USA

[d]Alzheimer's Disease Research Center, Mayo Clinic, Rochester, MN, USA

[e]Campaign to Prevent Alzheimer's Disease by 2020 (PAD2020), Rockville, MD, USA

[f]Layton Aging and Alzheimer Center, Oregon Health Sciences University, Portland, OR, USA

[g]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[h]Cognitive Neurology and Alzheimer's Disease Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Abstract

The demand for rapidly administered, sensitive, and reliable cognitive assessments that are specifically designed for identifying individuals in the earliest stages of cognitive decline (and to measure subtle change over time) has escalated as the emphasis in Alzheimer's disease clinical research has shifted from clinical diagnosis and treatment toward the goal of developing presymptomatic neuroprotective therapies. To meet these changing clinical requirements, cognitive measures or tailored batteries of tests must be validated and determined to be fit-for-use for the discrimination between cognitively healthy individuals and persons who are experiencing very subtle cognitive changes that likely signal the emergence of early mild cognitive impairment. We sought to collect and review data systematically from a wide variety of (mostly computer-administered) cognitive measures, all of which are currently marketed or distributed with the claims that these instruments are sensitive and reliable for the early identification of disease or, if untested for this purpose, are promising tools based on other variables. The survey responses for 16 measures/batteries are presented in brief in this review; full survey responses and summary tables are archived and publicly available on the Campaign to Prevent Alzheimer's Disease by 2020 Web site (http://pad2020.org). A decision tree diagram highlighting critical decision points for selecting measures to meet varying clinical trials requirements has also been provided. Ultimately, the survey questionnaire, framework, and decision guidelines provided in this review

*Corresponding author. Tel.: 401-444-4117. psnyder@lifespan.org.

should remain as useful aids for the evaluation of any new or updated sets of instruments in the years to come.

Keywords

Cognition; Neuropsychological assessment; Alzheimer's disease; Mild cognitive impairment; Clinical trials

1. Introduction

The diagnosis of (probable) Alzheimer's disease (AD) [1] in the living patient has been based almost entirely on clinical examination, with postmortem neuropathologic confirmation of the disease made only on occasion and typically in the context of research [2]. Clinicians have been largely limited to diagnosing probable AD after ruling out other potential causes for cognitive decline and confirming impairments in function or activities of daily living. With the limitations of currently available and well-validated biomarkers or symptomatic therapies at the turn of the 21st century, patients must still present with observable cognitive decline to meet diagnostic criteria for mild cognitive impairment (MCI), and with concomitant functional deficits for a diagnosis of AD, either of which would then lead to consideration for pharmacologic intervention.

However, by the time our patients present to clinic with demonstrable memory and other cognitive dysfunction, they have most probably suffered irreversible brain tissue damage and loss [3,4]. Although currently marketed medications may improve memory functioning (or slow the rate of deterioration) [5], they do not offer disease-altering properties or neuroprotective benefits [5–7]. Although the majority of clinical trials for new therapeutics have relied on the study of patients with mild-to-moderate probable AD, it seems likely that neuroprotective therapies applied during the earliest stages of disease would lead to better clinical outcomes for individuals—before much irreversible damage has occurred [8]. To achieve this goal, cognitive markers are needed that reliably identify individuals who are in the symptomatic predementia phase of AD or are presymptomatic but at high risk for experiencing clinical onset of MCI within several years, as well as markers that have the ability to measure and to track (with high sensitivity) very subtle cognitive changes over time in these same individuals [9–11].

Several large prospective studies are attempting to characterize the early disease course in persons who are either classified as healthy elderly, MCI, or early AD, by following these cohorts with regular clinical, neuroimaging, and laboratory-based biomarkers of the disease. Using a natural history approach, and following participants longitudinally from middle-to-late adulthood through old age, these studies are contributing much to our understanding of how individual genetic, biological, and physiological mechanisms, combined with social and environmental influences, combine to influence the rate of disease occurrence and progression. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI-1), initially started as a 5-year trial based largely in the United States, emphasizes the study of the relationships between changes in cognition, brain structure and function, and biomarkers (www.adni-info.org). Aiming to cover the progression of the disease, from the earliest

symptoms, before noticeable cognitive changes, through autopsy, ADNI-1 had enrolled 229 healthy controls, 405 participants with MCI, and 188 participants with AD as of May 2008 (www.adni-info.org). With extensive emphasis on wet laboratory biomarkers (e.g., cerebrospinal fluid measures of Aβ, genetic markers) and neuroimaging (e.g., structural magnetic resonance imaging [MRI], positron emission tomography [PET]), the ADNI protocol has a strong focus on the development and refinement of biological measures or predictors of incipient cognitive impairment. The cognitive tasks chosen for use in this study are largely pencil-and-paper measures that require direct administration and recording by trained personnel. Alternate forms of the tasks were not used. Recent internal reviews of ADNI data have revealed substantial (and yet expected) rates of human errors in the recording and transcription of cognitive data, thus requiring continual quality checking and cleaning of the database. In addition, most of these measures were not designed to specifically assess the earliest signs of cognitive impairment or AD; however, the cognitive functions that they assay are known to change with dementia.

A second example of a current, large, prospective study, aiming to assess the role of physical and environmental influences on cognitive health, and particularly changes in cognitive health during the aging process, is the Australian Imaging, Biomarkers, and Lifestyle (AIBL) study. This study aims to recruit a minimum of 200 participants with AD, as defined by the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria [1], 100 participants with MCI [12], and 700 participants without cognitive impairment, including participants with at least one *APOE* ε4 allele, participants without a copy of the apolipoprotein E (*APOE*) ε4 allele, and participants with subjective memory complaints [13]. Participants are asked to complete a neuropsychological battery consisting of 1 to 2 hours of traditional paper-and-pencil measures (e.g., Boston Naming Test [BNT], California Verbal Learning Test - II [CVLT-II]), and a subset in each group are administered the basic version of the CogState brief computerized battery. However, although the use of both standard and computerized measures provides for a comprehensive assessment, it also requires additional time and labor, thus making the final assessment burdensome, particularly for such a large study [13]. A specific benefit of using repeated computerized assessment in the AIBL study is the ability to examine rate-of-change (RCh) for a subgroup of the study participants. A recent linear mixed modeling analysis of this AIBL subgroup found a significantly different RCh, over the initial 4-month period of this study, on a measure of visual learning for participants with MCI as compared with cognitively healthy control participants, with patients with MCI showing greater relative decline (P. Maruff, personal communication, 2011) [14].

As the field continues to advance, newer interventions will depend on the assessment of a slowing of cognitive change as evidence of their efficacy. Therefore, we must remain committed to the continued development and refinement of cognitive measures, particularly with respect to their ability to accurately and reliably identify and track very early cognitive change. Evans et al [9] have stated that cognitive change does not begin in old age, but is rather an ongoing developmental process that varies in rate from one individual to another and from one cognitive domain to another. Thus, it is possible that imperceptible change

occurs very early in adulthood, beginning an accelerating cascade and requiring that longitudinal studies begin much earlier than is typically the case at present.

In addition to their ability to detect change, measures must also be evaluated in terms of their administration limitations, including length of administration, necessary involvement of trained administrators and/or healthcare providers, and the introduction of human error during administration, data entry, and/or data scoring. For example, many of the traditional neuropsychological measures used in research require extensive testing time, which in turn increases the potential for patient fatigue and boredom, necessitates trained administration, and often uses a pencil-and-paper approach, all of which increase both human error and administration time. Consequently, new cognitive measures must look to maximize recent technological advances (e.g., Internet-based administration, automated audit trails, real-time quality controls) while also preserving psychometric rigor. This suggests that it is important to consider a shift toward computer-administered measures in both the clinical and research domains, which may make assessments less time- and energy-intensive, thereby allowing for more frequent and less burdensome assessments to provide better monitoring of change/decline. The use of computerized testing will not replace the neuropsychologist's expertise in translating psychometric findings into practical recommendations for further workup and follow-up, but will instead facilitate the ability to obtain more reliable and less burdensome test results.

The Campaign to Prevent Alzheimer's Disease by 2020 (PAD2020) initiative has recently partnered with the Alzheimer's Association in an effort to advance several large projects aimed at enabling more rapid and efficient discovery of treatments for AD, such as a large multinational registry trial of older adults. For any such new, massive, longitudinal registry trial, PAD2020 is focused on identifying individuals at the very earliest stages of disease, ideally before cognitive impairment has been recognized. To support this effort, one of the "enabling technologies" is to identify a validated set of brief, reliable, and easily delivered cognitive measures that, together, would be a "cognitive toolbox" to be made widely available for point-of-care use (for frequent monitoring) by primary care providers and specialists alike. Although the specific cognitive measures have not yet been selected, those to be included in the PAD2020 toolbox will include measures that are intended for *frequent assessment of change over time*, in individual subjects, across several facets of cognition that are known to change over the aging process and also with early disease.

The PAD2020 Comparative Survey, described in this report, was designed to aid in the identification of measures that will provide the most comprehensive combination of reliable cognitive assessment brevity, demonstrated construct and predictive validity, and ease of use. A relatively recent review of computer-based cognitive measures used in aging began to address this need for a systematic comparison across commercially available assessment batteries [15]. Wild et al focused on a subset of available literature published through 2007, but they did not include measures that were in development or not widely published [15]. The current survey provides a wider compilation of information, including the review of a broader range of marketed products that may be of value to clinical researchers in this rapidly evolving field. In an effort to expand the number of measures relevant to early cognitive change, we have intentionally oversampled the range of marketed cognitive

measures and encouraged the scientists associated with each measure to provide information that may or may not be published in the public domain. This led to the inclusion of some measures in the early stages of development that may not yet have published data relevant to aging but are certainly potential contenders in the field. In addition, even for those tests that have solid literature support, this survey and review provided a forum for detailed explanation and synthesis that may be lacking in the previously published data.

2. Methods

The comprehensive survey was constructed by one of the present study authors (P.J.S.), in consultation with other content experts. In designing the survey, the focus was on a thorough assessment of the relevant psychometric properties and methodologic considerations, in a standardized manner, with ample opportunity for each test/battery author to provide his/her best supporting evidence and documentation. The survey thoroughly assessed each measure in terms of general measure conceptualization and development, availability of normative data, specificity of the tool for assessing cognitive decline in older adults, psychometric properties (i.e., sensitivity, specificity, reliability, validity), data collection and storage methods, and convergent validity with other cognitive measures, disease-related biomarkers, and instrumental activities of daily living (IADLs; Appendix).

2.1. Collection of survey data

In an effort to oversample from available cognitive measures that may have utility in identifying early cognitive change associated with MCI or AD, we completed a comprehensive review of published measures, Internet searches, and consultation with experts in the field of aging and dementia to generate a list of potential test authors and/or vendors. A survey was mailed to the primary author or known contact for each of 32 identified measures in January 2010 (see Table 1 for a complete list of measures). Follow-up e-mails with the survey and cover letter explaining the purpose of the survey were sent in February 2010. Reminder e-mails were sent as the deadline of June 4, 2010 approached, and a follow-up e-mail was sent to individuals who had not responded by June 14, 2010. The primary author or contact for select measures with potential relevance to aging and dementia, as determined by the primary author of this review, was contacted in July 2010 through telephone for additional follow-up. Subsequent e-mail contact was maintained with those individuals who indicated that they were interested in participating in the survey. Ultimately, we received responses on 16 measures by September 2010.

2.2. Number of publications: Survey responders versus nonresponders

Each test or battery was identified as either a "responder" indicating that a representative author, vendor or test publisher completed a survey addressing the test's/battery's utility for detecting age-related change and early dementia, or a "non-responder" indicating that no survey had been returned. The number of peer-reviewed publications relevant to each test/ battery was summed using the following guidelines: (1) three separate PubMed searches were conducted using the following terms " "Name of test/battery" AND "dementia"; "Name of test/battery" AND "mild cognitive impairment"; "Name of test/battery" AND "Last name of primary author," (2) a search was conducted on the company Web site, when

available, for additional references pertaining either to dementia, AD, or MCI, and (3) a review of all bibliographic citations contained in all returned surveys.

To count any individual publication in this tally for each test or battery, each article needed to meet several criteria: (1) the study must have used the test/battery as an endpoint in a population of individuals with MCI or dementia, or for monitoring of healthy elderly for cognitive decline, (2) only studies involving participants with neurologic impairments (excluding studies involving drug-induced impairments) were included, (3) studies of dementia only within a specific patient population (e.g., Down's syndrome) were not included, and (4) studies were required to be published in English.

2.3. Review and analyses of survey responses

After the survey responses were returned to us for review, all responses were subjected to random spot-checking to compare individual item responses with claims made in the published, peer-reviewed literature, to ensure veracity of the information provided. Approximately 25% of all individual item responses were double-checked in this manner by one of the authors of the present study (C.E.J.) over a 4-month period. In addition, all returned surveys, summary tables, and the decision tree diagrams were sent, on two separate occasions, to all coauthors who served as an editorial team to ensure fairness and accuracy in reporting as best as possible. Finally, all survey responses have been posted (with only minor stylistic editing to ensure consistent format across documents) in a public archive at the PAD2020 Web site (http://www.pad2020.org) [16] for review.

As mentioned previously, survey data were analyzed for several predefined criteria, including sensitivity and specificity of the measure for identifying individuals with MCI, administrative ease of use, and convergence with biomarkers and ADLs/IADLs, among others. In addition, after receiving and reviewing the responses, other general information, such as previous use in a drug trial, and information specific to the measure were also included in the final report. The reported sensitivity (the proportion of the participant population with cognitive impairment or dementia who were correctly identified by the measure) and specificity (the proportion of healthy participants correctly identified as such) of each test/battery, with respect to its ability to discriminate healthy elderly from those with MCI, were reviewed for all instruments for which sufficient supporting literature was available (for companies/products without such data, we asked for information on sensitivity/specificity with respect to discriminating healthy elderly individuals from those with diagnosed AD).

2.4. Decision tree diagram

To facilitate future reviews of cognitive assessment measures specific to cognitive decline associated with aging, MCI, and AD, we have summarized our findings in a decision tree diagram. Based on *a priori* requirements that may differ from one planned use to another (e.g., dependence on computer administration and/or scoring, previous use in a drug trial, published evaluation of convergent validity with a disease-related biomarker), a clinical investigator may use this decision tree to streamline his/her search for measures that will meet such criteria.

As we generated the diagram, we encountered several decisions regarding how to label certain assessment measures. The first decision we faced was: under what circumstances did the computer need to be used to consider a measure to be "computer-based?" Although some measures allowed for or even required computer scoring (in the case of the *MCI Screen [MCIS]*), we felt that the nature of "computer-based" was more accurately captured by those measures that *required* computer administration. Consequently, if a measure was not administered on the computer, it was not considered "computer-based." However, to apply the label "web-based," we were most interested in whether the data collected, either on the computer or through a human rater, could be uploaded onto a database (e.g., for automated scoring) through the Internet. Therefore, measures that were not necessarily considered "computer-based," such as the MCIS, could be considered "web-based" if data from the measure needed to be uploaded through the Internet for analysis.

3. Results

3.1. Number of publications: Survey responders versus nonresponders

Sixteen measures were identified as "responders," indicating that a survey response from a primary author associated with the measure was received. Likewise, 16 measures were identified as "non-responders" (Fig. 1). A general trend toward fewer publications related specifically to early cognitive decline, MCI, or early AD was identified for nonresponders, with the majority of nonresponder measures having between zero and three publications that were specifically related to aging and dementia (10 of 16 nonresponders had zero relevant publications, based on our search criteria). Notable exceptions to this are the Computerized Neuropsychological Test Battery (NTB), with five identified publications in aging and dementia, and the Cambridge Neuropsychological Test Automated Battery, with 46 identified publications. It should be noted that the primary scientist associated with one measure (Neurobehavioral Evaluation System) identified as a nonresponder is deceased and there were no other scientists in a position to complete this survey on his behalf. In addition, another nonresponder represented a company that has recently closed, and this battery is not available at this time.

3.2. Survey responses

The 16 survey responses that were received and reviewed are included in this report. A table listing all questionnaire items contained in the survey is provided in Appendix. Some respondents included additional information that was not specifically requested. A considerable amount of additional information was also collected and compiled for each measure (e.g., normative data, convergence/divergence with other cognitive measures, and convergence with biomarkers). These data were too extensive to include in full in this manuscript; however, tables summarizing the information are included, along with each of the full survey responses, on the PAD2020 Web site (http://pad2020.org) [16]. Table 2 also provides a summary of several key characteristics for each measure.

3.2.1. Automated Cognitive Test—The Automated Cognitive Test (ACT; Specialty Automated, Inc., Boca Raton, FL) comprises the Automatic Trail Making Tests (ATMTs; ATMT Part A, ATMT Part B), Automatic Clock Drawing Test, and Automatic Military

Clock Drawing Test. Designed to assess the same aspects of cognition as their paper-and-pencil counterparts, the ACT components offer a computer-based method of measuring attention, executive functioning, calculation, impulsivity, memory, visual attention, and visuospatial skills.

Primarily, the ACT has been used as a screening measure for U.S. military personnel and for older adults seeking to retain their driver's licenses. This instrument is administered exclusively on the computer, can be self-administered, and does not require a lengthy examination time, making it useful in the clinic or laboratory. In fact, it can also be used for at-home monitoring through the Internet or with a DVD version. We were not able to obtain any data regarding the sensitivity and specificity of these instruments for the identification of individuals with MCI or AD.

3.2.2. Automated Neuropsychological Assessment Metrics—The Automated Neuropsychological Assessment Metrics v4.0 (ANAM; University of Oklahoma, Norman, OK) is the culmination of extensive United States Department of Defense-sponsored research focused on automating well-established laboratory and clinical tests of neuropsychological functioning. The General Neuropsychological Screening battery tests psychomotor speed, psychomotor efficiency, attention, working memory, visuospatial memory, delayed memory, verbal reasoning, spatial processing, and executive functioning.

The length of ANAM batteries ranges from between 20 and 25 minutes for a traumatic brain injury battery to 45 minutes for a General Neuropsychological Screening battery. The flexibility to select and "build" a battery composed of ANAM measures offers clinicians and researchers the opportunity to streamline the length of the battery. Levinson et al report 93.8% correct classification between healthy controls and patients diagnosed with AD using a discriminant function analysis [26].

ANAM has been used for cognitive assessment within several governmental agencies, universities, and research institutions for screening, neuropsychological testing, and assessment of military personnel with regard to return to duty. To a lesser degree, the ANAM battery has also been used by pharmaceutical companies and in sports-related decisions regarding return to play. Normative data exist for college students, predeployment military personnel, high school athletes, college athletes, and a group of healthy individuals ranging in age from young adulthood to late life. Some ANAM tests are available in Spanish, German, and Russian. ANAM has several ongoing projects related to aging, cognitive impairment, and dementia. However, the use of this measure in the aging/dementia field is relatively new and, although it shows great promise, there is a relative paucity of published research.

Despite the limited number of published results using the ANAM with older adults experiencing cognitive decline, MCI, or AD, the extensive history of this measure and its long focus on computerized assessment are certainly strengths. The lack of matched alternate forms may make repeated testing difficult, although the authors report that the tool generates quasi-random sets of stimuli for each testing session, making multiple testing sessions possible.

3.2.3. Central Nervous System Vital Signs Brief Clinical Evaluation Battery—

The Central Nervous System Vital Signs (CNS-VS; CNS Vital Signs, LLC., Morrisville, NC) Brief Clinical Evaluation Battery is designed to adapt several traditional neuropsychological measures into computerized versions that will allow for relatively rapid and efficient administration. The Verbal Memory Recognition and Visual Memory Recognition tests are both adaptations of the Rey Auditory Verbal Learning Test (RAVLT) and the Rey Visual Design Learning Test, respectively. Computer-adapted version of the Finger Tapping Test (psychomotor speed) from the Hal-stead–Reitan battery, Symbol Digit Coding (psychomotor speed), Stroop Test (simple and complex reaction time), Shifting Attention Test (executive function), and Continuous Performance Test (CPT; sustained attention) are also included in this battery. Individual performance scores and domain scores (memory, psychomotor speed, and complex attention) are provided immediately after test completion.

The CNS Vital Signs Brief Clinical Evaluation Battery is self-administered on the computer and requires approximately 30 minutes for completion. This battery is available in more than 50 languages and, although it does not have matched alternate forms, the battery presentation is randomized across all tests, a feature that the authors claim allows for serial administration. A study including 37 patients with MCI, 54 patients with mild dementia, and 89 healthy controls examined the sensitivity and specificity of the CNS Vital Signs Brief Clinical Evaluation Battery [17]. When artificially setting the sensitivity for discriminating patients with MCI from healthy controls to 90%, specificity for identifying the healthy participants ranged between 65% (Complex Attention domain score and Verbal Memory) and 85% (Shifting Attention Test, number correct) [17]. When assessing the measure's sensitivity and specificity for discriminating between healthy controls and demented patients, and again holding sensitivity at 90%, specificity ranged between 50% (Visual Memory) and 81% (Symbol Digit Coding) [17]. A separate study examined the sensitivity and specificity of this battery in a group of 322 patients between the ages of 55 and 94 years at one site and 102 participants at a second site [25]. This study reported sensitivity for the Memory domain for differentiating patients with MCI patients from those with dementia ranging between 54% and 90%, depending on the cutoff standard score. Specificity ranged between 75% and 95%. Sensitivity for the Verbal Memory task ranged between 48% and 85%, with specificity between 74% and 92%. Finally, for the Visual Memory task, sensitivity ranged between 39% and 76%, with specificity between 79% and 96% [25]. The data from the study by Gualtieri and Johnson [17] suggest that the CNS Vital Signs battery is relatively successful at distinguishing between healthy individuals and those diagnosed with dementia, whereas the data reported by Gualtieri and Boyd [25] indicate that this measure lacks the predictive power to discriminate between those with mild impairment (MCI) and those with more advanced cognitive decline (early AD).

Although the CNS Vital Signs Brief Clinical Evaluation Battery has been used in clinical practice and research with a variety of clinical disorders, the extent to which it has been used in studies on dementia is relatively restricted. Certainly, the ease of use, minimal need for human administrators, and potential for Internet-based remote administration offer several strengths, particularly with regard to its utility in a large clinical trial. However, with its limited psychometric data demonstrating sufficiently high sensitivity for detecting early

cognitive decline and a lack of sufficient well-matched alternate forms, the CNS Vital Signs Brief Evaluation Battery may not have enough evidence supporting its utility in a large longitudinal aging trial.

3.2.4. Cognitive Drug Research Computerized Assessment System for Dementia—The Cognitive Drug Research Computerized Assessment System for Dementia (COGDRAS-D; United BioSource Corp., Bethesda, MD) is based on more than 25 years of development and pharmaceutical industry experience, and it incorporates several computer-based measures to comprehensively assess cognition. Attention, concentration, and vigilance are measured by a simple reaction time task, choice reaction time task, and digit vigilance task (attention for target digit presented within a string of digits). Verbal and visuospatial working memory and executive control are assessed with spatial and numerical working memory tasks. Verbal and visual episodic and declarative memory are measured with word presentation, immediate word recall, delayed word recall, word recognition, picture presentation, and picture recognition tasks.

This battery requires approximately 20 to 25 minutes for completion and is available in more than 50 languages. At least 30 matched parallel forms exist for all of the battery tests; however, more alternate forms may be available depending on the test and language. Although sensitivity and specificity data were not provided, the authors did provide data indicating that the COGDRAS-D is capable of discriminating between minimal cognitive impairment and the "worried well," namely individuals with no current changes in daily living but who do present to clinic with subjective complaints of memory change or impairment [18,27], and dementia with Lewy bodies versus Parkinson's disease dementia versus AD dementia versus vascular dementia [28–32]. The COGDRAS-D is currently being used in several longitudinal studies on aging, as well as in a host of other large research studies and clinical trials. The extensive normative data and use in assessment for a variety of clinical disorders, including disorders of age-related changes, make the COGDRAS-D a relatively unique tool. In addition, benefits of the COGDRAS-D include the potential for Internet-based testing, ease of use, and long history of acceptance with the Food and Drug Administration and other regulatory agencies.

3.2.5. CogState MCI/AD battery—The CogState battery (CogState, Ltd., Melbourne, Victoria, Australia) was designed with an emphasis on creating cross-culturally valid tasks requiring minimal verbal instructions and responses. The use of a computer interface and minimal human assistance needed for administration make this battery generally accepted by participant populations of varied clinical domains, cultures, and ages. The CogState MCI/AD Battery incorporates the following tasks: detection task (simple reaction time), identification task (choice reaction time), one-back task (working memory), one card learning task (visual learning), Groton Maze Learning task (executive function), continuous paired associate learning task (visual associate learning), and the international shopping list task (verbal learning and memory).

The CogState MCI/AD Battery requires approximately 20 to 25 minutes for completion and is available in 71 languages. Owing to the nature of the task designs, there are an extensive number of well-matched parallel forms for each of the tasks because of large stimulus set

sizes. Sensitivity for discriminating participants with mild memory decline from healthy control participants, over several repeated assessments to monitor for subtle change, was 94%, and specificity for this same group was 100% [33]. More recently, a longitudinal study of community-dwelling adults (aged, 50 years) conducted using serial evaluations found that any decline on the CogState visual learning accuracy measure, over a 12-month period, is highly associated with cerebral amyloid accumulation on amyloid PET imaging [19].

Each of the CogState battery measures has reasonable face validity, but it should be noted that a recent factor analytic study in older adults found that, collectively, these tasks load onto two factors—"learning efficiency" and "problem solving/accuracy"[34]. The CogState MCI/AD Battery has several strengths, including ease of use, minimal human supervision for administration, data entry or scoring, a relatively strong publication record in aging research, and excellent psychometric properties (e.g., limited interference by practice effects). In addition, CogState also offers Internet-based administration, allowing for remote testing, and it is the only measure that we reviewed that has the current capability for real-time quality assurance and data reporting. This aspect of the CogState data collection system may make this measure uniquely suited for the challenges of data collection in large studies for which these endpoints might be chosen for either active monitoring by data safety monitoring boards or remote monitoring of cognitive changes in the home setting.

3.2.6. Community Screening Instrument for Dementia—Developed in response to a need for a cognitive measure capable of assessing dementia symptoms in non-English speakers with little formal education in Canada, the Community Screening Instrument for Dementia (CSI-D; Indiana University School of Medicine, Indianapolis, IN) has to date been used in incidence and prevalence studies around the world. Using an interview-style format, the CSI-D briefly assesses short-term memory, abstract thinking, language, language comprehension, motor/visuomotor control, calculation, and orientation. In addition, the CSI-D also includes an interview with a close family member and collection of medical history; blood pressure, height, and weight measurements; and lifestyle factors.

The CSI-D has been translated into 30 languages. The total CSI-D assessment requires approximately 55 minutes for completion, with 15 minutes for the cognitive assessment, 10 minutes for the relative interview, and 30 minutes for medical examination procedures. Data for sensitivity and specificity related to MCI were not provided; however, data for the area under the receiver operating curve (AUC) were available for separating healthy elderly from those with dementia, with the AUC ranging between 0.74 and 0.93 across five research sites [35].

3.2.7. Computer-Administered Neuropsychological Screen for MCI—The Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment (CANS-MCI; Screen, Inc., Seattle, WA) was developed specifically to meet the needs of primary care physicians, and it includes an assessment of cognition, mood, health history and risk factors, substance use, and driving capabilities. The assessment of cognition includes measures of free and guided recall, delayed free and guided recognition, primed picture naming, word-to-picture matching, design matching, clock hand placement, and the Stroop Test.

For individuals with a high school education or less, the CANS-MCI showed sensitivity of 100% and specificity of 100%, indicating that it correctly identified all participants as either meeting criteria for MCI or as a healthy control. For individuals with 13 or more years of education, the CANS-MCI showed sensitivity of 100% and specificity of 84.8%, with an AUC of 0.96 [36].

The CANS-MCI battery requires 22 to 27 minutes for completion. Reports are completed by neuropsychology test technicians. The CANS-MCI does not have matched alternate forms, although there are different pictures used in the naming test for patients who are drivers versus non-drivers. The CANS-MCI is country- and language-specific and is available for the United States (English and Spanish), Canada (English and French), United Kingdom (English), Brazil (Portuguese), The Netherlands, Belgium, and Argentina (Spanish). Images, text, and audio are designed specifically for each country. Overall, the CANS-MCI has several strengths, including brevity, language and cultural specificity, and good evidence of sensitivity and specificity for MCI versus healthy controls. However, the lack of matched alternative forms, lack of automated quality assurance/quality checking, lack of unmodifiable audit trails, as well as a paucity of previous use and experience in drug development research all raise concerns about the use of this measure in either a large clinical trial or naturalistic longitudinal epidemiologic trial.

3.2.8. Computer Assessment of MCI—The Computer Assessment of Mild Cognitive Impairment (CAMCI; Psychology Software Tools, Inc., Pittsburgh, PA) is a battery of tests intended for use in conjunction with other neuropsychological tools. The measures in the CAMCI include paper-and-pencil tests modified for computer presentation and a uniquely developed virtual reality task [20]. The CAMCI is currently only available in English and is intended for English-speaking populations in the United States and Canada. Presently, there are no matched alternate forms; however, the Psychology Software Tools Corporation recently received a grant to fund the development of other forms. In a study comparing the CAMCI with the Mini-Mental State Examination, the CAMCI was found to show good sensitivity (86%) and specificity (94%) for the identification of putative MCI within the sample [20]. The CAMCI's ease of administration, sensitivity, and specificity for MCI, and unique incorporation of a virtual reality assessment measure designed for "real-world" simulation are highlights of this battery. However, the requirement that the entire battery must be completed for data reporting, lack of alternate parallel forms, and need for interpretation by a qualified healthcare provider (i.e., inability to be used over the Internet or for at-home monitoring), as well as the continued development and refinement that the CAMCI is currently undergoing, suggest that this measure may not yet be ideal for a large national trial requiring frequent reassessment.

3.2.9. Computer Self Test—The Computer Self Test (CST; University of Tennessee Medical Center, Knoxville, TN) was recently adapted from a paper-and-pencil version of the same battery, and although it is still being validated as a battery for the assessment of early cognitive change, it does show promise as a tool for future use in research and clinical trials. The CST requires about 10 minutes for completion, and it includes the following computer-based tests: Clock Drawing test (visuospatial and executive processing), memory for three

words (working memory), categorical fluency (verbal fluency), identification of the 5 months before December in reverse order (attention), and orientation to time. Each of the subtests is timed and participants are instructed to work as quickly as possible; this information is used as a proxy for processing speed and may be used for comparison with control samples and for statistical comparison. Of note, many of these tasks do not require that the participant generate the full response by himself/herself. Rather, participants are either given multiple-choice options or the computer automatically generates the response after the participant enters in a minimum amount of information (e.g., after typing the first three letters of a word, the remainder of the word is automatically completed).

Information regarding other available test languages was not provided; however, a review of the literature suggests that this measure is only available in English. To date, the CST has only been tested with patients with memory impairment associated with dementia. The CST has demonstrated a sensitivity of 96% for patients with cognitive impairment (collapsing across the range of early-to-severe AD) and was capable of correctly classifying 91% of patients into one of six experimental groups (control, MCI, early AD, mild-moderate AD, moderate-severe AD, and severe AD) [21]. Without further substantial literature to support its utility in identifying early cognitive changes, it is only possible to qualify the CST's strengths in administration ease and Internet accessibility.

3.2.10. Florida Brief Memory Scale—The Florida Brief Memory Scale (FBMS; University of Miami Miller School of Medicine, Miami Beach, FL) comprises the Semantic Interference Test, a task that includes three trials of the Fuld Object Memory Evaluation and distracter trials, plus a bag of 10 targets that are semantically related to the 10 objects of the Object Memory Evaluation. The battery also assesses verbal memory, using the immediate and delayed recall of passages and designs from the Wechsler Memory Scale (WMS, using the Mayo Older Americans Normative Studies [MOANS] normative data tables), and verbal fluency using the category fluency (animals and vegetables) and letter fluency tasks. Visuospatial functioning is assessed using Block design from the Wechsler Adult Intelligence Scale - III (WAIS-III), and attention and executive functioning are evaluated using the Stroop Color Word Test and Trail Making Test (parts A and B). On occasion, the BNT, the Conner's CPT-2, and the Wisconsin Card Sorting Task are also used to assess language, attention, and executive functioning, respectively.

The Florida Brief Memory Screen has been used in the longitudinal assessment of older adults living in the community as well as of patients recruited from an Alzheimer's Disease Research Center [22]. The authors for the FBMS did not provide responses to all of the questions included in the survey, including time for completion of the measure, sensitivity and specificity for the identification of MCI, and alternative test languages. Therefore, without much of these data, it is difficult to fully assess the utility of the FBMS as a cognitive measure appropriate for use in a large clinical trial. However, considering the psychometric properties and administration demands of many of the measures included in the FBMS, such as the WMS, WAIS-III, BNT, and Wisconsin Card Sorting Task, we can assume that, as a battery, the FBMS lacks the potential for frequent repeated use.

3.2.11. GrayMatters Assessment System—The GrayMatters Assessment System (Dementia Screening, Inc., Abilene, TX) was designed for use in outpatient medical practices. As such, it was designed to be self-administered, self-scoring, and "user-friendly" for older adults. Instructions are presented on the computer screen as well as through the speakers, and patients respond on a touch screen. On test completion, data are encrypted and automatically entered into a multivariate prediction equation, with the report providing a score reflective of the patient's probability of MCI and associated recommendations.

The GrayMatters assessment includes a Delayed Alternation Task, to assess executive functioning, and a Visual Delayed Recognition Task (VDR), to assess visual memory. The Delayed Alternation Task utilizes a set-switching paradigm, requiring participants to deduce the "rules" of the task based on the feedback that they receive. The VDR includes an initial presentation of pictures of objects. Patients are then required to correctly identify which pictures they have previously seen.

In a sample of 419 participants, the GrayMatters Assessment System correctly identified 91% of patients with early AD (Mini-Mental State Examination score, <25) and 80% of unimpaired participants (unpublished). Administration requires approximately 20 minutes, although some patients may take more than 30 minutes. Currently, one alternate form for the VDR is available; however, other alternate forms are presently being standardized. The GrayMatters Assessment System is available in Spanish, although this version of the system is not presently standardized. The GrayMatters Assessment System has been tested and normed using older adults determined to be cognitively "healthy" based on normal performance on the WMS-III, Trail Making Tests A and B, and Boston Verbal Fluency Test. Because it was created for use in physician offices, validation studies have been completed with community-dwelling older adults, with just a handful of publications available at the present time.

3.2.12. IntegNeuro and WebNeuro—IntegNeuro (Integneuro, Inc., New York, NY) is administered by computer touchscreen interface and, along with its Web-formatted version, WebNeuro, measures a variety of cognitive and emotional tasks, including RAVLT (verbal memory recall and recognition), California Verbal Learning Test (verbal memory recall and recognition), digit span (forward and reverse; auditory memory), Corsi Blocks (visual memory), Conner's CPT (sustained attention), Test of Variables of Attention (sustained attention), N-back (sustained attention), Trail Making Test A and B (attention switching), finger tapping (psychomotor functioning), simple decision-making tasks (choice reaction time), Stroop Task (verbal interference), Spot the Real Word (reading/verbal skills), Controlled Oral Word Test (word generation), Verbal Fluency (word generation), Austin Maze (executive functioning), Go/No-go tasks (inhibition, executive functioning), Penn Emotion Test (explicit emotion identification), and an implicit emotion recognition task. The IntegNeuro/WebNeuro battery takes 40 to 50 minutes for completion and is available with up to 30 parallel forms, depending on the specific test and language. This battery is available in eight language versions, including American English, Australian English, U.S. Spanish, French, Dutch, Hebrew, German, and Chinese.

IntegNeuro has been used with several clinical populations, including MCI and AD, schizophrenia, individuals with anorexia, Attention Deficit Hyperactivity Disorder (ADHD), and individuals with sleep apnea. Considerable research has also examined the correlation between performance on the IntegNeuro/WebNeuro system and measures of brain function (e.g., electroencephalography, event-related potentials, and MRI). Overall, the IntegNeuro/ Web-Neuro system has several benefits, including assessment of numerous domains, availability of parallel forms for repeated assessment, and clinical applicability in the identification of individuals with cognitive decline related to dementia. However, without sensitivity and specificity data, it is difficult to address the utility of this measure in accurately identifying early dementia.

3.2.13. MCI Screen—The MCIS (Medical Care Corp., Irvine, CA) is essentially a novel scoring approach for the CERAD Word-List Learning Test that maintains the core structure of a 10-item list-learning task. The MCIS aims to increase the utility of this well-known measure of episodic memory by standardizing administration using a computer interface, creating nine parallel versions of the word list, using a static instead of shuffled word presentation order, adjusting for test anxiety using a distracter task, and adjusting for covariates such as age, education, and gender. The MCIS is essentially a basic word-list learning task (three immediate recall trials, one delayed recall trial, and one recognition trial) assessing attention and working memory. The MCIS is administered by a human rater and scored using an online interface that guides the administrator through testing. It also includes statistical analysis of the data and provides an automatic report summarizing the participant's performance for three domains: (1) overall impression (normal, impaired, borderline), (2) Memory Performance Index (map of the participant's scores from the overall impressions domain on a scale of range 0–100), and (3) longitudinal report (table summarizing performance in attention, working memory, meta-memory, delayed free recall, delayed recognition, and unrehearsed delayed free recall).

The MCIS requires approximately 10 minutes for completion and is available in English, Spanish, and Japanese, although the Spanish translation has not yet been fully validated. Because the MCIS is a list-learning task, patients with hearing or speech impairments may not be well suited for this measure, although there is evidence that patients with hearing loss can be assessed using flash cards to aid in the learning process [37]. The MCIS has demonstrated good sensitivity and specificity. Sensitivity for identifying MCI as compared with healthy controls is 95% and specificity 88%. Sensitivity for identifying MCI/mild dementia as compared with healthy controls was 97% and specificity was 88%. Finally, sensitivity for mild dementia as compared with healthy controls was 96% and specificity was 99% [38].

The MCIS is a useful tool for identifying early cognitive change associated with dementia, benefiting from the extensive validation of the word-list learning task for this population. The administrative standardization, capability for either in-person or telephone administration, automatic data scoring and report generation, and Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and the Food and Drug Administration compliance are additional strengths for this measure. Of potential concern

may be the MCIS's lack of assessment of other cognitive domains, outside of verbal attention and both immediate and delayed verbal memory.

3.2.14. Mindstreams—The Mindstreams computerized battery (NeuroTrax, Inc., Newark, NJ) takes a standard neuropsychological approach to assessment across cognitive domains, relying on the following tasks: finger tapping (psychomotor skills), verbal memory (memory), nonverbal memory (memory), Go/No-go (attention, executive function), Stroop Interference test (executive function), visuospatial processing, verbal naming (verbal function), rhyming, staged information processing (information processing speed), problem solving (abstract reasoning), and the Catch game (attention, executive function). The Mindstreams MCI Battery includes these same measures, with the omission of the problem solving, finger tapping, and staged information processing speed tasks. The report that accompanies Mindstream test completion includes detailed scores, graphs, and summaries highlighting test performance. Study data are also provided in an Excel export, simplifying the data preparation phase for research and clinical trial use.

The Mindstreams Global Assessment Battery is approximately 45 to 60 minutes in length, whereas the MCI Battery requires approximately 30 minutes to administer. There are three alternate forms of the Mindstreams tests and these are available in English, Spanish, Russian, and Hebrew. Mindstreams has been used with several populations outside of aging and dementia, including, but not limited to, ADHD, people with high cardiovascular risk, postmenopausal women, and persons with schizophrenia.

The verbal and nonverbal memory tests from the Mindstreams battery discriminate between healthy elderly and MCI, with an AUC of 0.7 to 0.9 [23]. The verbal memory test best distinguished healthy elderly from MCI (Cohen's d =1.23) and from mild dementia (Cohen's d =0.60). However, the nonverbal memory test best discriminated MCI converters to dementia from nonconverters (Cohen's d = 0.54). The Mindstreams Global Assessment Battery memory, executive function, visuospatial, and verbal indices, combined with subjective cognitive complaint and Lawton IADL score at baseline, accurately classified 89% of individuals with MCI as either converters to dementia or non-converters (sensitivity = 79%, specificity = 95%) [39].

The strengths behind Mindstreams, including administration, inclusion of a variety of cognitive variables, history of involvement in clinical trials, and fairly strong discriminability, make it a useful tool for research and clinical use. However, the limited number of alternative forms and required presence of a trained administrator may limit its utility within very large clinical or epidemiologic studies.

3.2.15. Neuropsychological Test Battery—The NTB (CogState, Ltd., Melbourne, Victoria, Australia) [24] comprises standard neuropsychological measures, including: the Controlled Oral Word Association Test (executive functioning), Category Fluency Test (executive functioning), Digit Span from WMS-R (working memory), Visual Paired Associates Test from WMS-R (immediate and delayed visual memory), Verbal Paired Associates Test from WMS-R (immediate and delayed verbal memory), and RAVLT (episodic memory).

Despite the considerable use of these measures, and inclusion of them separately in studies of early cognitive change, there have been no published studies assessing sensitivity and specificity for MCI for this specific battery. The dependence of this battery on paper-and-pencil measures, the need for a trained clinical psychologist or psychometrician, as well as an administration time of 50 to 75 minutes, all serve to make this battery somewhat cumbersome to incorporate into large clinical trials. The lack of alternate parallel forms also complicates repeated assessment. The NTB has been translated into 18 languages and has been used successfully in several clinical trials. Ultimately, despite the extensive normative data and recent history of use in clinical trials, as well as some measure of acceptance by regulatory authorities, the administration requirements of this battery and lack of specific utility in the identification of early cognitive decline do not support its use in very large epidemiologic or registry trials.

3.2.16. National Institutes of Health Toolbox—The National Institutes of Health (NIH) Toolbox Assessment of Neurological and Behavioral Function (with cognitive, emotion, sensory, and motor components) is currently under development, sponsored by the NIH, and there are no published normative data on this battery specifically for older or cognitively impaired populations. The measures included in the NIH Toolbox Cognitive Battery are list sorting (working memory), dimensional change card sort (executive functioning), flanker task (visual attention and executive inhibition), picture sequence memory (episodic memory), pattern comparison (processing speed), and vocabulary comprehension (auditory single word comprehension).

The NIH Toolbox Cognitive Battery (Feinberg School of Medicine, Northwestern University, Chicago, IL) was designed to cover the full range of normal ability from 3 to 85 years of age. A study has been completed validating the cognitive battery against neuropsychological "gold standards" in a sample of 476 individuals in this age range [40]. Each subtest shows good convergent and divergent construct validity, test-retest reliability, and expected influences of age and education. However, despite its currently limited empirical support, there are several design aspects that may make this set of tools quite useful for the identification of early cognitive change.

The version of the NIH Toolbox cognitive component being used in ongoing normative studies is expected to require 45 minutes to complete. However, following this phase of research, it is expected that some measures will be removed, thus resulting in a desired total administration time of 30 to 35 minutes for the adult version. The NIH Toolbox is also presently being translated into Spanish, but it will not include alternative parallel forms. Administration requires the presence of an examiner to ensure compliance with test instructions and continued attention to the computer. The computer presents instructions visually and orally.

3.3. Decision tree diagrams

The decision trees in Figs. 2 and 3 visually summarize how the 16 tests/batteries compare with respect to several selected key differentiating features, and will aid in decision making as and when future tests are developed and validated in years to come. These decision trees

are parallel in design, with one diagram comparing computer-based measures (Fig. 2) and the other comparing paper/pencil/other measures (Fig. 3). For a complete description of the information included within the decision tree diagrams, please see the summary tables along with the individual survey responses on the PAD2020 Web site (http://www.pad2020.org) [16].

4. Discussion

This article provides a comprehensive review and comparison of the current state of the majority of commercially available measures that have been promoted as being sensitive and appropriate for the assessment of early cognitive decline associated with dementia. It is based on a structured survey returned by 16 of the 32 authors/publishers contacted and a thorough review of the existing literature on these tests and batteries. The survey was driven by the goal of identifying measures capable of assessing cognitive changes before noticeable decline suggestive of MCI or early AD. A strength of this survey is the use of self-report by the scientists associated with the development of these measures; using this approach allowed them to provide a thorough review of their measures (including the provision of information that may not be published but is certainly relevant to assessing the utility of a given instrument). In addition, the survey's focus on detailed properties of each measure, normative data, psychometric properties, and administrative information yielded practical information that influences test selection for individual study purposes. This unique approach involving self-report data collection, combined with the careful review process that accompanied each survey response and the full transparency of all source documents from each author/vendor, distinguishes this effort from other similar previous attempts to survey this content area [15,41].

There are several limitations to the scope of this review. First, some measures were not included because of a nonresponse to the survey. Second, this survey focused exclusively on measures of cognition and not on measures of IADLs or other aspects of functioning (e.g., psychological symptoms, occupational function) that typically undergo concomitant changes in the preclinical and early stages of AD. In addition, the measures included in this review were limited to performance-based tasks and did not include self-reported or other reported subjective rating scales or questionnaires. Although not included in this review, measures such as the informant-dependent questionnaire Everyday Cognition [42] offer important information for predicting cognitive impairment and certainly complement the data provided by the tests included in this review. Finally, there are additional considerations related to selecting any battery that have not been extensively covered in this review, mostly with regard to the proprietary nature of these instruments. For example, proprietary tests currently present a challenge to the research community for several reasons, including license fees and other related costs, inability to modify test paradigms for specific research questions because of unavailability of source code, and nontransparent scoring algorithms that may lead to difficulty in replicating previous findings.

One surprising finding, evident to the authors as we compiled the information obtained for this review, is the real paucity of cross-sectional and/or longitudinal studies to directly compare performance on a given cognitive test/battery with changes in known disease-

related biomarkers (e.g., structural MRI, amyloid imaging PET, cerebrospinal fluid measures of Aβ). However, several survey responders indicated interest in future research with such biomarkers, and this would certainly be highly warranted, considering the rapidly growing roles that these markers will have in the diagnosis and tracking of AD progression for individual patients [43].

Overall, the measures available for early detection of cognitive decline may vary on several critical points, such as computer versus paper-and-pencil administration, the need for specialized training for test administration, general ease of administration, the number of tests and/or aspects of cognition assessed, and the speed with which data are quality checked and made available to the investigator or clinician. However, for each of these fundamental differences, there are several similarities among measures, particularly for those measures designed for, and most often used in, large research studies and clinical trials. Hence, there is no single recommended "gold standard" battery but, rather, a subset of very good products to choose from, based on individual study needs (Figs. 2 and 3).

Ultimately, the information provided in this review is rather time-limited, in that these measures and the literature accompanying them will most definitely change in the coming years. Furthermore, new cognitive measures that are less multimodal and more tightly bound to the underlying disease-related functional neuropathology of AD are already being developed [44] and may lead to more precise tracking of very early cognitive changes in the initial stages of the disease. However, the actual questions included in this survey, and provided in Appendix, will remain germane even when the current information contained in this review becomes outdated. Thus, the benefit of this review, and the associated decision tree diagrams, is that new measures can be benchmarked and compared on the basis of the very same criteria (provided in Appendix) with those tests included in this review as decisions are made for future clinical research and practice.

Acknowledgments

References

1. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadian EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology. 1984; 34:939–44. [PubMed: 6610841]

2. Khachaturian ZS. Diagnosis of Alzheimer's disease. Arch Neurol. 1985; 42:1097–105. [PubMed: 2864910]

3. Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010; 9:119–28. [PubMed: 20083042]

4. Jack CR Jr, Wiste HJ, Vemuri P, Weigand SD, Senjem ML, Zeng G, et al. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. Brain. 2010; 133:3336–48. [PubMed: 20935035]

5. Lopez OL, Schwam E, Cummings J, Gauthier S, Jones R, Wilkinson D, et al. Predicting cognitive decline in Alzheimer's disease: an integrated analysis. Alzheimers Dement. 2010; 6:431–9. [PubMed: 21044773]

6. Jack CR Jr, Petersen RC, Grundman M, Jin S, Gamst A, Ward CP, et al. Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI. Neurobiol Aging. 2008; 29:1285–95. [PubMed: 17452062]

7. Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med. 2005; 352:2379–88. [PubMed: 15829527]

8. Cummings JL, Doody R, Clark C. Disease-modifying therapies for Alzheimer disease: challenges to early intervention. Neurology. 2007; 69:1622–34. [PubMed: 17938373]

9. Evans DA, Grodstein F, Loewenstein D, Kaye J, Weintraub S. Reducing case ascertainment costs in US population studies of Alzheimer's disease, dementia, and cognitive impairment. Alzheimers Dement. 2011; 7:94–109. [PubMed: 21255747]

10. Harrison J, Maruff P. Measuring the mind: assessing cognitive change in clinical drug trials. Expet Rev Clin Pharmacol. 2008; 1:471–3.

11. Nestor PJ, Scheltens P, Hodges JR. Advances in the early detection of Alzheimer's disease. Nat Med. 2004; 10:S34–41. [PubMed: 15298007]

12. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol. 1999; 56:303–8. [PubMed: 10190820]

13. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr. 2009; 21:672–87. [PubMed: 19470201]

14. Kaldi A, Ellis KA, Darby D, Ames D, Martins R, Masters CL, et al. Determining rate of decline in pre-AD and AD AIBL participants using a brief computerized battery. Alzheimers Dement. 2010; 6:S341–2.

15. Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimers Dement. 2008; 4:428–37. [PubMed: 19012868]

16. PAD2020.org [Internet]. Rockville, MD: PAD2020; 2011. [updated April 7, 2011]. Available at: http://www.pad2020.org

17. Gualtieri CT, Johnson LG. Neurocognitive testing supports a broader concept of mild cognitive impairment. Am J Alzheimers Dis Other Dement. 2005; 20:359–66.

18. Nicholl CG, Lynch S, Kelly CA, White L, Simpson PM, Wesnes KA, et al. The cognitive drug research computerized assessment system in the evaluation of early dementia—is speed of the essence? Int J Ger Psychiatry. 1995; 10:199–206.

19. Darby D, Pietrzak RH, Fredrickson J, Woodward M, Moore L, Fredrickson A, Sach J, Maruff P. Intra-individual cognitive decline using a brief computerized cognitive screening test. Alzheimers Dement. 2011 in press.

20. Saxton J, Morrow L, Eschman A, Archer G, Luther J, Zuccolott A. Computer assessment of mild cognitive impairment. Postgrad Med. 2009; 121:177–85. [PubMed: 19332976]

21. Dougherty JH, Cannon RL, Nicholas CR, Hall L, Hare F, Carr E, et al. The Computerized Self Test: an interactive, Internet accessible cognitive screening test for dementia. J Alzheimers Dis. 2010; 20:185–95. [PubMed: 20164591]

22. Loewenstein DA, Acevedo A, Small BJ, Agron J, Crocco E, Duara R. Stability of different subtypes of mild cognitive impairment among the elderly over a 2- to 3-year follow-up period. Dement Geriatr Cogn Disord. 2009; 27:418–23. [PubMed: 19365121]

23. Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, Chertkow H. Validity of a novel computerized cognitive battery for mild cognitive impairment. BMC Geriatr. 2003; 3:4. [PubMed: 14594456]

24. Harrision J, Minassian SL, Jenkins L, Black RS, Koller M, Grundman M. A neuropsychological test battery for use in Alzheimer disease clinical trials. Arch Neurol. 2007; 64:1323–9. [PubMed: 17846273]

25. Gualtieri, CT.; Boyd, AF. Sensitivity and specificity of computerized testing for MCI. International Congress for Alzheimer's Disease; 2008; Chicago, IL.

26. Levinson D, Reeves D, Watson J, Harrison M. Automated neuropsychological assessment metrics (ANAM) measures of cognitive effects of Alzheimer's disease. Arch Clin Neuropsychol. 2005; 20:403–8. [PubMed: 15797175]

27. Wesnes, K.; Brooker, H.; Newhouse, P.; Levin, E.; White, H.; Coderre, E., et al. The profile of cognitive impairment in MCI. International Society for CNS Clinical Trials and Methodology; 2009; San Diego, CA.

28. Ballard CG, O'Brien JT, Gray A, Cormack F, Ayre G, Rowan EN, et al. Attention and fluctuating attention in patients with dementia with Lewy bodies and Alzheimer disease. Arch Neurol. 2001; 58:977–82. [PubMed: 11405813]

29. Ballard CG, Walker M, O'Brien JT, Rowan EN, McKeith IG. The characterization and impact of 'fluctuating' cognition in dementia with Lewy bodies and Alzheimer's disease. Int J Geriatr Psychiatry. 2001; 16:494–8. [PubMed: 11376465]

30. Walker MP, Ayre GA, Ashton CH, Marsh VR, Wesnes KA, Perry EK, et al. A psychophysiological investigation of fluctuating consciousness in neurodegenerative dementias. Hum Psychopharm Clin. 1999; 14:483–9.

31. Walker MP, Ayre GA, Cummings JL, Wesnes KA, McKeith IG, O'Brien JT, et al. Quantifying fluctuation in dementia with Lewy bodies, Alzheimer's disease, and vascular dementia. Neurology. 2000; 54:1616–25. [PubMed: 10762503]

32. Walker MP, Ayre GA, Perry EK, Wesnes KA, McKeith IG, Tovee M, et al. Quantification and characterization of fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease. Dement Geriatr Cogn Disord. 2000; 11:327–35. [PubMed: 11044778]

33. Darby, D.; Maruff, P.; Collie, A.; Weaver-Cargin, J.; Masters, C.; Currie, J. Reliable detection of asymptomatic longitudinal cognitive decline in healthy community dwelling volunteers. American Geriatrics Society Meeting; 2004; Las Vegas, NV.

34. Pietrzak RH, Maruff P, Mayes LC, Roman SA, Sosa JA, Snyder PJ. An examination of the construct validity and factor structure of the Groton Maze Learning Test, a new measure of spatial working memory, learning efficiency, and error monitoring. Arch Clin Neuropsychol. 2008; 23:433–45. [PubMed: 18448309]

35. Hall KS, Gao S, Emsley CL, Ogunniyi AO, Morgan O, Hendrie HC. Community Screening Interview for Dementia (CSI 'D'); performance in five disparate study sites. Int J Geriatr Psychiatry. 2000; 15:521–31. [PubMed: 10861918]

36. Tornatore, JB.; Hill, E.; Laboff, JA.; Fogel, B. Preliminary screening for Mild Cognitive Impairment (MCI): using the CANS-MCI in primary care to determine need for imaging. International Conference of Alzheimer's Disease; 2004; Philadelphia, PA.

37. Brand N, Jolles J. Learning and retrieval rate of words presented auditorily and visually. J Gen Psychol. 1985; 112:201–10. [PubMed: 4056765]

38. Shankle WR, Romney AK, Hara J, Fortier D, Dick M, Chen J, et al. Method to improve the detection of mild cognitive impairment. PNAS. 2005; 102:4919–24. [PubMed: 15781874]

39. Simon, ES.; Goldstein, FC.; Lah, JJ.; Doniger, GM.; Levey, AI.; Dwolatzky, T. Computerized cognitive assessment predicts conversion to dementia within two years: Interim analysis of a prospective study of MCI. International Conference on the Prevention of Dementia; 2007; Washington, DC.

40. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. NIH Toolbox for the Assessment of Behavioral and Neurological Function: cognition domain instruments. Neurology. (in press).

41. Tierney MC, Lerner MA. Computerized cognitive assessment in primary care to identify patients with suspected cognitive impairment. J Alzheimers Dis. 20:823–832. [PubMed: 20413868]

42. Farias ST, Mungas D, Reed BR, Cahn-Weiner D, Jagust W, Baynes K, et al. The measurement of everyday cognition (ECog): scale development and psychometric properties. Neuropsychology. 2008; 22:531–44. [PubMed: 18590364]

43. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011; 7:270–9. [PubMed: 21514249]

44. Festa EK, Heindel WC, Ott BR. Dual-task conditions modulate the efficiency of selective attention mechanisms in Alzheimer's disease. Neuropsychologia. 2010; 48:3252–61. [PubMed: 20621109]

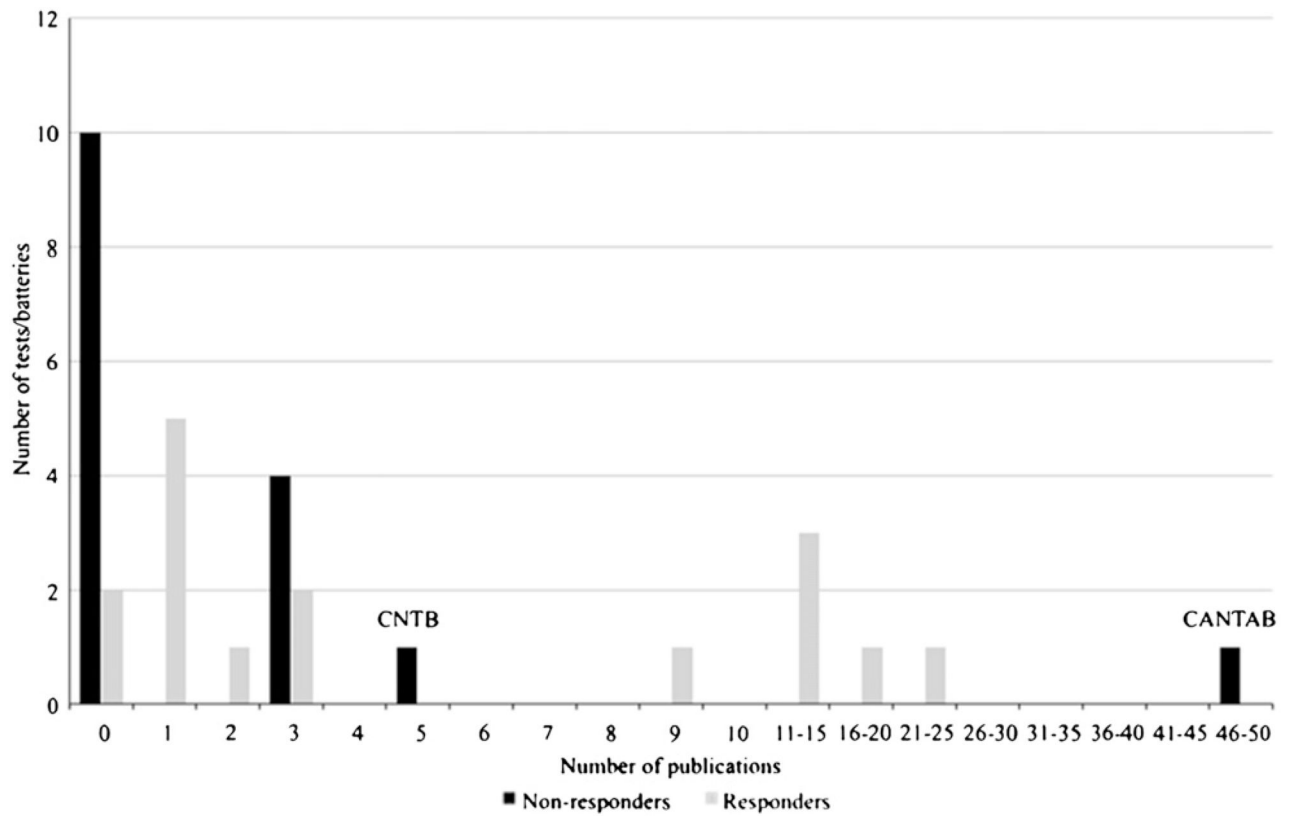**Fig. 1.**
Number of publications pertaining to older adults and transition to mild cognitive impairment for survey responders versus nonresponders.
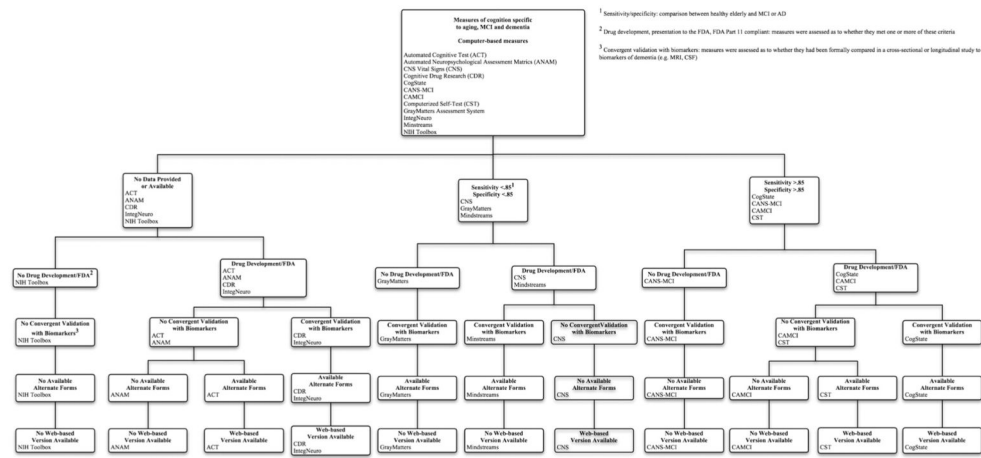
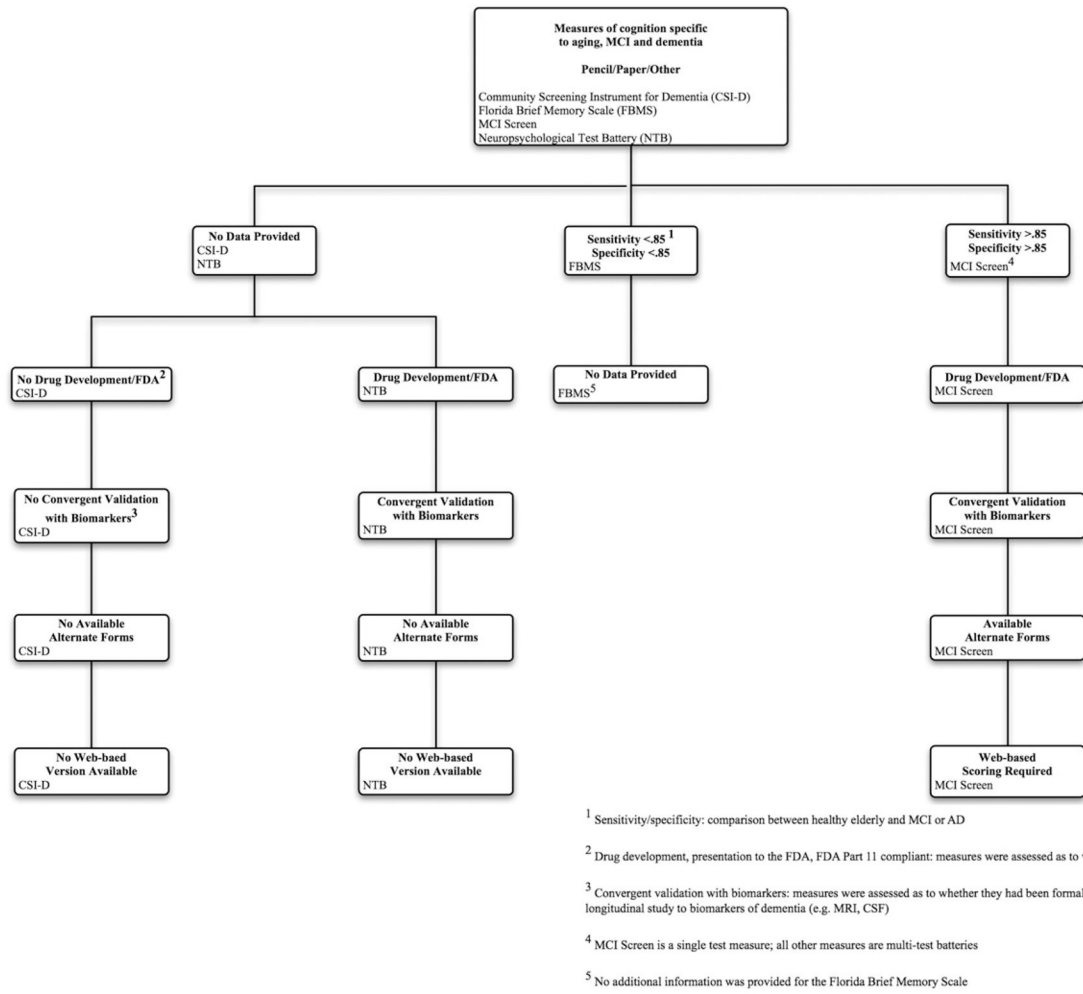**Fig. 2.**
Decision tree diagram for computer-based measures.

Fig. 3.
Decision tree diagram for paper/pencil/other measures.

Table 1

Cognitive tests/batteries selected for review

Completed surveys returned	No responses provided by primary author or contact
ACT (Straus)	ADCS Home-based Assessment Battery (Sano)
ANAM (Gililand)	BACS (Keefe)
CNS Vital Signs Neurocognitive Assessment Battery (Boyd)	CAMDEX-R (Huppert)
COGDRAS-D (Wesnes, Brooker, Edgar)	CANTAB (Blackwell)
CogState, Ltd. (Darby, Maruff, and others)	Cognometer (Buschke)
CSI-D (Hall, Hendrie, and others)	CogScreen (Kay)
CANS-MCI (Hill)	CMINDS (O'Halloran)
CAMCI (Eschman et al)	Computerized Neurocognitive Scan (Gur)
CST (Cannon, Dougherty et al)	CNTB (Cutler and Veroff)
FBMS (Loewenstein et al)	ELSMEM (Storandt)
GrayMatters Assessment System (Brinkman)	Interactive Voice Response (Gelenberg and Mundt)
IntegNeuro/WebNeuro (Fallahpour et al)	MCT (Buschke)
MCI Screen (Fortier)	Microcog (Toomin)
Mindstreams (Doniger et al)	Neurobehavioral Evaluation System (Baker)
NTB (Harrison)	NeurocogFX (Elger)
NIH Toolbox (Nowinski)	NAB (Stern)

Abbreviations: ACT, Automated Cognitive Test; ANAM, Automated Neuropsychological Assessment Metrics; COGDRAS-D, Cognitive Drug Research Computerized Assessment System for Dementia; CSI-D, Community Screening Instrument for Dementia; CANS-MCI, Computer-Administered Neuropsychological Screen for Mild Cognitive Impairment; CAMCI, Computer Assessment of Mild Cognitive Impairment; CST, Computerized Self Test; FBMS, Florida Brief Memory Scale; MCI, Mild Cognitive Impairment; NTB, Neuropsychological Test Battery; BACS, Brief Assessment of Cognition in Schizophrenia; CAMDEX-R, Cambridge Mental Disorders of the Elderly Examination—Revised; CANTAB, Cambridge Neuropsychological Test Automated Battery; CMINDS, Computerized Multiphasic Interactive Neurocognitive DualDisplay System; CNTB, Computerized Neuropsychological Test Battery; MCT, Memory Capacity Test; NAB, Neuropsychological Assessment Battery; ADCS, Alzheimer's Disease Cooperative Study; ELSMEM, Executive, Linguistic, Spatial and Memory Abilities Battery.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Summary of key characteristics for cognitive tests/batteries

	ACT	ANAM	CNS vital signs	COGDRAS-D	CogState	CSI-D	CANS-MCI	CAMCI	CST	FBMS	GrayMatters	Integneuro/Webneuro	MCI Screen	Mindstreams	NTB	NIH Toolbox
Total administration time (minutes)	8–20	45	25–30	20–25	15	55	25–30	20–25	10–15	Not provided	20–30	40	10	15–60	50–75	3–35
Self-administration	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	No
Sensitivity	N/A	93.8% HC vs. AD [17]	54%–90% MCI vs. Demented [18]	N/A	94% HC vs. mild memory decline [19]	N/A	100% HC vs. MCI (12 years education or less) [20] 100% HC vs. MCI (13+ years education) [20]	86% HC vs. MCI [21]	96% HC vs. early-severe AD [22]	Not provided	91% HC vs. early AD (unpublished)	N/A	95% HC vs. MCI [23] 97% HC vs. MCI/mild dementia [23] 96% HC vs. mild dementia [23]	79% Converters vs. nonconverters to dementia (plus subjective CC and IADL score) [24]	N/A	N/A
Specificity	N/A	N/A	65%–85% HC vs. MCI [25] 50%–81% HC vs. Demented [25] 75%–95% MCI vs. Demented [18]	N/A	100% HC vs. mild memory decline [19]	N/A	100% HC vs. MCI (12 years education or less) [20] 84.8% HC vs. MCI (13+ years education) [20]	94% HC vs. MCI [21]	N/A	Not provided	80% HC vs. early AD (unpublished)	N/A	88% HC vs. MCI [23] 88% HC vs. MCI/mild dementia [23] 99% HC vs. mild dementia [23]	95% Converters vs. nonconverters to dementia (plus subjective CC and IADL score) [24]	N/A	N/A
Web-based	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Not provided	No	Yes	Yes*	No	No	No
Drug development[†]	No	Yes	Yes	Yes	Yes	No	No	Yes	No	Not provided	No	Yes	Yes	Yes	Yes	No
FDA or other regulatory body[‡]	No	No	No	Yes	Yes	No	No	No	No	Not provided	No	No	Yes	Yes	Yes	No
FDA part 11 compliance[§]	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Not provided	No	Yes	Yes	Yes	Yes[¶]	Not provided

Abbreviations: HC, healthy control; AD, Alzheimer's disease; cc, cognitive complaint; IADL, instrumental activities of daily living; N/A, not available.

*
Web-based scoring required.

[†]
Previous known use in drug development research.

[‡]
Results from the test/battery have been presented to the FDA or other regulatory body.

[§]
Author/vendor claims full FDA 21 CFR Part 11 compliance.

[¶]
If CogState DataPoint system is used for data storage.

Appendix

Comprehensive survey of cognitive tests/batteries

Section	Topic	Question(s)
Section 1	General description of measure and measure development	Please provide a brief (one page maximum) description/introduction to your test(s) and/or cognitive battery. In this summary, please describe when and how your measures initially conceived and developed. What theoretical and/or experimental models were relied upon in the conceptual development of your product?
		Is your test or battery in the public domain, or is it proprietary. If proprietary, please describe how individual researchers or study sponsors may reach you (or your company) to obtain access to your instrumentation.
Section 2	Normative data	Please specify any important publications that describe normative data (for both healthy controls and early dementia populations, if available) for your test(s) and/or battery. Please provide additional documents, tables, or booklets that will assist in understanding how extensive your normative data are. Both cross-sectional and longitudinal (serial) normative data should be included where available. These may be sent electronically, as supplemental files. Alternatively, you may provide appropriate URL addresses to access these materials online.
Section 3	Sensitivity of tool for measuring cognitive change related to MCI/AD	Please describe how your test, or any specific portions of your battery (whichever is applicable), has been particularly tailored to be sensitive for identifying, and/or tracking change-over-time in, older adults at risk for MCI/early AD. Please provide all references describing studies that support this claim.
		Please provide your most updated listing of references for peer-reviewed journal publications that document use of your test or test battery (or any specific portions of your full battery) in healthy aging and the earliest stages of dementia (pre-MCI the transition to MCI).
		Has your test or battery ever been compared, in a formal prospective study or retrospective review, with any other similar marketed product (and/or cognitive test in the public domain) with regard to sensitivity and specificity issues in an aging or dementing population? Please describe and provide specific references.
Section 4	Convergence with biomarkers	Please describe (and provide references) for any cross-sectional design studies that directly compare change on your measure(s) with change in other related biomarkers of early disease progression (such as, but not limited to, changes in CSF A-Beta, A-Beta:pTau ratio, structural imaging, FDG-PET, amyloid imaging) in healthy elderly and the transitional stage to early MCI.
		Please describe (and provide references) for any within-subjects longitudinal design studies that directly compare change on your measure(s) with change in other related biomarkers of early disease progression (such as, but not limited to, changes in CSF A-Beta, A-Beta:pTau ratio, structural imaging, FDG-PET, amyloid imaging) in healthy elderly and the transitional stage to early MCI.
Section 5	Convergence with ADLs/IADLs	Do you know of any studies in which improvement, or progressive decline, in cognitive functioning, as measured by your test or battery, compared favorably with concomitant improvement or decline in other functional areas/functional health status (e.g., ADLs/IADLs)? Please describe.
Section 6	Psychometric properties	Please specify any publications (or unpublished analyses) that specifically address content, criterion (e.g., "predictive validity" and specifically, sensitivity and specificity), construct, and face validity issues, specifically regarding the use of your product with an aging or dementing population (with a particular focus on the transition from healthy aging to MCI). Please specify any publications (or unpublished analyses) that specifically address issues of convergent and divergent validities, specifically regarding the use of your product with an aging or dementing population.
		Please specify any publications (or unpublished analyses) that specifically address issues of reliability (test-retest, internal consistency) and the susceptibility of specific portions of your battery to practice effects. We are specifically interested in these issues as they pertain to an older adult population or dementing patient population.
		Please specify any publications (or unpublished analyses) that specifically address how intercorrelated the various subtests of your battery seem to be. That is, have you conducted any tests of correlated coefficients, as part of any larger studies, to look at whether your various subtests are measuring separate and differentiable cognitive functions? If this question is not applicable for your test(s) or measure(s), please indicate this.

Section	Topic	Question(s)
Section 7	Data collection and storage	Does your test platform require a computer for administration and/or scoring, or any other specific device (e.g., telephone administration)? If your product is computer-based, does your technology require a specific operating system platform to run on?
		Does your test or battery require a human rater to administer stimuli or record results (e.g., the recording of verbal responses during free-recall of items on a memory test)? If so, can this be accomplished remotely (by telephone, for instance), or is an in-person examination required? Please describe.
		Are there any specific computer hardware requirements (special equipment) for your system, such as special push-buttons, a special track-ball, etc.? If so, what options are available (purchase, rental during study period, etc.)?
		Are there any specific computer software requirements (special programs) for your system, such as task administration platforms like "E-Prime?"
		Please describe how individuals, who would administer your battery at an independent testing site, would be trained. How long (in terms of minutes/hours) and labor intensive would such training be for technicians who would administer your battery? Do these individuals need to be clinical neuropsychologists or have a similar level of expertise?
		Please describe the time demands (administration time in minutes) for your test, or each subtest of your recommended battery, if it were included in an experimental protocol. What are the time demands required for the administration of the full battery (if applicable)?
		Does your test or battery have matched alternate forms that have been validated and determined to be fit for use in clinical and/or research assessment? Please describe. How were these verified to be or what is the rationale behind describing the alternate forms as being equivalent?
		Please explain how missing data are accounted for, coded, and scored for your test or within subtests in your battery.
		How would study data be stored and transmitted? Specifically, what computer databasing options are there for storage and data transmission? Is data accessible in a real-time (or close to real-time) manner?
		Can your battery be used for at-home monitoring, either through the Internet or another technology that allows for assessment outside of the laboratory, clinic, or research site? Please describe.
		Does your data collection and storage system allow for the provision of real-time alerts regarding potentially worrisome cognitive change, for individual subjects, for use by caregivers, family members, or Data Safety Monitoring Boards (the latter being in the context of pharmaceutical trials)?
Section 8	QA/QC, use by the FDA, FDA Title 21 CFR Part 11 compliant	Does your technology allow for any automation of quality assurance/quality checking (QA/QC) in real-time, and "behind the scenes?" That is, please describe any automated QA/QC procedures to ensure that any discrepant values are communicated back to study sites (during a clinical trial), or to end-users for error-checking and resolution. If this capability is not available for your system, what are your standard operating procedures for managing QA/QC during the context of a clinical trial?
		To what extent has your test, battery or specific portions of your battery been accepted by the FDA and other health regulatory authorities (as part of an NDA) as a valid measure for any particular disease entity?
		Are unmodifiable audit trails created each time that the database is accessed? Does your test battery meet FDA Title 21 CFR Part 11 standards regarding the storage and transmission of electronic data? Has your company ever been audited, on-site, by a major pharmaceutical corporation? Please specify.
Section 9	Assessment of patient populations	Are there any physical, cognitive, or psychological restrictions that may limit successful test administration?
		Does your battery include features that make it particularly favorable or unfavorable for the assessment with different subpopulations? For example, are there aspects of culture or language that would impact how individuals complete your battery? How many separate languages have your test(s) been translated into for use in clinical research?