Outcomes Assessment in Clinical Trials of Alzheimer's Disease and its Precursors: Readying for Short-term and Long-term Clinical Trial Needs

by HOLLY POSNER, MD, MS; ROSIE CURIEL, PsyD; CHRIS EDGAR, PhD; SUZANNE HENDRIX, PhD; ENCHI LIU, PhD; DAVID A. LOEWENSTEIN, PhD; GLENN MORRISON, MSc, PhD; LESLIE SHINOBU, PhD; KEITH WESNES, BSc, PhD, FSS, CPsychol, FBPsS; and PHILIP D. HARVEY, PhD

Dr. Posner is with Pfizer Inc., New York, New York; Drs. Curiel, Loewenstein, and Harvey are with the University of Miami Leonard Miller School of Medicine, Department of Psychiatry and Behavioral Sciences, Miami, Florida; Dr. Edgar is with Roche, Roche Products Ltd, Hertfordshire, United Kingdom; Dr. Hendrix is with Pentara Corporation, Salt Lake City, Utah; Dr. Liu is with Prothena Biosciences, Inc., South San Francisco, California; Dr. Morrison is with Lumos Labs, Inc., San Francisco, California; Dr. Shinobu is with Decibel, Therapeutics, Inc., Cambridge, Massachussetts; and Dr. Wesnes is with Wesnes Cognition Ltd., Streatley on Thames and Department of Psychology, Northumbria University, Newcastle, United Kingdom.

ABSTRACT

An evolving paradigm shift in the diagnostic conceptualization of Alzheimer's disease is reflected in its recently updated diagnostic criteria from the National Institute on Aging-Alzheimer's Association and the International Working Group. Additionally, it is reflected in the increased focus in this field on conducting prevention trials in addition to improving cognition and function in people with dementia. These developments are making key contributions towards defining new regulatory thinking around Alzheimer's disease treatment earlier in the disease continuum. As a result, the field as a whole is now concentrated on exploring the next-generation of cognitive and functional outcome measures that will support clinical trials focused on treating the slow slide into cognitive and functional impairment.

With this backdrop, the International Society for CNS Clinical Trials and Methodology convened semi-annual working group meetings which began in spring of 2012 to address methodological issues in this area. This report presents the most critical issues around primary outcome assessments in Alzheimer's disease clinical trials, and summarizes the presentations, discussions, and recommendations of those meetings, within the context of the evolving landscape of Alzheimer's disease clinical trials.

INTRODUCTION

Alzheimer's disease (AD) is a fatal neurodegenerative disease that causes progressive cognitive, neuropsychiatric, and functional deterioration of memory and of self. Given that more than 30 million people worldwide are affected by AD, and that this number is growing dramatically,[1–3] the search for effective treatments to prevent its onset, significantly delay its progression, or

otherwise positively intervene in the disease course is an international research priority.[3–5]

It has been over 10 years since the last of five drugs was approved for the treatment of AD.[6–13] All five approved therapies target neurotransmitter systems and enhance the function of surviving neuronal circuitry. Although worthwhile, their effectiveness has provided modest symptomatic improvement but has not reversed prior decline or slowed the progression of the underlying disease processes. Efforts have turned toward identifying a disease-modifying therapy for AD, but these efforts have yet to yield a successful intervention.[14–21] The lack of success in developing a disease-modifying therapy for AD may arise, in part, from methodological imprecision in the conduct of clinical trials. Recent advances have occurred to reduce this imprecision by 1) improving the diagnostic accuracy of subjects enrolled into trials, especially through the incorporation of biomarkers; 2) moving to secondary prevention trials to evaluate therapies; and 3) developing more sensitive clinical instruments, including composite, computerized, and performance-based measures to assess disease progression and treatment effects.

In past clinical trials, the diagnosis of AD was based largely on the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria[22] that relied on clinical assessments of cognitive and functional deficits and the concurrent exclusion of other etiologies of dementia. These criteria had important utility in helping to identify a relatively homogeneous cohort of patients to be studied, but were susceptible to imprecision since up to 20 percent of those with a clinical diagnosis of AD did not have evidence of AD neuropathology (by current criteria) on post-mortem examination[23–25] or on positron emission tomography (PET) imaging of brain amyloid burden during life.[26–29] Furthermore, the criteria did not allow AD to be diagnosed until symptoms were well established, presumably at a point when disease pathology was advanced.

It is now widely accepted that pathologic changes in AD begin decades before the emergence of the constellation of clinical symptoms that define dementia onset.[3,30–32] There is also an emerging consensus that intervention at the earliest stages of AD, prior to dementia onset and widespread irreversible neuronal damage, will be the most feasible scientifically and socioeconomically, and will have the most impact.

Technical advances in imaging and fluid biomarkers have furthered our understanding of the underlying disease processes, and now allow AD to be identified with some certainty before the onset of obvious clinical symptoms. These advances have been incorporated into new diagnostic criteria proposed by the National Institute on Aging-Alzheimer's Association (NIA-AA)[33–36] and by the International Working Group (IWG).[37–39] Both sets of criteria hold that AD encompasses a continuum from preclinical to dementia, and they both aim to improve diagnostic specificity early in the continuum through the use of AD biomarkers. The criteria differ in that those proposed by the IWG capture both the prodromal and the more advanced stages of dementia within the same diagnostic framework, whereas those proposed by the NIA-AA classify three phases of disease: the first identified by the presence of biomarker abnormalities before the emergence of clinical symptoms,[36] the second by the presence of both mild clinical symptoms and biomarker abnormalities,[35] and the third by the presence of AD dementia.[22] Application of either of these new criteria in therapeutic trials[40,41] opens up the possibility for more accurate identification of patients for treatment. In the interim and since the initial IWG criteria, further research in biomarkers associated with stages of AD (i.e., markers of progression vs. diagnosis) has prompted recent publication of the IWG-2 criteria.[42]

As aspects of current practices in trial design and conduct improve, as outlined above, the regulatory framework around approval of new AD therapies with different mechanisms of action than the currently approved neurotransmitter-based therapies requires re-thinking. Additional work needs to be done in anticipation of a novel AD therapy that shows clear clinical benefit, including, whenever possible, performance-based assessments that underscore real-world functional benefit. Our companion paper[43] considers the state of the art regarding potential performance-based measures of functional capacity.

REGULATORY FRAMEWORK

Regulators, recognizing the pressing need for a shift in paradigm(s) when evaluating interventions for preclinical and predementia disease, have drafted guidance for AD drug approval. Historically, approval for drugs to treat AD has depended on dual primary endpoints assessing cognition and functioning; however, given the increasing emphasis on preclinical and prodromal AD trials, new definitions of what constitutes "functionally meaningful benefit" and "global decline" must be reexamined. Furthermore, as studies move into earlier intervention, the current divide between clinical trial design and registration requirements for "symptomatic" and "disease-modifying" therapies are likely to close.

By United States law, a clinical trial must establish clinical benefit to be the basis for an approval. Customarily, a clinically meaningful benefit is defined as a favorable, statistically robust effect on how a patient feels, functions, and/or survives.[44–46] This "clinical meaningfulness" could be reflected, for example, in an established "minimally important difference" between groups on a measure of functional ability/disability and/or a "responder" definition that is based on prior evidence that a certain degree of improvement on the selected scale represents a meaningful improvement for an individual. When the primary outcome does not directly measure daily function, additional evidence for a relationship between the impact of the drug on the treatment target and how the patient feels or functions is needed. Alternatively, a co-primary functional (e.g. ADCS-ADL) or global endpoint (e.g. ADCS-CGIC) could be used in addition to a cognitive outcome measure as is historically the case for AD trials.

Draft guidance was recently issued in the United States offering the US Food and Drug Administration's (FDA) current thinking on the development of drugs for early-stage disease.[47] While their view on trial endpoints remains unchanged for

dementia, they suggest that trials with individuals with early AD or late mild cognitive impairment (MCI) can theoretically use a single primary outcome measure if it can substantiate a treatment effect on both cognition and function, though no drug has been approved on such an outcome. The draft guidance newly acknowledges that the co-primary endpoint approach is impractical for predementia trials because of limited or relatively absent functional deficits. An effect solely on cognition may also be insufficient in preclinical cohorts, as detecting changes from no cognitive impairment or a relatively weak signal remains challenging. Nevertheless, in the case where functional deficits are essentially absent (e.g., early MCI or preclinical AD), an improvement on a cognitive measure may be sufficient for possible accelerated approval under Subpart H (i.e., an effect on an intermediate clinical endpoint that is reasonably likely to predict ultimate clinical benefit[48] [i.e., an effect on function]). This route would then require confirmatory post-marketing studies to establish clinical benefit.

The European Medical Agency (EMA) has also considered new guidance for trials in AD with concept papers, entitled "Need" and "No need" issued at the end of 2013.[49,50] The "No need" paper made the case that it is premature to rewrite the current, broadly worded guidance that accepts data-driven discussions and agreements with the EMA on a program by program basis. The "Need" paper holds that some clinical trials already focus on "prodromal" populations, and highlights the increasing emphasis on strategies aimed at earliest diagnosis, prognostication, and enrichment of clinical trials. The "Need" concept paper (as in the preceding 2009 guidance) explicitly acknowledges the "asymptomatic" disease stage. Version 2 of the "Discussion Paper on the Clinical Investigation of Medicines for the Treatment of Alzheimer's Disease and Other Dementias," released in October 2014, moved this discussion forward, and a new version of the EU draft AD guidance was released in January 2016.[51]

The 2016 draft of the EMA guidance recognizes potential challenges with a cognition-function co-primary requirement for pAD/MCI due to AD, including ceiling effects with existing tools and impact of compensation strategies. It also notes that creation of more sensitive and specific tools or novel cognition-function composites may be possible solutions, while stressing the need to assess both cognition and function in a comprehensive manner and demonstrate clinical relevance of effects. For preclinical AD, multiple approaches are described, such as diagnosis of dementia, significant cognitive decline, and change in cognitive function, including the use of more sensitive novel tools. Issues of a lack of tool validation, use of responder definitions and time-to-event analyses to support relevance, feasibility issues around trial duration and drop-out, and lack of reliable surrogates/need for lengthy follow-up to confirm relevant cognitive changes, are also discussed.

There is no doubt that regulatory conversations are moving the field in a much needed direction. Evolving therapeutic strategies that sequentially move from established regulatory paths for AD dementia to novel strategies for MCI due to AD/prodromal AD, and now to early MCI, take more direct aim at targeting the disease at its various stages. Building on this progress, consensus discussions focused around new principles for accepting the earliest predictive features of the AD in its preclinical state become crucial.

HEALTH ECONOMICS AND PAYERS

Consideration should also be given to the evidentiary needs of payers (i.e., coverage bodies and health technology assessment bodies) to support reimbursement. Currently, this includes using endpoint and statistical analyses to address the following questions: Is a new product better or safer than an existing product? How does its value compare to existing products? These considerations will continue to apply to therapeutic strategies aimed at AD. However, the next phase of AD drug development has prevention as its goal, and there are no precedents for judging value. In this context, what constitutes "meaningful benefit" is being debated across the research field and has yet to be resolved. Issues include determining the potential impacts on patient-reported outcomes, including quality of life, and on ethical, legal, and social issues.[52] Finally, the field needs to be prepared to address the likely question from third party payers of "How do we value therapeutics that prevent the slide from 'normal function' to clinical decline?" It is likely that some clarity will ensue with regard to health economic criteria once there are new successful clinical trials for the prevention of Alzheimer's disease and the totality of the evidence base can be evaluated.

COGNITIVE OUTCOMES: HISTORY AND PSYCHOMETRICS

The most frequently used cognitive outcome measure in clinical drug trials for mild to moderate AD is the Alzheimer's Disease Assessment Scale-Cognitive, which was designed in 1984 to assess deficiencies in episodic memory and a number of other cognitive domains (e.g., language, orientation, praxis) that are universally affected by AD disease.[53,54] While the ADAS-Cog has proven utility in clinical trials of cholinesterase inhibitors in demented patients and has become standard in the field, there is concern that the test is not particularly sensitive to cognitive changes in very mild AD, MCI, or preclinical AD, and will not be appropriate for trials in these cohorts.[55–59] Indeed, in several recent clinical trials of cholinesterase inhibitors and other novel interventions in patients with MCI, the ADAS-Cog did not differentiate between treatment and placebo effects,[15–20] despite the introduction of supplementary items and use of 11-, 12-, 13- and 14-item versions.

Psychometric evaluation has shown that total scores on the ADAS-Cog are very low in MCI, often in the range of 11 to 12 (standard deviation [SD]=~4.0) out of a potential range 0 to 70 or 0 to 80 (potential range 0–70, 11-item; 0–80, 12-item including Delayed Word Recall).[60–62] This contrasts with average scores of 25 to 28 in mild-to-moderate AD cohorts entering treatment trials.[62,63] Although the ADAS-Cog performs adequately in terms of scaling assumptions, reliability, and validity, in the mild AD and MCI cohorts, the presence of large floor effects in mild AD and MCI cohorts imposes serious limitations on the instrument's ability to

detect change.[64] Furthermore, the response categories for some of the items do not work as intended, and this contributes to the inability of several subtests to discriminate subtle changes in cognition. The limited utility of the ADAS-Cog in mild AD, MCI, and less impaired states was further confirmed by a Rasch analysis[65] of the experimental measurement paradigm, Andrich.[66] Taken together, these findings indicate that the ADAS-Cog is not optimal for measuring cognitive decline in clinical trials targeting individuals in mild AD, MCI, or preclinical states.

Similar problems with floor and ceiling effects and inability to detect cognitive change in MCI and preclinical AD also occur with two other measures that are commonly used in AD clinical trials: the Mini-Mental State Exam (MMSE) and the Clinical Dementia Rating (CDR) scale sum of boxes (CDR-SB). In a large database of individuals with MCI (n=2,551) or various forms of dementia (n=4,796), MCI cases with a CDR global score of 0.5 had a mean CDR-SB score of 1.30 (SD=1.16) (out of a possible total of 18) versus a CDR-SB score of 0.11 (SD=0.36) in healthy elderly controls. In the same sample, MMSE scores were 27.2 (SD=2.3) for MCI cases (out of a possible total of 30) compared to 28.9 (SD=1.3) for the healthy elderly controls.[67] These findings suggest that the MMSE and CDR-SB lack sensitivity to reliably differentiate individuals with MCI from healthy individuals and could indicate they are not optimal outcome measures in clinical trials involving MCI or preclinical AD, though they do show sensitivity to decline in longer-term studies in these patients.[59,68] Indeed, such issues can be handled in current study design by sample size and trial duration, and improving outcomes should have the added benefit of increasing trial efficiency.

Discussion of this issue is expanded considerably by Harvey et al,[43] but the take-home point from the perspective of outcomes measurement and clinical trial design is that different assessments than the ADAS-COG, MMSE, and CDR-SB, and which are likely performance-based, will be required for clinical trials targeting individuals with MCI or preclinical AD.

PSYCHOMETRIC REQUIREMENTS FOR DETECTING CHANGE IN MINIMALLY IMPAIRED SAMPLES

There are now many more statistical, psychometric, and methodological tools available to validate potential cognitive outcome measures than those originally used to validate the ADAS-cog.[54] Input from statisticians and psychometricians is critical at all stages of scale development as new measurement strategies are developed for clinical trials across the expanded AD spectrum. This input should include 1) ensuring content validity (this may constitute establishing a sound conceptual basis in the case of a cognitive test, or ecological validity on the case of a performance-based outcome, and where concept elicitation is not feasible), 2) verifying the best question-answer combinations to suit the aspect of cognition or clinical function being assessed, 3) calculating sample size, 4) selecting models to evaluate various aspects of the scale, and 5) selecting among the various options for the final assessment tool and its accompanying analysis guidelines. Statisticians will need to work closely with neuropsychologists, epidemiologists, psychometricians, and clinical trial methodologists to develop outcomes that best reflect changing cognitive and functional status of the individuals assessed, and will need to contribute to understanding the clinical meaningfulness of these outcome measures.[69,70]

A key consideration that has been poorly addressed by traditional psychometric methods of scale development is the need to ensure that scales are capable of measuring the range of cognitive performance that will be exhibited by the subjects at study entry and over time. This includes avoiding floor and ceiling effects and ensuring that there are enough possible responses within the metric in order that change can be visualized while minimizing possible false effects resulting from normal fluctuations in cognition. This latter consideration is particularly important when participants are selected for being apparently healthy (i.e., without clinically notable cognitive deficits at the time of study entry).

A central feature of outcomes assessments in clinical trials will be the ability to accurately identify very subtle declines and separate these from the expected stability in performance that would accrue from successful treatment in a prevention trial. In order to identify stability (expected in the case of successfully treated participants in the active treatment group) and decline (expected in some proportion of the placebo group), a scale needs to cover a wide range of functioning with gradations that enable precise measurement across the spectrum of disease severity, including those in the apparently healthy range. An important complication here is that some participants in these studies are not likely to develop a cognitive disorder (i.e., not all participants receiving placebo treatment will worsen). Thus, change will be detectable in only a subset of placebo-treated patients.

STATISTICAL STRATEGIES TO INCREASE SENSITIVITY IN EXISTING MEASURES

As the above discussion makes clear, detection of change in treatment trials of prodromal and preclinical AD has a number of highly specific requirements. Thus, it may take time to develop completely new scales to measure subtle changes in very early AD. However, treatment development efforts are ongoing even though existing measures are challenged in terms of detecting potential change in mildly impaired populations. To address these challenges, alternative strategies have been adopted to examine existing measures in nontraditional ways. Some options used to improve the usefulness of existing scales are 1) the creation of alternate and empirically derived composite outcomes that provide robust and sensitive combinations of existing items and 2) use of item-level analysis in the form of item-response theory (IRT) and Rasch analysis to understand the contribution of scale items that best assess an underlying latent variable (in this case, cognition or function). These strategies are not mutually exclusive, and multiple methods can be applied to the same set of items to identify the most robust way to identify

treatment-related changes. Added benefits may be that these more robust outcome measures will reduce the estimated sample sizes for clinical trials and otherwise improve the efficiency of the analysis of trial data.[55,64,65,71,72] However, a clear conceptual basis is needed prior to undertaking such work (e.g., a belief that items may be combined on the basis of best estimating a new unidimensional underlying construct such as disease progression). Without such clarity, concerns may persist regarding clinical meaningfulness and interpretation of such scores and difficulty in understanding what is being measured.

One approach that has been used to create composite endpoints is to identify and combine a reasonable set of items based on their face validity given our knowledge of the disease state and the concept we are attempting to estimate or measure (e.g., disease progression). Again, issues around content validity and conceptual basis should be considered here as mentioned above. Confirmatory factor analysis is often used to eliminate redundant items and to ensure that there are items that represent different levels of function within each domain of interest. Item selection can be challenging when cross-sectional data are used because 1) the disease is degenerative, and elements of functioning that represent the core disease process change over time, and 2) there is substantial measurement error because individuals with the disease manifest day-to-day variability in cognitive and functional performance, particularly when more severely impaired. The first point can be addressed by extracting items for the composite from longitudinal studies so that decline over time can be included as a method of confirming which items are relevant to disease progression (depending on the length of the trial from which the items were extracted and the expected amount of decline). The second point can be addressed by measuring the same domain with more than one assessment tool (i.e., with a construct validity approach) to potentially reduce measurement error. Once appropriate items are identified, they are combined in some way (e.g., summing across standardized scores for each item) to form the composite.

Another approach to the development of composite outcomes is to use principal component analysis (PCA) with all available items and let the results determine which items best represent a domain, as indicated by a high correlation with one of the factors identified by the model. These factors represent the empirical consensus of the items; thus, domains with more items are likely to emerge as primary factors in the model and under-defined domains may be error variance. As is true for the face-validity approach, different results might be obtained from a PCA depending on whether cross-sectional or longitudinal data are used in the analysis. Using baseline measures could yield a PCA solution with a first principal factor that includes both items that are sensitive to decline and items that are insensitive to decline. In contrast, the first principal component of an analysis of change scores across domains should reflect the group of items most sensitive to change over time regardless of their baseline factor structure.

A third approach to composite development focuses on maximizing sensitivity to decline by using the mean to SD ratio (MSDR) of change scores.[57] MSDR change scores are similar to standard scores in that scales with different measurement metrics can be combined. This approach is quite amenable to large-scale combination approaches, and original analyses were based on an exhaustive search (i.e., brute force) method where data from multiple clinical trials with multiple outcome measures were combined into a single database. The MSDR approach can also be implemented directly with a reduced rank regression model with time as the outcome. Development of a partial least-squares regression model that represents a compromise between principle components regression and reduced rank regression models allows simultaneous consideration of both the time factor and factors identified through convergence of the items.

Several companies have launched trials with new composite outcome measures that were developed through a combination of empirical (based on sensitivity of individual items) and construct validity (based on opinions of experienced clinicians and neuropsychologists) approaches.[58,71,74] Edland et al[71] have taken a different tactic by attempting to optimize existing outcome measures. Using data from the Alzheimer's Disease Neuroimaging Initiative, these investigators modeled data from a scale for activities of daily living (ADL), and reset the scoring paradigm using an IRT approach. They have shown that sample size estimates using this optimized outcome are reduced by approximately 17 to 20 percent.[71] This approach makes the assumption that the targeted population will have a similar range of multivariate functioning as the "training set" (i.e., the ADNI cohort) used to adjust the scale. It also requires a very large database to serve as the training dataset. In addition, power calculations based on measures optimized in this manner include terms for the fixed and random variance, both of which are determined by the choice of outcome measure and the target population of the trial. Additional work with this measure using data from a clinical trial identified a learning effect during the first six months of the trial and suggested the need for a single-blind, run-in phase for three months. Including this run-in phase while not changing the overall 18-month length of the study reduced retesting effects and halved the sample size needed to show the same amount of decline on the outcome assessment.

Another way to possibly extract increased sensitivity from existing measures is to reanalyze existing trial data to gain a better understanding of the sources of variability in cognitive outcome measures that may mask true drug effects. One specific direction, for example, might be to remove items that contribute to error variance in outcome tests such as the ADAS-Cog and reassess drug-versus-placebo effects over time. It may be the case, for example, that subjective ratings of language and memory made by the examiner on the ADAS-Cog are susceptible to rater-to-rater and/or test session-to-test session variability that add noise to the signal of change over time. These subjective ratings are scored on a 0 to 5-point scale and can account for about 25 percent of the ADAS-Cog total score. It has

been shown by Hobart et al[64,65] that the response categories for these subjective ADAS-Cog subtests could be made dichotomous without loss of information. It also possible that a version of the ADAS-Cog that includes only those items that are well-targeted to a population with mild cognitive changes (i.e., excluding items aimed at more severe illness) would provide a clearer signal of the cognitive impact of an intervention in early AD.

Regardless of the statistical approach used to improve sensitivity of outcome measures, consideration should be given to an evaluation of how well the assessment tool measures the latent trait under study. In the case of cognitive, functional, or clinical decline in AD, the measurement tool chosen should be capable of measuring the range of performance that will likely be exhibited by the people entering the specific study and the expected changes that will follow intervention. The gradation of measurement for each item should be sufficient to show any clinically meaningful change for every patient and, as such, should not be sparse.

DISCUSSION AND FUTURE DIRECTIONS

The design of prevention trials poses an enormous challenge and risk for sponsors. As clinical trials focus more on prevention of AD or intervention in very early disease stages, the assessment tools used to detect change will need to be matched to the status of the participants in the study. Furthermore, the psychometric properties of these assessment tools will need to be optimized for the study population. Newer statistical methods and novel cognitive and functional measures, as discussed by Harvey et al,[43] hold great promise in this regard as we move forward. While the majority of risk in any trial lies in whether the drug under study has the desired effect, failure to use appropriate tools for assessing cognitive and functional change adds the risk of missing interventions that work.

The use of unmodified outcome measures designed for more severe stages of illness in trials holds as great a risk as imprudent drug selection. Careful conceptualization of treatment targets and state-of-the-art understanding of the characteristics of the very early stages of AD are required to make progress. Regulatory guidance is being updated in concert with new developments in understanding treatment targets and populations of interest. While novel ways of using existing instruments hold some promise for detection of change in populations with very early AD, this effort will have to be paired with developing new assessment strategies.

ACKNOWLEDGMENT

REFERENCES

1. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement*. 2007;3:186–191.
2. Colantuoni E, Surplus G, Hackman A, et al. Web-based application to project the burden of Alzheimer's disease. *Alzheimers Dement*. 2010;6:425–428.
3. Winblad B, Amouyel P, Andrieu S, et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. *Lancet Neurol*. 2016;15:455–532.
4. National Institute of Aging—US Department of Health and Human Services [website]. Obama administration presents national plan to fight Alzheimer's disease. Newsroom. May 15, 2012. https://www.nia.nih.gov/newsroom/2012/05/obama-administration-presents-national-plan-fight-alzheimers-disease Accessed December 2013.
5. National Institutes of Health—US Department of Health and Human Services [website]. Research portfolio online reporting tools (RePORT). NIH categorical spending, estimates of funding for various research, condition, and disease categories (RCDC). February 10, 2016. https://report.nih.gov/PFSummaryTable.aspx. Accessed February 2017.
6. Forest Pharmaceuticals. Namenda XR (memantine HCl) patient information [online]. September 2014. https://www.allergan.com/assets/pdf/namendaxr_pi. Accessed February 2017.
7. Drugs.com [website]. Tacrine hydrochloride [discontinued]. https://www.drugs.com/monograph/tacrine-hydrochloride.html Accessed February 2017.
8. Novartis. Exelon patch prescribing information [online]. November 2016. https://www.pharma.us.novartis.com/sites/www.pharma.us.novartis.com/files/exelonpatch.pdf. Accessed February 2017.
9. Pfizer. Aricept (donepezil HCl) tablet, 5mg, 10mg, 23mg, prescribing and patient information [online]. February 2016. http://www.aricept.com/ Accessed February 2017.
10. Janssen. Razadyne ER (galantamine HBr) extended release capsules [online]. September 2016. https://www.janssenmd.com/pdf/razadyne/PI-Razadyne-RazadyneER.pdf. Accessed February 2017.
11. Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev*. 2006;(1):CD005593.
12. Knapp MJ, Knopman DS, Solomon PR, et al. A 30-week, randomized, controlled trial of high-dose tacrine in patients with Alzheimer's disease. The Tacrine Study Group. *JAMA*. 1994;271:985–991.
13. McShane R, Areosa Sastre A, Minakaran N. Memantine for dementia. *Cochrane Database Syst Rev*. 2006;(2):CD003154.
14. Coric V, Salloway S, van Dyck CH, et al. Targeting prodromal Alzheimer disease with avagacestat: a randomized clinical trial. *JAMA Neurol*. 2015;72:1324–1333.
15. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther*. 2014;6:37.
16. Doody RS, Farlow M, Aisen PS, Alzheimer's disease cooperative study

data A, publication C. Phase 3 trials of solanezumab and bapineuzumab for Alzheimer's disease. *N Engl J Med.* 2014;370:1460.

17. Doody RS, Thomas RG, Farlow M, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. *N Engl J Med.* 2014;370:311–321.

18. Salloway S, Sperling R, Brashear HR. Phase 3 trials of solanezumab and bapineuzumab for Alzheimer's disease. *N Engl J Med.* 2014;370:1460.

19. Salloway S, Sperling R, Fox NC, et al. Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *N Engl J Med.* 2014;370:322–333.

20. Schneider LS, Mangialasche F, Andreasen N, et al. Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. *J Intern Med.* 2014;275:251–283.

21. Turner RS, Thomas RG, Craft S, et al. A randomized, double-blind, placebo-controlled trial of resveratrol for Alzheimer disease. *Neurology.* 2015;85:1383–1391.

22. McKhann G, Drachman D, Folstein M, et al. 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;34:939–944.

23. Lim A, Tsuang D, Kukull W, et al. Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series. *J Am Geriatr Soc.* 1999;47:564–569.

24. Mayeux R, Saunders AM, Shea S, et al. Utility of the apolipoprotein E genotype in the diagnosis of Alzheimer's disease. Alzheimer's Disease Centers Consortium on Apolipoprotein E and Alzheimer's Disease. *N Engl J Med.* 1998;338:506–511.

25. Ranginwala NA, Hynan LS, Weiner MF, White CL, 3rd. Clinical criteria for the diagnosis of Alzheimer disease: still good after all these years. *Am J Geriatr Psychiatry.* 2008;16:384–388.

26. Pontecorvo MJ, Mintun MA. PET amyloid imaging as a tool for early diagnosis and identifying patients at risk for progression to Alzheimer's disease. *Alzheimers Res Ther.* 2011;3:11.

27. Clark CM, Pontecorvo MJ, Beach TG, et al. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study. *Lancet Neurol.* 2012;11:669–678.

28. Johnson KA, Sperling RA, Gidicsin CM, et al. Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. *Alzheimers Dement.* 2013;9:S72–83.

29. Liu E, Schmidt ME, Margolin R, et al. Amyloid-beta 11C-PiB-PET imaging results from 2 randomized bapineuzumab phase 3 AD trials. *Neurology.* 2015;85:692–700.

30. Jack CR, Jr., Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurology.* 2013;12:207–216.

31. Andrews KA, Modat M, Macdonald KE, et al. Atrophy rates in asymptomatic amyloidosis: implications for Alzheimer prevention trials. *PLoS One.* 2013;8:e58816.

32. Bateman RJ, Xiong C, Benzinger TL, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med.* 2012;367:795–804.

33. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:263–269.

34. Jack CR, Jr., Albert MS, Knopman DS, et al. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:257–262.

35. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:270–279.

36. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7:280–292.

37. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology.* 2007;6:734–746.

38. Dubois B, Feldman HH, Jacova C, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* 2010;9:1118–1127.

39. Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurology.* 2014;13:614–629.

40. Knopman DS, Jack CR, Jr., Wiste HJ, et al. Short-term clinical outcomes for stages of NIA-AA preclinical Alzheimer disease. *Neurology.* 2012;78:1576–1582.

41. Vos SJ, Xiong C, Visser PJ, et al. Preclinical Alzheimer's disease and its outcome: a longitudinal cohort study. *Lancet Neurol.* 2013;12:957–965.

42. Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimers Dement.* 2016;12:292–323.

43. Harvey PD, Cosentino S, Curiel R, et al. Performance-based and observational assessments in clinical trials across the Alzheimer's disease spectrum. *Innov Clin Neurosci.* 2017;14(1–2):30–39

44. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA.* 1999;282:790–795.

45. New drug, antibiotic and biological drug product regulations: accelerated approval. Proposed Rule. 57. Federal Register, 1992: 13234–13242.

46. Katz R. Biomarkers and surrogate markers: an FDA perspective. *NeuroRx.* 2004;1:189–195.

47. United States Food and Drug Administration. Guidance for industry. Alzheimer's disease. Developing drugs for the treatment of early stage disease (draft guidance). February 2013. https://www.fda.gov/downloads/Drugs/

…/Guidances/UCM338287.pdf. Accessed February 2017.

48. Kozauer N, Katz R. Regulatory innovation and drug development for early stage Alzheimer's disease. *N Engl J Med.* 2013;368:1169–1171.

49. European Medicines Agency [website]. Concept paper on no need for revision of the guideline on medicinal products for the treatment of Alzheimer's disease and other dementias. 15 March 2012. http://www.ema.europa.eu/ docs/en_GB/document_library/Scientific _guideline/2012/03/WC500124534.pdf. Accessed February 2017.

50. European Medicines Agency [website]. Concept paper on need for revision of the guideline on medicinal products for the treatment of Alzheimer's disease and other dementias. 24 October 2013. http://www.ema.europa.eu/docs/en_GB/ document_library/Scientific_guideline/20 13/10/WC500153464.pdf. Accessed February 2017.

51. European Medicines Agency [website]. Draft guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease and other dementias. 28 January 2016. http://www.ema.europa.eu/ docs/en_GB/document_library/Scientific _guideline/2016/02/WC500200830.pdf. Acessed February 2017.

52. Green Park Collaborative. Center for Medical Technology Policy [website]. Evidence Guidance Document: Alzheimer's Disease. April 2013. http://www.cmtpnet.org/docs/resources/ GPC_Evidence_Guidance_Document_-_Alzheimers_Disease.pdf Accessed April 29, 2016.

53. Mohs RC, Rosen WG, Davis KL. The Alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. *Psychopharmacol Bull.* 1983;19:448–450.

54. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry.* 1984;141:1356–1364.

55. Hendrix SB. Measuring clinical progression in MCI and pre-MCI populations: enrichment and optimizing clinical outcomes over time. *Alzheimers Res Ther.* 2012;4:24.

56. Karin A, Hannesdottir K, Jaeger J, et al. Psychometric evaluation of ADAS-Cog and NTB for measuring drug response. *Acta Neurol Scand.* 2014;129:114–122.

57. Langbaum JB, Hendrix SB, Ayutyanont N, et al. An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimers Dement.* 2014;10:666–674.

58. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement.* 2013;9:S21–S31.

59. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry.* 2016;87(9):993–999. Epub 2016 Mar 23.

60. Skinner J, Carvalho JO, Potter GG, et al. The Alzheimer's Disease Assessment Scale-Cognitive-Plus (ADAS-Cog-Plus): an expansion of the ADAS-Cog to improve responsiveness in MCI. *Brain Imaging Behav.* 2012;6:489–501.

61. Grundman M, Petersen RC, Ferris SH, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol.* 2004;61:59–66.

62. Sano M, Raman R, Emond J, et al. Adding delayed recall to the Alzheimer Disease Assessment Scale is useful in studies of mild cognitive impairment but not Alzheimer disease. *Alzheimer Dis Assoc Disord.* 2011;25:122–127.

63. Doraiswamy PM, Bieber F, Kaiser L, et al. The Alzheimer's Disease Assessment Scale: patterns and predictors of baseline cognitive performance in multicenter Alzheimer's disease trials. *Neurology.* 1997;48:1511–1517.

64. Hobart J, Cano S, Posner H, et al. Putting the Alzheimer's cognitive test to the test. I: traditional psychometric methods. *Alzheimers Dement.*
2013;9:S4–9.

65. Hobart J, Cano S, Posner H, et al. Putting the Alzheimer's cognitive test to the test. II: Rasch Measurement Theory. *Alzheimers Dement.* 2013;9:S10–20.

66. Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11:571–585.

67. O'Bryant SE, Lacritz LH, Hall J, et al. Validation of the new interpretive guidelines for the clinical dementia rating scale sum of boxes score in the national Alzheimer's coordinating center database. *Arch Neurol.* 2010;67:746–749.

68. Coley N, Andrieu S, Jaros M, et al. Suitability of the Clinical Dementia Rating-Sum of Boxes as a single primary endpoint for Alzheimer's disease trials. *Alzheimers Dement.* 2011;7:602–610 e602.

69. Balsamo M, Giampaglia G, Saggino A. Building a new Rasch-based self-report inventory of depression. *Neuropsychiatr Dis Treat.* 2014;10:153–165.

70. Borsboom D. The attack of the psychometricians. *Psychometrika.* 2006;71:425–440.

71. Ard MC, Galasko DR, Edland SD. Improved statistical power of Alzheimer clinical trials by item-response theory: proof of concept by application to the activities of daily living scale. *Alzheimer Dis Assoc Disord.* 2013;27:187–191.

72. Mungas D, Crane PK, Gibbons LE, et al. Advanced psychometric analysis and the Alzheimer's Disease Neuroimaging Initiative: reports from the 2011 Friday Harbor conference. *Brain Imaging Behav.* 2012;6:485–488.

73. Vellas B, Bateman R, Blennow K, et al. Endpoints for pre-dementia AD trials: a report from the EU/US/CTAD task force. *J Prev Alzheimers Dis.* 2015;2:128–135.

74. Ard MC, Raghavan N, Edland SD. Optimal composite scores for longitudinal clinical trials under the linear mixed effects model. *Pharm Stat.* 2015;14:418–426. ■