Perspective

FDA position statement "Early Alzheimer's disease: Developing drugs for treatment, Guidance for Industry"

Marwan N. Sabbagh[a,*], Suzanne Hendrix[b], John E. Harrison[c,d,e]

[a]Cleveland Clinic, Lou Ruvo Center for Brain Health, Las Vegas, NV, USA
[b]Pentara, Salt Lake City UT, USA
[c]Principal Consultant at Metis Cognition Ltd, Kilmington, UK
[d]Associate Professor at the Alzheimer Center, VUmc, Amsterdam, The Netherlands
[e]Visiting Professor Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London

Abstract
Despite billions of dollars invested in clinical trials to develop novel therapeutics for Alzheimer's disease, no approved treatments have been developed in the past 15 years. In that span, new classes of drugs have been developed and tested, including monoclonal antibodies, γ-secretase modulators, γ-secretase inhibitors, BACE inhibitors, RAGE inhibitors, nicotinic agonists, 5HT6 antagonists, and others. The one constant for all of these clinical trials programs is the use of the ADAS-cog as the primary scale to determine efficacy. The question that needs to be considered is whether it is the target engagement of the drug or the clinical trial measure testing the efficacy. The FDA put out a new position statement in 2018 informing the field on possible considerations for demonstrating efficacy to open the path for approval. Here, we propose and comment on a variety of approaches that are alternatives to the ADAS for FDA-specified stage 3 and 4 Alzheimer's disease. These novel outcomes are being validated in current clinical trials and could be used as efficacy measures moving forward.
© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

In the draft statement released by the Food and Drug Administration (FDA) titled "Early Alzheimer's Disease: Developing Drugs for Treatment, Guidance for Industry" [1], the FDA provided much-needed guidance regarding drug approval paths for pharmacological agents being developed for Alzheimer's disease (AD). Important in their draft guidelines is the articulation of a three-stage system for classifying early AD, reflecting the pathobiology of AD, and replacing the terms preclinical and prodromal. In stage 1, biomarkers are abnormal, but people have no cognitive complaints or detectable clinical decline, even on sensitive tests (preclinical). In stage 2, subtle cognitive effects, but no functional deficits, appear (preclinical). In stage 3, people begin to have problems with some daily tasks measurable with instruments sensitive to AD stage 3 (prodromal), which corresponds with mild cognitive impairment due to AD, whereas the first two stages are preclinical. Stage 4 refers to symptomatic dementia with demonstrable cognitive and functional impairment and was specifically not covered in the guidance.

The efficacy outcomes for stages 1 and 2 are summarized in the report and commented extensively on elsewhere, most notably discussed at the Alzheimer's Disease Research Summit (AlzForum) on March 1, 2018. In these presymptomatic stages, a path forward can be found for drug approval with biomarker improvement sufficient to indicate a treatment success in stage 1 and

improvement on individual or composite neuropsychological test scores in stage 2.

Because many studies combine patients from stage 3 with very early stage 4 patients (usually Mini-Mental State Examination [MMSE] $\geq$ 23), and since functional deficits are detectable before a diagnosis of dementia, blurring the line between stage 3 and early stage 4, relevant approaches for stage 3 are also relevant for this combination stage. For stage 3 (early symptomatic AD—including MCI) and very early dementia, the FDA left the regulatory approval requiring functional improvement as a threshold of efficacy but expressed a willingness to consider a combination of outcome measure and functional and cognitive aspects. Some have proposed the possibility of isolating functional measures alone in stage 3 as meaningful outcomes. The guidance also allows approval based on a functional endpoint alone, which is a new possibility. This commentary focuses on stages 3 and 4. Furthermore, to avoid complexities that might be argued, we define stages 3 and 4 as early symptomatic and symptomatic AD dementia as amyloid-positive disease.

For most of the past 2 decades, the focus therapeutically has been on what is presently defined as stage 3/4. The current position statement, though encouraging for stages 1 and 2 of AD, continues to pose challenges for stage 3/4. A key issue concerns the selection of appropriate outcome metrics, and an interest in reassessing clinical outcomes is beginning to emerge [2].

Unfortunately, drugs approved for AD have the lowest success rate (99.6% failure rate from 2004-2009) circumstances partially due to variance in outcomes, highly heterogeneous patient populations, and a high standard for success (co-primary endpoints) [3]. Given this low success rate, developing drugs to treat AD seems to be a discouraging venture. Despite this limitation, however, numerous drugs and targets are in development for AD [4,5].

2. Reasons to consider revamping outcomes for drugs developed to address stages 3 and 4

Fundamentally, the methodology we use for measuring efficacy of drugs is reliance on a cognitive/psychometric measure (Alzheimer's Disease Assessment Scale-cognitive subscale; Mohs, 1984) and a functional measure (Clinical Dementia Rating scale, Alzheimer's Disease Cooperative Study-Activities of Daily Living) [6,7]. This approach should be reconsidered for several reasons. Reliance is excessively heavy on ADAS-cog as a measure of clinical target engagement or efficacy. The one commonality of the failed semagacestat, solanezumab, bapineuzumab, intepirdine, latrepirdine, idalopirdine, and verubecestat is the use of the ADAS-cog as the primary outcome measure in mild-to-moderate AD dementia populations. All drugs showed target engagement before the phase III randomized clinical trials, in the case of idalopirdine, a significant positive effect of treatment on the ADAS-cog, though note the

cohort in which proof of concept was observed was made up of exclusively moderate-stage patients (MMSE range of 12-19) [8]. This fact is consistent with the observation that the ADAS-cog is most sensitive in moderate-stage patients, but not those in the mild stage [9]. In fact, the ADAS-cog suffers from variability caused by various sources, including an increased number of sites, increased number of languages, and a high rater turnover. The failure to replicate the phase II study may in part be due to the later selection of mild-to-moderate–stage patients [10]. These drugs span different classes and include symptomatic and disease course altering drugs.

There are a number of possible reasons for the lack of efficacy detection with the ADAS-cog. The extrapolation is that the placebo group should decline in a projected manner and by proxy, if the treated group has slower decline, this signals treatment efficacy. In essence, much of the perceived effect of any given drug is influenced by the performance of the placebo. ADAS-cog has excessive variance. There is variance due to patient heterogeneity, and there is variance due to measurement error. Our comments are focused on the latter. Variations in forms, administration procedures, and scoring rules, along with rater turnover and intrarater drift, may decrease the reliability of the instrument. A survey of possible variants of the ADAS-cog was administered to 26 volunteer raters at a clinical trials meeting. Results indicate notable protocol variations in the forms used, administration procedures, and scoring rules [11]. In addition, not all domains are tested in all cases. Another study detected a total of 108 errors were made by 80.6% of the 72 raters and concluded that most experienced raters made at least one error that may affect ADAS-cog scores and clinical trials outcomes. These errors may undermine detection of medication effects by contributing both to a biased point estimate and increased variance of the outcome [12]. Variations in form have been shown to make a difference of nearly a full point on the ADAS-cog [13]. There has been a significant push by vendors (e.g., Bracket, MedAvante) to standardize raters and assessment capabilities across the sites. New scoring methods to improve sensitivity of the ADAS-cog have been proposed [14,15].

Because amyloid has not been shown to consistently correlate with progression [16], removal of amyloid or blocking production of amyloid may not move clinical cognitive measures, as was demonstrated in the solanezumab, bapineuzumab, and verubecestat trials. In recent solanezumab trials, sample sizes increased dramatically, providing sufficient power to detect miniscule effect sizes. Individual patient trajectories cannot be estimated reliably due to the large within-patient variability. However, analysis of patient populations over time within a designated stage of disease is an ideal approach for identifying composite measures that are reliable, minimally influenced by floor or ceiling effects, sensitive to change over time, and contribute to disease progression. This

approach was used to derive the ADCOMS composite, which was used successfully in the BAN-2401 study (ADCOMS described in detail below).

Estimating trajectory has been explored as a proxy estimate of clinical efficacy. Progression of normal subjects is a good way to identify item combinations that consistently change with the disease. A cross-sectional difference between normal and AD subjects or amyloid-negative and amyloid-positive subjects does not always identify the disease trajectory that changes. It may be necessary to correct for normal aging as well, particularly, if the treatment being assessed is not expected to affect cognitive decline separate from an AD-specific mechanism. Treatments with broader neuroprotective effects may benefit from using an outcome that includes both AD-specific and aging-related cognitive decline. Deriving composite scores with this type of latent variable modeling uses the degenerative nature of the disease as an advantage. The derivation identifies symptoms that are common to patients in a particular stage and then selects the combination that has the highest signal-to-noise ratio over time (mean to standard deviation ratio of the change score). Implicit in this approach is the assumption that the true disease process must have the largest signal and smallest noise in the majority of patients, so the goal is to approximate that latent trajectory with a linear combination of the clinical symptoms.

Approaches to identifying composite scores (e.g., Pre-Alzheimer's cognitive composite and Catch-Cog) that relay on maximizing the cross-sectional differences between normal controls and stage 3 patients or between amyloid-negative and amyloid-positive patients may result in similar composite scores. These approaches rely on assumptions about the relationship between cross-sectional and longitudinal data that often do not hold, resulting in less than optimal composites that need further refining to successfully measure longitudinal change.

All trials for stage 3 in the present circumstances are above the standard of care because there is a prevailing opinion that standard of care is not disease modifying. However, no recently developed drugs, such as intepirdine, idalopirdine, and latrepirdine, have shown an additive effect. Consequently, the threshold of approval remains high. Although skepticism regarding approved drugs, such as donepezil, rivastigmine, galantamine, and memantine, calls into question the robust efficacy of these potential therapeutic interventions, evidence exists to show that prolonged use of the approved drugs slows progression somewhat [17,18]. Thus, the additive effects above approved treatments have been difficult to achieve. True monotherapy trials without standard of care have been difficult to enroll and perform for the past 10 years. The rate of decline on the ADAS-cog has changed over the past 2 decades from 7 points per year to 4.5 points per year, although the standard deviation has remained constant at 6-7 points. The rate of change on ADAS is likely dependent on stage of the disease with a faster change

over time in moderate than in prodromal-early dementia. Furthermore, the ADAS-cog is not an infallible testing instrument as is susceptible to some shortcomings, such as errors caused by incomplete understanding of its logistics and problems related to regulatory processes intrinsic to its administration and implementation. Consequently, to power studies sufficiently, sample sizes and study durations needed to increase. Requiring a larger sample size increases variability due to the addition of sites, and longer study durations result in higher variability of scores at the end of the study.

The standard for efficacy for AD drugs is greater in AD than in other disease states partly due to reliance on a highly variable metric. In diabetes studies, efficacy is demonstrated by alteration of the HbA1C [19] or tumor marker regression and survival [20]. Biomarker evidence of efficacy, tantamount to disease regression, as seen in cancer and diabetes, has already been demonstrated in AD [21,22].

3. Suggested changes for outcomes to address stage 3

There is growing consensus for revising clinical trial outcomes for AD trials [2]. What might different outcomes look like? Here are some possibilities. First, consider any cognitive or functional measure prespecified as approvable. There is a precedent of use of the Severe Impairment Battery as an approvable instrument in more advanced dementia [23]. This fact suggests that alternatives to ADAS have been successfully deployed in clinical trials. Cognitive outcomes are more sensitive to change than functional outcomes and therefore may change by clinically irrelevant amounts and still be significant. Statistically significant versus clinically meaningful is hotly debated. The FDA has not supported single cognitive endpoints in the past for dementia trials except where the drug had been approved already (e.g., donepezil). It is speculative to consider if regulatory approval might be open to a single functional endpoint. Because decline of function is closer to the disease trajectory, and our measures of functional change tend to be more stable (partly due to looking back over 2 weeks to assess them), is making them often better powered than cognitive outcomes?

The search for alternative measures of cognition to the ADAS-cog has spawned a number of initiatives. One method, which we will call the "first principles," selects measures based on the process of first identifying the cognitive domains of interest and then selecting cognitive tests that have demonstrated reliability, validity, and sensitivity [24]. Reliability includes a consideration of test administration (interrater and intrarater reliability) as well as reliability in the context of temporal issues, such as stability and test-retest reliability. Validity in this approach has been focused on the cognitive content, that is, the extent to which selected measures index the domains of cognition known to be compromised in the earliest stages of the

disease process. Sensitivity issues have been largely considered in the context of the statistical characteristics of test data, including range restrictions, that is, the critical issues of "floor" and "ceiling" effects [25]. However, floor and ceiling effects need to be taken in context. "Internal responsiveness" refers to the ability to detect changes in disease progression, and "external responsiveness" refers to the ability to detect changes due to treatment effects. The floor and ceiling effects change the internal responsiveness, and the "assay sensitivity" seems to be the same as external responsiveness. A further dimension of test sensitivity has been "assay sensitivity" in which context test selectors have focused on evidence that the chosen measures have proven capable of capturing treatment effects [26]. Examples of positive "assay sensitivity" include the use of the Control Oral Word Association Test in studies of encenicline, the Digit Symbol Substitution Test (DSST) in studies of galantamine [27], and the Neuropsychological Test Battery in studies of AN1792 and donepezil [28,29]. The DSST has been shown as a single item to measure trajectory in MCI with higher sensitivity than any other single instrument [27].

A second approach, though one with some shared concerns to the first principles method, is driven by historic data. A concern with the first principles approach has been the absence of longitudinal AD patient data from which cognitive change trajectories can be computed. This approach has tended to focus on large study data sets, such as ADNI, and to select those measures which show the greatest sensitivity to change over time and with the least variance (high signal to noise as discussed previously). Many of the measures selected for these large study data set analyses tend to be reliable, sensitive, and valid, but test selection is by definition restricted to tests for which data are available. A recent derivation shows the conditions under which combining outcomes will result in a better performing outcome [30], offering the possibility of identifying optimal composites based on estimated correlations and performance metrics, perhaps from different studies. This approach could increase the tests considered for inclusion in a composite score [31].

One example of this approach was an analysis of historic ADNI data conducted by Raghavan et al. [32]. This group constructed a number of possible composite measures and evaluated their potential for meeting the twin criteria of capturing change and low variance. A summary of the constructed measure is shown in Table 1.

The ADAS-3 is a common element of other validation endeavors, including a recent analysis of possible enrichment strategies [33] and the "Catch-Cog" initiative [34]. This latter initiative has its origins in the proof of concept studies of cognitive efficacy in studies of AD patients receiving encenicline. A component of the Statistical Analysis Plan was provision of the analysis of a composite measure comprising the ADAS-3 combined

Table 1
Permutations of cognitive components used in composite scales

Composite	Cognitive components	Functional components
CC1	ADAS-3,* AVLT-I, and MMSE	N/A
CC2	ADAS-3 and cognitive subscale of CDR	N/A
CFC1	ADAS-3, AVLT-I, and MMSE	FAQ
CFC2	ADAS-3 and cognitive subscale of CDR	FAQ
CFC3	Cognitive subscale of CDR	FAQ

Abbreviations: ADAS-3, Alzheimer's Disease Assessment Scale-3; CDR, Clinical Dementia Rating scale; MMSE, Mini-Mental State Examination.

*ADAS-3 comprises the ADAS-cog Word Recall, Word Recognition, and Orientation subscales.

with data from the Control Oral Word Association Test and Complex Figure Test. A statistically significant positive effect of treatment was observed using this composite, which yielded a treatment effect size of >0.4. Effect size (Cohen's D) is proving to be a useful *lingua franca* for considering treatment effects across trials employing different tests but usually targeted on the same cognitive domains [33]. When Cohen's D is used as an effect size, it is affected by both the size of the effect and the variability of the measure used to capture the effect. If two treatments slow disease progression by the same amount (say 50%) and one uses a variable outcome measure and the other is half as variable, then the Cohen's D effect sizes may be 0.4 and 0.2 and this ends up being a comparison of responsiveness of the cognitive test. If the treatment used in the study with the more variable outcome has twice the disease slowing effect of the other treatment (say 100% slowing), then the effect sizes would be identical (0.40), although one treatment slowed disease by 100% and one by 50% because the difference in disease-slowing effect hides the differences in responsiveness of the outcome.

Thus, internal responsiveness is used to compare outcomes. If two populations are in the same stage, then the progression over time should be similar, and the mean to standard deviation ratio of the change scores could be compared between studies as a way of identifying more internally responsive outcomes with the expectation that a disease-modifying treatment would affect all aspects of disease progression similarly (same % slowing). Having a progressive disease gives us a gold standard to compare against time [34].

The Catch-Cog composite features all five of these measures plus DSST and Digit Span [34]. The rationale for including the ADAS-3 measures was multifactorial, but criteria included the absence of ceiling effects on these measures in early-stage patients; the extensive experience of using these measures in the AD clinical trial community; and in contrast to much of the ADAS-cog, the availability of parallel forms of the Word recall and Word Recognition subtests, as well as some inherent variability of correct response for the Orientation subtest.

To increase the breadth of assessment of the Catch-Cog assessment, we included the executive function component the Neuropsychological Test Battery (Control Oral Word Association Test, Digit Span, and the Complex Figure Test) and also selected the DSST based on its good psychometric properties and previous evidence of efficacy detection in early-stage AD [35]. The recent draft guidance has emphasized the importance of employing sensitive cognitive measures [36]. A further issue for the "historic data" approach is whether tests selected from a substantially larger assessment will perform in the same way when the context is changed. There is a rich and well-established literature on this issue showing interactions between tests such as proactive interference. Some researchers are collecting tests in the original context rather than reducing the test battery to only include the ones needed for the primary outcome. A number of new assessment validation studies drawn from historic data analyses are ongoing [37].

Other recent development of instruments, such as the integrated Alzheimer's disease rating scale [38] and AD Composite Score [39], combines cognitive and functional outcomes together. ADCOMS combines elements of the ADAS-cog, CDR, and MMSE (70% CDR, 30% ADAS/MMSE) that have been shown to change most responsively over time in a prodromal population [36]. ADCOMS has been successfully implemented as a measure in the phase II BAN-2401 clinical trial [40]. It has been shown to measure progression in MCI better than ADAS-cog or CDR alone.

Some alternatives to the functional aspects of the CDR are being proposed. The functional aspects of the CDR are not measured directly and lack granularity. Proposed alternatives include instrumental Activities of Daily Living to replace the function of the CDR. One suggestion is to use the Amsterdam instrumental Activities of Daily Living to more accurately separate statistical significance from clinical relevance [41].

Next, consider hybrid models that combine biomarker signal with any prespecified cognitive or functional signal. Specifically, if the biomarker changes occur in the manner predicted by the IP or intervention, coupled to any clinical outcome (prespecified), albeit mild, this would meet threshold for efficacy. Models that combine clinical and biomarker outcomes to predict progression are beginning to emerge [42]. In this scenario, aducanumab and BAN-2401 would have met a threshold of efficacy. A caveat to combining clinical and biomarker outcomes is that the biomarker should be tied to disease progression rather than to a specific mechanism of action because this approach is most effective when treatment effects and responsiveness to disease progression are similar for the clinical outcome and the biomarker outcome. Biomarkers that are more downstream in the disease process are preferable to biomarkers that may change independently of clinical outcomes.

One way to combine biomarker and clinical outcomes is with a global statistical test as originally proposed by O'Brien in 1984 and further extended over the next 30 years. Global statistical test has been widely used in clinical research on stroke, [18,19] dermatology, [20] multiple sclerosis, [21] asthma, [22] rheumatoid arthritis, [23] and more recently in Parkinson's disease. It allows combining outcomes with different scales by summing or averaging the percentiles or z-scores for each outcome, getting a score for each subject, and then analyzing those scores. The assumption is that all of the included outcomes are measuring different aspects of one underlying disease that is the true measurement target.

Trajectory-based analyses that use composite scores that approximate a latent progression variable offer the best statistical power for identifying effective treatment interventions. Although time-to-event analysis and survival analyses are more meaningful clinically, they are generally harder to detect statistically with the exception of rare cases in which only a few patients have measurable decline and those patients also have observed events. Statistical analysis of a quantitative outcome is generally more powerful than a time-to-event type of outcome because every patient can contribute to the average score with a continuous outcome, and only those with events contribute to the power with a time-to-event endpoint. There is some reassurance that events are meaningful, but combining clinical relevance and statistical significance into a single outcome may reduce our chances for detecting small but real effects that may be clinically meaningful in some circumstances or may be additive with other small effects resulting in clinically meaningful combination effects.

Time-to-event analyses have been done in AD and have been informative. In a prospective, 54-week, double-blind, placebo-controlled, survival to endpoint study, patients with AD were required to have at entry an MMSE score of 12 to 20; a Clinical Dementia Rating of 1 or 2; and capability of performing 8 of 10 instrumental activities of daily living and 5 of 6 basic activities of daily living. Patients (n = 431) were randomized to placebo or donepezil (5 mg/day for 28 days, 10 mg/day thereafter). Outcome measures were the AD Functional Assessment and Change Scale, the Mini–Mental State Examination, and Clinical Dementia Rating scale. At each visit, investigators determined whether predefined criteria for clinically evident decline in functional status had been met. Patients who met the endpoint criteria were discontinued per protocol. Donepezil extended the median time to clinically evident functional decline by 5 months versus placebo. The probability of patients treated with donepezil remaining in the study with no clinically evident functional loss was 51% at 48 weeks, compared with 35% for placebo. The Kaplan-Meier survival curves for the two treatment groups were different ($P = .002$, log-rank test). It is possible to speculate why this was informative. It was performed in the most progressive stage of AD, with a symptomatic (and possibly disease modifying also) treatment and therefore had some power to spare. In the early stages, we are not likely to have enough power for this

type of success. In addition, the time-to-event endpoints require frequent check-in by patients to have a truly continuous outcome that may have comparable power to a continuous outcome measuring disease trajectory.

In summary, for the FDA guidance regarding measuring efficacy endpoints in their defined stages 3 and 4, the reliance on the ADAS as a measure of efficacy or the imperative to have benefit on cognitive and functional endpoints continues to imperil determination of efficacy. Many of these measures are blunt instruments that are dependent on informant interview or a correct administration of a test. Because there are multiple tests in development including Catch-Cog, integrated Alzheimer's disease rating scale, ADCOMS, and others, there is a new opportunity to consider alternative outcome measures. Furthermore, novel analytical plans might also be informative rather than trajectory-based assessments of efficacy. Finally, combination of disease-specific biomarkers with a single clinical, cognitive, or functional endpoint might be sufficient to demonstrate efficacy. Novel endpoint considerations might pave the way for more drugs to be approved.

RESEARCH IN CONTEXT

1. Systematic review: For citations used in the perspective, references were gathered from PubMed. Also one citation was from a link from AlzForum.

2. Interpretation: Traditional clinical trial measures need to be reconsidered after dozens of drug trial failures—specifically, a reconsideration of the ADAS. There are many alternative outcome measures being developed and many of these outcome measures are proposed and discussed here including consideration of analytical methodologies.

3. Future directions: Alternatives to the ADAS that are being developed and discussed here need additional validation. Does it make sense to require clinical and functional efficacy measures? Could composite measures that include function and cognition be coalesced into single endpoints that might be approvable? Options such as ADCOMS are explored. Also, alternative analytical methods are discussed and explored.

References

[1] U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER),

Center for Biologics Evaluation. Early Alzheimer's Disease: Developing Drugs For Treatment, Guidelines for Industry, https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596728.pdf, 2018.

[2] Posner H, Curiel R, Edgar C, Hendrix S, Liu E, Loewenstein DA, et al. Outcomes assessment in clinical trials of Alzheimer's disease and its precursors: readying for short-term and long-term clinical trial needs. Innov Clin Neurosci 2017;14:22–9. eCollection 2017 Jan-Feb.

[3] Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline" few candidates, frequent failures. Alzheimers Res Ther 2014;6:37.

[4] Khatchaturian ZS. Prospects for effective treatment of the Dementia-Alzheimer syndrome: A renewed odyssey in search of the Magic Elixir. J Prev Alzheimers Dis 2017;4:215–7.

[5] Cummings J, Lee G, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2018. Alzheimers Dement (N Y) 2018; 4:195–214.

[6] Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–24.

[7] Galasko D, Bennett D, Sano M, Ernesto C, Thomas R, Grundman M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Discord 1997;11(Suppl 2):S33–9.

[8] Wilkinson D, Windfield K, Colding-Jorgensen E. Safety and efficacy of idalopirdine, a 5-HT6 receptor antagonist, in patients with moderate Alzheimer's disease (LADDER): a randomised, double-blind, placebo-controlled phase 2 trial. The Lancet Neurol 2014;13:1092–9.

[9] Cano SJ, Posner HB, Moline ML, Hurt SW, Swartz J, Hsu T, et al. The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. J Neurol Neurosurg Psychiatry 2010;81:1363–8.

[10] Keppel Hesselink JM. Idalopirdine (LY483518, SGS518, Lu AE 58054) in Alzheimer disease: never change a winning team and do not build exclusively on surrogates. Lessons Learned from Drug Development Trials. J Pharmacol Clin Res 2016;2 https://doi.org/10.19080/JPCR.2017.02.555580.

[11] Connor DJ, Sabbagh MN. Administration and scoring variance on the ADAS-Cog. J Alzheimers Dis 2008;15:461–4.

[12] Schafer K, De Santi S, Schneider LS. Errors in ADAS-cog administration and scoring may undermine clinical trials results. Curr Alzheimer Res 2011;8:373–6.

[13] Hendrix S, Ellison N. Are Alzheimer's drug failures due to inactive compounds or are we doing something wrong? Alzheimers Dement 2017;13:P617.

[14] Wouters H, van Gool WA, Schmand B, Lindeboom R. Revising the ADAS-cog for a more accurate assessment of cognitive impairment. Alzheimer Dis Assoc Disord 2008;22:236–44.

[15] Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK. Alzheimer's Disease Neuroimaging Initiative New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. Alzheimers Res Ther 2015;7:64.

[16] Nelson PT, Alafuzoff I, Bigio EH, Bouras C, Braak H, Cairns NJ, et al. Correlation of Alzheimer's disease neuropathologic changes with cognitive status: a review of the literature. J Neuropathol Exp Neurol 2012;71:362–81.

[17] Rountree SD, Atri A, Lopez OL, Doody RS. Effectiveness of antidementia drugs in delaying Alzheimer's disease progression. Alzheimers Dement 2013;9:338–45.

[18] Atri A, Rountree SD, Lopez OL, Doody RS. Validity, significance, strengths, limitations, and evidentiary value of real-world clinical data for combination therapy in Alzheimer's disease: comparison of efficacy and effectiveness studies. Neurodegener Dis 2012; 10:170–4.

[19] Blonde L, Chava, Dex T, Lin J, Nikova E, Goldberg R. Predictors of outcomes in patients with type 2 diabetes in the lixisenatide GetGoal clinical trials. Diabetes Obes Metab 2017;19:275–83.

[20] Ellis L, Bernstein D, Voest E, Berlin JD, Sargent D, Cortazar P, et al. American Society of clinical oncology perspective: raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 2014;32:1277–80.

[21] Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature 2016;537:50–6.

[22] Rinne JO, Brooks DJ, Rossor MN, Fox NC, Bullock R, Klunk WE, et al. 11C-PiB PET assessment of change in fibrillar amyloid-beta load in patients with Alzheimer's disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending-dose study. Lancet Neurol 2010;9:363–72.

[23] Reisberg B, Doody R, Stöffler A, Schmitt F, Ferris S, Möbius HJ, Memantine Study Group. Memantine in moderate-to-severe Alzheimer's disease. N Engl J Med 2003;348:1333–41.

[24] Ritchie K, Ropacki M, Albala B, Harrison JE, Kay J, Kramer J, et al. Recommended cognitive outcomes in pre-clinical Alzheimer's disease: consensus statement from the European Prevention of Alzheimer's Dementia (EPAD) Project. Alzheimer's Demen 2017; 13:186–95.

[25] Mortamais M, Ash JA, Harrison JE, Kaye J, Kramer J, Randolph C, et al. Detecting cognitive changes in preclinical Alzheimer's disease: a review of its feasibility. Alzheimer's Demen 2016;13:468–92.

[26] Harrison JE, Lam RW, Baune BT, McIntyre RS. Selection of cognitive tests for trials of therapeutic agents. The Lancet Psychiatry 2016; 3:499.

[27] Winblad B, Gauthier S, Scinto L, Feldman H, Wilcock GK, Truyen L, et al., GAL-INT-11/18 Study Group. Safety and efficacy of galantamine in subjects with mild cognitive impairment. Neurology 2008; 70:2024–35.

[28] Gilman S, Koller M, Black RS, Jenkins L, Griffith SG, Fox NC, et al., AN1792(QS-21)-201 Study team. Clinical effects of Abeta immunization (AN1792) in patients with AD in an interrupted trial. Neurol 10 2005;64:1553–62.

[29] Black RS, Harrison J, Li D, Yang L, Booth K, Barlas S. A comparison of Neuropsychological Test Battery (NTB) and ADAS-cog based on Alzheimer's disease clinical trials. Alzheimer's Demen 2009;5:P457.

[30] Liu-Seifert H, Andersen SW, Lipkovich I, Holdridge KC, Siemers E. A novel approach to delayed-start analyses for demonstrating disease-modifying effects in Alzheimer's disease. PLoS One 2015;10:e0119632.

[31] Raghavana, Samtania MN, Farnum M, Yag E, Novak G, Grundman M, Narayan V, Di Bernado A, Alzheimer's Disease Neuroimaging Initiative. The ADAS-Cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. Alzheimers Demen 2013;9:S21–31.

[32] Podhorna J, Krahnke T, Shear M, Harrison JE. Alzheimer's Disease Assessment Scale – Cognitive subscale variants in mild cognitive impairment and mild Alzheimer's disease. Alzheimer's Res Ther 2016;8:1–13.

[33] Cummings J, Scheltens P, McKeith I, Blesa R, Harrison J, Bertolucci PH, et al. Effect size analyses of Souvenaid in patients with Alzheimer's disease. Alzheimer's Res Ther 2017; 55:1131–9.

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

[35] Jutten RJ, Harrison JE, de Jong F, Aleman A, Ritchie CW, Scheltens P, et al. A composite measure of cognitive and functional progression in Alzheimer's disease: Design of the Capturing Changes in Cognition study. Alzheimer's Demen Translational Res Clin Interventions 2017;3:130–8.

[36] Harrison JE. Cognition comes of age: Comments on the new FDA draft guidance for early Alzheimer's disease. Alzheimer's Res Ther:61, https://doi.org/10.1186/s13195-018-0386-7.

[37] Harrison, JE and Hendrix, S. The assessment of cognition in translational medicine: A contrast between the approaches used in Alzheimer's disease and Major Depressive Disorder. In: Nomikos G (eds). Translational Medicine in CNS Drug Development. Handbook of Behavioral Neuroscience, volume 25 Amsterdam: Elsevier (In press).

[38] Wessels AM, Siemers ER, Yu P, Andersen SW, Holdridge KC, Sims JR, et al. A combined measure of cognition and function for Clinical Trials: the integrated Alzheimer's disease rating scale (iADRS). J Prev Alzheimers Dis 2015;2:227–41.

[39] Wang J, Logovinsky V, Hendrix SB, Stanworth SH, Perdomo C, Xu L, et al. ADCOMS: A composite clinical outcome for prodromal Alzheimer's disease trials. J Neurol Neurosurg Psychiatry 2016;87:993–9.

[40] https://www.alz.org/aaic/releases_2018/AAIC18-Wed-3-30-pm.asp.

[41] Scheltens P, Sikkes SA. The sensitivity to change over time of the Amsterdam IADL Questionnaire(©). Alzheimer's Demen 2015; 11:1231–40.

[42] van Maurik IS, Zwan MD, Tijms BM, Bouwman FH, Teunissen CE, Scheltens P, et al., Alzheimer's Disease Neuroimaging Initiative. Interpreting biomarker results in individual patients with mild cognitive impairment in the Alzheimer's biomarkers in daily practice (ABIDE) project. JAMA Neurol 2017;74:1481–91.