The Journal of Prevention of Alzheimer's Disease - JPAD
Volume 6, Number

The 2018 Revised FDA Guidance for Early Alzheimer's Disease: Establishing the Meaningfulness of Treatment Effects

C.J. Edgar[1], G. Vradenburg[2], J. Hassenstab[3]

1. Cogstate Ltd, London, UK; 2. UsAgainstAlzheimer's and Alzheimer's Disease Patient and Caregiver Engagement (AD PACE), Washington, DC, USA; 3. Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA

Corresponding Author: : Chris J. Edgar, Cogstate Ltd, London, UK, cedgar@cogstate.com

J Prev Alz Dis
Published online June 28, 2019, http://dx.doi.org/10.14283/jpad.2019.30

Abstract

The present report reviews the revised 2018 FDA guidance for early AD, with an emphasis on meaningfulness of clinical outcome assessments (COAs). A radical shift is evident in the importance given to establishing the meaningfulness of COAs in the 2018 draft versus the 2013 draft. The implications of this shift include the assertion that cognition is clinically meaningful, but that a persuasive effect on cognition, depending upon disease stage of the participants in the trial, is one that is of enough magnitude, established across multiple relevant domains, and can be supported by biomarkers reflecting underlying AD pathological changes. Meaningfulness is established through an understanding of the conceptual relevance of what is being measured and magnitude of any treatment effect. Precedent exists within other FDA guidance and independent good practices publications as to how meaningfulness may be assessed e.g. via evaluation of content validity and concepts such as minimally important difference. Additionally, FDA is developing a series of methodological Patient Focused Drug Development (PFDD) documents to provide further guidance on this topic, which are aimed at addressing gaps in methodology and recommended best practice. Importantly, application of PFDD approaches to AD is behind that in other areas and there is limited published content validity for COAs and a lack of supportive qualitative research. Initiatives to build robust conceptual models of AD and develop novel direct measures of meaningful health outcomes will have a significant impact on measurement of efficacy in clinical trials and on payer determinations of beneficiary value. Greater recognition of what is meaningful from the perspective of the patient and caregiver will inform regulatory reviews and determinations for payment and coverage of treatments.

Key words: Cognition, function, clinical relevance, patient focused drug development, regulatory guidance.

Introduction

FDA first published draft guidance on "Alzheimer's Disease: Developing Drugs for the Treatment of Early Stage Disease" in 2013. This guidance made mention of the co-primary approach at the AD dementia stage, where a functional or global assessment would "ensure the clinical meaningfulness of a cognitive benefit that may be observed." Challenges for early disease were related to mild or absent functional impairment, for which solutions might include integrated cognition-function assessment (e.g. CDR-Sum of Boxes), cognition assessment alone, or time-to-dementia. The terms 'meaningful' or 'meaningfulness' were used twice, once in relation to the co-primary approach and once in relation to a biomarker effect. In 2018, a revision was published "Early Alzheimer's Disease: Developing Drugs for Treatment" (1). Notably, with respect to clinical outcomes assessments (COAs) the revised draft guidance does not mention any example assessments, but now uses the terms 'meaningful' or 'meaningfulness' 27 times, suggesting an important shift in focus. The use of these terms in the revised draft guidance can be broken down into two different contexts: that of conceptual relevance ('is what is being measured meaningful?') and that of magnitude of effect ('is the size of a treatment effect sufficient to confer a benefit?') [Figure 1]. Importantly, the guidance also introduces a clinical staging framework and clarifies the focus as Stages 1-3. Stage 1 includes patients with characteristic pathophysiologic changes of AD but no evidence of clinical impact; Stage 2 includes patients with characteristic pathophysiologic changes of AD and subtle detectable abnormalities on sensitive neuropsychological measures, but no functional impairment; Stage 3 includes patients with characteristic pathophysiologic changes of AD, subtle or more apparent detectable abnormalities on sensitive neuropsychological measures, and mild but detectable functional impairment; and Stage 4 includes patients with overt dementia. This guidance does not discuss definitions of or methods for establishing conceptual relevance and meaningful magnitude of effect. However, precedent exits within other FDA guidance and publications as to how this may be addressed.

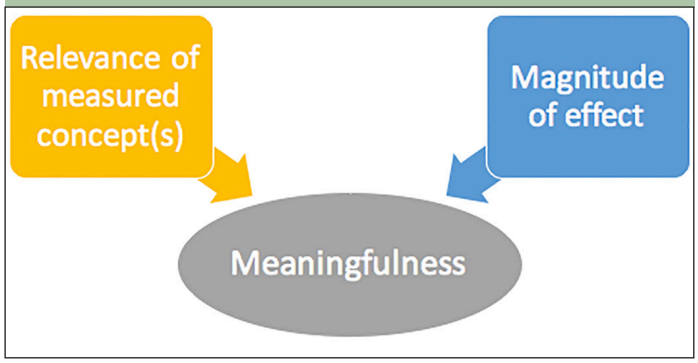
FDA is currently developing a series of four methodological PFDD guidance documents to address collection and submission of patient experience data and other relevant information from patients and caregivers for medical product development and regulatory decision making. This includes a new patient experience data table to be reviewed as a part of new drug applications. This table includes multiple types of suitable data including

Received February 13, 2019
Accepted for publication April 1, 2019

that from COAs, qualitative studies in patients and caregivers, PFDD stakeholder meetings, and survey, natural history and patient preference studies. A key component of this work is the development and validation of COAs as measures of treatment benefit. In the 2009 FDA PRO guidance, two important issues are discussed which are the need to establish content validity i.e. "the extent to which the instrument measures the concept of interest" and the need to define a clinically meaningful magnitude of change. These two issues are considered important to all COA types by FDA i.e. patient-reported outcome (PRO), clinician-reported outcome (ClinRO), observer-reported outcome (ObsRO) and performance-based outcome (PerfO) assessments. Each of these will be discussed here in relation to the revised draft early Alzheimer's disease (AD) guidance.

The present report will review the revised 2018 FDA draft guidance for early AD, with an emphasis on the meaningfulness of COAs and the implications for COA development and validation.

Figure 1. Establishing the meaningfulness of a treatment effect



Relevance of measured concept(s)

FDA revised draft guidance for early AD mentions meaningfulness in the context of conceptual relevance in several places e.g. "cognition is meaningful, but when measured using conventional approaches with sensitive tools directed at particular domains, the meaningfulness of measured changes may not be apparent." suggesting that both the domains measured (concept) as well as the ability of sensitive tools to measure small effects (magnitude), should be considered. Also, the need to ensure coverage of important cognitive concepts is expressed e.g. "cognitive changes of particular character, perhaps defined by magnitude or breadth of effect(s), may represent clinically meaningful benefit." suggesting again that both breadth of measurement across multiple domains (concept) as well as magnitude, are important to clinical meaningfulness This suggests the importance of conceptual relevance or content validity i.e. ensuring important measurement concepts are captured ('breadth' of assessment); distinguishing between direct, interpretable measures of important health outcomes and indirect ('conventional') measures.

Content validity (ensuring the breadth of relevant concepts for measurement)

Establishing concepts of interest (COIs) for measurement is foundational for COA development (2). COIs can be identified via literature review and qualitative research in patients, caregivers and clinicians. This work may be used to build a conceptual model of a disease or condition, or a conceptual framework for a given COA to ensure content validity i.e. that important measurement concepts are captured. To date, relatively little qualitative research has been conducted in people with early (predementia) AD, and their families and caregivers, with no explicit published conceptual model(s). However, published research has suggested there are potential gaps in existing measurements including concepts such as "situational lapses," "burdensome coping strategies," "slowness," and modern instrumental activities of daily living (iADLs) such as cell phone, or email use (3–5). As one would predict given the limited amount of qualitative research conducted in AD, there are relatively few COAs based on qualitative insights, or with well described conceptual frameworks, exceptions being e.g. the C-PATH Cognition Working Group PRO (6), and Amsterdam iADL questionnaire (7). Recently, a first conceptual model for the dementia stage of the disease has been published (8), which has been used to evaluate the conceptual relevance of four COAs in mild-moderate AD (ADAS-Cog, ADCS-ADL, NPI, and Dependence Scale). Importantly, this work concluded that these "assessment measures do not appear to capture the concepts most relevant to/important to patients with mild/mild-moderate AD."

To address this gap and the lack of established conceptual models across the spectrum of AD, a patient and caregiver-led collaboration of industry, academics, government agencies and advocates, the Alzheimer's Disease Patient and Caregiver Engagement (AD PACE), has been formed. The aim is to understand what matters most to individuals across the spectrum of the AD lived-experience (including individuals with underlying AD pathology who are asymptomatic or have mild cognitive impairment), matching FDA PFDD initiatives and policies, and eventually informing clinical development programs, regulatory submissions, payer value models, coverage and payment determinations, and research on care and services (https://www.usagainstalzheimers.org/networks/ad-pace).

Direct versus indirect measures of important health outcomes

Although good practice discussions have suggested PerfO development should utilize qualitative insights from patients and caregivers (9), published evidence indicates cognitive tests (cognitive PerfO assessments)

have not employed robust qualitative data in their development. Often such tests are not intended as direct measures of meaningful health aspects and the test activities are not a part of a person's usual normal life. Thus, the meaning of a score is not intrinsically known and must be established during validation (2). For example, the widely used Digit Symbol Substitution Test is not an activity of daily life and the meaning of a score or score change in number of symbols substituted is not directly interpretable. However, data show that performance is strongly correlated with real world functional outcome and functional capacity and such data may then be used to support meaningfulness and score interpretation. Many cognitive tests and test batteries are based on empirical models arising from disciplines within the cognitive neurosciences (neuropsychology, cognitive and experimental psychology, psychopharmacology etc.). Within this conceptual model, cognitive function is viewed as: common to all people not a sign or symptom unique to a given disease or condition; composed of concepts not readily isolated, quantified, reported, or observed (i.e. not best known to the patient); and most reliably measured by objective tests. Cognitive assessments may be developed based on face validity, and theoretical and quantitative models using empirical evidence of impairment in different domains.

Application to novel composite outcomes

Several novel composite outcomes have been proposed for early AD and these have broadly been developed and/or validated as either integrated assessments of cognition and function for MCI due to AD/prodromal AD (Stage 3), or as cognition only assessments for preclinical AD (Stages 1 and 2). Examples of these include ADCOMS (10) for Stage 3 and the ADCS-PACC (11) for Stages 1 and 2. Such assessments may be further subdivided in respect of their conceptual basis as empirically driven, theory driven, or a combination (12). ADCOMS and ADCS-PACC have been differentiated as being empirically driven and theory driven respectively. For ADCOMS, statistical modeling within target datasets was used to select and weight items for "sensitivity to clinical decline." The theory-driven approach for the PACC initially selected "4 measures that are well established as showing sensitivity to decline in prodromal and mild dementia, and with sufficient range to detect early decline in the preclinical stages of the disease" based on a literature review. Thus, they could be considered close in conceptual basis, though making use of different methodologies. Importantly, none of these composites has been based on a predefined conceptual model or framework or used qualitative patient-caregiver insights in the development and selection of items, with all incorporating 'conventional' cognitive test items that are indirect measures of meaningful health outcomes. Though there has been some attempt to retrospectively confirm the content validity of the ADCOMS using qualitative data (15), the use of statistical modelling to select and weight items and the incorporation of cognitive tasks, which are not part of usual normal life, suggests an indirect measure for which the meaning of scores must be established (13). Indeed, the EU/US/CTAD Task Force in discussing current prevention trials argues that cognitive changes are "possibly the best "biomarker" for AD trials." Thus like imaging or fluid biomarkers, cognitive measures also have the potential to be developed and validated as intermediate or surrogate clinical trial outcomes (12). As reported in this journal, a study is now underway named iMAP to assess the meaningfulness of two cognitive composites (RBANS and APCC) in preclinical disease, and will evaluate this via ability to predict clinically meaningful differences as determined by diagnosis of MCI or dementia due to AD and changes in Clinical Dementia Rating Global Scores [CDR Global] and Clinical Dementia Rating Sum of Boxes [CDR-SOB]) (14).

Magnitude of effect

Several techniques exist for the estimation of meaningful effect, including response thresholds for individual patients and change or difference thresholds for groups of patients. Multiple terms have been employed to describe these approaches including minimally important difference (MID) or minimally clinically important difference (MCID), and different individual patient (e.g. minimum detectable change (MDC), clinically important responder (CIR) and group estimates (e.g. minimum detectable difference (MDD), clinically important difference (CID) estimates. The most well-established of these techniques are anchor-based and distribution-based estimates, though other techniques such as exit interviews and vignettes might be employed. Additionally, data regarding patient and caregiver preferences and priorities in respect of magnitude of effect may be derived from quantitative stated preference methods, or other approaches suited to the population under study (15).

Anchor- and distribution-based approaches

Anchor-based approaches to determining meaningful within-patient change involve the use of an external reference with already established relevance. The most commonly used of these are 'global transition questions,' examples of which are patient or clinician global impression (PGI and CGI) ratings. Mean change in the target scale for the group, which was e.g. "minimally improved" or "minimally worse" on a CGI of change, would be used as one estimate of the minimally important difference (MID). Another approach is the 'clinical anchor,' also described as 'known groups' where

there is an accepted difference in clinical status that may be used as an anchor (16) or biological parameters with established clinical interpretation such as hemoglobin levels (17). In AD, the most well-established are the various forms of clinical staging of the disease. Dividing the disease into clinically defined stages based on severity of cognitive and functional impairment, or related concepts such as functional dependence has been widely employed in diagnosis, management and treatment. Staging criteria and instruments have also been used as clinical trial outcomes, including in time-to-event designs. Whilst there is clear face validity to the relevance of delay, or prevention of e.g. MCI or dementia, the low frequency/long time to progression has made this a challenging endpoint. Closely related to this, it is also apparent that in applying stage progression as an anchor, estimates may be relatively large, representing several standard deviations of change (18). Given this and the paucity of other anchors in available data sets, clinician judged change has more often been used (19).

Distribution-based, or internal estimates utilize statistical properties of the measures themselves and of these the most common are effect size metrics e.g. the standard deviation (SD) and the standard error of measurement (SEM) that incorporates some measure of scale reliability e.g. test re-test or Cronbach's α as a measure of internal consistency reliability.

Other approaches

More recently, approaches have been proposed that may serve as alternatives to or supplement anchor and distribution-based methods. Examples of these include bookmarking/standard-setting and scale-judgment. In bookmarking/standard-setting, patients and experts are presented with clinical vignettes of a disease in order to reach a consensus on thresholds supportive of meaningful change (20). This may also involve the use of modern psychometric approaches such as Rasch in order to support the generation of the vignettes based on empirical evidence for a relationship between item level changes and the total score. Another approach is the scale-judgment method, in which panels of judges evaluate pairs of completed tests to determine whether the difference indicated by responses before and after an intervention constitute a meaningful change (21). Though beyond the scope of this article, it is notable that Goal Attainment scaling presents a potentially useful methodology in AD in respect of relevance and magnitude, since the achievement of self-selected goals has inherent face validity with respect to both relevance of the concept and magnitude of effect e.g. (22, 23).

Key messages

- Both cognition and function represent potentially meaningful health outcomes
  - o Indeed, there may be considerable conceptual overlap between the two
- No clinical outcome assessment tool should be viewed as inherently meaningful in all contexts, irrespective of whether it is intended to measure cognition or function
- Many traditional cognitive tests (cognitive PerfO) may be indirect measures of meaningful health outcomes i.e. the test itself is not an activity that is a part of daily life
  - o Indirect measures may still be meaningful, but the steps to establish relevance and interpretation may differ from direct measures
  - o Indirect measures might also be developed and validated as intermediate or surrogate outcomes
- Meaningfulness has two key elements
  - o Relevance of the concepts being measured
  - o Magnitude of any treatment effect
- Methodologies exist to establish the meaningfulness of COAs via qualitative methods such as assessment of content validity; and quantitative methods such as assessment of meaningful change and difference via anchor- and distribution-based approaches
- Traditional or 'gold-standard' COAs developed for the dementia stage of AD and prior to emerging good practice recommendations and PFDD guidance may lack established clinical meaningfulness in early AD

Conclusions

A key component of PFDD is the meaningfulness of COAs. This has two components: the concept being measured and whether this is relevant to patients, caregivers, and clinicians; and the size of any treatment effect. In order to conclude that a treatment benefit has been observed, it is critical to establish that both a meaningful concept has been measured and a meaningful magnitude of treatment effect has been achieved. Initiatives to build robust conceptual models via qualitative research, the development of novel direct measures of meaningful health outcomes, and the validation of indirect measures as intermediate or surrogate outcomes, will have a significant impact on measurement in clinical trials for AD over the coming months and years. Greater recognition of what is meaningful from the view of the patient and caregiver will inform not only regulatory reviews but will also be used to inform other aspects of drug development, as well as determinations for payment and coverage.

Funding: No funding was received in support of this work.

Acknowledgments: C.J. Edgar, G. Vradenburg, and J. Hassenstab authored this manuscript. B. Hauber of RTI Health Solutions and D. Lappin of FaegreBD Consulting contributed to review of this manuscript.

Conflict of interest: C.J. Edgar is a full-time employee of Cogstate Ltd.

References

1. FDA. Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry.; 2018. https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm%0Ahttps://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.
2. Walton MK, Iii JHP, Hobart J, et al. Clinical Outcome Assessments: Conceptual Foundation — Report of the ISPOR Clinical Outcomes Assessment – Emerging Good Practices for Outcomes Research Task Force. Value Heal. 2015;18:741-752. doi:10.1016/j.jval.2015.09.2863
3. Buckley RF, Ellis KA, Ames D, et al. Phenomenological characterization of memory complaints in preclinical and prodromal alzheimer's disease. Neuropsychology. 2015. doi:10.1037/neu0000156
4. De Vriendt P, Gorus E, Cornelis E, Velghe A, Petrovic M, Mets T. The process of decline in advanced activities of daily living: A qualitative explorative study in mild cognitive impairment. Int Psychogeriatrics. 2012. doi:10.1017/S1041610211002766
5. Banningh LJW, Vernooij-Dassen M, Rikkert MO, Teunisse JP. Mild cognitive impairment: Coping with an uncertain label. Int J Geriatr Psychiatry. 2008. doi:10.1002/gps.1855
6. Gordon MF, Lenderking WR, Duhig A, et al. Development of a patient-reported outcome instrument to assess complex activities of daily living and interpersonal functioning in persons with mild cognitive impairment: The qualitative research phase. Alzheimer's Dement. 2016. doi:10.1016/j.jalz.2015.04.008
7. Sikkes SAM, De Lange-De Klerk ESM, Pijnenburg YAL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimer's Dement. 2012. doi:10.1016/j.jalz.2011.08.006
8. Hartry A, Aldhouse NVJ, Al-Zubeidi T, Sanon M, Stefanacci RG, Knight SL. The conceptual relevance of assessment measures in patients with mild/mild-moderate Alzheimer's disease. Alzheimer's Dement Diagnosis, Assess Dis Monit. 2018. doi:10.1016/j.dadm.2018.07.006
9. Richardson E, Burnell J, Adams HR, et al. Developing and Implementing Performance Outcome Assessments: Evidentiary, Methodologic, and Operational Considerations. Therapeutic Innovation and Regulatory Science. 2018.
10. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: A composite clinical outcome for prodromal Alzheimer's disease trials. J Neurol Neurosurg Psychiatry. 2016. doi:10.1136/jnnp-2015-312383
11. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. JAMA Neurol. 2014. doi:10.1001/jamaneurol.2014.803
12. Aisen P, Touchon J, Amariglio R, et al. EU/US/CTAD Task Force: Lessons Learned from Recent and Current Alzheimer's Prevention Trials. J Prev Alzheimer's Dis. 2017. doi:10.14283/jpad.2017.13
13. Edgar CJ. Use of Qualitative Data to Support Content Validity of Performance-Based Cognitive Outcome Assessments. Ther Innov Regul Sci. 2017;51(6):671-671. doi:10.1177/2168479017719934
14. A. Graf, V. Risson, A. Gustavsson, V. Bezlyak, A. Caputo, P.N. Tariot, J.B. Langbaum, C. Lopez Lopez VV. Assessment of Clinical Meaningfulness of Endpoints in the Generation Program by the Insights to Model Alzheimer's Progression in Real Life (iMAP) Study. J Prev Alzheimers Dis. 2019;6(2):85089.
15. Mansfield C, Bullok KE, Fuhs J V., et al. The patient voice: Exploring treatment preferences in participants with mild cognitive concerns to inform regulatory decision making. Alzheimer's Dement. 2017. doi:10.1016/j.jalz.2017.07.650
16. Phillips GA, Wyrwich KW, Guo S, et al. Responder definition of the Multiple Sclerosis Impact Scale physical impact subscale for patients with physical worsening. Mult Scler J. 2014. doi:10.1177/1352458514530489
17. Cella D, Nichol MB, Eton D, Nelson JB, Mulani P. Estimating clinically meaningful changes for the functional assessment of cancer therapy - Prostate: Results from a clinical trial of patients with metastatic hormone-refractory prostate cancer. Value Heal. 2009. doi:10.1111/j.1524-4733.2008.00409.x
18. Edgar CJ, Le Scouiller S, Delmar P, Mackey H. Estimating clinically relevant change and difference thresholds for the clinical dementia rating: Sum of boxes (CDR-SB) for early Alzheimer's disease. Alzheimer's Dement. 2017.
19. Schrag A, Schott JM. What is the clinically relevant change on the ADAS-Cog? J Neurol Neurosurg Psychiatry. 2012. doi:10.1136/jnnp-2011-300881
20. Cook KF, Victorson DE, Cella D, Schalet BD, Miller D. Creating meaningful cut-scores for Neuro-QOL measures of fatigue, physical functioning, and sleep disturbance using standard setting with patients and providers. Qual Life Res. 2015. doi:10.1007/s11136-014-0790-9
21. Thissen D, Liu Y, Magnus B, et al. Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method. Qual Life Res. 2016. doi:10.1007/s11136-015-1058-8
22. Rockwood K, Stolee P, Howard K, Mallery L. Use of goal attainment scaling to measure treatment effects in an anti-dementia drug trial. Neuroepidemiology. 1996. doi:10.2478/jee-2014-0027
23. Rockwood K, Howlett SE, Hoffman D, Schindler R, Mitnitski A. Clinical meaningfulness of Alzheimer's Disease Assessment Scale–Cognitive subscale change in relation to goal attainment in patients on cholinesterase inhibitors. Alzheimer's Dement. 2017. doi:10.1016/j.jalz.2017.02.005