PEER REVIEWED

Challenges of Measuring Clinical Meaningful Changes in Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that profoundly impacts patients and public health. It manifests through cognitive decline such as memory loss and functional impairment, leading to a loss of independence and the need for full-time care.[1] The number of patients with dementia worldwide is expected to exceed 150 million by 2050.[2] The disease not only affects individuals but also significantly impacts families, caregivers, and society. Family members often become primary caregivers leading to emotional, physical, and financial strains.

Clinical trials in AD represent a critical frontier in the search for effective treatments and interventions, however, the field faces several challenges and complexities.[1] As an example, with increased intervention targeting early-stage AD, the choice of specific and sensitive clinical endpoints to capture subtle cognitive and functional changes, and understanding their applicability in clinical practice, are becoming critical.[3] Currently, there is a lack of definitive evidence of how trial outcome measures correlate with changes in disease progression and treatment response which creates ambiguity around their clinical relevance.[3,4] Addressing this uncertainty would benefit patients, caregivers, primary care providers, and regulators, by improving our comprehension of how trial endpoints relate to everyday clinical assessments.

Presently, clinical trials in AD employ a variety of assessment tools, with considerable variability in the endpoints selected. This review will cover what constitutes clinically meaningful changes in early-stage AD, the most frequently used assessment tools, and discuss those measures in relation to disease progression and treatment efficacy.

Common Outcome Measures

AD clinical trial endpoints have been primarily tailored to track symptom changes in patients with moderate to severe AD dementia. However, as the focus has shifted towards developing therapeutics for mild cognitive impairment (MCI) due to AD and early-stage AD, there is a growing need to develop and validate clinical endpoints that can accurately detect changes during the earlier stages of the disease.[4]

Several outcome measures are commonly used to assess treatment efficacy and disease progression. These measures include cognitive assessments, such as standardised cognitive tests, the Mini-Mental State Examination (MMSE), the Montreal Cognitive Assessment (MoCA) and the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) that are used to evaluate changes in cognition over time.[4,5,6] To capture the sensitivity of cognitive change in MCI due to AD, additional tools are used in clinical trials such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), the Neuropsychological Test Battery (NTB) and the Preclinical Alzheimer's Cognitive Composite (PACC).[4,7] Assuming that functional changes in AD are due to cognitive decline, functional scales to assess activities of daily living (ADLs) and instrumental activities of daily living (IADLs) are crucial to provide insight into the impact of treatment and cognitive decline on patients' ability to perform everyday tasks independently. Both the Amsterdam Instrumental Activities of Daily Living (A-IADL) and the Alzheimer's Disease Cooperative Study – Activities of Daily Living Scale for use in Mild Cognitive Impairment (ADCS-ADL-MCI) have regularly been used in early symptomatic AD trials to assess instrumental activities of daily living known to decline in early stages of AD due to the high level of cognitive ability associated with these tasks.[4,8]

The US Food and Drug Administration (FDA)[9] recognises the value of composite assessments (both cognitive and function in a single scale) for the evaluation of patients with MCI due to AD. The Clinical Dementia Rating (CDR) is a semi-structured interview that was identified in the 2013 FDA guidance[10] as a suitable tool and has served as a primary or secondary endpoint in multiple AD clinical trials. Recent studies[3,4,6,11] have demonstrated the value of the CDR as an anchor measure to determine minimal clinically important differences (MCID). MCID provide a way of measuring changes in disease symptoms that go beyond just statistical significance. An MCID is defined as the smallest change in an outcome measure that results in a noticeable change in a patient's life.[12]

The CDR global score (GS) is used to quantify stage severity of AD and MCI and ranges from 0 to 3 (0= none; 0.5 = questionable; 1 = mild; 2= moderate and 3 = severe). The CDR sum of boxes (CDR-SB) is a continuous measure of AD and MCI stage severity and ranges from 0 to 18.[13] The CDR GS has been used as an anchor to estimate meaningful changes for other scales, but the CDR SB seems more efficient at capturing what is considered minimally important change.[3,4]

Andrews *et al.,*[14] used the data from the National Alzheimer's Coordinating Center Uniform Data Set (UDS) to estimate MCID for commonly used cognitive and functional assessment tools in AD, such as the MMSE, CDR-SB and the Functional Activities Questionnaire (FAQ). Their findings showed that as AD progressed, the MCID estimate increased. Specifically, a 1-3-point decrease in the MMSE, a 1-2-point increase in the CDR-SB, and a 3-5-point increase in FAQ were associated with meaningful clinical decline, depending on the severity of the disease.[14] Other composite scales that have shown sensitivity to change in AD are the Alzheimer's Disease Composite Score (ADCOMS) and the integrated Alzheimer's Disease Rating Scale (iADRS)[4,15,16]

Assessing Clinically Meaningful Changes

Assessing clinically meaningful changes in AD trials requires robust methodologies that capture changes considered relevant to patients, caregivers, and clinicians. Several approaches have been proposed, including distribution-based and anchor-based methods. The distribution-based methods rely on statistical properties of the data distribution to determine what constitutes a meaningful change in a clinical outcome measure.[4,17] These methods provide a quantitative approach to interpreting the clinical significance of changes observed in clinical trials or studies.

Common approaches include effect size calculations to measure the magnitude of change in a clinical outcome relative to the

variability of the measurements within the study population. It is typically calculated as the mean change divided by the standard deviation of the baseline scores.[17] An MCID based on effect size indicates how many standard deviations of change are considered clinically meaningful.[17,18] The anchor-based approaches rely on external criteria or anchors that are considered clinically meaningful by patients, caregivers or clinicians.[4,19] These approaches establish a connection between changes observed in a specific outcome measure and their perceived clinical significance. They offer a patient-centered perspective and enhance the interpretability of changes observed in a clinical trial. By aligning outcome measures with the patient's and clinician's perspectives of meaningful change, these approaches provide valuable insights into treatment efficacy and the impact of interventions on patients' lives. They complement distribution-based methods and contribute to a comprehensive understanding of what constitutes a clinically important change in AD.[18,19]

The FDA recommends using these anchor-based methods, along with empirical cumulative distribution methods, to determine a threshold or range of thresholds that signify a meaningful within-patient change score for the target outcome or the derived endpoint for the patient population.[9,18] Anchoring measures serve as external benchmarks to identify patients who have experienced a meaningful change in their condition, with the outcome score change assessed in these patient groups.

Current State of Affairs

There has been a shift towards developing treatments that target the underlying pathology of AD, such as amyloid-beta and tau protein aggregation and AD clinical trials are increasingly focusing on disease-modifying treatments aimed at slowing or halting disease progression.[4,11,12] Identification of these treatment effects on validated outcome measures is essential for regulatory approval. However, defining and measuring what constitutes a clinically meaningful benefit from these trials to various stakeholders, including clinicians and patients, needs further clarification.

While reviewing the definitions of clinically meaningful changes it is crucial to discuss misinterpretation regarding thresholds for meaningful within-patient progression and thresholds for determining meaningful group-level differences. In fact, a critical aspect of understanding clinical trial outcomes is distinguishing "between-group differences" from "within-individual change."[11,12] Measuring group-level differences is considered an appropriate statistical approach in parallel-group AD trials, but these estimates depend not only on changes observed within individuals but on several trial design factors, including sample size among others.[20] Thus, group-level differences do not directly indicate likely treatment effects at the individual level. This is mostly relevant for the early stages of AD trials, where disease progression is slow and even highly effective treatments may yield relatively small effects.[20,21] Consequently, large sample sizes and longer duration trials are often required to achieve sufficient statistical power to detect significant group-level differences during these stages.[3,4,12]

Utilising validated thresholds that represent the MCID in clinical outcome measures would help to objectively evaluate the clinical significance of trial results. However, MCID thresholds for most AD trial outcomes have not been established. For instance, when evaluating treatment efficacy based on the absolute point difference in change from baseline on a scale with a large range, small differences may be perceived as modest effects.[20] Conversely, the same magnitude of difference on a scale with a narrower range may be seen as large effects. Moreover, researchers interpreting these differences should consider the scale's range, reflecting the expected change over time for a specific cohort, which is often distinct from the full-score range of the scale. Without these reference points, it is challenging to assess the clinical relevance of small, yet statistically significant, differences between groups.[3,20,21]

Anti-amyloid monoclonal antibody therapies, which have recently garnered significant attention due to the FDA approval of lecanemab and donanemab, have demonstrated slowing of cognitive decline as measured by clinical endpoints such as the CDR-SB among others. In the Clarity AD phase 3 trial,[22] the change from baseline at 18 months in the CDR-SB (primary endpoint) was less with lecanemab than with placebo and the other secondary endpoints (ADAS-Cog 14, ADCS-MCI-ADLI) were in the same direction, favouring lecanemab. However, the drug-placebo difference did not meet the boundary of definitive MCID and exceeded the boundary of no MCID established by Andrews *et al.*[14,22]

In the TRAILBLAZER-ALZ 2 phase 3 trial, the results showed that donanemab slowed the rates of cognitive and functional decline in participants with early AD.[23] Additionally, a point change of -5 on the iADRS and +1 on the CDR-SB for those with MCI, or -9 on the iADRS and +2 on the CDR-SB for those with mild AD at consecutive visits from baseline, were also considered meaningful.

Furthermore, the CDR-SB is a pivotal endpoint in an Alzheimer's disease trial providing critical insights into the efficacy of treatments like lecanemab, donanemab and aducanumab.[22,23,24] Clinically meaningful changes in the CDR-SB scores might reflect real-world benefits emphasising the potential of these therapies to slow disease progression.

Future Directions

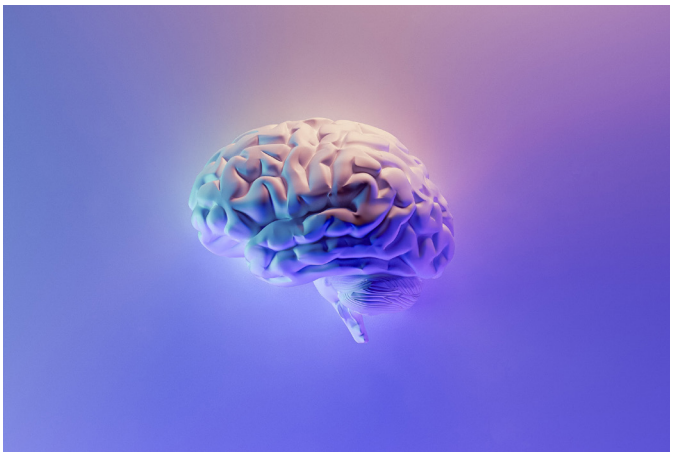
Despite significant advances and recent approvals of treatments aimed at modifying the course of AD, the clinical trials in the field remains challenging. AD encompasses a spectrum of clinical presentations, including early-onset and late-onset forms, as well as notable variations in disease progression and underlying pathology. Therefore, designing effective clinical trials for AD requires taking into consideration several key factors, including selecting appropriate outcome measures, defining clinically meaningful endpoints, and optimising trial duration and sample size.

The modest decrease in cognitive decline observed following treatment with lecanemab has reignited discussions on what constitutes "clinically meaningful" change. Recent literature[3,4,12,20] has explored this topic, involving scientists from various fields who agreed that meaningfulness should be evaluated individually rather than based solely on group differences. There is an emphasis on the need to consider the effects of treatments over time and the importance of outcome measures that reflect patients' priorities.

While clinical trials report statistical differences, they do not determine the minimum change noticeable to patients, caregivers, or physicians, which makes determining clinical meaningfulness at the participant or patient level challenging. Previous attempts to quantify meaningful change suggested a 1- to 2-point worsening on the CDR-SB scale,[14] but applying this threshold to consider small benefits seen with emerging therapies seems insufficient. Different studies have reported similar conclusions, with the minimal meaningful difference varying by disease stage and suggesting the need for longer trials to assess disease-modifying therapies adequately.

Researchers also highlight the limitations of current measurement scales, advocating for scales that capture a broader range of relevant changes based on patients' and caregivers' perspectives.[3,4,6,9,12] Therefore, multiple anchors, such as self-reported and partner-reported measures should be explored to gather comprehensive evidence to interpret a clinically meaningful within-patient score change, and these anchors should be used at comparable time points as the target outcome.[9,8,25,26]

It is also important to consider that patient-reported symptoms may become less predictive of objective cognitive impairment as the disease progresses due to loss of the patient's insight into their symptoms (*i.e.*, anosognosia), therefore the use of study-partner symptoms may become more predictive.[26] Additional research is required to comprehend the impact of anchor agreement on MCID estimation in the context of AD severity.



RESOURCES

1. Kumar A, Sidhu J, Lui F, et al. Alzheimer Disease. [Updated 2024 Feb 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan
2. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022; 7: e105–25
3. Borland E, Edgar C., Stomrud E, et al. Clinically relevant changes for cognitive outcomes in preclinical and prodromal cognitive stages. Implications for Clinical Alzheimer Trials. Neurology. 2022; 99(11) e1142-e1153. https://doi.org/10.1212/WNL.0000000000200817
4. Cohen S, Cummings J, Knox S, Potashman M, Harrison J. Clinical trial endpoints and their clinical meaningfulness in early stages of Alzheimer's disease. J Prev Alzheimer's Disease. 2022; 9 (3): 507-522. Doi:10.14283/jpad.2022.41
5. 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: change over time and the effect of enrichment strategies. Alzheimer's Res Ther 2016; 8:8; doi:10.1186/s13195-016-0170-5
6. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2020. Alzheimer's Dement.2020;6(1):12050; doi:10.1002/trc2.12050
7. Karantzoulis S, Novitski J, Gold M, Randolph C. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): utility in detection and characterization of mild cognitive impairment due to Alzheimer's disease. Arch Clin Neuropsychol. 2013; Dec; 28 (8): 837-44.doi 10.1093/arclin/act057
8. Pedrosa H, De Sa A, Guerreiro M, et al. Functional evaluation distinguishes MCI patients from healthy elderly people—the ADCS/MCI/ADL scale. J Nutr Health Aging 2010;14(8):703-9; doi:10.1007/s12603-010-0102-1
9. Early Alzheimer's Disease: Developing Drugs for Treatment, Guidance for Industry. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER); 2018.
10. Early Alzheimer's Disease: Developing Drugs for Treatment, Guidance for Industry. U.S. Department of Health and Human Services, Food and Drug Administration, CDER, CBER; 2013.
11. Stojanovic M, Mikula, C, John S, Kiselica A. Clinical importance in Alzheimer's disease: effects of anchor agreement and disease severity. Aging Clinical and Exp Research. 2024; 36: 5.doi.org/10.1007/s40520-023-02643-0
12. Petersen RC, Aisen PS, Andrews JS, et al. Expectations and clinical meaningfulness of randomized controlled trials. Alzheimer Dement. 2023; 19:2730-2736
13. McDougall F, Edgar C, Mertes M, et al. Psychometric properties of the clinical dementia rating - Sum of boxes and other cognitive and functional outcomes in a prodromal Alzheimer's disease population. J Prev Alzheimer's Dis. 2021;8(2):151-160.
14. Andrews JS, Desai U, Kirson NY, et al. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. Alzheimer's Dement.2019;5:354-363.
15. Wessels AM, Siemers ER, Yu P, et al. A combined measure of cognition and function for clinical trials: the integrated Alzheimer's Disease Rating Scale (iADRS). J Prev Alzheimer's Dis 2015;2(4):227-41; doi:10.14283/jpad.2015.82
16. 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):993-9; doi:10.1136/jnnp-2015-312383
17. McLeod LD, Cappelleri JC, Hays RD. Best (but oft-forgotten) practices: expressing and interpreting associations and effect sizes in clinical outcome assessments. American Journ Clin Nutrition. 2016; 103(3). DOI:10.3945/ajcn.115.120378
18. Edgar CJ, Vradenburg G, Hassenstab J. The 2018 revised FDA guidance for early Alzheimer's disease: establishing the Meaningfulness of treatment effects. J Prev Alzheimer's Dis. 2019;6(4):223-227.
19. Zhang Yu, Xiaoyu X, Huang Y. The anchor design of anchor-based method to determine the minimal clinically important difference: a systematic review. Health and Quality of Life Outcomes. 2023; 21:74 https://doi.org/10.1186/s12955-023-02157-3
20. Tarawneh R, Pankratz V. The search for clarity regarding "clinically meaningful outcome" in Alzheimer disease clinical trials: CLARITY-AD and beyond. Alzheimer's Research & Therapy. 2024; 16:37. https://doi.org/10.1186/s13195-024-01412-z
21. Rentz DM, Wessels AM, Annapragada AV, et al. Building clinically relevant outcomes across the Alzheimer's disease spectrum. Alzheimer's Dement. 2021;7e12181. https://doi.org/10.1002/trc2.12181.
22. van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. N Engl J Med. 2023; 388:9-21. https://doi.org/ 10.1056/NEJMoa2212948
23. Sims JR, Zimmer JA, Evans C, et al. Donanemab in Early Symptomatic Alzheimer Disease. The TRAILBLAZER-ALZ2 Randomized Clinical Trial. JAMA. 2023;330(6):512-527. doi:10.1001/jama.2023.13239
24. Day G, Scarmeas, N, et al. Aducanumab use in symptomatic Alzheimer disease evidence in focus. Neurology. 2022; 98:619-631. Doi:10.1212/WNL.0000000000200176
25. Kiselica AM, Johnson E, Benge JF. How impaired is too impaired? Exploring futile neuropsychological test patterns as a function of dementia severity and cognitive screening scores. J Neuropsychol. 2021; 15:410–427
26. Nosheny RL, Jin C, Neuhaus J, et al. Study partner-reported decline identifies cognitive decline and dementia risk. Ann Clin Transl Neurol. 2019; 6:2448–2459. https:// doi. org/ 10. 1002/ acn3. 50938

Catarina Cunha

Catarina Cunha is a Senior Clinical Scientist at WCG Clinical Research Solutions, overseeing clinical trials in neurology, focusing on conditions such as Alzheimer's disease, Parkinson's disease, Huntington's disease, progressive supranuclear palsy, and frontotemporal dementia. She has an MS in Clinical Psychology and is currently pursuing a PhD in Clinical Neuropsychology.