

POSITION PAPER

Low Sensitivity of Neuropsychological Scales Hinder Detection of Potential Benefit of Treatments in Alzheimer's Disease: A Position Paper

Simona Luzzi¹  | Julie S. Snowden²

¹Unit of Cognitive and Behavioral Neurology, Neurologic Unit, Department of Experimental and Clinical Medicine, Polytechnic University of Marche, Azienda Ospedaliero Universitaria Delle Marche, Ancona, Italy | ²Division of Neuroscience, Faculty of Biology and Mental Health, University of Manchester, Manchester, UK

Correspondence: Simona Luzzi (s.luzzi@staff.univpm.it)

Received: 3 October 2025 | **Revised:** 13 February 2026 | **Accepted:** 17 March 2026

Keywords: ADAS-cog | Alzheimer's disease | functional scales | minimal clinical important difference | neuropsychological scales

ABSTRACT

Background: Despite the advent of Disease Modifying Therapies (DMTs) for Alzheimer's Disease (AD), the approval and commercialization of anti-amyloid monoclonal antibodies has been slow and contentious, particularly in Europe. The primary source of debate is the discrepancy between robust biological effects—namely, effective β -amyloid clearance—and modest clinical improvements, which, although statistically significant, often fail to reach the minimal clinically important difference (MCID) compared to placebo.

Methods: This paper highlights a confounding factor in the interpretation of the results of clinical trials: limited sensitivity of neuropsychological outcome measures. These tools, developed in the 1980s and only marginally updated, are not suited to detect subtle but meaningful cognitive changes in early disease stages.

Results: The ADAS-Cog, the most commonly used cognitive endpoint, suffers from a substantial ceiling effect, impairing its ability to capture cognitive decline over short durations in prodromal populations. Likewise, functional scales such as the CDR-SB are inherently insensitive in mild cognitive impairment (MCI), as functional independence is, by definition, preserved. Moreover, the use of composite multidomain scales with high baseline scores may mask domain-specific improvements, further limiting a drug's capacity to reach MCID thresholds.

Conclusion: Methodological limitations risk undervaluing the therapeutic impact of treatment, particularly in trials targeting early or preclinical phases where changes are subtle and domain-specific. Urgent reconsideration of outcome measures is necessary to ensure accurate assessment of clinical efficacy and to avoid prematurely discarding potentially beneficial therapies.

1 | Background and Objective

The optimal therapeutic approach to Alzheimer's Disease (AD) remains controversial. The role of monoclonal antibodies is a case in point. Antibodies have been shown to remove beta amyloid from the brain effectively [1, 2], yet clinical benefit seems to be negligible [3–6]. This has led some authors to rethink the

pathophysiological model of AD and consider beta amyloid to be an innocent bystander [7, 8], that is, a protein that does not strictly intervene in the biological pathway of the disease. This hypothesis, if confirmed, would lead to a total rethinking of the diagnostic aspects of AD. Beta amyloid, is currently considered, according to current criteria [9, 10], to be crucial for the biological diagnosis of AD. If it does not actively intervene in

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *European Journal of Neurology* published by John Wiley & Sons Ltd on behalf of European Academy of Neurology.

the pathophysiological process, it should not be considered a biomarker in the etymological sense, necessitating alternative diagnostic markers.

Central to the resolution of debates of this kind is the capacity of clinical trials to detect change effectively. One needs to be certain that an apparent absence of effect of therapy is not merely an artifact of inadequacies of the tools by which change is measured. In this position paper we will argue that neuropsychological batteries currently in use are neither sufficiently sensitive nor adaptable to the current disease paradigm, leading to the potential for misinterpretation.

We begin by describing proposed ideal clinical characteristics of neuropsychological tools used in pharmacological trials and then consider the limits of the most commonly used cognitive and functional scales, the repercussions on the concept of minimal clinically important difference (MCID) and suggestions for future research.

2 | Optimal Neuropsychological Battery for Exploring Drug Efficacy

According to Wessels et al. [11], an outcome measure/scale should have the following characteristics:

1. Construct validity: it captures core cognitive and functional features of the disease.
2. Performance range: it is sensitive both to the earliest stage and progression of disease. Ceiling and floor effects (the task is too easy yielding top scores at baseline or too difficult and scores are zero) obscure the potential for detection of change and benefit from therapy.
3. Granularity: it can detect and measure subtle changes, allowing discrimination between cognitively stable subjects and those showing progression.
4. Objectivity: it is not dependent on subjective evaluation by the patient, caregiver, or examiner. Cognitive test scores must correspond to an objective value, such as the number of correct responses in a naming test.
5. Avoidance of rater bias: the measure avoids the potential for biases related to factors such as gender, age, education, and cultural background.

3 | Cognitive Scales in Practice: The ADAS-Cog

The Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) is the most widely used cognitive assessment instrument in clinical trials. It was developed more than 40 years ago, the standardization being based on 27 people with AD and 28 healthy controls (Rosen et al. [12]). The battery has served as the gold standard for assessing the efficacy of pharmacological and other therapeutic interventions. Its uniform use carries the advantage of permitting direct comparisons across studies.

There are, nevertheless, problems. In the past, patients enrolled in clinical trials were typically in the moderate-severe stages of

disease [13–16]. People with mild cognitive impairment (MCI) were excluded. The contemporary situation is radically different. The therapeutic focus is now on MCI and prodromal phases of disease when therapy is thought most likely to be beneficial. The ADAS-Cog is insensitive to changes in cognition in that population [17–20]. When the different tasks composing the battery are analyzed in relation to their sensitivity (see Kueper et al. [20] for a detailed review), it is observed that many of them show a marked ceiling effect (see Table 1). Accordingly, multiple modified versions have been proposed [17–23]. These have involved inclusion of additional tests, to explore (a) previously untapped cognitive domains and (b) specific cognitive domains such as memory more comprehensively. These efforts to improve the performance of the ADAS-Cog are laudable. Yet, the multiplicity of versions contributes to confusion regarding the cognitive endpoint.

Furthermore, notable protocol variations in the forms used, administration procedures, and scoring rules add variance and reduce reliability and sensitivity to treatment effects [24].

3.1 | Methodological Limitations of ADAS-Cog: Content of Battery

3.1.1 | Pitfall 1: Limited Cognitive Domains

The focus of the original ADAS-Cog is on verbal memory and language (Table 1). It neglects other cognitive domains affected in AD: executive functions, nonverbal memory, perceptuospatial skills ideomotor praxis. A battery should mirror the cognitive profile of disease.

3.1.2 | Pitfall 2: Limited Exploration Within Cognitive Domain

The “Word recall” subtest is designed to tap memory, a core characteristic of AD. The test requires the subject to listen to a list of 10 words and then immediately recall as many of the words as possible. As such, it makes demands principally on working memory rather than long-term episodic (hippocampal) memory, a core feature of AD. Similarly, the “word recognition” task requires the subject to read 12 words aloud and then, when presented with the 12 target words intermingled with 12 distractor words, to indicate whether they have seen each word before. Again, yes/no recognition is tested immediately only, so it does not adequately reflect the nature of the memory deficit in AD. A battery should adequately address the components of a cognitive domain affected by disease.

3.1.3 | Pitfall 3: Level of Difficulty: Tasks “Too Easy”

Some items applicable for moderate-severe dementia are simply too easy to perform for a subject in prodromal or mild phase AD and with a cultural background different from that of 40 years ago [25]. A couple of examples are as follows.

3.1.3.1 | Naming Objects and Fingers. The subject is asked to name 12 objects and the five fingers for a total of 17

TABLE 1 | Adas-Cog 11: Characteristics and limitations.

	Task	Description	Score range	Modality input/output	Cognitive domain	Ceiling effect ^a
Performance-based	Q1. Word recall	10 words read by the examiner have to be recalled for 3 trials	0–10	Verbal	Memory	No
	Q2. Naming objects and fingers	12 objects and the fingers of the dominant hand have to be named	0–4	Verbal	Language	Severe
	Q3. Commands	A set of command ranging each from one to five steps have to be executed	0–5	Verbal/motor	Language	Severe
	Q4. Constructional praxis	Four geometrical figures have to be copied	0–5	Verbal/Non verbal	Visuospatial	From mild to severe
	Q5. Ideational praxis	A five step task involving the actions need to send a letter has to be performed	0–5	Verbal/motor	Language	Severe
	Q6. Orientation	The subject is asked his name, 6 questions of orientation in time and one in place	0–8	Verbal	Memory	From mild to severe
	Q7. Word Recognition	Three trials are administered; in each trial the subjects has to recognize 12 target word previously read among 12 word distractors. The score is the mean number of incorrect responses across the three trials	0–12	Verbal	Memory	No
Rater-based	Q8. Language	The examiner assesses the subject ability to speak correctly during a 10 min free conversation	0–5	Verbal	Language	Severe
	Q9. Comprehension of spoken language	The examiner assesses the subject ability to understand correctly during a 10 min free conversation	0–5	Verbal	Language	Severe
	Q10. Word finding difficulties	The examiner assesses the subject difficulty to find words during a 10 min free conversation	0–5	Verbal	Language	Severe
	Q11. Remembering test instructions	The examiner scores the number of times he has to remind instructions to the subject during the Word Recognition Task	0–5	Verbal	Memory	Severe

Note: Ceiling levels are referred to as “severe” where mean error scores for a task are <0.5 and median scores (where reported) are 0. They are described as “mild” where mean errors are <1 but >0.5 and median error scores are 1. Abbreviation: Q, question.

^aCeiling effects, defined as error scores clustering around or approaching 0, are based on data derived from a series of studies of the ADAD-cog cited in a review by Kueper et al. [20].

items. Although Rosen et al. [12] reports that these constitute high (4), medium (4) and low (4) frequency items, they are easy words for a subject in the mild phase. Raghavan et al. [13] compared this test with the Boston Naming test, a naming test commonly used in clinical practice, in four groups of subjects for which a follow-up of at least 24 months was available: normal subjects, MCI non-convert, MCI converter and AD. Whilst the *Naming of objects and fingers* test showed a clear severe ceiling effect in all the four groups the Boston Naming Test was able to discriminate between the four groups (see figures 1 and 2 of Raghavan et al. [13] for details).

3.1.3.2 | Ideational Praxis. This task comprises a series of simple commands (fold a sheet of paper, insert it into an envelope on which you write your address). It is usually carried out without problems by subjects in the early stage of the disease. Ceiling effects are marked [20, 21]. Moreover, this task may no longer be ecologically valid, as younger generations may have limited experience with sending physical letters, owing to evolving cultural norms and technological advancements. One of the fundamental principles of cognitive assessment [26] is to adapt the complexity of the test to the demographic characteristics of the subject and to the general cognitive level. The assessment should be personalized, adapting the tests to the subject and not vice versa.

3.2 | Methodological Limitations: Scoring

1. Inconsistencies between the attribution of the score in the different subtests are evident. The “Word recall” and “Word recognition” tests have a score ranging from 0 to 10 and from 0 to 12 respectively, determined by the number of errors made by the patient. The score for the “Orientation” test ranges from 0 to 8, based on the number of errors. For other tests, the score ranges from 0 (absence of deficit) to 5 (deficit of maximum severity). The scoring system leads to heterogeneity in the quality of the data: in some tests it constitutes raw numerical data and in others a scale of interval or ordinal type.
2. Inconsistencies are also present within scales. For example, in the *Naming objects and fingers* a score of 0 is applied when the subject incorrectly names 0, 1 or 2 objects (i.e., allows up to 2 errors for an optimal score) whereas in the execution of commands a score of 0 applies only when no errors are made.
3. Some scales are ordinal, despite attempts to give them an interval-scale quality. For example, in the evaluation of spoken language skills, 2 points are attributed to a “mild deficit”, defined as the subject having difficulty making himself understood <25% of the time and 3 points to a “medium deficit” when the subject has difficulty 25%–50% of the time. Ordinal scales, where a magnitude is not necessarily equal between the different ranks, may lead to a distortion of the data [27].
4. Part of the battery, relating to language and memory, is dependent upon the judgment of the examiner. Shortcomings are (a) judgments are acquired on two cognitive domains only, (b) scores are subjective and (c) they

are based on informal aspects of performance such as free conversation, which is difficult to standardize across subjects.

3.3 | The Optimal Battery: How Does the ADAS-Cog Measure Up?

If the ADAS-Cog is judged in relation to proposed ideal characteristics (see Section 2), four of the five criteria are inadequately respected:

1. Construct validity: Tests selected to explore the cognitive status of a subject should be based on the prototypical cognitive profile of disease. In mild AD, attention needs to focus specifically on those cognitive domains typically impaired at the earliest stages of disease. The first structures involved in the degenerative process in classical AD are temporo-mesial cortices [28]. The predominant impairment is long-term episodic memory, or hippocampal memory.

Tests need to involve long-term, delayed recall, and recognition performed after a several minutes interval. Delayed memory testing is not present in the original version of the ADAS-Cog. One test exploring delayed recall has been added to some versions, such as the ADAS-Cog 13 [29].

Parieto-frontal associative cortex and posterior cingulate are involved early, particularly in people with younger onset. Patients experience problems with working memory, selective and divided attention [30]. In the original version of the ADAS-Cog, attentive tasks are not present. One test exploring attention, *Digit cancellation*, has been added to some versions [29, 31].

Studies that have compared modified versions of the ADAS-Cog with the original highlight the importance of addressing these distinct cognitive domains.

Padhorna et al. [18] compared the original 11-item test, with abbreviated 3 and 5-item tests, and an extended 13-item version. Memory items contributed most to overall score. “In particular *Delayed Word Recall* was impaired early and contributed largely to the overall score of the ADAS-Cog variants containing this item (i.e., the 13- and 5-item variants)” [18]. Notably, the sensitivity to detect change of the ADAS-Cog 5, containing only 5 subtests, exceeded that of the ADAS-Cog 11, with eleven. Similar results have been shown in other studies [20, 32].

Why should this be? It is simply because ADAS-Cog 5 contains two tests: *Delayed word recall* and *Digit cancellation* that explore respectively long-term memory and attention affected in prodromal AD.

These findings are reinforced by a further analysis in Padhorna's study [18]. They calculated the difference across the four ADAS-Cog variants (3, 5, 11, 13) by means of the signal-to-noise ratio (SNR), an estimated mean change from baseline divided by the corresponding standard deviation. The strongest evidence for a difference occurred for the comparisons between 13 and 11

item versions and between 5 and 3-item versions, attributable to the inclusion of delayed word recall and digit cancellation only in ADAS-Cog 13 and 5. Such findings have implications for predictions in clinical trials. A beneficial effect of therapy should be manifest by change in performance on tests specifically of long-term memory and attention.

2. Performance range: One of the strongest points commonly underlined in the literature about the ADAS-Cog is its large score range. The total score range varies from 0 to 70 in the original version, the ADAS-Cog 11; further versions have a higher score range (i.e., ADAS-Cog 13 used in evaluating lecanemab and donanemab varies from 0 to 85). If patients are tested at different stages of disease (from mild to severe dementia) scores spread across the total score range. Nevertheless, the score range does not translate into a sufficient range of difficulty. Ceiling effects occur in many components (Table 1).
3. Granularity: ADAS-Cog is not granular. The main arguments in favor of the granularity of the ADAS-Cog are that the battery is able to detect change over the entire course of the disease from the mild to the severe stages [12, 33], by virtue of its wide score range. This argument has little relevance for pharmacological studies, in which neuropsychological tests need to detect real differences from control over the limited time frame of the clinical study. Pharmacological treatments have a very limited duration of about 18–24 months; in contrast, the duration of diseases is several years with an extremely varied cognitive profile.

The argument of granularity was broadly applicable historically when patients were recruited for clinical trials in the moderate–severe phase of the disease and, by definition, subjects with MCI were excluded. If we go back to results of the ADAS-Cog in pioneering work [33] we can observe that a subject with an average MMSE of 23 (a value at the inferior limits of enrollment in a clinical trial) corresponded to an average ADAS-Cog score of 16.77. This score is at the upper limit of scores achieved on the ADAS-Cog by subjects included in clinical trials nowadays, by virtue of the high frequency of ceiling effects on subtests, making the theoretical advantage of the wide score range of the ADAS-Cog illusory.

This point is crucial. There is a gap between the theoretical global range score (0–70) and the “effective score” we measure in clinical trials.

As evaluated clearly in the Podhorna work [18] on 4 variants of the ADAS-Cog (version 3, 5, 11 and 13) the analysis performed on a wide number of MCI and mild AD subjects revealed that there was “no decline in mean ADAS-Cog 11 score in subjects with MCI (worsening of 0.9 on a 70-point scale over 24 months and 1.9 points over 36 months). The change from baseline was modest in patients with mild AD (worsening of 3.5 points over 12 months and of 8.3 points over 24 months)”.

Their population was classified as MCI/mild AD “enriched” or “not enriched” based on whether or not subjects showed positive biological markers for AD.

Similarly, they found minimal decline over 24 months and minimal impact of enrichment in other ADAS-Cog versions (3-, 5-, and 13-item). Interestingly, changes were also modest (4.35 points over 12 months on ADAS-Cog 13) in subjects with mild AD.

When they explored change over time by means of the SNR, the numerically largest SNRs in MCI were seen for the ADAS-Cog 5, but the 95% confidence intervals were broad, suggesting little substantial difference between test versions [18].

Similar findings occurred in mild AD, the numerically highest SNRs being seen for the 5-item and 13-item versions, but again the 95% confidence intervals were broad and overlapping.

It is interesting to note that Padhorna [18] found comparable results, in terms of magnitude of sensitivity to change over time, in the MMSE, considered a test of global cognition. The problem is that the MMSE is a basic screening test whereas the ADAS-Cog is designed to be a more sophisticated battery exploring major cognitive domains in depth. Thus, from a practical standpoint, the ADAS-Cog fares no better in detecting change than a standard screening test.

4. Objectivity: ADAS-Cog is burdened by the presence of subjective judgments of the examiner (20 points out of 70). The limited objectivity is compounded by significant variance in intra-rater experience with the ADAS-Cog administration procedures, scoring rules, and materials used in clinical trials, thereby undermining reliability [24].

The methodological limitations allow us to conclude that the ADAS-Cog does not provide an adequate measure of potential change in the earliest stages of AD and consequently it is not a suitable tool for evaluating the potential effects of a drug on cognition in MCI and mild AD.

Paradoxical findings in the research literature support this view. For example, in a study of biological and cognitive markers across clinical stages of AD, from preclinical to frank dementia [34] baseline evaluation showed that CSF Aβ1-42, CSF tau levels, ventricular volume, ADAS-Cog scores, and composite executive scores all distinguished asymptomatic-stage AD patients from healthy controls. Yet, at 4 year follow-up, whereas the decline in ventricular volume ($p=0.05$) and Clinical Dementia Rating scale sum of boxes (CDR-SB) ($p=0.04$) exceeded the decline observed in the controls, no difference in the cognitive scale (ADAS-Cog) was detected.

CDR-SB is a scale used to assess functional abilities in six domains: memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. The score is calculated by summing the individual domain scores. It is demonstrated to be a valid instrument for staging of AD [35].

Such data are puzzling because functional activities declined, even though these are expected to be normal in asymptomatic and MCI stage disease, by definition and according to diagnostic criteria, whereas cognitive impairment did not. Change in function is consequent upon cognition and not the reverse.

In a similar vein, studies of the effects of lecanemab and donanemab [36] that compared CDR-SB with ADAS-Cog14 (in lecanemab) and MMSE (in donanemab) reported “functional scale slopes diverge, confirming that the slopes of the scales based on caregiver input diverge, while the slopes of the objective cognitive variables collected by the patients are parallel” [34].

When analyses were conducted in the two Phase 3 studies of solanezumab [37] it was shown that 87% of the treatment effect on function was related to improved cognition, and only 13% to a direct effect on function.

If functional autonomy is dependent and consequent upon a cognitive deficit, it is illogical and not ecological that functional scales show a gradual loss of functioning in an individual who seems not to show change on a cognitive scale. The more conservative and logical conclusion is that the cognitive scale is not sensitive enough to capture the minimal variations in cognitive status that can cause subtle modification of functioning in daily life activities.

Any other potential interpretation is implausible; why would a subject in the prodromal or mild stage of AD show a functional alteration in the absence of an underlying cognitive deficit?

4 | Functional Scales: We Still Stay in the Past

It is not our intention to focus in detail on the functional scales in current use since they prove capable of detecting differences, even if minimal. For functional scales, an ideological change has taken place over time, admitting minimal variations as sufficient to demonstrate effectiveness [38].

According to the US Food and Drug Administration (FDA) requirements [39] to demonstrate drug effectiveness changes should be clinically meaningful. Changes in clinical trials of drugs in people with dementia and in particular AD need to demonstrate efficacy using both a measure of cognition and a functional scale. Several functional scales have been used to explore functioning; these include the Clinician Interview-Based Impression of Change plus Carer Interview (CIBIC-plus) and the Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (ADCS-CGIC) [40, 41], and the most frequently used the Clinical Dementia Rating Scale-Sum of Boxes (CDR-SB), which demonstrates clinically meaningful changes [38, 42].

Functional scales have been used successfully for the development of cognitive enhancers in mild-to-moderate stages of disease [13]. Nevertheless, in the very mildest phases of disease, functional scales have proved insensitive to minimal change [17].

In consequence, the FDA [43] has, over time, refocused on the importance of cognition (“FDA may consider other approaches, including endpoints based on cognitive assessments or surrogate endpoints, which may allow for shorter trial durations as a basis for approval in the earliest stages of AD”).

As with cognitive scales, low sensitivity of a functional scale is automatically translated into low efficacy of the drug. For example, in relation to lecanemab and donanemab it is said: “The absolute effect size (Cohen's *d* coefficient) at 18 months for the scale measuring the clinical performance of participants (Clinical Dementia Rating-Sum of Boxes, CDR) is 0.21 in the lecanemab study and 0.23 in the donanemab study, that is, small, and similar to symptomatic drugs” [36].

We re-emphasize that symptomatic drugs have been tested in moderate-severe phase disease where the rate of progression and therefore of accumulation of disability is different from the mild or prodromal phase. Simply considering differences between placebo and drug, without taking into account the rate of progression, leads to potentially misleading interpretations. In other words, a difference in minimal disability in the mild phase can mean turning back the clock by months or even years (how long does that differential disability between placebo and drug accumulate in the mild phase?), while in the moderate-advanced phase, where the rate of accumulation of disability is high, it can mean much less time—months or weeks.

In light of the changes in the scenario concerning pharmacological trials, we should ask ourselves whether including as a primary endpoint a functional scale that modifies and changes only in relation to cognitive impairment and alters only in the symptomatic phase is still an acceptable principle or not.

Our opinion is that whereas it may be acceptable in mild AD, in prodromal and preclinical disease it is not.

5 | Combined Scales

Combined scales are composed of both cognitive and functional scales. The most widely used is the integrated Alzheimer's Disease Rating Scale (iADRS). In its original version [31] it combines scores from 2 widely accepted measures, the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog 14) and the Alzheimer's Disease Cooperative Study—instrumental Activities of Daily Living (ADCS-iADL). In further studies ADAS-Cog 14 has been replaced by the ADAS-Cog 13 [44].

To combine measures into an integrated scale requires manipulation of data.

Because higher scores on the ADAS-Cog14 reflect worse performance, whereas higher scores on the ADCS-iADL reflect better performance, the ADAS-Cog score is multiplied by (−1). To anchor the ADAS-Cog at 0, a constant (90) is added. The iADRS score is then computed as the sum of the transformed ADAS-Cog14 and the ADCS-iADL, as shown in the formula below:

$$\text{iADRS score: } [1 - (\text{ADAS-Cog 14}) + 90] + \text{iADL}$$

The iADRS score ranges from 0 to 146 with lower scores indicating worse performance.

Aside from the inherent practical difficulty of combining scales, the appropriateness of summing measures of cognition and function is questionable. Function is dependent upon cognition, which means that constituents of the overall scale are not equally independent.

Another limitation of combined scales is that the functional component is necessarily applied to all subjects, including those with MCI who have no functional alteration. Indeed, in a study of AD and MCI using the scale, it was concluded that: “no items measuring function had a significant loading in this dataset, suggesting that in this MCI data set, the observed variability at the baseline was primarily driven by cognition and not by function” [11].

6 | Minimal Clinical Important Difference (MCID): Why Do Drugs Not Reach the Target?

In research, effects are traditionally measured by statistical significance. When considering drug effects, however, it is recognized that in clinical practice, a statistically significant result does not necessarily translate into demonstrable clinical change. For this reason, there has been a gradual shift from statistical significance to clinical relevance in interpreting results from clinical trials [27].

The minimal clinically important difference (MCID) is the smallest change in a scale measuring cognition or function that is noticeable by the patient or their caregiver [45, 46]. It is known also as *minimal important difference*, *minimal clinically relevant change*, or *minimum detectable change* [47]. It is estimated to represent 7% of the range of the scale employed [48]. Moderate and large changes constitute between 11.5%–13.7% and 12.3%–21.0% respectively [48]. For some scales, MCID is available in the literature.

Although the concept of clinical relevance is easily understood, from a factual point of view, translating it into a value is complicated since many variables come into play.

The most important and difficult issue is how to quantify changes considered clinically relevant.

There are some methodological limitations that must be considered in relation to the concept of MCID, defined by Draak as “pitfalls”. Examples of these are the use of ordinal scales that are then processed as if they were interval scales, and the possibility of ignoring small but clinically relevant changes because they fall within the variation of a measurement, especially when the variation of a measurement has a wide range [27].

Moreover, MCID is a dynamic concept that varies according to factors such as the severity of illness, social status, and subjective perception [27]. Socioeconomic status might affect a person's subjective perception of symptom severity, expectations of outcome, and perceived impact on performance of daily activities.

Yet, there is a widespread tendency to use fixed MCID cut-offs [45].

MCID varies too on the treatment tested. We fully support the suggestion provided by Draak [27]: “So, when examining a new treatment, one cannot just blindly copy an MCID determined in a different study with different treatments”.

There are two key aspects to take into consideration when evaluating drug efficacy that could explain, at least in part, the difficulty for experimental drugs to reach the MCID threshold in the mild or prodromal phase of AD. The first relates to the characteristics and limits of the cognitive and functional scales used; the second to the temporal profile of disease.

6.1 | First Point: The Structure of Cognitive Scales Contributes to Their Failure to Reach the MCID

The exhaustive review performed by Ebell [45] of cognitive or functional scales used to evaluate the efficacy of monoclonal antibodies (see Table 1 for details) suggests that the MCID value varies for cognitive tests from a minimum of 1–3 in the case of the MMSE (range 0–30) to a maximum of 4 for the ADAS-Cog 14 (range 0–90). For functional scales, it ranges from a minimum MCID score of 1–2 for the CDR-SB (range 0–18) to 7 for the DAD (range 0–100). Of all the reviewed scales, the iADRS had the highest total score of 146 with an MCIB of 8.8.

Given that the value of the MCID is related to a scale's total maximum score, it follows that the higher the scale's overall total score, the higher the MCID minimum score of variation in subjects' performance judged to be significant.

We reported above that ADAS-Cog 5 is more or almost as sensitive compared to lengthier test versions because it focuses on relevant cognitive domains and removes items associated with a ceiling effect, which do not add information. The corollary is that the ADAS-Cog 5 score range is more limited, from 0 to 45, yielding a MCID of 3.15%. This contrasts with 6.3% in the 90-point scale of ADAS-Cog 14. The substantial difference raises the question of whether drug studies would reach the MCID when lower thresholds are required, specifically, where tasks avoid the inclusion of non-informative items. The latter serve only as a burden, raising the MCID level and making the detection of MCID more difficult.

The iADCS-iADL is a case in point. It combines ADAS-Cog13 or 14 with iADRS. The inclusion of “ceiling level” tasks and of functional items into an overall score would elevate the MCID and mask the drug effect. Furthermore, as clearly shown the functional scale is not sensitive in the MCI subgroup [11, 17]. When results are considered overall (i.e., examining both MCI patients and mild AD patients enrolled in the study) the functional scale serves as a hindrance in the MCI subgroup where it adds points to the denominator but is not useful in identifying functional variations.

Of course, our hypothesis should be confirmed by a reanalysis of data, which should include the possibility of extrapolating only the results obtained in patients treated with the drug and placebo respectively in the selected ADAS-Cog subtests and from here verifying whether the results obtained allow the drug to reach the MCID.

6.2 | Second Point: Speed of Progression Cannot Be Ignored

Different factors may influence the presence, rate, and type of progression of AD. These include concomitant systemic disease and medication, lifestyle factors such as physical exercise, cognitive and social stimulation, and leisure activities, and environmental factors such as toxins and pollution [49].

Progression of AD occurs in a nonlinear manner over the spectrum of disease severity, and the overall disease trajectory appears sigmoid shaped. This applies both to biomarker changes and clinical presentation [11, 50–54]. In progressive diseases such as AD, the magnitude of MCID varies across the different stages of the disease [47]. Andrews et al. [47] found that the proportions of visits characterized as having a clinically meaningful decline (according with the clinician's opinion) increased from 16% among the normal cognition cohort to 99% among those with moderate–severe AD dementia.

The non-linear trajectory has implications for drug trials [55]. It is inherently more difficult to demonstrate drug efficacy in the earliest phases of disease when natural progression is slow. The average duration of clinical trials of 18 months is too short to elicit meaningful clinical differences between drug and placebo.

7 | Atypical Variants and Early-Onset AD

A crucial consideration, in the selection of cognitive batteries, is the clinical heterogeneity of AD. In non-amnesic variants [56] (e.g., logopenic variant, posterior cortical atrophy, frontal variant, and “motor” or corticobasal variant) classical memory impairment is not the predominant presenting feature. Standard batteries, such as ADAS-Cog, risk failing to capture adequately the salient deficits: in language, visual skills, behavior or praxis. Moreover, when patients perform poorly this may be for distinct underlying reasons, such as working memory deficits. Atypical presentations of AD are commonly associated with early-onset disease [57]. Patients often show faster disease progression so do not align well with cohorts enrolled in published trials of classical amnesic AD. This problem is further compounded by the fact that atypical or early-onset variants—sometimes with a hereditary basis—are typically excluded from clinical trials [1, 2, 4] due to their divergence from the traditional model. As a result, not only are appropriate cognitive batteries lacking from the outset, but there are also limited trial data available for these populations.

8 | Escaping From the Past: Proposals

The continued use of the ADAS-Cog is mainly justified by its historical role in enabling objective and reproducible comparisons across dementia trials. However, as noted previously, several compelling reasons argue for its abandonment. It was developed in the context of symptomatic treatments, making comparisons with disease-modifying drugs inappropriate [58]. The proliferation of multiple ADAS-Cog versions with different tasks and scoring systems undermines the reliability of cross-study comparisons [20]. Contemporary trial populations differ both clinically and demographically from those enrolled decades ago, rendering

direct comparisons across disease stages and cohorts meaningless. Collectively, these limitations highlight the urgent need for a paradigm and methodological shift in future research.

8.1 | Primum Movens Is Cognition

Let us escape from the tautological and erroneous assumption that cognitive aspects are late and insidious, impossible to identify early. The real problem is that neuropsychological batteries are insufficiently sensitive and adequate to capture the early and salient aspects of the disease.

Although functional aspects are undoubtedly important, it is the cognitive changes that must be set as the primary endpoint, because it is the cognitive alterations that are primary: they cause disordered function and not the reverse.

The long-held model of AD is that pathophysiological biomarkers of disease emerge long before cognitive symptoms [59, 60]. This is an oversimplification. Cognitive measures, if sensitive, can show alterations from the earliest stages of the disease in parallel with beta amyloid biomarkers [17].

A detailed systematic review [61] exploring neuropsychological differences in early preclinical AD showed that amyloid-positive individuals showed poorer performance in the domains of global cognitive function, memory, language, visuospatial ability, processing speed, and attention/working memory/executive functions when compared to amyloid-negative people and reinforced the view that neuropsychological measures are differentially sensitive to different stages of preclinical AD [61].

Cross-sectional studies [62] performed using appropriate cognitive assessments from populations with autosomal dominant AD have demonstrated changes in cognition 15 years before disease onset.

Another point that should be addressed is whether it is worth continuing to use cognitive tests that are generated and used exclusively for research purposes rather than using cognitive tests that are validated in common daily practice. The use of the latter has advantages, particularly practical value. The current situation is that once drugs are licensed the scales used in trials are typically no longer applied preventing comparisons between the experimental and clinical situations, relevant and fundamental to obtaining solid feedback about the real efficacy of the drugs.

A final consideration concerns the meaning and utility that cognitive tests used as outcome measures should provide in pharmacological trials [54]. Since we have just entered the era of disease-modifying drugs and are far from curative therapies, the only real measure that cognitive tests can offer, at present, is to determine whether the drug is able to slow the progression of the disease and to vary its course over time. To this end, it is more useful to use the score not so much in terms of change in the score, but along the timeline, that is, to allow us to understand how much the drug is able to “put the disease back on the timeline” compared to a placebo [54].

Recently, several authors have attempted to design new neuropsychological test batteries that may be more useful in the early

TABLE 2 | Core cognitive domains in AD variants.

Examples of sensitive and specific neuropsychological tests widely used in the clinical practice					References
Domains	Subdomains				
Primary domains impaired in the classical amnesic variant	Memory	Verbal Long term episodic memory	RAVLT; FCSRT	[67, 68]	
			Babcock story recall	[65]	
	Attention		Logical Memory from Wechsler Memory Scale	[63, 64]	
		Visual long term episodic memory	ROCF immediate recall	[68]	
		Selective attention	Attentive Matrices; Cancellation tasks	[65, 67]	
Secondary domains impaired in the classical amnesic variant or primarily impaired in non-classical amnesic AD	Executive functions	Divided attention	TMT	[67]	
		Sustained attention	Digit symbol substitution	[64]	
		Global assessment	Frontal Assessment Battery	[65]	
		Logical abilities	Raven's Progressive Matrices	[64, 67]	
	Language	Naming	Boston Naming Test	[67]	
		Letter fluency	Graded Naming Test	[67]	
	Visuoperceptual and visuospatial abilities	Semantic fluency	FAS	[65]	
			Animal, vegetable and fruit	[66]	
		Semantic knowledge	Animals	[66]	
		Visuoperception	Object function recognition	[68]	
	Spatial orientation	VOSP battery—silhouettes	[68]		
		Judgment line orientation	[64]		

and preclinical stages of Alzheimer's disease. Examples include the Preclinical Alzheimer Cognitive Composite (PASS) [63] and the Alzheimer's Prevention Initiative Composite Cognitive Test (APCC) [64].

There is also evidence that certain cognitive tests can predict conversion to dementia in patients with MCI [65–69] with sensitivity reaching more than 90% and excellent specificity [68] with some neuropsychological measures revealing an overall accuracy of >90% [68]. We propose, as one possible way forward—based on current literature and the pathophysiological model of Alzheimer's disease—a set of clinical tests commonly used in daily practice that explore core cognitive domains (Table 2). The tests selected in the table are derived from studies in which sensitivity and specificity were investigated in detail and were found to be excellent. These tools could also be employed in the context of pharmacological trials, thereby ensuring that the assessment is sensitive and specific, consistent with the cognitive impairment model of the disease, and ecologically valid due to its alignment with routine clinical practice.

New testing procedures grounded in neuroscience research need to be developed, leveraging technological advances and enhanced neuropsychological expertise among clinicians specializing in dementia. In the context of disease-modifying drugs, for instance, clinical trials in prodromal and preclinical Alzheimer's disease require rapid and frequent assessments of cognition and functional abilities, as well as the identification of cognitive markers that are sensitive, reliable, and reproducible in real-world settings. Alternatively, the advent of new technologies offers the opportunity to move away from traditional paper and pencil tests to assess cognition. Computer-based tests facilitate the recording of speed of response, likely more sensitive to minimal change than scores based on correct/incorrect alone. Moreover, they facilitate frequent repeated testing, crucial for pharmacological trials. Technological advances have elicited novel data, leading to better understanding of the functions of brain regions affected in AD. A test of virtual reality navigation [70], for example, revealed larger spatial navigation errors in people with MCI who are biomarker-positive for AD compared to those who are biomarker negative. Crucially, the task proved more sensitive than traditional “gold standard” neuropsychological tests, highlighting the importance of spatial navigation changes as an early symptom of AD. Moreover, the study demonstrates the potential of computer technology for the development of cognitive tests that have ecological validity, relevant to daily life.

In summary, new testing procedures grounded in neuroscience research need to be developed, leveraging technological advances and enhanced neuropsychological expertise among clinicians specializing in dementia. In the context of disease-modifying drugs, clinical trials in prodromal and preclinical Alzheimer's disease require rapid and frequent assessments of cognition and functional abilities, as well as the identification of cognitive markers that are sensitive, reliable, and reproducible in real-world settings.

9 | Conclusion

Test batteries currently available do not adequately capture the minimal and domain-specific changes of the earliest stages of

AD. They are no longer adequate for use in experimental trials of anti-dementia drugs.

There is an urgent need to move toward a new era in cognitive testing, to allow improved detection of change and to avoid the prospect of false negative trial results: a potentially useful drug is ruled out because of inadequate neuropsychology. Improved cognitive tests might also help to resolve fundamental controversies about the underlying pathophysiology of Alzheimer's disease.

Author Contributions

Simona Luzzi: conceptualization, data curation, writing – original draft, investigation, writing – review and editing. **Julie S. Snowden:** conceptualization, writing – original draft, writing – review and editing, investigation, data curation.

Funding

The authors have nothing to report.

Acknowledgement

Open access publishing facilitated by Università Politecnica delle Marche, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. C. H. van Dyck, C. J. Swanson, P. Aisen, et al., “Lecanemab in Early Alzheimer's Disease,” *New England Journal of Medicine* 388, no. 1 (2023): 9–21, <https://doi.org/10.1056/NEJMoa2212948>.
2. J. R. Sims, J. A. Zimmer, C. D. Evans, et al., “Donanemab in Early Symptomatic Alzheimer Disease,” *JAMA* 330, no. 6 (2023): 512, <https://doi.org/10.1001/jama.2023.13239>.
3. S. F. Ackley, S. C. Zimmerman, W. D. Brenowitz, et al., “Effect of Reductions in Amyloid Levels on Cognitive Change in Randomized Trials: Instrumental Variable Meta-Analysis,” *BMJ* 372 (2021): n156, <https://doi.org/10.1136/bmj.n156>.
4. S. Budd Haeberlein, P. S. Aisen, F. Barkhof, et al., “Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease,” *Journal of Prevention of Alzheimer's Disease* 9, no. 2 (2022): 197–210, <https://doi.org/10.14283/jpad.2022.30>.
5. A. Granzotto and S. L. Sensi, “Once Upon a Time, the Amyloid Cascade Hypothesis,” *Ageing Research Reviews* 93 (2024): 102161, <https://doi.org/10.1016/j.arr.2023.102161>.
6. M. T. Kurkinen, “Donanemab: Not Two Without a Third,” *Advances in Clinical and Experimental Medicine* 32, no. 10 (2023): 1085–1087, <https://doi.org/10.17219/acem/172673>.
7. N. K. Robakis, “Are A β and Its Derivatives Causative Agents or Innocent Bystanders in AD?,” *Neurodegenerative Diseases* 7, no. 1–3 (2010): 32–37, <https://doi.org/10.1159/000266476>.
8. G. Logroscino, D. Urso, and R. Savica, “Descriptive Epidemiology of Neurodegenerative Diseases: What Are the Critical Questions?,”

- Neuroepidemiology* 56, no. 5 (2022): 309–318, <https://doi.org/10.1159/000525639>.
9. C. R. Jack, D. A. Bennett, K. Blennow, et al., “NIA-AA Research Framework: Toward a Biological Definition of Alzheimer’s Disease,” *Alzheimer’s & Dementia* 14, no. 4 (2018): 535–562, <https://doi.org/10.1016/j.jalz.2018.02.018>.
 10. C. R. Jack, J. S. Andrews, T. G. Beach, et al., “Revised Criteria for Diagnosis and Staging of Alzheimer’s Disease: Alzheimer’s Association Workgroup,” *Alzheimer’s & Dementia* 20, no. 8 (2024): 5143–5169, <https://doi.org/10.1002/alz.13859>.
 11. A. M. Wessels, E. B. Denney, S. A. Dowsett, S. P. Dickson, and S. B. Hendrix, “Meaningful Clinical Changes in Alzheimer Disease Measured With the iADRS and Illustrated Using the Donanemab TRAILBLAZER-ALZ Study Findings,” *Neurology Clinical Practice* 13, no. 2 (2023): e200127, <https://doi.org/10.1212/CPJ.0000000000200127>.
 12. W. G. Rosen, R. C. Mohs, and K. L. Davis, “A New Rating Scale for Alzheimer’s Disease,” *American Journal of Psychiatry* 141, no. 11 (1984): 1356–1364, <https://doi.org/10.1176/ajp.141.11.1356>.
 13. S. L. Rogers, R. S. Doody, R. C. Mohs, and L. T. Friedhoff, “Donepezil Improves Cognition and Global Function in Alzheimer Disease: A 15-Week, Double-Blind, Placebo-Controlled Study. Donepezil Study Group,” *Archives of Internal Medicine* 158, no. 9 (1998): 1021–1031, <https://doi.org/10.1001/archinte.158.9.1021>.
 14. M. Rösler, R. Anand, A. Cicin-Sain, et al., “Efficacy and Safety of Rivastigmine in Patients With Alzheimer’s Disease: International Randomised Controlled Trial,” *BMJ* 318, no. 7184 (1999): 633–638, <https://doi.org/10.1136/bmj.318.7184.633>.
 15. P. N. Tariot, P. R. Solomon, J. C. Morris, P. Kershaw, S. Lilienfeld, and C. Ding, “A 5-Month, Randomized, Placebo-Controlled Trial of Galantamine in AD,” *Neurology* 54, no. 12 (2000): 2269–2276, <https://doi.org/10.1212/WNL.54.12.226>.
 16. B. Reisberg, R. Doody, A. Stöffler, F. Schmitt, S. Ferris, and H. J. Möbius, “Memantine in Moderate-to-Severe Alzheimer’s Disease,” *New England Journal of Medicine* 348, no. 14 (2003): 1333–1341, <https://doi.org/10.1056/NEJMoa013128>.
 17. P. S. Aisen, “Cognitive/Clinical Endpoints for Pre-Dementia AD Trials,” *Journal of Prevention of Alzheimer’s Disease* 2 (2015): 82–84, <https://doi.org/10.14283/jpad.2015.62>.
 18. J. Podhorna, T. Krahne, M. Shear, and J. E. Harrison, “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 Research & Therapy* 8, no. 1 (2016): 8, <https://doi.org/10.1186/s13195-016-0170-5>.
 19. A. Karin, K. Hannesdottir, J. Jaeger, et al., “Psychometric Evaluation of ADAS-Cog and NTB for Measuring Drug Response,” *Acta Neurologica Scandinavica* 129, no. 2 (2014): 114–122, <https://doi.org/10.1111/ane.12153>.
 20. J. K. Kueper, M. Speechley, and M. Montero-Odasso, “The Alzheimer’s Disease Assessment Scale–Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review,” *Journal of Alzheimer’s Disease* 63, no. 2 (2018): 423–444, <https://doi.org/10.3233/JAD-170991>.
 21. N. Raghavan, M. N. Samtani, M. Farnum, et al., “The ADAS-Cog Revisited: Novel Composite Scales Based on ADAS-Cog to Improve Efficiency in MCI and Early AD Trials,” *Alzheimer’s & Dementia* 9, no. 1S (2013): S21–S31, <https://doi.org/10.1016/j.jalz.2012.05.2187>.
 22. J. Skinner, J. O. Carvalho, G. G. Potter, 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 and Behavior* 6, no. 4 (2012): 489–501, <https://doi.org/10.1007/s11682-012-9166-3>.
 23. M. Sano, R. Raman, J. Emond, et al., “Adding Delayed Recall to the Alzheimer Disease Assessment Scale Is Useful in Studies of Mild Cognitive Impairment but Not Alzheimer Disease,” *Alzheimer Disease & Associated Disorders* 25, no. 2 (2011): 122–127, <https://doi.org/10.1097/WAD.0b013e3181f883b7>.
 24. D. J. Connor and M. N. Sabbagh, “Administration and Scoring Variance on the ADAS-Cog,” *Journal of Alzheimer’s Disease* 15, no. 3 (2008): 461–464, <https://doi.org/10.3233/JAD-2008-15312>.
 25. J. R. Kunst and A. Mesoudi, “Decoding the Dynamics of Cultural Change: A Cultural Evolution Approach to the Psychology of Acculturation,” *Personality and Social Psychology Review* 29, no. 2 (2025): 111–144, <https://doi.org/10.1177/10888683241258406>.
 26. M. D. Lezak, D. B. Howieson, and D. W. Loring, *Neuropsychological Assessment*, 5th ed. (Oxford University Press, 2012).
 27. T. H. P. Draak, B. T. A. de Greef, C. G. Faber, and I. S. J. Merckies, “The Minimum Clinically Important Difference: Which Direction to Take,” *European Journal of Neurology* 26, no. 6 (2019): 850–855, <https://doi.org/10.1111/ene.13941>.
 28. A. Wuestefeld, A. Pichet Binette, D. van Westen, et al., “Medial Temporal Lobe Atrophy Patterns in Early-Versus Late-Onset Amnesic Alzheimer’s Disease,” *Alzheimer’s Research & Therapy* 16, no. 1 (2024): 204, <https://doi.org/10.1186/s13195-024-01571-z>.
 29. R. C. Mohs, D. Knopman, R. C. Petersen, et al., “Development of Cognitive Instruments for Use in Clinical Trials of Antidementia Drugs: Additions to the Alzheimer’s Disease Assessment Scale That Broaden Its Scope,” *Alzheimer Disease & Associated Disorders* 11 (1997): S13–S21.
 30. S. Weintraub, A. H. Wicklund, and D. P. Salmon, “The Neuropsychological Profile of Alzheimer Disease,” *Cold Spring Harbor Perspectives in Medicine* 2, no. 4 (2012): a006171, <https://doi.org/10.1101/cshperspect.a006171>.
 31. A. M. Wessels, E. R. Siemers, P. Yu, et al., “A Combined Measure of Cognition and Function for Clinical Trials: The Integrated Alzheimer’s Disease Rating Scale (IADRS),” *Journal of Prevention of Alzheimer’s Disease* 2 (2015): 227–241, <https://doi.org/10.14283/jpad.2015.82>.
 32. R. Ihl, S. Ferris, P. Robert, B. Winblad, S. Gauthier, and F. Tenenigkeit, “Detecting Treatment Effects With Combinations of the ADAS-Cog Items in Patients With Mild and Moderate Alzheimer’s Disease,” *International Journal of Geriatric Psychiatry* 27, no. 1 (2012): 15–21, <https://doi.org/10.1002/gps.2679>.
 33. P. M. Doraiswamy, L. Kaiser, F. Bieber, and R. L. Garman, “The Alzheimer’s Disease Assessment Scale: Evaluation of Psychometric Properties and Patterns of Cognitive Decline in Multicenter Clinical Trials of Mild to Moderate Alzheimer’s Disease,” *Alzheimer Disease and Associated Disorders* 15, no. 4 (2001): 174–183, <https://doi.org/10.1097/00002093-200110000-00003>.
 34. D. Bertens, D. L. Knol, P. Scheltens, and P. J. Visser, “Temporal Evolution of Biomarkers and Cognitive Markers in the Asymptomatic, MCI, and Dementia Stage of Alzheimer’s Disease,” *Alzheimer’s & Dementia* 11, no. 5 (2015): 511–522, <https://doi.org/10.1016/j.jalz.2014.05.1754>.
 35. S. E. O’Bryant, “Staging Dementia Using Clinical Dementia Rating Scale Sum of Boxes Scores,” *Archives of Neurology* 65, no. 8 (2008): 1091, <https://doi.org/10.1001/archneur.65.8.1091>.
 36. T. Daly, K. P. Kepp, and B. P. Imbimbo, “Are Lecanemab and Donanemab Disease-Modifying Therapies?,” *Alzheimer’s & Dementia* 20, no. 9 (2024): 6659–6661, <https://doi.org/10.1002/alz.14114>.
 37. H. Liu-Seifert, E. Siemers, K. Price, et al., “Cognitive Impairment Precedes and Predicts Functional Impairment in Mild Alzheimer’s Disease,” *Journal of Alzheimer’s Disease* 47, no. 1 (2015): 205–214, <https://doi.org/10.3233/JAD-142508>.

38. M. M. Williams, M. Storandt, C. M. Roe, and J. C. Morris, "Progression of Alzheimer's Disease as Measured by Clinical Dementia Rating Sum of Boxes Scores," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 9, no. 1 Suppl (2013): S39–S44, <https://doi.org/10.1016/j.jalz.2012.01.005>.
39. FAO, "Guidance for Industry: Alzheimer's Disease: Developing Drugs for the Treatment of Early-Stage Disease," US Department of Health and Human Services, Food and Drug Administration, Washington, DC (2013), <https://www.fda.gov/media/85614/download>.
40. L. S. Schneider and J. T. Olin, "Clinical Global Impressions in Alzheimer's Clinical Trials," *International Psychogeriatrics* 8, no. 2 (1996): 277–288; discussion 288–290, <https://doi.org/10.1017/s1041610296002645>.
41. L. S. Schneider, J. T. Olin, R. S. Doody, et al., "Validity and Reliability of the Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change. The Alzheimer's Disease Cooperative Study," *Alzheimer Disease and Associated Disorders* 11 (1997): S22–S32, <https://doi.org/10.1097/00002093-199700112-00004>.
42. R.-C. Tzeng, Y.-W. Yang, K.-C. Hsu, H.-T. Chang, and P.-Y. Chiu, "Sum of Boxes of the Clinical Dementia Rating Scale Highly Predicts Conversion or Reversion in Predementia Stages," *Frontiers in Aging Neuroscience* 14 (2022): 1021792, <https://doi.org/10.3389/fnagi.2022.1021792>.
43. FAO, "Guidance for Industry: Early Alzheimer's Disease: Developing Drugs for Treatment," US Department of Health and Human Services, Food and Drug Administration, Washington, DC (2018), <https://www.fda.gov/regulatoryinformation/search-fda-guidance-documents/alzheimers-disease-developing-drugs-treatment-guidance-industry>.
44. A. M. Wessels, M. Belger, J. A. Johnston, et al., "Demonstration of Clinical Meaningfulness of the Integrated Alzheimer's Disease Rating Scale (iADRS): Association Between Change in iADRS Scores and Patient and Caregiver Health Outcomes," *Journal of Alzheimer's Disease* 88, no. 2 (2022): 577–588, <https://doi.org/10.3233/JAD-220303>.
45. M. H. Ebell, H. C. Barry, K. Baduni, and G. Grasso, "Clinically Important Benefits and Harms of Monoclonal Antibodies Targeting Amyloid for the Treatment of Alzheimer Disease: A Systematic Review and Meta-Analysis," *Annals of Family Medicine* 22, no. 1 (2024): 50–62, <https://doi.org/10.1370/afm.3050>.
46. C. J. Lansdall, F. McDougall, L. M. Butler, et al., "Establishing Clinically Meaningful Change on Outcome Assessments Frequently Used in Trials of Mild Cognitive Impairment due to Alzheimer's Disease," *Journal of Prevention of Alzheimer's Disease* 10, no. 1 (2023): 9–18, <https://doi.org/10.14283/jpad.2022.102>.
47. J. S. Andrews, U. Desai, N. Y. Kirson, M. L. Zichlin, D. E. Ball, and B. R. Matthews, "Disease Severity and Minimal Clinically Important Differences in Clinical Outcome Assessments for Alzheimer's Disease Clinical Trials," *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 5, no. 1 (2019): 354–363, <https://doi.org/10.1016/j.trci.2019.06.005>.
48. R. Jaeschke, J. Singer, and G. H. Guyatt, "Measurement of Health Status," *Controlled Clinical Trials* 10, no. 4 (1989): 407–415, [https://doi.org/10.1016/0197-2456\(89\)90005-6](https://doi.org/10.1016/0197-2456(89)90005-6).
49. R. Duara and W. Barker, "Heterogeneity in Alzheimer's Disease Diagnosis and Progression Rates: Implications for Therapeutic Trials," *Neurotherapeutics* 19, no. 1 (2022): 8–25, <https://doi.org/10.1007/s13311-022-01185-z>.
50. A. A. Tahami Monfared, S. Fu, N. Hummel, et al., "Estimating Transition Probabilities Across the Alzheimer's Disease Continuum Using a Nationally Representative Real-World Database in the United States," *Neurology and Therapy* 12, no. 4 (2023): 1235–1255, <https://doi.org/10.1007/s40120-023-00498-1>.
51. F. de Oliveira Silva, J. V. Ferreira, J. Plácido, et al., "Stages of Mild Cognitive Impairment and Alzheimer's Disease Can Be Differentiated by Declines in Timed Up and Go Test: A Systematic Review and Meta-Analysis," *Archives of Gerontology and Geriatrics* 85 (2019): 103941, <https://doi.org/10.1016/j.archger.2019.103941>.
52. M. Davis, T. O'Connell, S. Johnson, et al., "Estimating Alzheimer's Disease Progression Rates From Normal Cognition Through Mild Cognitive Impairment and Stages of Dementia," *Current Alzheimer Research* 15, no. 8 (2018): 777–788, <https://doi.org/10.2174/156720501566180119092427>.
53. P. N. Tariot, M. Boada, K. L. Lanctôt, et al., "Relationships of Change in Clinical Dementia Rating (CDR) on Patient Outcomes and Probability of Progression: Observational Analysis," *Alzheimer's Research & Therapy* 16, no. 1 (2024): 36, <https://doi.org/10.1186/s13195-024-01399-7>.
54. M. Jin and J. M. Noble, "What's in It for Me? Contextualizing the Potential Clinical Impacts of Lecanemab, Donanemab, and Other Anti- β -Amyloid Monoclonal Antibodies in Early Alzheimer's Disease," *Eneuro* 11, no. 9 (2024): ENEURO.0088-24.2024, <https://doi.org/10.1523/ENEURO.0088-24.2024>.
55. J. Cummings, "Meaningful Benefit and Minimal Clinically Important Difference (MCID) in Alzheimer's Disease: Open Peer Commentary," *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 9, no. 3 (2023): e12411, <https://doi.org/10.1002/trc2.12411>.
56. J. Graff-Radford, K. X. X. Yong, L. G. Apostolova, et al., "New Insights Into Atypical Alzheimer's Disease in the Era of Biomarkers," *Lancet Neurology* 20, no. 3 (2021): 222–234, [https://doi.org/10.1016/S1474-4422\(20\)30440-3](https://doi.org/10.1016/S1474-4422(20)30440-3).
57. D. Putcha, Y. Katsumi, A. Touroutoglou, et al., "Heterogeneous Clinical Phenotypes of Sporadic Early-Onset Alzheimer's Disease: A Neuropsychological Data-Driven Approach," *Alzheimer's Research & Therapy* 17, no. 1 (2025): 38, <https://doi.org/10.1186/s13195-025-01689-8>.
58. G. Marucci, M. Buccioni, D. D. Ben, C. Lambertucci, R. Volpini, and F. Amenta, "Efficacy of Acetylcholinesterase Inhibitors in Alzheimer's Disease," *Neuropharmacology* 190 (2021): 108352, <https://doi.org/10.1016/j.neuropharm.2020.108352>.
59. C. R. Jack, D. S. Knopman, W. J. Jagust, et al., "Hypothetical Model of Dynamic Biomarkers of the Alzheimer's Pathological Cascade," *Lancet Neurology* 9, no. 1 (2010): 119–128, [https://doi.org/10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6).
60. C. R. Jack, D. S. Knopman, W. J. Jagust, et al., "Tracking Pathophysiological Processes in Alzheimer's Disease: An Updated Hypothetical Model of Dynamic Biomarkers," *Lancet Neurology* 12, no. 2 (2013): 207–216, [https://doi.org/10.1016/S1474-4422\(12\)70291-0](https://doi.org/10.1016/S1474-4422(12)70291-0).
61. S. Duke Han, C. P. Nguyen, N. H. Stricker, and D. A. Nation, "Detectable Neuropsychological Differences in Early Preclinical Alzheimer's Disease: A Meta-Analysis," *Neuropsychology Review* 27, no. 4 (2017): 305–325, <https://doi.org/10.1007/s11065-017-9345-5>.
62. R. J. Bateman, C. Xiong, T. L. S. Benzinger, et al., "Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease," *New England Journal of Medicine* 367, no. 9 (2012): 795–804, <https://doi.org/10.1056/NEJMoa1202753>.
63. M. C. Donohue, R. A. Sperling, D. P. Salmon, et al., "The Preclinical Alzheimer Cognitive Composite," *JAMA Neurology* 71, no. 8 (2014): 961, <https://doi.org/10.1001/jamaneurol.2014.803>.
64. J. B. Langbaum, N. N. Ellison, A. Caputo, et al., "The Alzheimer's Prevention Initiative Composite Cognitive Test: A Practical Measure for Tracking Cognitive Decline in Preclinical Alzheimer's Disease," *Alzheimer's Research & Therapy* 12, no. 1 (2020): 66, <https://doi.org/10.1186/s13195-020-00633-2>.
65. C. Cazzolli, M. Chierici, M. Dallabona, C. Guella, and G. Jurman, "Neuropsychological Tests and Machine Learning: Identifying Predictors of MCI and Dementia Progression," *Aging Clinical and Experimental Research* 37, no. 1 (2025): 79, <https://doi.org/10.1007/s40520-025-02962-4>.

66. M. R. Ayers, J. Bushnell, S. Gao, et al., "Verbal Fluency Response Times Predict Incident Cognitive Impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 14, no. 1 (2022): e12277, <https://doi.org/10.1002/dad2.12277>.
67. T. Pereira, F. L. Ferreira, S. Cardoso, et al., "Neuropsychological Predictors of Conversion From Mild Cognitive Impairment to Alzheimer's Disease: A Feature Selection Ensemble Combining Stability and Predictability," *BMC Medical Informatics and Decision Making* 18, no. 1 (2018): 137, <https://doi.org/10.1186/s12911-018-0710-y>.
68. S. Belleville, C. Fouquet, C. Hudon, H. T. V. Zomahoun, and J. Croteau, "Neuropsychological Measures That Predict Progression From Mild Cognitive Impairment to Alzheimer's Type Dementia in Older Adults: A Systematic Review and Meta-Analysis," *Neuropsychology Review* 27, no. 4 (2017): 328–353, <https://doi.org/10.1007/s11065-017-9361-5>.
69. G. H. Weissberger, J. V. Strong, K. B. Stefanidis, M. J. Summers, M. W. Bondi, and N. H. Stricker, "Diagnostic Accuracy of Memory Measures in Alzheimer's Dementia and Mild Cognitive Impairment: A Systematic Review and Meta-Analysis," *Neuropsychology Review* 27, no. 4 (2017): 354–388, <https://doi.org/10.1007/s11065-017-9360-6>.
70. D. Howett, A. Castegnaro, K. Krzwicka, et al., "Differentiation of Mild Cognitive Impairment Using an Entorhinal Cortex-Based Test of Virtual Reality Navigation," *Brain* 142, no. 6 (2019): 1751–1766, <https://doi.org/10.1093/brain/awz116>.