Clinically Relevant Changes for Cognitive Outcomes in Preclinical and Prodromal Cognitive Stages
Implications for Clinical Alzheimer Trials

Emma Borland, MD, Chris Edgar, PhD, Erik Stomrud, MD, PhD, Nicholas Cullen, BS, Oskar Hansson, MD, PhD, and Sebastian Palmqvist, MD, PhD

Neurology® 2022;99:e1142-e1153. doi:10.1212/WNL.0000000000200817

Abstract

Background and Objectives
Identifying a clinically meaningful change in cognitive test score is essential when using cognition as an outcome in clinical trials. This is especially relevant because clinical trials increasingly feature novel composites of cognitive tests. Our primary objective was to establish minimal clinically important differences (MCIDs) for commonly used cognitive tests, using anchor-based and distribution-based methods, and our secondary objective was to investigate a composite cognitive measure that best predicts a minimal change in the Clinical Dementia Rating—Sum of Boxes (CDR-SB).

Methods
From the Swedish BioFINDER cohort study, we consecutively included cognitively unimpaired (CU) individuals with and without subjective or mild cognitive impairment (MCI). We calculated MCIDs associated with a change of ≥0.5 or ≥1.0 on CDR-SB for Mini-Mental State Examination (MMSE), ADAS-Cog delayed recall 10-word list, Stroop, Letter S Fluency, Animal Fluency, Symbol Digit Modalities Test (SDMT) and Trailmaking Test (TMT) A and B, and triangulated MCIDs for clinical use for CU, MCI, and amyloid-positive CU participants. For investigating cognitive measures that best predict a change in CDR-SB of ≥0.5 or ≥1.0 point, we conducted receiver operating characteristic analyses.

Results
Our study included 451 cognitively unimpaired individuals, 90 with subjective cognitive decline and 361 without symptoms of cognitive decline (pooled mean follow-up time 32.4 months, SD 26.8, range 12–96 months), and 292 people with MCI (pooled mean follow-up time 19.2 months, SD 19.0, range 12–72 months). We identified potential triangulated MCIDs (cognitively unimpaired; MCI) on a range of cognitive test outcomes: MMSE −1.5, −1.7; ADAS delayed recall 1.4, 1.1; Stroop 5.5, 9.3; Animal Fluency: −2.8, −2.9; Letter S Fluency −2.9, −1.8; SDMT: −3.5, −3.8; TMT A 11.7, 13.0; and TMT B 24.4, 20.1. For amyloid-positive CU, we found the best predicting composite cognitive measure included gender and changes in ADAS delayed recall, MMSE, SDMT, and TMT B. This produced an AUC of 0.87 (95% CI 0.79–0.94, sensitivity 75%, specificity 88%).

Discussion
Our MCIDs may be applied in clinical practice or clinical trials for identifying whether a clinically relevant change has occurred. The composite measure can be useful as a clinically relevant cognitive test outcome in preclinical AD trials.

From the Clinical Memory Research Unit (E.B., E.S., N.C., O.H., S.P.), Department of Clinical Sciences, Lund University; Department of Neurology(E.B.), Skåne University Hospital, Malmö, Sweden; Department of Clinical Science (C.E.), Cogstate, London, United Kingdom; and Memory Clinic (E.S., O.H., S.P.), Skåne University Hospital, Malmö, Sweden.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article. The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.
The minimal clinically important difference (MCID) is defined as the smallest change on a measure that is reliably associated with a meaningful change in a patient’s clinical status, function, or quality of life. It is important to decide the smallest change in an outcome that constitutes a clinically meaningful change—that is, MCID—to interpret whether, for example, the treatment effect measured using a cognitive test is clinically relevant or whether a change in cognitive testing during a clinical follow-up represents a clinically meaningful change in cognition. MCIDs are thus necessary to make accurate clinical decisions and to design clinical trials with the statistical power to detect an effect equal to or greater than the MCID.

In the 2018 US Food and Drug Administration (FDA) guidance for clinical trials in early AD, the guidance introduced a clinical staging framework for AD stages 1–3. Stage 1 includes individuals with abnormal biomarkers without cognitive complaints or detectable decline even on sensitive tests. Stage 2 includes individuals with subtle cognitive effects without functional deficits, and stage 3 includes individuals beginning to have difficulties with daily tasks. To presume a drug has a clinically beneficial effect for individuals in stage 2, the agency states that a pattern of beneficial effects on neuropsychological assessments is more persuasive if seen on multiple tests and that if only seen on 1 assessment, it needs to show a large magnitude of effect to be persuasive of a beneficial effect. However, for many cognitive assessments, the magnitude that corresponds to a clinically meaningful effect compared with that for placebo is unknown.

Several methods exist for estimating a meaningful clinical effect, among which the most well-established are anchor-based and distribution-based estimates. Anchor-based approaches to determine meaningful within-patient change involve the use of an external reference with an already established relevance. Distribution-based, or internal estimates, use statistical properties of the measures themselves, and of them, the most common are effect size metrics—for example, the SD and the SEM that incorporate some measure of scale reliability (e.g., test-retest or Cronbach α as a measure of internal consistency reliability).

In addition to establishing clinically relevant cutoffs for test changes, it is also important to determine which tests best represent clinically relevant changes. The preclinical Alzheimer cognitive composite (PACC) has earlier been proposed as an outcome measure sensitive for early cognitive changes in AD (stages 1 and 2). The PACC was initially created by selecting 4 well-established cognitive tests that are sensitive to detecting change/worsening in prodromal and mild dementia and with sufficient range to also detect early decline in preclinical stages of disease. However, the PACC was established purely based on the presumed sensitivity to detect changes and not whether the changes were clinically meaningful. We propose that by developing and validating cognitive composites and test batteries using predictive validity for a clinically important change incorporating anchor-based approaches, more relevant outcomes may be developed than by focusing on within-measure change/worsening, which is distribution-based alone.

The aims of this study were (1) to establish cutoffs for cognitive test changes for use to conclude whether a meaningful magnitude of treatment effect has been achieved and (2) to investigate which single and combinations of cognitive test differences best corresponds to a clinically meaningful decline. In addition to examining the second aim in cognitively unimpaired (CU) participants and participants with mild cognitive impairment (MCI), we also examined this in Aβ-positive CU participants because this is a group of special interest in present and future clinical AD trials.

**Methods**

**Population**

The participants in the study were consecutively included from the prospective Swedish BioFINDER study (biofinder.se), and participants for this study were enrolled from July 6, 2009, to March 4, 2015. The population consisted of 451 CU individuals and 292 people classified as experiencing mild cognitive impairment (MCI). In the CU group, 90 individuals experienced subjective cognitive decline, and 361 people were cognitively healthy controls based on a structured assessment. The individuals were followed up longitudinally (for CU pooled mean for all different tests 32.4 months (pooled SD 26.8, range 12–96 months), MCI pooled mean 19.2 months (pooled SD 19.0, range 12–72 months), with a mean number of data points of 1588 for CU and 727 for MCI.

MCI was defined according to the performance on a comprehensive neuropsychological battery, as previously described. All cognitively unimpaired individuals had a Clinical Dementia Rating—Sum of Boxes at inclusion of 0. Participants with MCI were excluded after converting to major
neurocognitive disorder. Participants were assessed by physicians well experienced in dementia disorders, underwent a physical examination, MRI scan, lumbar puncture, and cognitive assessments, and were rated with the CDR. Participants experiencing cognitive symptoms at baseline (subjective cognitive disease or MCI) were followed up annually, while participants without cognitive symptoms at baseline were examined every second year by physicians.

Cognitive Tests
Eight cognitive tests were examined in this study, covering the cognitive domains of executive function, attention, episodic and semantic memory, and visuospatial function. Participants were examined with the Mini-Mental State Examination (MMSE), the Alzheimer Disease Assessment Scale (ADAS) 10-word delayed recall, Letter S Fluency, Animal Fluency, Stroop Color and Word Test (Stroop), Trailmaking Tests A and B, and Symbol Digit Modalities Test. Further explanation of tests, what they assess, and how points are counted are described in eMethods.

Clinical Dementia Rating
Clinical Dementia Rating (CDR) is an ordinal scale with scores of 0–3 points used to quantify the functional effect of cognitive impairment (0 = none, 0.5 = questionable, 1 = mild, 2 = moderate, and 3 = severe) in domains (box scores) of memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Participants in this study were rated on the CDR by a thorough review of patient records where dementia experts assessed cognitive symptoms and activities of daily living (ADL) in comprehensive semistructured interviews with patients and informants, including informant-based questionnaires of ADL (the functional activities questionnaire) and cognitive symptoms (the CIMP-QUEST) (supported if necessary by cognitive test results). The CDR scale provides a quantitative index of impairment, referred to as the Sum of Boxes or CDR-SB (range of 0–18), and may also be scored in the CDR global severity stage score (range 0–3) using an algorithm. An increase in the CDR-SB score has been identified as having both face and predictive validity to identify people who are later diagnosed with probable AD or another dementia. For predementia AD, a single box score increment of either 0.5 or 1.0 has been proposed to capture efficacy and clinical relevance for early AD. A change of 1 in CDR-SB could be either a change of 0.5 points in 2 boxes or a change of 1 point in 1 box.

Determining Amyloid Positivity
The procedure and analysis of CSF followed the Alzheimer Association Flowchart for CSF biomarkers. We used the ratio for Aβ42/40 that we establish acquire through CSF analysis. CSF Aβ42 and Aβ40 were analyzed using the Roche Elecsys CSF immunoassays (NeuroToolKit) on all participants. The cutoff for Aβ42/40 was established with mixture modeling statistics and set at 0.066.

Statistical Analysis
The psychometric criterion reliable change index (RCI) is used to evaluate whether a change over time of an individual score is considered statistically significant. RCI provides a CI that represents the predicted changes that would occur if a patient’s test score does not change significantly from one assessment to another. We calculated the RCI for all 8 test differences with a 90% CI (the most common CI for an RCI) for CU participants and participants with MCI. This is performed for tests separately using the following equation:

\[
SE_{\text{meas}} = SD \sqrt{1-r}
\]

\[
SE_{\text{diff}} = \sqrt{2 \times (SE_{\text{meas}})^2}
\]

\[
RCI = \pm SE_{\text{diff}} \times 1.64
\]

SD = Standard deviation of the test score at baseline
r = Pearson coefficient for test results in cognitively stable individuals
SE_{\text{meas}} = standard error of measurement
SE_{\text{diff}} = standard error of differentiation

Estimates of effect size (ES) are useful for determining the magnitude or size of an effect, the relative contribution of different factors or the same factor in different circumstances, and the power of an analysis. ES is defined as a mean difference in score divided by standard deviation of baseline scores. An ES of 0.5 is generally considered a moderate clinically significant change, whereas an ES of 0.2 is considered a small change and 0.8 a large change. The standardized response mean (SRM) is an effect size used to measure the responsiveness of outcome measures (the ability to detect change over time), defined as mean difference in score divided by SD of the change from previous visit score. ES and SRM were calculated for all test changes in CU and MCI participants. Experts have previously defined a clinically meaningful cognitive decline as a decline in cognitive function of 0.5 or more SDs from baseline cognitive scores, which we present in our results.

For the anchor-based approach, we analyzed mean differences in cognitive test scores anchored to differences in CDR-Sum of Boxes (CDR-SB) scores. For the CU individuals, we used a change of CDR-SB ≥0.5 points and for the MCI group a change of ≥1 point as anchors to represent a clinically meaningful change. We calculated the mean, SD, and ES for changes in the cognitive tests separately for meaningful decline (CDR-SB difference of ≥0.5 and ≥1) and no meaningful decline (CDR-SB difference of 0).

According to previously described methods, MCIDs are recommended to be triangulated (calculated of the arithmetic mean) to produce a final MCID for each mean value. Triangulation integrates results from ratings with clinical changes,
Results

Baseline characteristics are summarized in Table 1. Four hundred fifty-one people were included in the cognitively unimpaired (CU) group (mean number of data points for each test 1,588) and 292 people in the MCI group (mean number of data points 727). The mean (SD) age of the CU group was 72.9 (5.5) years and that of the MCI group 71.1 (5.5) years. 116 (16%) people in the CU group were amyloid positive. Pooled mean time between test results for CU participants was 32.4 months (pooled SD 26.8), for MCI 19.2 months (pooled SD 19.0), and for CU amyloid positive 32.5 months (pooled SD 27.2).

Table 2 summarizes the RCIs, that is, the 90% CIs where a change above that interval indicates a significant change in the test. In summary, for MMSE, ADAS, and Animal Fluency, we found similar RCIs for the CU and MCI groups, whereas in Stroop, Letter S, Symbol Digit, TMT A and B, we found larger RCIs for the MCI group than the CU group.

Correlation coefficients between the change in test score and change in CDR-Sum of Boxes (CDR-SB) are summarized in eTable 1 (links.lww.com/WNL/C168). We found significant correlations between CDR-SB and all cognitive tests, however weak to moderate (correlation coefficients −0.2 to 0.6 depending on test). We calculated triangulated MCIDs for each test to derive cut-offs for cognitive test differences that are both empirically and clinically relevant for deciding whether a change in each test represents a meaningful change. We listed estimated MCIDs for group-based clinical worsening or progression in CU and MCI participants using the anchor-based (mean change [SD], ES, and SEM) and distribution-based methods (Pearson for test-retest and RCI) in Table 2 and Figure 1 and triangulated MCIDs for CU and MCI participants in Table 3. Supplementary data for MCIDs (1/2 SD of baseline and SRM) are summarized in eTable 2 (links.lww.com/WNL/C168).

Next, we examined how accurately differences in test scores could estimate a minimal clinically relevant change using logistic regression models (see Figure 2 and eTable 3, links.lww.com/WNL/C168). A clinical change was defined as discrimination between the progression of CDR-SB of ≥0.5 vs CDR SB change of 0 for CU and 0 vs ≥1 for MCI participants. For CU (amyloid-positive and amyloid-negative) participants, the best predicting univariate test was ADAS delayed recall (AUC 0.75) and the worst Animal Fluency (AUC 0.54). For MCI participants, MMSE was the best predicting test (AUC 0.75) and TMT A the least (AUC 0.61), taken in account that Animal Fluency, Letter S, and TMT B were excluded from the group because of lack of follow-up data. Univariate analyses are summarized in eTables 4–6 (links.lww.com/WNL/C168). When combining the tests, we found that the best model (the lowest AIC) for CU individuals was a combination of age and changes in ADAS delayed recall, MMSE, and TMT B, which
produced an AUC of 0.79 (95% CI 0.72–0.86) for identifying a clinical change (discrimination between the progression of CDR-SB of ≥0.5 vs CDR SB change of 0). For MCI participants, the best predicting model included changes in Stroop, MMSE, and age, which had an AUC of 0.82 (95% CI 0.76–0.88). For the amyloid-positive CU group, we found the best predicting composite cognitive measure included changes in ADAS delayed recall, Stroop, Symbol Digit, and TMT B, including gender in the model. We performed a stepwise backward selection, removing the least significant variables, starting with Stroop because this was not significant (p = 0.12). After removing Stroop, we found the best predicting composite cognitive measure included ADAS delayed recall, MMSE, symbol digit modalities test, TMT B, and gender. This produced an AIC of 130.5 and AUC of 0.87 (95% CI 0.79–0.94, sensitivity 75%, specificity 88%).

**Discussion**

We have established minimally clinically important differences (MCIDs) for group-based worsening in test scores for 8 commonly used cognitive tests to help guide clinicians and researchers on clinically relevant cognitive decline with repeated assessments. We investigated changes in cognitive tests longitudinally with distribution-based and anchor-based methods. The distribution-based MCIDs were generally higher (i.e., a larger test change required to indicate an MCID) than the anchor-based MCIDs, showing the importance of using clinical measures of importance (such as CDR) according to the population. We found the best predicting model for a clinical change included differences in test results in ADAS delayed recall, MMSE, and TMT B for cognitively unimpaired (CU), Stroop, MMSE, and age for MCI, and...
**Table 2** Changes in Clinical Outcome Scores That Indicate Clinically Meaningful Differences

<table>
<thead>
<tr>
<th>Clinical outcome</th>
<th>Cognitively unimpaired</th>
<th>MCI</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No meaningful decline</td>
<td>Meaningful decline (CDR-SB diff ≥0.5)</td>
<td>Meaningful decline (CDR-SB diff ≥1)</td>
</tr>
<tr>
<td><strong>MMSE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1099</td>
<td>148</td>
<td>54</td>
</tr>
<tr>
<td>Mean change (SD, 95% CI)</td>
<td>−0.2 (2.0, −0.4 to −0.1)</td>
<td>−1.6 (4.0, −2.2 to −0.9)</td>
<td>−3.7 (5.5, −5.2 to −2.2)</td>
</tr>
<tr>
<td>ES</td>
<td>−0.2</td>
<td>−0.5</td>
<td>−0.8</td>
</tr>
<tr>
<td>SEM</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.263</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCI (±)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ADAS delayed recall</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1093</td>
<td>144</td>
<td>50</td>
</tr>
<tr>
<td>Mean change (SD, 95% CI)</td>
<td>1.1 (1.7, −0.0 to 0.2)</td>
<td>1.5 (2.3, 1.1 to 1.8)</td>
<td>1.5 (2.7, 0.8 to 2.3)</td>
</tr>
<tr>
<td>ES</td>
<td>0.0</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>SEM</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.588</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCI (±)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stroop</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>906</td>
<td>95</td>
<td>35</td>
</tr>
<tr>
<td>Mean change (SD, 95% CI)</td>
<td>0.2 (6.1, −0.2 to 0.6)</td>
<td>6.1 (21.7, 1.7 to 10.5)</td>
<td>8.1 (31.9, −2.9 to 19.1)</td>
</tr>
<tr>
<td>ES</td>
<td>0.0</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>SEM</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.704</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCI (±)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Animal fluency</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>966</td>
<td>128</td>
<td>52</td>
</tr>
<tr>
<td>Mean change (SD, 95% CI)</td>
<td>−0.6 (5.1, −0.9 to −0.2)</td>
<td>−1.4 (5.5, −2.4 to −0.5)</td>
<td>−2.4 (6.7, −4.3 to −0.5)</td>
</tr>
<tr>
<td>ES</td>
<td>−0.1</td>
<td>−0.2</td>
<td>−0.4</td>
</tr>
<tr>
<td>SEM</td>
<td>3.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.658</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCI (±)</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Letter S</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>969</td>
<td>129</td>
<td>53</td>
</tr>
<tr>
<td>Mean change (SD, 95% CI)</td>
<td>1.0 (5.2, −0.3 to 0.3)</td>
<td>−1.3 (4.7, −2.2 to −0.5)</td>
<td>−2.6 (4.5, −3.9 to −1.4)</td>
</tr>
<tr>
<td>ES</td>
<td>0.0</td>
<td>−0.3</td>
<td>−0.5</td>
</tr>
<tr>
<td>SEM</td>
<td>3.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
ADAS delayed recall, MMSE, symbol digit modalities test, TMT B, and gender for amyloid-positive CU individuals.

The novelty of this study is that we present MCIDs for several cognitive tests that, to our knowledge, has not been studied before, which could be used in future clinical AD trials for establishing clinical meaningful treatment effect for treatments seeking to prevent or slow disease progression. Besides, this study has the advantage of presenting triangulated data for MCIDs representing clinical changes, statistical estimates, and qualitative data from clinicians using CDR ratings. When triangulating MCIDs, we integrate results from ratings of clinical changes from ES (based on clinical changes measured with CDR) and statistical estimates (SEM). To our knowledge, previous studies have not investigated which tests best predict a cognitive decline using anchor-based methods, and in this study, we present this for CU individuals, individuals with MCI, and specifically amyloid-positive

Table 2 Changes in Clinical Outcome Scores That Indicate Clinically Meaningful Differences

<table>
<thead>
<tr>
<th>Clinical outcome</th>
<th>Cognitively unimpaired</th>
<th>MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No meaningful decline</td>
<td>Meaningful decline (CDR-SB diff ≥0.5)</td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.675</td>
<td>0.773</td>
</tr>
<tr>
<td>RCI (±)</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Symbol digit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>978</td>
<td>119</td>
</tr>
<tr>
<td>Mean change</td>
<td>-0.9 (5.4, -1.3 to -0.6)</td>
<td>-3.3 (6.3, -4.5 to -2.2)</td>
</tr>
<tr>
<td>ES</td>
<td>-0.1</td>
<td>-0.4</td>
</tr>
<tr>
<td>SEM</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.809</td>
<td>0.778</td>
</tr>
<tr>
<td>RCI (±)</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>TMT A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>989</td>
<td>119</td>
</tr>
<tr>
<td>Mean change</td>
<td>1.0 (15.7, 0.0 to 2.0)</td>
<td>12.4 (29.6, 7.0 to 17.8)</td>
</tr>
<tr>
<td>ES</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>SEM</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td>Pearson r</td>
<td>0.659</td>
<td>0.679</td>
</tr>
<tr>
<td>RCI (±)</td>
<td>24</td>
<td>34</td>
</tr>
<tr>
<td>TMT B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>926</td>
<td>104</td>
</tr>
<tr>
<td>Mean change</td>
<td>1.7 (34.5, -0.5 to 3.9)</td>
<td>24.7 (56.2, 13.8 to 35.7)</td>
</tr>
<tr>
<td>ES</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>SEM</td>
<td>23.9</td>
<td></td>
</tr>
<tr>
<td>Pearson r&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.708</td>
<td>0.620</td>
</tr>
<tr>
<td>RCI (±)</td>
<td>55</td>
<td>57</td>
</tr>
</tbody>
</table>

Abbreviations: CDR-SB = clinical dementia rating—sum of boxes; ES = estimates of effect size; N = no. of follow-up points; MCI = mild cognitive impairment; MMSE = Mini-Mental State Examination; RCI = reliable change index; TMT = Trailmaking Test.

<sup>a</sup> Reliable Change Index.
<sup>b</sup> Pearson r for test-retest scores.

That is, a participant can contribute with several test/CDR-SB differences. For example, 1 with assessments at baseline, 2 years, and 4 years will contribute with 2 data points (baseline to 2 years and 2 years to 4 years).
CU participants, which is the target population of several large ongoing AD trials.

The present findings are important because there is no previous consensus on MCIDs for cognitive test outcomes in AD trials; yet FDA specifically highlights that a clinically meaningful improvement on cognitive test scores should be shown before approval of the drug. In the EMERGE study, the population receiving high-dose treatment with the antiamyloid treatment aducanumab reported a statistically significant reduced decline of 0.6 points on the MMSE between placebo and aducanumab groups favoring aducanumab. However, using our MMSE MCID (1.7 points) would render this mean change clinically insignificant. In the TRAILBLAZER-ALZ study for Donanemab including individuals with MCI-mild AD, they found a difference on MMSE of 0.64 between the placebo and Donanemab cohorts, which again would not be clinically significant.

(A) Cognitively unimpaired. (B) Mild Cognitive impairment. For MMSE, Animal Fluency, Letter S and Symbol Digit all mean difference scores are negative (see Table 2). ES = estimates of effect size.
The CDR global has been used as an external anchor to establish meaningful change estimates for other scales. It is important to note that the CDR global is less sensitive to minimal changes at the group level. Much smaller changes may indeed not only be sensitive to cognitive changes over time as shown previously but also represent a clinically meaningful change. We did not find that changes in Animal Fluency were accurate in estimating a minimal meaningful decline. Overall, we found the best combined model of changes in cognitive tests with logistic regression models and found for amyloid-positive CU individuals the best model combined differences in cognitive test results in ADAS delayed recall, MMSE, TMT B, and Symbol digit with patient’s gender, which includes all 3 cognitive domains. We suggest that this technique could be used to develop other clinically relevant cognitive composites and test batteries for use in predementia populations, using broader cognitive test batteries to find the best model for predicting a cognitive change.

Previous studies have shown similar or larger MCIDs for MMSE compared with our results between 1 and 4 points (see eTables 8 and 9, links.lww.com/WNL/C168). Previous studies have also shown a correlation of at least 0.3–0.35 between the change score and the anchor. In our study, Animal Fluency (CU only), Letter S (CU and MCI), and TMT B (MCI only) had correlations <0.3 with CDR-SB as the anchor.

To our knowledge, no previous study has investigated which combination of cognitive tests best estimate clinically relevant yet minimal worsening in CDR-SB. The most frequently used cognitive composite is the PACC, which originally included the Free and Cued Detective Reminding Test, logical memory, Digit Symbol Substitution Test (equivalent to the symbol digit test used in this study), and the MMSE. Later, modified PACC versions have included animal fluency, TMT B, symbol digit, and/or ADAS delayed recall. Using our model selection approach, we could confirm that a combination of TMT B, ADAS delayed recall, Symbol Digit, and MMSE indeed not only are sensitive to cognitive changes over time as shown previously but also represent a clinically meaningful change. We did not find that changes in Animal Fluency were accurate in estimating a minimal meaningful decline. Overall, we found the best combined model of changes in cognitive tests with logistic regression models and found for amyloid-positive CU individuals the best model combined differences in cognitive test results in ADAS delayed recall, MMSE, TMT B, and Symbol digit combined with patient’s gender, which includes all 3 cognitive domains. We suggest that this technique could be used to develop other clinically relevant cognitive composites and test batteries for use in predementia populations, using broader cognitive test batteries to find the best model for predicting a cognitive change.
Figure 2 ROC Curves From Logistic Regression Models Estimating Minimally Clinical Difference in CU, MCI, and Amyloid positive CU groups.

(A) CU group (N=451). (B) MCI group (N=292). (C) Amyloid positive CU group (N=116). Performance of the best model (defined as lowest AIC including predictors with $p < 0.1$) and individual tests are shown in receiver operating curves (ROC). Further results are presented in eTables 3-6. Outcomes in A and C are a clinical change in Clinical Dementia Rating Sum of Boxes (CDR-SB) ≥0.5 and in C a clinical change in CDR-SB ≥1 point.
an anchor, this should not largely affect MCID estimates. In addition, any progression occurs slower and less frequent in CU participants, which is why the study was designed to have less frequent follow-ups for controls. Another limitation is that MCI participants can potentially fluctuate and revert to CU. Unfortunately, we have not classified participants as CU or MCI at follow-ups; however, if using a CDR-SB of 0 points as a proxy for being CU, only 2.0%–4.7% reverted from CDR-SB>0 to 0 in the MCI group (eTable 7, links.lww.com/WNL/C168), why it should have very little relevance to the results. Furthermore, the purpose of the present estimates is to elucidate what may be a meaningful change in cognitive and clinical status; irrespective of whether such changes remain stable or represent continuous progression of underlying disease, they are still relevant to how people feel and function. A limitation to our calculated MCIDs is that Stroop violates the expectation of ordered ES (partly due to lower N [N = 35] and an increase in SD), and it should therefore be interpreted with caution. This sample-dependent nature is a challenge to the use of ES in general. An alternative would have been to use a CDR-SB difference of ≥0.5 in all cases as minimal and defined as the smallest difference a clinician is able to observe and score; however, we reason that the magnitude of the change in score could then potentially be too small to be clinically meaningful. In general, this is why we seek to use both anchor and distribution in generating estimates and not just the latter and give priority to the anchor.

Our triangulated MCIDs for cognitive test measures could potentially be applied in clinical practice to evaluate whether a clinical progression has occurred since last visit or whether the patient has remained stable. However, further work would be needed to define cutoffs representing possible scores on the instruments, as opposed to aggregate, group-level changes. The results from the logistic regression models (Figure 2 and eTable 3, links.lww.com/WNL/C168) suggests the suitable tests depending on setting (CU, MCI, or amyloid-positive CU) and Table 3 cutoffs that indicate that a meaningful change in the test has occurred. However, in clinical practice, MCIDs need to be rounded up to the nearest higher integer to evaluate differences. This selection of tests and identified cutoffs should however be validated in independent and more diverse populations with wider age range and education level. The MCIDs can also help to identify treatment benefits in clinical trials of therapies for early AD, and as we have reported earlier, several new studies on pharmaceutical treatments for AD have found significant changes in cognitive outcomes but may not be clinically relevant.

**Study Funding**

The Swedish Research Council (2016-00906, #2018-02052), the Knut and Alice Wallenberg foundation (2017-0383), the Marianne and Marcus Wallenberg foundation (2015.0125), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson disease) at Lund University, the Swedish Alzheimer Foundation (AF-939932, #AF-940046), the Swedish Brain Foundation (FO2021-0293, #FO2020-0271), The Parkinson foundation of Sweden (1280/20), the Konung Gustaf Vs och Drottning Victorias Fimurarestitfelse, the Skåne University Hospital Foundation (2020-O000028, #2020-O000028), Regionalt Forskningsstöd (2020-0314, Projekt-2020-0383), and the Swedish federal government under the ALF agreement (2018-Projekt0279, #2018-Projekt0226).

**Disclosure**

E. Borland, N. Cullen, and E. Stomrud have no conflicts of interest to disclose. C. Edgar is a full-time employee at Cogstate Ltd. E. Stomrud and N. Cullen have no conflicts of interest to disclose. O. Hansson has acquired research support (for the institution) from Roche, GE Healthcare, Biogen, AVID Radiopharmaceuticals, and Euroimmun. In the past 2 years, he has received consultancy/speaker fees (paid to the institution) from Biogen and Roche. S. Palmqvist has served on scientific advisory boards and/or given lectures in symposia sponsored by Roche, Biogen, and Geras Solutions. Go to Neurology.org/N for full disclosures.

**Publication History**

Received by Neurology December 14, 2021. Accepted in final form April 19, 2022. Submitted and externally peer reviewed. The handling editor was Linda Hershey, MD, PhD, FAAN.

**Appendix Authors**

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emma Borland, MD</td>
<td>Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden; Department of Neurology, Skåne University Hospital, Malmö, Sweden</td>
<td>Drafting/revision of the article for content, including medical writing for content; study concept or design; analysis or interpretation of data</td>
</tr>
<tr>
<td>Chris Edgar, PhD</td>
<td>Department of Clinical Science, Cogstate, London, United Kingdom</td>
<td>Drafting/revision of the article for content, including medical writing for content; study concept or design</td>
</tr>
<tr>
<td>Erik Stomrud, MD, PhD</td>
<td>Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden; Memory Clinic, Skåne University Hospital, Malmö, Sweden</td>
<td>Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data</td>
</tr>
<tr>
<td>Nicholas Cullen, BS</td>
<td>Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden</td>
<td>Drafting/revision of the article for content, including medical writing for content; analysis or interpretation of data</td>
</tr>
<tr>
<td>Oskar Hansson, MD, PhD</td>
<td>Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden; Memory Clinic, Skåne University Hospital, Malmö, Sweden</td>
<td>Drafting/revision of the article for content, including medical writing for content; study concept or design</td>
</tr>
<tr>
<td>Sebastian Palmqvist, MD, PhD</td>
<td>Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden; Memory Clinic, Skåne University Hospital, Malmö, Sweden</td>
<td>Drafting/revision of the article for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data</td>
</tr>
</tbody>
</table>
Clinically Relevant Changes for Cognitive Outcomes in Preclinical and Prodromal 
Cognitive Stages: Implications for Clinical Alzheimer Trials
Emma Borland, Chris Edgar, Erik Stomrud, et al.
Neurology 2022;99:e1142-e1153 Published Online before print July 14, 2022
DOI 10.1212/WNL.0000000000200817
This information is current as of July 14, 2022

Updated Information & Services
inclusion high resolution figures, can be found at:
http://n.neurology.org/content/99/11/e1142.full

References
This article cites 41 articles, 6 of which you can access for free at:
http://n.neurology.org/content/99/11/e1142.full#ref-list-1

Citations
This article has been cited by 1 HighWire-hosted articles:
http://n.neurology.org/content/99/11/e1142.full##otherarticles

Subspecialty Collections
This article, along with others on similar topics, appears in the
following collection(s):
All Cognitive Disorders/Dementia
http://n.neurology.org/cgi/collection/all_cognitive_disorders_dementia
Neuropsychological assessment
http://n.neurology.org/cgi/collection/neuropsychological_assessment

Errata
An erratum has been published regarding this article. Please see next page or:
/content/100/11/545.1.full.pdf

Permissions & Licensing
Information about reproducing this article in parts (figures,tables) or in
its entirety can be found online at:
http://www.neurology.org/about/about_the_journal#permissions

Reprints
Information about ordering reprints can be found online:
http://n.neurology.org/subscribers/advertise

Neurology® is the official journal of the American Academy of Neurology. Published continuously since
1951, it is now a weekly with 48 issues per year. Copyright Copyright © 2022 The Author(s). Published by
Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.. All rights reserved. Print
ISSN: 0028-3878. Online ISSN: 1526-632X.
Disputes & Debates: Editors’ Choice

Editors’ Note: Association of Amount of Weight Lost After Bariatric Surgery With Intracranial Pressure in Women With Idiopathic Intracranial Hypertension

In their multicenter randomized controlled clinical trial of patients with idiopathic intracranial hypertension (IIH), Mollan et al. sought to determine the amount of weight loss necessary to achieve physiologic remission (intracranial pressure [ICP] ≤ 25 cm CSF). The investigators not only found that greater weight loss was achieved with bariatric surgery over community weight management but patients who underwent surgery experienced greater and more rapid reduction in ICP. Furthermore, only patients in the surgical arm achieved a fall in ICP to ≤ 25 cm, which was associated with a mean weight loss of approximately 24% from their baseline weight. Dr. Ramsamy and colleagues highlight that surgery was not associated with statistically significant improvement in visual outcomes or headache, which are the most disabling symptoms of IIH. In response, Dr. Mollan and their coinvestigators note the study was powered to demonstrate a treatment effect for the primary outcome of ICP reduction (as a biomarker of disease remission), rather than for secondary outcomes regarding visual or other symptoms of IIH. That said, the investigators admit there remains no consensus definition for clinical remission in IIH. Dr. Brenner also comments on the potential adjuvant use of caffeine in weight loss for patients with IIH and the benefits of optic nerve sheath fenestration when vision loss occurs. The investigators cite their original manuscript (published in JAMA Neurology), which reported a rapid and significant fall in ICP within 2 weeks of bariatric surgery. It is possible that acute hormonal changes after surgery such as a rapid rise in gut neuropeptide glucagon-like peptide-1 may have played a larger role in ICP reduction than weight loss alone.

James E. Siegler, MD, and Steven Galetta, MD
Neurology® 2023;100:542. doi:10.1212/WNL.0000000000207118

Reader Response: Association of Amount of Weight Lost After Bariatric Surgery With Intracranial Pressure in Women With Idiopathic Intracranial Hypertension

Mollan et al. report that bariatric surgery was superior to community weight management (CWM) regarding weight lost, intracranial pressure (ICP) reduction, disease remission, and quality of life in women with idiopathic intracranial hypertension (IIH). They recommend for clinicians to have low thresholds of referral for bariatric surgery. However, surgery did not have a greater effect on papilledema grade, visual function, or headache disability than CWM. More specifically, there was no difference in visual acuity, contrast sensitivity, perimetric mean deviation, or optic nerve and the retinal nerve fiber layer. Given that loss of vision is the most severe complication in IIH, and headache is the principal symptom, not

Author disclosures are available upon request (journal@neurology.org).
highlighting these supplementary data in either the initial or follow-up paper seems to be a crucial omission. In particular, defining disease remission only in opening pressure or ICP reduction could be misleading.

In addition, although physical functioning and general health seemed to improve more in the surgery arm of the study (though we note no correction for multiple comparisons), this was not the case in emotional wellbeing, anxiety and depression, social functioning, or pain. Bariatric surgery, despite its potential risks, does have proven long-term health benefits. However, its role in IIH treatment remains far from clear.


Author Response: Association of Amount of Weight Lost After Bariatric Surgery With Intracranial Pressure in Women With Idiopathic Intracranial Hypertension

Susan P. Mollan (Birmingham, England) and Alex J. Sinclair (Birmingham, England)
Neurology® 2023;100:543. doi:10.1212/WNL.00000000000207120

We thank Ramsamy et al. for their comments. Idiopathic intracranial hypertension (IIH) is a systemic metabolic disease and has an established association with obesity. IIH has a detrimental effect on quality of life because most recipients experience headache.

There are few IIH randomized clinical trials (RCTs), and the choice of a primary endpoint to determine a successful outcome varies between intracranial pressure (ICP) and visual function. The IIH Weight Trial hypothesis was based on which method of weight loss would be superior for sustained ICP reduction and weight loss and was powered accordingly. Ramsamy et al.’s critique may be based on their assumption that the trial was powered for secondary outcomes, which it was not, or their preference for visual function as the only outcome measure. We chose the diagnostic criteria threshold for ICP to determine remission. There is no consensus definition for clinical remission in IIH. The results were reported according to the prespecified statistical analysis plan, and hence, this represents a low risk for reporting bias.

When designing RCTs, choosing the most relevant primary and secondary trial outcomes is key and involves many stakeholders. Weight loss methods in IIH were determined by patient partners and experts and form part of the top 10 research priorities for IIH, which this trial addresses.

I read the article by Mollan et al. regarding weight loss and intracranial hypertension (IH) with interest. Gastric bypass results in decreased dicarbonyl stress (DS). DS-related compounds, such as methylglyoxal, are associated with the breakdown of the blood-brain barrier, as seen in IH through the glycation of biomolecules and the production of advanced glycation end products (AGEs). Inhibiting semicarbazide-sensitive amine oxidase, which causes DS, with caffeine may lead to decreased weight gain and AGEs.

Rapid weight loss through gastric surgery may alleviate the complications of IH; however, fenestration of the optic nerve sheath to avoid visual loss should be considered when there is progressive visual loss. Medical management of obesity may have more application in the future with continued progress in this field.


Copyright © 2023 American Academy of Neurology

**CORRECTIONS**

Clinically Relevant Changes for Cognitive Outcomes in Preclinical and Prodromal Cognitive Stages
Implications for Clinical Alzheimer Trials
Neurology® 2023;100:545. doi:10.1212/WNL.0000000000206876

In the Research Article entitled “Clinically Relevant Changes for Cognitive Outcomes in Preclinical and Prodromal Cognitive Stages: Implications for Clinical Alzheimer Trials” by Borland et al., in the Cognitive Tests section, the sentence “Further explanation of tests, what they assess, and how points are counted are described in eMethods.” should include the following link to the eMethods: links.lww.com/WNL/C167. The publisher regrets the omission.

Reference

Juvenile Myoclonic Epilepsy 25 Years After Seizure Onset
A Population-Based Study
Neurology® 2023;100:545. doi:10.1212/WNL.0000000000206877

In the Research Article entitled “Juvenile Myoclonic Epilepsy 25 Years After Seizure Onset: A Population-Based Study” by Camfield et al., the second sentence of the Discussion section of the Abstract should read, “All seizure types in juvenile myoclonic epilepsy (JME) resolved in 26%, and for 13%, only myoclonus persisted.” In addition, the third sentence of the Discussion section should read, “All seizure types remit in approximately 26%, and in 13%, only myoclonic seizures persist for up to 22 years after stopping AEDs.” The authors regret the errors.

Reference