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

Author(s):
Emma Borland, MD\textsuperscript{1,2}; Chris Edgar, PhD\textsuperscript{3}; Erik Stomrud, MD, PhD\textsuperscript{1,4}; Nicholas Cullen, BS\textsuperscript{1}; Oskar Hansson, MD, PhD\textsuperscript{1,4}; Sebastian Palmqvist, MD, PhD\textsuperscript{1,4}

Corresponding Author:
Emma Borland, emma.borland@med.lu.se

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.

Neurology® Published Ahead of Print articles have been peer reviewed and accepted for publication. This manuscript will be published in its final form after copyediting, page composition, and review of proofs. Errors that could affect the content may be corrected during these processes.
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 as 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 \( \geq 0.5 \) or \( \geq 1.0 \) on CDR-SB for 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. For investigating cognitive measures that best predict a change in CDR-SB of \( \geq 0.5 \) or \( \geq 1.0 \) point we conducted ROC 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, 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
AIC 130.5, 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 if
a clinically relevant change has occurred. The composite measure can be useful as a clinically
relevant cognitive test outcome in preclinical AD trials.

Introduction
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”\textsuperscript{1}. It is important to decide the smallest change in an outcome that
constitutes a clinically meaningful change – i.e. MCID – to interpret whether e.g. the
treatment effect measured using a cognitive test is clinically relevant or if a change in
cognitive testing during clinical follow-up represents a clinically meaningful change in
cognition. MCIDs are thus necessary for making accurate clinical decisions and to design
clinical trials with the statistical power to detect an effect equal to or greater than the MCID\textsuperscript{2}. 
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 are more persuasive if seen on multiple tests and that if only seen on one assessment it needs to show a large magnitude of effect to be persuasive of a beneficial effect. However, for many cognitive assessments, the magnitude which corresponds to a clinically meaningful effect compared to placebo is unknown.

Several methods exist for estimating a meaningful clinical effect, the most well-established of these 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, utilize statistical properties of the measures themselves, and of these the most common are effect size metrics - e.g. the standard deviation (SD) and the standard error of measurement (SEM) that incorporates some measure of scale reliability (e.g. test re-test or Cronbach’s α as a measure of internal consistency reliability).

In addition to establishing clinically relevant cut-offs for test changes, it is also important to determine which tests best represent clinically relevant changes. To our knowledge, few previous studies have investigated which tests best estimate a clinically relevant worsening. 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)\(^6\). The PACC was initially created by selecting four 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\(^6\). 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 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 cut-offs for cognitive test changes for use to conclude if 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), respectively, we also examined this in A\(\beta\)-positive CU participants since this is a group of special interest in present and future clinical AD trials\(^7\).

**Methods**

**Population**

The participants in the study were consecutively included from the prospective Swedish BioFINDER study (http://biofinder.se), and participants for this study were enrolled from July 6\(^{th}\) 2009 to March 4\(^{th}\) 2015. The population consisted of 451 cognitively unimpaired (CU) individuals and 292 people classified as having mild cognitive impairment (MCI). In the CU group, 90 individuals had subjective cognitive decline and 361 people were cognitively
healthy controls based on a structured assessment. The individuals were followed 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\textsuperscript{8}. All cognitively unimpaired 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 and underwent a physical exam, MRI scan, lumbar puncture, cognitive assessments and were rated with the CDR. Participants experiencing cognitive symptoms at baseline (SCD or MCI) were followed annually, while participants without cognitive symptoms at baseline were examined every second year by physicians.

**Cognitive tests**

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

Clinical Dementia Rating (CDR) is an ordinal scale with scores of 0-3 points used to quantify the functional impact of cognitive impairment (0 = none, 0.5 = questionable, 1 = mild, 2 = moderate and 3 = severe) in domains (‘box scores’) of memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care\(^9,10\). Participants in this study were rated on the CDR by thorough review of patient records where dementia experts assessed cognitive symptoms and activities of daily living (ADL) in comprehensive semi-structured interviews with patients and informants, including informant-based questionnaires of ADL (the functional activities questionnaire)\(^11\) and cognitive symptoms (the CIMP-QUEST)\(^12\) (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 stages score (range 0-3) using an algorithm. An increase in the CDR-SB score has been identified as having both face and predictive validity for identifying people who are later diagnosed with probable AD or another dementia\(^13\). 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\(^14\). A change of 1 in CDR-SB could be either a change of 0.5 points in two boxes, or a change of 1 point in one box.

**Determining amyloid positivity**

The procedure and analysis of cerebrospinal fluid (CSF) followed the Alzheimer’s Association Flow chart for CSF biomarkers\(^15,16\). We used the ratio for A\(\beta\)42/40 that we establish acquire through CSF analysis. CSF A\(\beta\)42 and A\(\beta\)40 were analysed using Roche Elecsys CSF immunoassays (NeuroToolKit) on all participants. The cut-off for A\(\beta\)42/40 was established with mixture modelling statistics\(^17,18\) and set at 0.066.
Statistics

The psychometric criterion reliable change index (RCI) is used to evaluate whether a change over time of an individual score is considered statistically significant\textsuperscript{19}. RCI provides a confidence interval (CI), which 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 eight test differences with a 90% CI (the most common CI for a RCI)\textsuperscript{19} for CU participants and participants with MCI, respectively. This is done for tests separately using the following equation:

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

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

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

\[SD = \text{Standard deviation of the test score at baseline}\]

\[r = \text{Pearson coefficient for test results in cognitively stable individuals}\]

\[SE_{\text{meas}} = \text{standard error of measurement}\]

\[SE_{\text{diff}} = \text{standard error of differentiation}\]

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\textsuperscript{20}. 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, while an ES of 0.2 is considered a small change and 0.8 a large change\textsuperscript{21,22}. 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\textsuperscript{23}. 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\textsuperscript{24-26}, which we present in our results.

For the anchor-based approach we analysed 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 $\geq 0.5$ points and for the MCI group a change of $\geq 1$ point as anchors to represent a clinically meaningful change. We calculated mean, SD, and ES for changes in the cognitive tests separately for meaningful decline (CDR-SB difference of $\geq 0.5$ and $\geq 1$) 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\textsuperscript{27}. Triangulation integrates results from ratings with clinical changes, statistical estimates and qualitative data from patients and/or clinicians to derive guidelines\textsuperscript{28}. A previously suggested method is to assign the anchor-based results a weight of two-thirds, and the distribution-based method a weight of one-third\textsuperscript{29}. The final triangulated MCID is then calculated as the mean of these three parts. Our anchor-based MCIDs are estimated from ES (based on clinical changes measured with CDR), and distribution-based MCIDs are based on statistical measures (SEM). Since an ES of 0.5 is generally considered a clinically significant change, we used the estimated anchor-based MCID with the ES closest to 0.5.

To examine which tests that best represented a clinically meaningful change, we analysed the cognitive tests as independent variables in logistic regression models with CDR-SB as dependent variable. For the CU group, the CDR-SB difference was dichotomized as either 0
(no clinical change) or ≥0.5-point change (smallest clinically relevant change). For the MCI group, we dichotomized with a larger CDR-SB difference as either 0 or ≥1 point, this excluded between 55-159 data points depending on test because of a CDR-SB difference of 0.5. The area under receiver operating characteristic (AUC) curve and sensitivity and specificity for each test difference were calculated using ROC analyses. Logistic regression models were performed on a subsample with complete data for the analysed cognitive tests (i.e., all logistic regression models were performed on the same population). To find the most optimal combination of test differences to estimate a cognitive change we examined all cognitive test changes in the model for CU to identify a model with the lowest Akaike Information Criterion (AIC). AIC accounts for the trade-off between model fit and sparsity (as few included biomarkers as possible) to protect against model overfitting and can be used as a tool for model selection. Lower AIC indicates a better model. To find the optimal combination for MCI we excluded Animal Fluency, Letter S and TMT B as these tests were only conducted every second year and therefore excluded many participants due to lack of complete cases. We added age at visit, education years and gender in the models. Predictors with a p value >0.1 were removed from the model.

**Standard Protocol Approvals, Registrations, and Patient Consents**

The study was approved by the regional ethical committee at Lund University, Lund, Sweden. All participants gave their written informed consent to participate in the study.

**Data availability**

Anonymized data will be shared by request from a qualified academic investigator for the sole purpose of replicating procedures and results presented in the article and as long as data transfer is in agreement with EU legislation on the general data protection regulation and
decisions by the Swedish Ethical Review Authority and Region Skåne, which should be regulated in a material transfer agreement.

**Results**

Baseline characteristics are shown in Table 1. 451 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 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 shows the RCIs, i.e. 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 CU and MCI groups, while in Stroop, Letter S, Symbol Digit, TMT A and B, we found larger RCIs for MCI compared to CU.

Correlation coefficients between the change in test score and change in CDR-Sum of Boxes (CDR-SB) are shown in eTable 1. 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 present estimated MCIDs for group-based clinical worsening or progression in CU and MCI using anchor-based (mean change (SD), ES, SEM) and distribution-based methods (Pearson for test-retest, RCI) in Table 2.
and Figure 1, and triangulated MCIDs for CU and MCI in Table 3. Supplementary data for MCIDs (1/2 SD of baseline and SRM) are presented in eTable 2.

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). A clinical change was defined as a discrimination between 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 negative), the best predicting univariate test was ADAS delayed recall (AUC 0.75) and the worst Animal Fluency (AUC 0.54). For MCI, 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 due to lack of follow-up data. Univariate analyses are presented in eTables 4-6. When combining the tests, we found that the best model (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 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 as 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, 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 eight 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- and anchor-based methods. The distribution-based MCIDs were generally higher (i.e., larger test change required to indicate a 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 ADAS delayed recall, MMSE, symbol digit modalities test, TMT B and gender for amyloid positive CU.

The novelty of the present 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 as well as 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 here we present this for CU individuals, individuals with MCI and specifically amyloid positive CU participants, which is the target population of several large ongoing AD trials.
The present findings are important because there is no prior 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\(^3\). Recent trials on treatment for AD have investigated changes in cognition comparing individuals receiving placebo with active treatment. In the EMERGE study\(^2\), the population receiving high-dose treatment with the anti-amyloid treatment aducanumab reported a statistically significant reduced decline of 0.6 points on the MMSE between placebo and aducanumab groups favouring aducanumab. However, using our MMSE MCID (1.7 points) would render this mean change clinically insignificant. In the TRAILBLAZER-ALZ\(^2\) study for Donanemab including individuals with MCI-mild AD, they found a difference on MMSE of 0.64 between placebo and Donanemab cohorts, which again would not be clinically significant.

Previous studies have shown similar or larger MCIDs for MMSE compared to our results between 1-4 points\(^{32-35}\), however we have not found previous estimated MCIDs for the other examined tests. One previous paper on MCID for MMSE suggested a 0.4 SD change from baseline for MMSE as MCID, corresponding to a MCID of 1.4 MMSE points\(^{33}\), close to results from another study showing a MCID for MMSE of 1.6 points for 0.4 SDs from baseline MMSE\(^{34}\), close to our calculated MCIDs (-1.5 MMSE points for CU-group and -1.7 for MCI-group). Another previous study showed an estimated MCID for MMSE of 1-3 points depending on disease severity, with larger results using only distribution-based approach similar to our study\(^32\). Yet another study has showed a far higher MCID of MMSE of 3.72 points\(^{35}\). We found a very low 0.5 SD of baseline MMSE (1.1 for CDR SB change of ≥0.5 in CU), which is partly caused by the inclusion criteria in the BioFINDER study for CU of MMSE score ≥28 points but does not explain why the MCI-group had the same results. The
estimated RCIs are much larger than MCID explained by the methodology with a large SD (1.65), being individual patient-based, differing from MCID as being minimal change at the group level. Much smaller changes may in fact be relevant as seen in our calculated MCIDs.

The CDR global has been used as an external anchor to establish meaningful change estimates for other scales. Whilst it has clinical validity as a meaningful change, progression from one stage to another represents a change which is much larger than what may be considered ‘minimally important’, which is why we have chosen to use the CDR-Sum of Boxes (CDR-SB) as an anchor for this study. Previous studies have also shown that to identify MCI, CDR-SB might be more accurate than global CDR. Studies have reported a high internal consistency for the CDR-SB across the AD spectrum with a low variability in mean changes, and that mean scores decline nearly linearly. In summary, we therefore chose to use CDR-SB as the anchor for determining clinically meaningful important differences in cognitive test results. In a recent study, it was shown that CDR-SB was not strongly correlated with the cognitive assessments MMSE or ADAS-Cog at baseline, however there was a moderate correlation between change in CDR-SB and ADAS-Cog13 (r=0.5) and MMSE (r=-0.4) at 2-year follow-up. The same study showed that both CDR-SB and MMSE had a strong responsiveness to change. In our study, we have seen significant correlations between CDR-SB and all cognitive tests, however weak to moderate (correlation coefficients -0.17 to 0.63 depending on test, see eTable 1). Previous studies have recommended 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\textsuperscript{6}. Later modified PACC versions have included animal fluency, TMT B, symbol digit and/or ADAS delayed recall\textsuperscript{41, 42}. 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 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 three cognitive domains\textsuperscript{6}. 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 neurocognitive test batteries to find the best model for predicting a cognitive change.

A potential limitation to the study is that the follow-up of participants is annual for MCI participants and every second year for most CU (annual for those with subjective cognitive symptoms at baseline), which might result in missing some fluctuation or decline in cognition in CU individuals. However, since the primary approach is based on an anchor this should not have a large impact on MCID estimates. Additionally, 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 MCI-group (eTable 7), 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 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 also give priority to anchor.

Our triangulated MCIDs for cognitive test measures could potentially be applied in clinical practice to evaluate if a clinical progression has occurred since last visit or if the patient has remained stable. However, further work would be needed to define cut-offs 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) suggests the suitable tests depending on setting (CU, MCI or amyloid positive CU) and Table 3 cut-offs 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 cut-offs 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 above,
several new studies on pharmaceutical treatments for AD have found significant changes in
cognitive outcomes but may not be clinically relevant.

**Key words:** MCID, MID, MDC, CID, minimal change, PACC, cognitive assessments,
clinical trials

References

1. Malec JF, Kean J, Monahan PO. The Minimal Clinically Important Difference for the
Mayo-Portland Adaptability Inventory. J Head Trauma Rehabil 2017;32:E47-E54.
2. Liu KY, Schneider LS, Howard R. The need to show minimum clinically important
3. FDA. Early Alzheimer’s Disease: Developing Drugs for Treatment Guidance for
Industry. [online].
Alzheimer's Disease: Establishing the Meaningfulness of Treatment Effects. J Prev
Sclerosis Impact Scale physical impact subscale for patients with physical worsening. Mult
Scler 2014;20:1753-1760.
7. Sperling RA, Rentz DM, Johnson KA, et al. The A4 study: stopping AD before
22. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care 2002;2:e15.
A. Mean change in cognitive tests in CU group

B. Mean change in cognitive tests in MCI group
A. Prediction of cognitive decline in CU group

- AUC 0.80 (95% CI 0.73-0.86), AIC 394.0: ADAS delayed recall, MMSE, TMT B, age
- AUC 0.75 (95% CI 0.67-0.82), AIC 412.4: ADAS delayed recall
- AUC 0.69 (95% CI 0.61-0.76), AIC 453.9: MMSE
- AUC 0.64 (95% CI 0.56-0.72), AIC 459.9: TMT B

B. Prediction of cognitive decline in MCI group

- AUC 0.82 (95% CI 0.76-0.88), AIC 218.9: MMSE, Stroop, age
- AUC 0.76 (95% CI 0.70-0.82), AIC 236.4: MMSE
- AUC 0.67 (95% CI 0.59-0.74), AIC 268.4: Stroop

C. Prediction of cognitive decline in CU amyloid-positive group

- AUC 0.87 (95% CI 0.79-0.94), AIC 130.5: ADAS delayed recall, TMT B, Symbol Digit, MMSE, gender
- AUC 0.77 (95% CI 0.66-0.88), AIC 143.9: ADAS delayed recall
- AUC 0.68 (95% CI 0.58-0.79), AIC 169.2: TMT B
- AUC 0.67 (95% CI 0.55-0.78), AIC 171.1: Symbol Digit
- AUC 0.65 (95% CI 0.53-0.77), AIC 173.2: MMSE
**Tables**

<table>
<thead>
<tr>
<th>Total N Participants</th>
<th>Cognitively unimpaired (Controls and SCD)</th>
<th>MCI</th>
<th>Aβ positive CU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>451 (361 and 90)</td>
<td>292</td>
<td>116</td>
</tr>
<tr>
<td>Mean follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- N data points</td>
<td>1588</td>
<td>727</td>
<td>402</td>
</tr>
<tr>
<td>- Pooled months (SD, range)</td>
<td>32.4 (26.8, 12-96)</td>
<td>19.2 (19.0, 12-72)</td>
<td>32.5 (27.5, 12-96)</td>
</tr>
<tr>
<td>Age (years, SD)</td>
<td>72.9 (5.5)</td>
<td>71.1 (5.5)</td>
<td>73.5 (5.6)</td>
</tr>
<tr>
<td>Men (%)</td>
<td>40.4%</td>
<td>57.9%</td>
<td>39.7%</td>
</tr>
<tr>
<td>Education (years, SD)</td>
<td>12.3 (3.4)</td>
<td>11.1 (3.4)</td>
<td>12.5 (4.2)</td>
</tr>
<tr>
<td>Baseline CDR-SB (SD)</td>
<td>0 (0)</td>
<td>1.4 (0.9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Baseline cognitive test scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- MMSE</td>
<td>29.0 (1.0)</td>
<td>27.0 (1.8)</td>
<td>29.1 (0.9)</td>
</tr>
<tr>
<td>- ADAS delayed recall</td>
<td>2.2 (1.9)</td>
<td>6.5 (2.3)</td>
<td>2.4 (2.0)</td>
</tr>
<tr>
<td>- Stroop</td>
<td>29.3 (8.0)</td>
<td>41.4 (24.6)</td>
<td>30.0 (7.5)</td>
</tr>
<tr>
<td>- Animal Fluency</td>
<td>21.6 (5.6)</td>
<td>14.2 (5.6)</td>
<td>20.7 (5.9)</td>
</tr>
<tr>
<td>- Letter S</td>
<td>15.3 (5.7)</td>
<td>10.8 (5.1)</td>
<td>15.4 (5.0)</td>
</tr>
<tr>
<td>- Symbol Digit</td>
<td>36.6 (8.5)</td>
<td>26.2 (8.9)</td>
<td>35.0 (8.3)</td>
</tr>
<tr>
<td>- TMT A</td>
<td>46.4 (17.6)</td>
<td>67.7 (33.8)</td>
<td>46.9 (18.3)</td>
</tr>
<tr>
<td>- TMT B</td>
<td>104.0 (51.8)</td>
<td>130.8 (32.0)</td>
<td>109.7 (50.8)</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Ischemic heart disease</td>
<td>7.1%</td>
<td>17.1%</td>
<td>8.6%</td>
</tr>
<tr>
<td>- Hypertension</td>
<td>39.2%</td>
<td>32.2%</td>
<td>40.5%</td>
</tr>
<tr>
<td>- Diabetes</td>
<td>8.2%</td>
<td>10.3%</td>
<td>10.3%</td>
</tr>
<tr>
<td>- Stroke or TIA</td>
<td>3.1%</td>
<td>15.1%</td>
<td>4.3%</td>
</tr>
<tr>
<td>Amyloid positive, N (%)</td>
<td>116 (16%)</td>
<td>84 (29%)</td>
<td>116 (100%)</td>
</tr>
</tbody>
</table>

Table 1. Demographics. Data are shown as mean (SD) if not otherwise specified. Abbreviations: SCD= subjective cognitive disease, N = number, SD = Standard deviations, CDR-SB = Clinical Dementia Rating - Sum of Boxes.

<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>MMSE</td>
<td>1099</td>
<td>148</td>
</tr>
<tr>
<td>- Mean change (SD, 95% CI)</td>
<td>-0.2 (2.0)</td>
<td>-1.6 (4.0)</td>
</tr>
<tr>
<td>- ES</td>
<td>0.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>- SEM</td>
<td>1.2</td>
<td>0.263</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>ADAS delayed recall</td>
<td>1093</td>
<td>144</td>
</tr>
<tr>
<td>- Mean change (SD, 95% CI)</td>
<td>0.1 (1.7)</td>
<td>1.5 (2.3)</td>
</tr>
<tr>
<td>- ES</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>- SEM</td>
<td>1.3</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Stroop</td>
<td>906</td>
<td>95</td>
</tr>
<tr>
<td>- Mean change (SD, 95% CI)</td>
<td>0.2 (6.1)</td>
<td>6.1 (21.7)</td>
</tr>
<tr>
<td>- ES</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>- SEM</td>
<td>4.2</td>
<td>0.704</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Animal Fluency</td>
<td>966</td>
<td>128</td>
</tr>
<tr>
<td>- Mean change (SD, 95% CI)</td>
<td>-0.6 (5.1)</td>
<td>-1.4 (5.5)</td>
</tr>
</tbody>
</table>

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.
<table>
<thead>
<tr>
<th>Test</th>
<th>Triangulated MCID for cognitively unimpaired</th>
<th>Triangulated MCID for MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE</td>
<td>-1.5</td>
<td>-1.7</td>
</tr>
<tr>
<td>ADAS delayed recall</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Stroop</td>
<td>5.5</td>
<td>9.3</td>
</tr>
<tr>
<td>Animal fluency</td>
<td>-2.8</td>
<td>-2.9</td>
</tr>
</tbody>
</table>

Table 2. Changes in clinical outcome scores that indicate clinically meaningful differences.

*Pearson’s r for test-retest scores. Abbreviation: N= Number of follow-up points. That is, a participant can contribute with several test/CDR-SB differences. For example, one with assessments at baseline, 2 year and 4 year will contribute with two data points (baseline to 2 year and 2 year to 4 year). ES = Estimates of effect size, SEM = standard error of the mean, RCI = reliable change index, SD = Standard deviations.
Table 3. Triangulated MCID test scores for cognitively unimpaired (CU) and MCI participants. These triangulated MCIDs show the cut-offs for changes in test scores that represents a MCID. The triangulated MCIDs were calculated by weighting the anchor-based method with two-thirds using the MCID closest to an ES of 0.5 (minimal change). That is: (anchor-based MCID + anchor-based MCID + distribution-based MCID) / 3. The anchor-based method constituted of mean change for a CDR change of $\geq 0.5$ points for CU and $\geq 1$ point for MCI, and distribution-based method constituted of SEM) as shown in Table 2. For example, the triangulated MCID for MMSE among CU was calculated the following way: $(-1.6 -1.6 -1.2) / 3 = -1.5$ (see eTables 8-9). For clinical practice, MCIDs need to be rounded up to the nearest higher integer to evaluate differences. Abbreviation: ES, effect size; MCID, minimal clinically important difference; SEM; Standard error of measurement.

<table>
<thead>
<tr>
<th>Test</th>
<th>CU MCID</th>
<th>MCI MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter S</td>
<td>-2.9</td>
<td>-1.8</td>
</tr>
<tr>
<td>Symbol digit</td>
<td>-3.5</td>
<td>-3.8</td>
</tr>
<tr>
<td>TMT A</td>
<td>11.7</td>
<td>13.0</td>
</tr>
<tr>
<td>TMT B</td>
<td>24.4</td>
<td>20.1</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 published online July 14, 2022
DOI 10.1212/WNL.0000000000200817

This information is current as of July 14, 2022

Updated Information & Services
including high resolution figures, can be found at:
http://n.neurology.org/content/early/2022/07/14/WNL.0000000000200817.full

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

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.