Published in final edited form as:

Clin Neuropsychol. 2019 April ; 33(3): 478–489. doi:10.1080/13854046.2018.1484518.

External validation of change formulae in neuropsychology with neuroimaging biomarkers: A methodological recommendation and preliminary clinical data

Kevin Duff¹, Kayla R. Suhrie¹, Bonnie C.A. Dalley¹, Jeffrey S. Anderson², and John M. Hoffman^{2,3}

¹Center for Alzheimer's Care, Imaging and Research, Department of Neurology, University of Utah

²Department of Radiology, University of Utah

³Huntsman Cancer Institute, University of Utah

Abstract

Objective: Within neuropsychology, a number of mathematical formulae (e.g., reliable change index, standardized regression based) have been used to determine if change across time has reliably occurred. When these formulae have been compared, they often produce different results, but “different” results do not necessarily indicate which formulae are “best.” The current study sought to further our understanding of change formulae by comparing them to clinically-relevant external criteria (amyloid deposition and hippocampal volume).

Method: In a sample of 25 older adults with varying levels of cognitive intactness, participants were tested twice across one week with a brief cognitive battery. Seven different change scores were calculated for each participant. An amyloid PET scan (to get a composite of amyloid deposition) and an MRI (to get hippocampal volume) were also obtained.

Results: Deviation-based change formulae (e.g., simple discrepancy score, reliable change index with or without correction for practice effects) were all identical in their relationship to the two neuroimaging biomarkers, and all were non-significant. Conversely, regression-based change formulae (e.g., simple and complex indices) showed stronger relationships to amyloid deposition and hippocampal volume.

Conclusions: These results highlight the need for external validation of the various change formulae used by neuropsychologists in clinical settings and research projects. The findings also preliminarily suggest that regression-based change formulae may be more relevant than deviation-based change formulae in this context.

Keywords

cognitive change; biomarkers; Alzheimer's disease

Introduction

Within neuropsychology, a number of mathematical formulae have been used to determine if change across time has reliably occurred. These formulae can be divided into two camps: deviation-based scores and regression-based scores. The deviation-based scores usually subtract a baseline score from a follow-up score and divide by some metric that reflects a standard deviation of change units. Examples of this deviation-based method include: standard deviation index (SDI), reliable change index (RCI) (Jacobson & Truax, 1991), RCI accounting for practice effects (RCI+PE) (Chelune, Naugle, Luders, Sedlak, & Awad, 1993), and RCI+PE as modified by Iverson (RCI+PEI) (Iverson, 2001). A simple discrepancy score would be another example of this method; however, it is not divided by any metric. The regression-based scores (McSweeney, Naugle, Chelune, & Luders, 1993) predict a follow-up score based on: 1) only a baseline score (i.e., “simple” standardized regression based score [SRB]) or 2) baseline score and other variables (i.e., “complex” SRB). The predicted follow-up score is compared to the observed follow-up score and then divided by some metric of change (often the standard error of estimate from the regression model). Both deviation- and regression-based methods (with the exception of the simple discrepancy score) yield a z-score, which reflects the amount of change seen on follow-up. Z-scores greater than +1.645 are typically interpreted as indicating improvement across time, whereas z-scores below -1.645 typically indicate decline across time.

A number of studies have compared the various change formulae in a range of healthy and patient samples. Some of these have found that the deviation- and regression-based formulae perform comparably (Barr & McCrea, 2001; Blasi et al., 2009; Frerichs & Tuokko, 2005; Heaton et al., 2001; Hinton-Bayre, 2011, 2016; Temkin, Heaton, Grant, & Dikmen, 1999). Others have shown notable discrepancies between the various change scores, usually with regression-based formulae better estimating cognitive change (Duff, Atkinson, et al., 2017; Estevis, Basso, & Combs, 2012; Levine et al., 2007; Maassen, Bossema, & Brand, 2009; Ouimet, Stewart, Collins, Schindler, & Bielajew, 2009; Tombaugh, 2005). However, “different” results do not necessarily indicate which formulae are “best.”

To our knowledge, only two studies have attempted to compare the various change scores to some clinically-relevant external criterion, which might shed light on the clinical utility of the change scores. In the first study, Heaton et al. (2001) examined RCI+PE, simple SRB, and complex SRB in two clinical samples where cognitive change was expected: patients with a recent traumatic brain injury who were expected to improve over time and patients with a new brain insult who were expected to decline over time. As hypothesized, the three change methods showed improvement in traumatic brain injury sample and decline in new insult group. However, the three change methods were largely comparable. In the second, Frerichs & Tuokko (2005) examined cognitive change in healthy older adults enrolled in a longitudinal study. These authors also calculated RCI+PE and simple and complex SRBs in their cohort, and they attempted to see which change method best predicted conversion to dementia (as external criterion). Like the first study, all three change scores were comparable in their prediction of this criterion.

Surprisingly, few other studies have followed the lead of Heaton et al. and Frerichs & Tuokko in the search for the “best” change formulae. Therefore, the current study sought to further our understanding of cognitive change methodology by comparing various change scores to clinically-relevant external criteria in a sample of older adults with varying levels of cognitive intactness. The external criteria were neuroimaging biomarkers used in studies of Alzheimer’s disease: hippocampal volume on magnetic resonance imaging (MRI) and amyloid deposition on positron emission tomography (PET). The current cohort of normal aging and prodromal Alzheimer’s disease was chosen as there is a growing body of literature that suggests that late life cognitive disorders show reduced practice effects (Calamia, Markon, & Tranel, 2012), and diminished practice effects may predict future cognitive decline and greater brain-related pathology (Duff et al., 2007; Duff, Foster, & Hoffman, 2014; Duff, Hammers, et al., 2017; Hassenstab et al., 2015; Machulda et al., 2017). Based on the existing literature, it was hypothesized that the regression-based scores (e.g., simple and complex SRBs) would be more strongly related to the neuroimaging biomarkers than the deviation-based scores (e.g., simple discrepancy, various RCIs). This study might serve as a model of a paradigm for external validation of change formulae in future work in this area.

Methods

Participants.

Twenty-five older adults (19 females/6 males; mean age = 77.5 years, SD = 6.5; mean education = 16.0 years, SD = 2.9) were enrolled in this study. These individuals were all recruited from senior centers and independent living facilities to participate in studies on memory and aging. These particular participants agreed to undergo multiple neuroimaging procedures, which were not part of the larger parent study. No other a priori selection criteria were used (e.g., demographics, performance on cognitive tests). Criteria for classifying individuals as intact or Mild Cognitive Impairment were consistent with those of Winblad et al. (2004). To classify individuals as cognitively intact, participants needed to demonstrate normal objective cognitive functioning (e.g., comparable premorbid intellect and current cognitive functioning) and report being functionally independent in activities of daily living, which was corroborated by a knowledgeable informant. To classify individuals as Mild Cognitive Impairment, participants needed to demonstrate deficits in objective cognitive functioning (e.g., premorbid intellect being notably higher than current cognitive functioning) and report being functionally independent in activities of daily living. Based on these criteria, the minority of these individuals were classified as cognitively intact (n = 8), with the remainder characterized as Mild Cognitive Impairment (n = 17), exhibiting at least an amnesic profile. Exclusion criteria for this study included: history of neurological disease known to affect cognition (e.g., stroke, head injury with loss of consciousness of >30 minutes, seizure disorder, demyelinating disorder, etc.); dementia based on DSM-IV criteria; current or past major psychiatric illness (e.g., schizophrenia, bipolar affective disorder); 30-item Geriatric Depression Score >15; history of substance abuse; current use of cognitive enhancers, antipsychotics, or anticonvulsant medications; history of radiation therapy to the brain; history of significant major medical illnesses, such as cancer or AIDS; and currently pregnant.

Procedures.

The local institutional review board approved all procedures and all participants provided informed consent before data collection commenced. As part of a larger study, all participants completed a neuropsychological battery during a baseline visit designed to characterize their functioning, and classify participants either as intact or as showing Mild Cognitive Impairment. This baseline battery included the Reading subtest of Wide Range Achievement Test – 4th Edition (to estimate premorbid intellect) and the Repeatable Battery for the Assessment of Neuropsychological Status (to assess current cognitive functioning). This battery also included two following tests of memory:

- Hopkins Verbal Learning Test - Revised (HVLT-R) (Brandt & Benedict, 1997) is a verbal memory task, in which an individual learns a list of 12 words across three learning trials, and recalls the words after a 20 – 25-minute delay. Only the Delayed Recall score (range = 0 – 12) was used in this study, and higher values indicate better memory. Form 6 of this test was used.
- Brief Visuospatial Memory Test - Revised (BVMt-R) (Benedict, 1997) is a visual memory task, in which an individual learns a set of 6 geometric designs across three learning trials, and recalls the designs after a 20 – 25-minute delay. Only the Delayed Recall score (range = 0 – 12) was used in this study, and higher values indicate better memory. Form 5 of this test was used.

The delayed recall trials on these two memory tests were chosen for multiple reasons, even though other indices of these tests provide valuable clinical information (e.g., immediate memory, learning over trials, recognition hits and false positives). First, delayed recall is one variable that has been widely used to characterize cognitive changes associated with prodromal and manifest Alzheimer's disease (Weissberger et al., 2017). Second, we have previously found that short-term practice effects on the delayed recall trials of these two memory tests are particularly sensitive in normal aging and amnesic Mild Cognitive Impairment (Duff et al., 2008; Duff et al., 2014; Duff et al., 2011). Third, we focused on these two variables to limit the number of statistical comparisons in our already under-powered study.

After approximately one week ($M = 7.1$ days, $SD = 0.9$), the HVLT-R and BVMt-R were repeated. The same form of each test was used to maximize practice effects. Raw scores on the delayed recall trials of HVLT-R and BVMt-R were used in all analyses. Seven change scores were calculated on each test: simple discrepancy score, SDI, RCI, RCI+PE, RCI+PEI, simple SRB, and complex SRB. These change scores were based on an independent sample of 167 non-demented older adults who were also administered this cognitive battery at baseline and one-week (Duff, 2014). More information about the calculation and use of these change formulae is provided elsewhere (Duff, 2012, 2014).

MRI was acquired on a Siemens Trio 3.0T scanner with a standard head coil (Siemens, Erlangen, Germany). The imaging protocol was a sagittal 3D magnetization prepared rapid acquisition gradient-echo (MPRAGE) T1-weighted acquisition (inversion time = 1000 ms, echo time = 2.08 ms, repetition time = 2400 ms, flip angle = 8 degrees, field of view = 224 mm, slice thickness = 0.7 mm, 256 slices). All MRI scans were batched processed on the

same workstation using FreeSurfer image analysis suite v5.3.0 (<http://surfer.nmr.mgh.harvard.edu/>) to estimate total intracranial and hippocampal volumes. Technical details are described previously (Fischl & Dale, 2000; Fischl et al., 2002; Fischl et al., 2004). Hippocampal volumes were normalized to total intracranial volume.

Amyloid PET was acquired as described previously (Duff et al., 2013). ^{18}F -Flutemetamol was produced under PET cGMP standards and the studies were conducted under an approved Federal Drug Administration Investigational New Drug application. Imaging was performed 90 minutes after the injection of 185 mBq (5 mCi) of ^{18}F -Flutemetamol. Emission imaging time was approximately 30 minutes. A GE ST PET/CT scanner was used for ^{18}F -Flutemetamol imaging in this study, which possessed the full width at half maximum spatial resolution at 5.0 mm. The field of view for reconstruction was set to 25.6 cm on the scanner to generate a pixel size of 2.0 mm X 2.0 mm (image matrix size 128 X 128). The native slice thickness was 3.27 mm. Volumes of interest were automatically generated by the CortexID Suite analysis software (GE Healthcare) with Z axis dimensions substantially larger than the slice thickness. ^{18}F -Flutemetamol binding was analyzed using a regional semi-quantitative technique described by Vandenberghe et al. (2010) and refined by Thurfjell et al. (2014). In this technique, semi-quantitative regional (prefrontal, anterior cingulate, precuneus/posterior cingulate, parietal, mesial temporal, lateral temporal, occipital, sensorimotor, cerebellar grey matter, and whole cerebellum) and composite standardized uptake value ratios (SUVRs) in the cerebral cortex were generated automatically and normalized to the pons using the CortexID Suite software (Lundqvist et al., 2013). This software uses a threshold z score of 2.0 to indicate abnormally increased regional amyloid burden that corresponds to a composite SUVR of 0.59 when normalized to the pons, providing a 99.4% concordance with visual assessment (Thurfjell et al., 2014). For ^{18}F -Flutemetamol amyloid imaging, there is no specific age-related “normal” level of binding in the CortexID Suite database to assess age-matched normality. Therefore, the study images were compared to the intrinsic software database control group as a whole to calculate the z-scores compared to clinically negative amyloid scans.

Hippocampal volume via MRI and a composite of SUVR via amyloid PET were chosen for multiple reasons. First, these two neuroimaging biomarkers have been widely used to characterize brain changes associated with prodromal and manifest Alzheimer’s disease (Jack et al., 2016; Weiner et al., 2017). Second, we have previously found that short-term practice effects are related to amyloid deposition in this cohort (Duff et al., 2014; Duff, Hammers, et al., 2017). Third, we focused on these two imaging variables to limit the number of statistical comparisons in our study, even though other related variables were available (e.g., MRI volumes of temporal and parietal lobes, regional amyloid deposition values).

Statistical analyses. In the primary analyses, Pearson correlations were calculated between the two neuroimaging biomarkers (hippocampal volumes and a composite of amyloid deposition) and the seven change scores on the HVLt-R and BVMT-R. In secondary analyses, Fisher r -to- z transformations were also used to compare correlations. An alpha value of 0.05 was used for these analyses. The primary analyses were hypothesis-driven, and

a more limited number of Fisher r -to- z tests were performed to assess differences in correlation size.

Results

Data was screened for missing values, univariate outliers, and normality (e.g., skewness, kurtosis), and assumptions were met for parametric statistical analyses.

Table 1 provides the cognitive test scores that were used to classify participants as intact or as having Mild Cognitive Impairment. As can be seen in these scores, those classified as intact had current cognitive scores that were comparable to their premorbid intellect. Those classified as having Mild Cognitive Impairment had current cognitive scores that were well below their premorbid intellect.

Table 2 provides the raw scores on the Delayed Recall trials of the HVLt-R and BVMT-R at the baseline and one-week visits, as well as the seven change scores. The mean bilateral hippocampal volumes normalized to intracranial volume was 0.005 (SD = 0.001, which equate to $M = 6,653.64 \text{ mm}^3$, $SD = 1,205.37$). On average, subjects classified as cognitively intact had significantly greater normalized hippocampal volumes than those classified as MCI (intact: $M = 0.005$, $SD = 0.001$; Mild Cognitive Impairment: $M = 0.004$, $SD = 0.001$, $t(23) = 2.33$, $p = 0.03$). The mean composite of SUVRs normalized to pons was 0.65 (SD = 0.18). Of the 25 scans, 52% were categorized as “positive” for ^{18}F -Flutemetamol uptake, using a cutoff of z -score greater than or equal to 2.0. On average, subjects classified as MCI had higher composite measures of amyloid than those classified as intact, although this difference did not reach statistical significance (intact: $M = 0.57$, $SD = 0.14$; Mild Cognitive Impairment: $M = 0.68$, $SD = 0.19$, $t(23) = -1.51$, $p = 0.15$).

Table 3 provides the correlations between the two neuroimaging biomarkers (hippocampal volumes and a composite of amyloid deposition) and the seven change scores on the HVLt-R. On the HVLt-R, hippocampal volumes were significantly related to the simple SRB ($r = 0.66$, $p < 0.001$) and complex SRB ($r = 0.65$, $p < 0.001$), but none of the other change scores (p 's > 0.05). Using Fisher r -to- z transformations, the correlations between hippocampal volumes and simple and complex SRBs were significantly larger than the correlations between hippocampal volumes and any of the other change scores (simple SRB: $z = -2.03$, $p = 0.02$; complex SRB: $z = -1.97$, $p = 0.02$). Amyloid deposition was only significantly related to the complex SRB for the HVLt-R ($r = -0.41$, $p = 0.04$). Again using Fisher r -to- z transformations, the correlations between amyloid deposition and simple and complex SRBs were significantly larger than the correlations between amyloid deposition and any of the other change scores (simple SRB: $z = 1.87$, $p = 0.03$; complex SRB: $z = 1.95$, $p = 0.02$).

Table 4 provides the correlations between the two neuroimaging biomarkers (hippocampal volumes and a composite of amyloid deposition) and the seven change scores on the BVMT-R. On the BVMT-R, hippocampal volumes were not significantly related to any of the change scores (p 's > 0.05). Using Fisher r -to- z transformations, the correlations between hippocampal volumes and simple and complex SRBs tended to be larger than the correlations between hippocampal volumes and any of the other change scores (simple SRB:

$z = -1.35, p = 0.08$; complex SRB: $z = -1.38, p = 0.08$). Amyloid deposition, however, was significantly related to the complex SRB ($r = -0.54, p = 0.006$), with a trend of being related to the simple SRB ($r = -0.39, p = 0.052$). Again using Fisher r -to- z transformations, the correlation between amyloid deposition and complex SRBs was significantly larger than the correlations between amyloid deposition and any of the other change scores (complex SRB: $z = 1.90, p = 0.02$), although the correlation between amyloid deposition and the simple SRB was also trending in the expected direction (simple SRB: $z = 1.30, p = 0.10$).

Discussion

Multiple studies have compared the various cognitive change formulae used in neuropsychology (Barr & McCrea, 2001; Blasi et al., 2009; Duff, Atkinson, et al., 2017; Estevis et al., 2012; Frerichs & Tuokko, 2005; Heaton et al., 2001; Hinton-Bayre, 2011, 2016; Levine et al., 2007; Maassen et al., 2009; Ouimet et al., 2009; Temkin et al., 1999; Tombaugh, 2005). Although some of these studies have identified differences in the different change scores, few have attempted to provide guidance about which formulae might be most useful compared to clinically-relevant external criteria (Frerichs & Tuokko, 2005; Heaton et al., 2001). Therefore, the current study sought to further our understanding of these change formulae by comparing seven commonly-used formulae to two neuroimaging biomarkers in a sample of older adults with varying levels of cognitive functioning. Results revealed that regression-based change formulae were more strongly related to hippocampal volume and amyloid deposition than deviation-based change formulae.

As reported in prior studies, we observed a wide range of change scores. For example, the RCI indicated that our subjects improved between one half and nearly one full standard deviation unit on the HVLT-R and BVMT-R across one week. Conversely, the two RCI+PE indices showed slight “declines” across time. (Note that “declines” largely reflects less improvement across one week than the independent sample on which these change formulae were developed, and not necessarily a worsen in test scores at follow-up.) Such reduced practice effects would be consistent with results in studies of patients with Mild Cognitive Impairment and Alzheimer’s disease (Calamia et al., 2012; Duff, Atkinson, et al., 2017; Duff et al., 2007; Duff et al., 2008). The two SRBs tended to evidence the largest “declines” in performance in this sample. One reason for this discrepancy between the RCI and the other change scores is that the traditional RCI does not account for practice effects, which might be particularly important with briefer retest intervals between cognitive testings. The SRBs may indicate larger changes than the two RCI+PE scores as the former control for baseline scores and demographic variables, where the latter do not. Regardless, differences between various change scores do not indicate which ones are more/less accurate. Comparison to an external criterion is necessary to further inform the clinical utility of the various change formulae.

When the deviation-based change scores were compared to the two neuroimaging biomarkers, they all yielded identical results. For example, as seen in Table 3, the correlation between hippocampal volume and the simple discrepancy score for the HVLT-R was 0.18. This was also the exact same correlation as between hippocampal volume and the HVLT-R’s SDI, RCI, RCI+PE, and RCI+PEI. In hindsight, this should not be a surprising result, as

each deviation-based score is a variation of the simple discrepancy score (Time 2 – Time 1). For the SDI, the simple discrepancy score is divided by the standard deviation at Time 1. For the RCI, the simple discrepancy score is divided by the standard error of the difference. For the two RCI+PE scores, the mean practice effect is subtracted from the simple discrepancy score and then divided by some version of the standard error of the difference. Although the denominator may slightly vary in these different change scores, the numerator always contains the simple discrepancy score. As such, if one of these deviation-based change scores correlates with an external criterion, then they all will, at exactly the same level. If this is the case, then it may not matter if you use the simple discrepancy score or any of the other more complicated variations of it when calculating change. The problem is that none of these deviation-based change scores significantly correlated with the external criteria used in this study.

In this sample and with these neuroimaging biomarkers, the regression-based change scores more strongly related to the external criteria. Additionally (and unlike all cases of the deviation-based change scores), these correlations went in the expected direction, with more “decline” being associated with smaller hippocampi on MRI and more amyloid deposits on PET. The apparent “better” sensitivity of the regression-based change scores might be expected. For example, simple SRBs use baseline cognitive scores (and complex SRB add demographic variables to the equation) to predict a follow-up score. The predicted follow-up score is subtracted from the observed follow-up score and divided by the standard error of the estimate of the regression equation to get an estimate of change. The inclusion of the baseline/demographic variables would intuitively appear to make them more accurate in assessing change, as the SRB models seem to tailor change much more specifically than the deviation-based models. For example, a predicted follow-up score from an SRB may be higher or lower depending on one’s baseline score and demographic variables, whereas RCI +PE and RCI+PEI correct for practice effects uniformly. Calamia et al. (2012) have shown that practice effects are not uniform, and that these effects vary by cognitive domain, test, retest interval, demographic variables, and neurological/psychiatric condition.

The current results cannot conclusively indicate if the simple or complex SRB is the “best” regression-based change formula, as they tended to similarly correlate with the HVLT-R and hippocampal volume and ¹⁸F-Flutemetamol uptake. For the BVMT-R, the complex SRB was statistically related to ¹⁸F-Flutemetamol uptake, whereas the simple SRB was not. For hippocampal volume, neither the simple nor complex SRB on the BVMT-R was related. Although one could interpret these findings to suggest that the complex SRB is preferred for BVMT-R and ¹⁸F-Flutemetamol uptake relationships, it seems wise to be cautious about that conclusion. To remind the reader, simple SRBs use baseline cognitive scores to predict follow-up scores, and complex SRB use baseline scores and demographic variables (e.g., age, education, gender) to predict follow-up scores. In this way, complex SRBs can be more challenging to compute and interpret. Ultimately, it will be the change formula (deviation-based or regression-based) that best relates to outcomes of interest that will be declared the “winner.”

Although this study does not “prove” that regression-based change methods are superior to deviation-based methods in all cases, it does preliminarily indicate that SRBs might be more

applicable than other change scores when examining brain pathology associated with Alzheimer's disease. More globally, this study highlights the need for more external validation of change scores as a way of informing the field as to which method should be used. In our convenience sample of non-demented older adults, we used two neuroimaging biomarkers relevant to Alzheimer's disease. However, other external criteria may be more appropriate in other clinical conditions. For example, return to work might be more applicable in patients with recent traumatic brain injuries. Lesion load on MRI might be more useful in patients with multiple sclerosis. Response to medication might be more informative in patients with depression or Attention Deficit Hyperactivity Disorder. We are not suggesting which external criteria should be used, but we are advocating for studies that compare the different change formulae to some clinically-relevant outcome, as a way of informing the field about which formulae are more useful in predicting change in a research and clinical context.

The current study has a number of notable limitations, which reduces its generalizability. The sample size was small, so these results should be viewed as preliminary and additional validation is needed. The sample was heterogeneous regarding their cognitive status. The analyses of our limited cognitive battery focused on measures of delayed recall. However, this was by design, as we wanted to use cognitive measures that would be most relevant to the chosen biomarkers. We did not correct for multiple comparisons in the statistical analyses. However, statistical significance was not the ultimate goal in this study. We sought to augment the very limited literature that externally validates cognitive change methods, and to provide a model for how this might be done in future studies.

References

- Barr WB, & McCrea M (2001). Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *J Int Neuropsychol Soc*, 7(6), 693–702. [PubMed: 11575591]
- Benedict RHB (1997). *Brief Visuospatial Memory Test-Revised*. Odessa, FL: Psychological Assessment Resources, Inc.
- Blasi S, Zehnder AE, Berres M, Taylor KI, Spiegel R, & Monsch AU (2009). Norms for change in episodic memory as a prerequisite for the diagnosis of mild cognitive impairment (MCI). *Neuropsychology*, 23(2), 189–200. [PubMed: 19254092]
- Brandt J, & Benedict RHB (1997). *Hopkins Verbal Learning Test-Revised*. Odessa, FL: Psychological Assessment Resources, Inc.
- Calamia M, Markon K, & Tranel D (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *Clin Neuropsychol*, 26(4), 543–570. [PubMed: 22540222]
- Chelune GJ, Naugle RI, Luders H, Sedlak J, & Awad IA (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7(1), 41–52.
- Duff K (2012). Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Arch Clin Neuropsychol*, 27(3), 248–261. [PubMed: 22382384]
- Duff K (2014). One-week practice effects in older adults: tools for assessing cognitive change. *Clin Neuropsychol*, 28(5), 714–725. doi: 10.1080/13854046.2014.920923 [PubMed: 24882553]
- Duff K, Atkinson TJ, Suhrie KR, Dalley BC, Schaefer SY, & Hammers DB (2017). Short-term practice effects in mild cognitive impairment: Evaluating different methods of change. *J Clin Exp Neuropsychol*, 39(4), 396–407. doi: 10.1080/13803395.2016.1230596 [PubMed: 27646966]

- Duff K, Beglinger L, Schultz S, Moser D, McCaffrey R, Haase R,...Paulsen J (2007). Practice effects in the prediction of long-term cognitive outcome in three patient samples: A novel prognostic index. *Arch Clin Neuropsychol*, 22(1), 15–24. [PubMed: 17142007]
- Duff K, Beglinger LJ, Van Der Heiden S, Moser DJ, Arndt S, Schultz SK, & Paulsen JS (2008). Short-term practice effects in amnesic mild cognitive impairment: implications for diagnosis and treatment. *Int Psychogeriatr*, 20(5), 986–999. [PubMed: 18405398]
- Duff K, Foster NL, Dennett K, Hammers DB, Zollinger LV, Christian PE,...Hoffman JM (2013). Amyloid deposition and cognition in older adults: the effects of premorbid intellect. *Arch Clin Neuropsychol*, 28(7), 665–671. [PubMed: 23817438]
- Duff K, Foster NL, & Hoffman JM (2014). Practice effects and amyloid deposition: preliminary data on a method for enriching samples in clinical trials. *Alzheimer Dis Assoc Disord*, 28(3), 247–252. doi: 10.1097/WAD.000000000000021 [PubMed: 24614265]
- Duff K, Hammers DB, Dalley BCA, Suhrie KR, Atkinson TJ, Rasmussen KM,...Hoffman JM (2017). Short-Term Practice Effects and Amyloid Deposition: Providing Information Above and Beyond Baseline Cognition. *J Prev Alzheimers Dis*, 4(2), 87–92. doi: 10.14283/jpad.2017.9 [PubMed: 28966919]
- Duff K, Lyketsos CG, Beglinger LJ, Chelune G, Moser DJ, Arndt S,...McCaffrey RJ (2011). Practice effects predict cognitive outcome in amnesic mild cognitive impairment. *Am J Geriatr Psychiatry*, 19(11), 932–939. [PubMed: 22024617]
- Estevis E, Basso MR, & Combs D (2012). Effects of practice on the Wechsler Adult Intelligence Scale-IV across 3- and 6-month intervals. *Clin Neuropsychol*, 26(2), 239–254. doi: 10.1080/13854046.2012.659219 [PubMed: 22353021]
- Fischl B, & Dale AM (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*, 97(20), 11050–11055. doi: 10.1073/pnas.200033797 [PubMed: 10984517]
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C,...Dale AM (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355. [PubMed: 11832223]
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH,...Dale AM (2004). Automatically parcellating the human cerebral cortex. *Cereb Cortex*, 14(1), 11–22. [PubMed: 14654453]
- Frerichs RJ, & Tuokko HA (2005). A comparison of methods for measuring cognitive change in older adults. *Arch Clin Neuropsychol*, 20(3), 321–333. [PubMed: 15797168]
- Hassenstab J, Ruvolo D, Jasielec M, Xiong C, Grant E, & Morris JC (2015). Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*, 29(6), 940–948. doi: 10.1037/neu0000208 [PubMed: 26011114]
- Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD, & Grant I (2001). Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychol*, 16(1), 75–91. [PubMed: 14590193]
- Hinton-Bayre AD (2011). Specificity of reliable change models and review of the within-subjects standard deviation as an error term. *Arch Clin Neuropsychol*, 26(1), 67–75. doi: 10.1093/arclin/acq087 [PubMed: 21147862]
- Hinton-Bayre AD (2016). Clarifying Discrepancies in Responsiveness Between Reliable Change Indices. *Arch Clin Neuropsychol*. doi: 10.1093/arclin/acw064
- Iverson GL (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. *Arch Clin Neuropsychol*, 16(2), 183–191. [PubMed: 14590186]
- Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB,...Dubois B (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, 87(5), 539–547. doi: 10.1212/WNL.0000000000002923 [PubMed: 27371494]
- Jacobson NS, & Truax P (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol*, 59(1), 12–19. [PubMed: 2002127]
- Levine AJ, Hinkin CH, Miller EN, Becker JT, Selnes OA, & Cohen BA (2007). The generalizability of neurocognitive test/retest data derived from a nonclinical sample for detecting change among two HIV+ cohorts. *J Clin Exp Neuropsychol*, 29(6), 669–678. [PubMed: 17691040]

- Lundqvist R, Lilja J, Thomas BA, Lotjonen J, Villemagne VL, Rowe CC, & Thurfjell L (2013). Implementation and validation of an adaptive template registration method for 18F-flutemetamol imaging data. *J Nucl Med*, 54(8), 1472–1478. doi: 10.2967/jnumed.112.115006 [PubMed: 23740104]
- Maassen GH, Bossema E, & Brand N (2009). Reliable change and practice effects: outcomes of various indices compared. *J Clin Exp Neuropsychol*, 31(3), 339–352. [PubMed: 18618359]
- Machulda MM, Hagen CE, Wiste HJ, Mielke MM, Knopman DS, Roberts RO,...Petersen RC (2017). [Formula: see text]Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. *Clin Neuropsychol*, 31(1), 99–117. doi: 10.1080/13854046.2016.1241303 [PubMed: 27724156]
- McSweeney A, Naugle RI, Chelune GJ, & Luders H (1993). “T Scores for Change”: An illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychologist*, 7(3), 300–312.
- Quimet LA, Stewart A, Collins B, Schindler D, & Bielajew C (2009). Measuring neuropsychological change following breast cancer treatment: an analysis of statistical models. *J Clin Exp Neuropsychol*, 31(1), 73–89. [PubMed: 18608651]
- Temkin NR, Heaton RK, Grant I, & Dikmen SS (1999). Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc*, 5(4), 357–369. [PubMed: 10349298]
- Thurfjell L, Lilja J, Lundqvist R, Buckley C, Smith A, Vandenberghe R, & Sherwin P (2014). Automated quantification of 18F-flutemetamol PET activity for categorizing scans as negative or positive for brain amyloid: concordance with visual image reads. *J Nucl Med*, 55(10), 1623–1628. doi: 10.2967/jnumed.114.142109 [PubMed: 25146124]
- Tombaugh TN (2005). Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. *Arch Clin Neuropsychol*, 20(4), 485–503. [PubMed: 15896562]
- Vandenberghe R, Van Laere K, Ivanoiu A, Salmon E, Bastin C, Triau E,...Brooks DJ (2010). 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: a phase 2 trial. *Ann Neurol*, 68(3), 319–329. [PubMed: 20687209]
- Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC,...Alzheimer’s Disease Neuroimaging, I. (2017). Recent publications from the Alzheimer’s Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimers Dement*, 13(4), e1–e85. doi: 10.1016/j.jalz.2016.11.007 [PubMed: 28342697]
- Weissberger GH, Strong JV, Stefanidis KB, Summers MJ, Bondi MW, & Stricker NH (2017). Diagnostic Accuracy of Memory Measures in Alzheimer’s Dementia and Mild Cognitive Impairment: a Systematic Review and Meta-Analysis. *Neuropsychol Rev*, 27(4), 354–388. doi: 10.1007/s11065-017-9360-6 [PubMed: 28940127]
- Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO,...Petersen RC (2004). Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *J Intern Med*, 256(3), 240–246. [PubMed: 15324367]

Table 1.

Cognitive test scores to classify participants.

Cognitive test	Intact	MCI
WRAT-4 Reading	111.1 (4.9)	108.2 (13.1)
RBANS Immediate Memory Index	111.7 (7.4)	85.3 (21.1) *
RBANS Visuospatial Constructional Index	101.0 (15.0)	89.3 (18.0)
RBANS Language Index	104.6 (5.0)	92.8 (10.0) *
RBANS Attention Index	113.1 (14.9)	96.6 (20.7)
RBANS Delayed Memory Index	108.8 (5.2)	80.8 (27.8) *
RBANS Total Scale	111.4 (10.9)	86.2 (20.3) *

Note. WRAT-4 = Wide Range Achievement Test, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status. All scores are age-corrected standard scores from their respective manuals. Means and standard deviations (in parentheses) in columns 2 and 3.

* = $p < 0.05$.

Table 2.

Cognitive test scores and change indices for the sample (n=25).

Cognitive score/change index	HVLT-R	BVMT-R
Baseline visit	5.92 (4.71)	5.08 (3.94)
One-week visit	7.44 (4.57)	6.96 (3.71)
Simple discrepancy	1.52 (1.76)	1.88 (1.62)
Standard deviation index	0.45 (0.52)	0.55 (0.48)
RCI	0.56 (0.65)	0.85 (0.73)
RCI+PE	-0.25 (0.65)	-0.19 (0.73)
RCI+PE Iverson	-0.28 (0.73)	-0.20 (0.77)
Simple SRB	-0.58 (1.26)	-0.29 (0.81)
Complex SRB	-0.69 (1.39)	-0.33 (0.92)

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, HVLT-R = Hopkins Verbal Learning Test – Revised, PE = practice effects, RCI = reliable change index, SRB = standardized regression based, Baseline and One-week visit scores are raw scores.

Table 3.

Correlations between change indices on HVLT-R Delayed Recall and biomarkers (n=25).

Change index	Hippocampal volume	Flutemetamol uptake
Simple discrepancy	0.18	0.16
Standard deviation index	0.18	0.16
RCI	0.18	0.16
RCI+PE	0.18	0.16
RCI+PE Iverson	0.18	0.16
Simple SRB	0.66 ^{*!}	-0.39 [!]
Complex SRB	0.65 ^{*!}	-0.41 ^{*!}

Note. HVLT-R = Hopkins Verbal Learning Test – Revised, PE = practice effects, RCI = reliable change index, SRB = standardized regression based

^{*} = $p < 0.05$ for the primary analyses with Pearson correlations

[!] = $p < 0.05$ (one-tailed) for the secondary analyses with Fisher r -to- z transformation.

Table 4.

Correlations between change indices on BVMT-R Delayed Recall and biomarkers (n=25).

Change index	Hippocampal volume	Flutemetamol uptake
Simple discrepancy	−0.15	−0.03
Standard deviation index	−0.15	−0.03
RCI	−0.15	−0.03
RCI+PE	−0.15	−0.03
RCI+PE Iverson	−0.15	−0.03
Simple SRB	0.25	−0.39
Complex SRB	0.26	−0.54 ^{*/!}

Note. BVMT-R = Brief Visuospatial Memory Test – Revised, PE = practice effects, RCI = reliable change index, SRB = standardized regression based

^{*} = $p < 0.05$ for the primary analyses with Pearson correlations

[!] = $p < 0.05$ (one-tailed) for the secondary analyses with Fisher r -to- z transformation.