Published in final edited form as:

Clin Neuropsychol. 2010 July ; 24(5): 779–792. doi:10.1080/13854041003627795.

Predicting cognitive change within domains

Kevin Duff¹, Leigh J. Beglinger², David J. Moser², and Jane S. Paulsen²

¹University of Utah, Department of Neurology

²University of Iowa, Department of Psychiatry

Abstract

Standardized regression based (SRB) formulas, a method for predicting cognitive change across time, traditionally use baseline performance on a neuropsychological measure to predict future performance on that same measure. However, there are instances in which the same tests may not be given at follow-up assessments (e.g., lack of continuity of provider, avoiding practice effects). The current study sought to expand this methodology by developing SRBs to predict performance on different tests within the same cognitive domain. Using a sample of 127 non-demented community-dwelling older adults assessed at baseline and after one year, two sets of SRBs were developed: 1. those predicting performance on the same test, and 2. those predicting performance on a different test within the same cognitive domain. The domains examined were learning and memory, processing speed, and language. Across both sets of SRBs, one year scores were significantly predicted by baseline scores, especially for the learning and memory and processing speed measures. Although SRBs developed for the same test were comparable to those developed for different tests within the same domain, less variance was accounted for as tests became less similar. The current results lend preliminary support for additional development of SRBs, both for same- and different-tests, as well as beginning to examine domain-based SRBs.

Keywords

Predicting cognition; standardized based regression

There are a variety of choices available to clinicians when trying to determine if a “real” and “meaningful” change in cognition has occurred in their patients. For example, there are several statistical formulas that produce confidence intervals of “normal” change, and scores falling outside this “normal” range are considered to reflect “real” change. Some of these formulas (e.g., Reliable Change Index (Jacobson & Truax, 1991; Zegers & Hafkenscheid, 1994) utilize psychometric properties of the test (e.g., reliability, standard deviation) to produce confidence intervals. Other formulas (e.g., Reliable Change Index adjusted for practice (Chelune, Naugle, Luders, Sedlak, & Awad, 1993) have also attempted to control for practice effects from repeated testing. Alternatively, standardized regression based (SRB) formulas are a third option for assessing if a “real” and “meaningful” change has occurred. SRBs, introduced by McSweeney and colleagues (McSweeney, Naugle, Chelune, & Luders, 1993), use multiple regression algorithms to predict follow-up test performances using baseline test performances and demographic variables. SRBs have some distinct advantages over other methods. For example, SRBs can account for baseline testing performance, regression to the mean, and demographic factors, whereas other Reliable Change Indexes often cannot (Barr, 2002). In

general, SRBs have demonstrated greater sensitivity than other methods of assessing change (Temkin, Heaton, Grant, & Dikmen, 1999).

SRB methodology has been applied to a number of patient samples within neuropsychology, including epilepsy (Hermann et al., 1996; McSweeney et al., 1993; Sawrie, Chelune, Naugle, & Luders, 1996), cardiac conditions (Bruggemans, Van de Vijver, & Huysmans, 1997), and concussion (Barr & McCrea, 2001). This same methodology has also been applied to samples of healthy adults (Attix et al., 2009; Duff et al., 2005; Duff et al., 2004; Duff, Schoenberg et al., 2008; Temkin et al., 1999) to develop “normal” change algorithms. Across multiple studies using SRBs, baseline cognitive performance on a test has consistently been found to be the best predictor of follow-up test performance. Demographic variables (e.g., age, education) tend to slightly improve prediction accuracy.

Despite the growing number of SRB studies, it is questionable how often clinicians utilize these formulas in determining if cognitive change has occurred. SRBs clearly pose some unique challenges in their development. For example, modest samples are needed to develop and validate these formulas (Crawford & Howell, 1998). SRBs can also be cumbersome to calculate by hand (although they are easily adapted to spreadsheet programs like Microsoft Excel). An additional barrier to using SRBs is that they are developed on a specific test, and it is unclear how they might apply to different tests within the same cognitive domain. For example, a clinician might see a patient who was initially tested with the Rey Osterreith Complex Figure Test (ROCFT). On retesting, should the clinician repeat the ROCFT, so the SRBs of Attix et al. (2009) can be used? If, for whatever reason, a different visual memory test is used, can the ROCFT SRBs still be used? To our knowledge, no one has specifically examined if SRBs are only applicable to a specific test, or if SRBs can be used for similar tests within the same cognitive domain (e.g., memory). If SRBs are applicable across multiple tests within a single domain, then they might be more useful for routine clinical practice.

The current study sought to expand SRB methodology by comparing SRBs developed to predict performance on the same test vs. different tests within the same cognitive domain. Since there is considerable shared method variance between similar neuropsychological measures from the same cognitive domain, it is reasonable to hypothesize that both same test SRBs and different test SRBs would significantly predict their respective follow-up scores.

Method

Participants

One hundred twenty-seven community-dwelling older adults participated in the current study, and these participants have been previously described (Duff, Beglinger et al., 2008). Briefly, these individuals were recruited from senior centers and independent living facilities to prospectively study practice effects in older adults. Their mean age was 78.7 (7.8) years and their mean education was 15.5 (2.5) years. Most were female (81.1%) and all were Caucasian. Premorbid intellect at baseline was average (WRAT-3 Reading: $M=108.4$ [6.0]). Depression was minimal (Geriatric Depression Scale: $M=4.0$ [3.3]). Fifty-three individuals (42% of sample) were classified with amnesic Mild Cognitive Impairment according to existing criteria (Petersen et al., 1999), and the remainder were classified as cognitively normal.

Procedures

All participants provided informed consent prior to participation, and all procedures were approved by the local Institutional Review Board. During a baseline visit, all participants completed a battery of neuropsychological tests that included the following: Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), Brief Visuospatial Memory Test

– Revised (BVMT-R), Hopkins Verbal Learning Test – Revised (HVLT-R), Trail Making Test Parts A and B (TMT-A and TMT-B), Symbol Digit Modalities Test (SDMT), Controlled Oral Word Association Test (COWAT), and Animal fluency. Based on the results of an exploratory factor analysis (see Results section), these cognitive tests were divided into the following three domains: 1) learning and memory, 2) processing speed, and 3) language, which are further described in Table 1. All tests were administered and scored as defined in the manual, with the exception of the RBANS Figure Copy and Figure Recall, which is more thoroughly described elsewhere (Duff et al., 2007). After one year, the battery was again repeated. Alternate forms were purposefully not utilized on re-evaluation, as the original study sought to examine practice effects.

Data Analyses

SRBs were calculated similar to the procedures described by McSweeney et al. (McSweeney et al., 1993). Briefly, stepwise multiple regression was used to predict a criterion variable (i.e., one year score) from a set of predictor variables (i.e., baseline score, age, education, gender, and MCI status). For example, baseline HVLT-R Total score, MCI status, and demographic variables were regressed on one year HVLT-R Total score. Baseline and one year scores were the raw test scores. Age was years old at baseline. Gender was coded as male = 0, and female = 1. Education was coded as years. MCI status was coded as intact = 0, and MCI = 1. Two separate sets of SRBs were generated for each cognitive domain: 1. baseline scores predicted one year scores on the same test (e.g., baseline HVLT-R Total predicting one year HVLT-R Total), and 2. baseline scores predicting one year scores on a different test within the same cognitive domain (e.g., baseline List Recall of the RBANS predicting one year HVLT-R Delayed Recall).

In a preliminary attempt to validate these SRBs, the resulting SRBs were then applied to the entire sample. Observed one-year scores were compared to predicted one-year scores with dependent t-tests and correlations. Additionally, for each cognitive variable, standard errors of measurement (SEM) were calculated based on the observed baseline and one-year scores. The SEM, when multiplied by 1.64, creates a 90% confidence interval around the score from that measure. Discrepancy scores between the observed one-year score and the predicted one-year score were generated for each individual in the entire sample, and the percentage of participants that fell outside of the respective 90% confidence interval was examined.

Results

Cognitive scores for the entire sample at baseline and one-year follow-up are presented in Table 2. An exploratory factor analysis (with varimax rotation) of the 14 cognitive variables identified three factors (eigenvalues/% variance: factor 1 = 6.4/45%, factor 2 = 1.7/12%, factor 3 = 1.1/8%). The first factor, which was labelled a “memory” factor, included: HVLT-R Total Recall, HVLT-R Delayed Recall, RBANS List Learning, RBANS List Recall, RBANS Story Memory, and RBANS Story Recall. The second factor, which was labelled a “processing speed” factor, included: RBANS Coding, SDMT, TMT-A, and TMT-B. The final factor, which was labelled a “language” factor, included: RBANS Picture Naming and COWAT. The remaining two cognitive variables, RBANS Semantic Fluency and Animal fluency, had moderate loadings on all three factors, but were ultimately included in the “language” factor due to content similarities. Predicting scores on the same test

As can be seen in the top six rows of Table 3, all one-year learning and memory scores were significantly predicted by their respective baseline scores (p 's < 0.001). For example, the one-year score on the HVLT-R Total Recall was significantly predicted by the baseline score on the HVLT-R Total Recall ($F[1,125]=89.12$, $p<0.001$, $R^2=0.42$). Similarly, all four one-year processing speed measures were significantly predicted by their respective baseline scores

(p 's<0.001, Table 4, top four rows). This pattern was also observed on the four language tasks (p 's<0.001, Table 5, top four rows). In many of these regression models, age also significantly contributed to the prediction of one-year follow-up scores. MCI status only significantly contributed to one of the models (Story Memory).

Predicting scores on different tests within the same domain

As can be seen in the lower portions of Table 3, we were able to significantly predict one-year learning and memory scores using baseline scores from other tests within the same domain (p 's<0.001). For example, the one-year score on the HVLT-R Total Recall was significantly predicted by the baseline score on the RBANS List Learning task ($F[2,124]=29.11$, $p<0.001$, $R^2=0.32$). Similarly, one-year processing speed measures were significantly predicted by baseline scores on other measures within that domain (p 's<0.001, Table 4). For example, the one-year score on the SDMT was significantly predicted by the baseline score on TMT-A ($F[2,122]=26.66$, $p<0.001$, $R^2=0.30$). This pattern was also observed for the four language tasks (p 's<0.05, Table 5). For example, the one-year score on the RBANS Semantic Fluency subtest was significantly predicted by the baseline score on animals ($F[3,119]=21.02$, $p<0.001$, $R^2=0.35$). As observed in the models predicting follow-up scores with the same baseline measure, age also significantly contributed to the prediction of one-year follow-up scores within the same cognitive domain, especially for the learning and memory measures. MCI status also significantly contributed to nine of the models when predicting scores within the same domain.

Preliminary validation of SRBs

The SRBs from Tables 3 - 5 were then applied to each individual in the entire sample, which yielded predicted one-year scores for each cognitive measure. The observed and predicted one-year scores are presented in Table 6, as are difference scores (i.e., observed - predicted), dependent t-test results, and correlations between observed and predicted scores. Confidence intervals at 90% were also generated using observed baseline and one-year scores for the entire sample, and the percentage of participants that fell outside those confidence intervals are presented in the final column of Table 6.

Discussion

To our knowledge, this is the first study to specifically examine SRBs developed for the same test vs. those developed for different tests within the same cognitive domain. If within-domain SRBs can be validated, then the clinical utility for this method of assessing change across time might be increased. Results of the current study are encouraging for these within-domain SRBs. Across three learning and memory measures, baseline scores significantly predicted one year scores, regardless of whether the baseline and one year scores were from the exact same test or different tests (mean R^2 for all memory SRBs = 0.40). This pattern was also observed across four measures of processing speed (mean R^2 for all SRBs = 0.44) and four measures of language functioning (mean R^2 for all SRBs = 0.20). This is consistent with the existing literature on same-test SRBs for other neuropsychological measures (Attix et al., 2009; Barr & McCrea, 2001; Duff et al., 2005; Duff et al., 2004; McSweeney et al., 1993; Sawrie et al., 1996; Temkin et al., 1999). For example, in their sample of 285 healthy adults, Attix et al. (2009) accounted for approximately 22 – 41% of the variance in their SRBs for the ROCFT Delayed Recall. Coincidentally, Attix et al. also examined an alternative-form SRB (i.e., different-test SRB), in which the RAVLT Form A was given at baseline and Form B was given at follow-up. In additional support of the concept of different-test SRBs, their alternate-form SRB was largely comparable to a same-test SRB (RAVLT-Total: $R^2 = 16\%$ and 18% , respectively). In a sample of 223 community-dwelling older adults, Duff et al. (2005) accounted for 43 – 52% of the variance in their SRBs for the List and Story subtests of the RBANS. The relative consistency

of SRBs across studies is encouraging. However, for all SRB studies (same-test and within-domain), there is considerable variance that is not accounted for, and future research should seek methods to improve the prediction of follow-up scores.

Some interesting comparisons are revealed both within and between subsets of SRBs. For example, across all three cognitive domains, the same-test SRBs (e.g., baseline List Recall predicting one year List Recall) tended to account for more variance than the different-test SRBs (e.g., baseline TMT-A predicting Coding): 45% vs. 31% of the variance, respectively (although the Fisher r to z transformation was not statistically significant [$p=0.06$]). This finding should not deter researchers from further developing different-test SRBs, but caution us that same-test SRBs are likely to yield better estimates of change across time. Differences were also observed between cognitive domains, with the memory and processing speed domain SRBs capturing much more variance than the language domain (memory = 39.7%, processing speed = 43.7%, language = 20.3% [processing speed vs. language $p=0.008$, memory vs. language $p=0.02$]). Some of the domain differences may relate to the domains themselves. For example, some domains appeared more heterogeneous (e.g., language tapped naming, phonemic fluency, and semantic fluency) than others (e.g., memory only tapped verbal learning and recall). If a broader range of memory measures were used (e.g., verbal and visual stimuli, single and multi-trial learning, recall and recognition formats), then cross-domain differences might have been smaller. Finally, differences were observed within cognitive domains. The more similar that baseline and one year tests were, the more variance accounted for in the different-test SRBs. For example, TMT-A, SDMT, and Coding are all processing speed tasks. The SRBs between SDMT and Coding, two tasks that are very similar, accounted for much more variance than the SRBs between TMT-A and Coding, two tasks that are less similar (57% and 38% of the variance, respectively [$p=0.02$]). Not only can these SRBs inform us about the amount of change observed across time, but they also seem to be able to inform us about the measures we use.

In a preliminary attempt to validate these SRBs, we compared observed one-year scores to predicted one-year scores. As seen in Table 6, the difference between observed and predicted one-year scores were quite small (e.g., averaging less than one raw score point across all measures). The minimal differences between observed and predicted one-year scores are supported by the dependent t -tests: only 5 of the 50 comparisons were statistically significant (2 of the 14 same-test SRBs and 3 of the 36 different-test SRBs). Of these 5 statistically significant differences, they exclusively involved two tests: Story Recall and Picture Naming. In the case of Picture Naming, the observed raw scores at baseline and one-year were near ceiling (i.e., 9.7 and 9.6 out of 10, respectively), and this may have limited the predictability of these scores with SRBs. This may provide additional evidence as to why the SRBs in the language domain tended to be less informative than SRBs in the processing speed and memory domains. Since Story Recall scores were further from their ceiling (i.e., 8.6 and 8.7 out of 12, respectively), this does not likely account for its relatively poorer performance. Correlations between observed and predicted one-year scores tended to be in the 0.50 – 0.70 range, although some of the language measures were lower (e.g., 0.02 – 0.40). These low correlations create some concern about the clinical utility of the SRBs in this language domain.

A final method for examining the preliminary validity of these SRBs is to see how many predicted scores fell outside of a 90% confidence interval (i.e., true score + error). If our SRBs (and the scores that were used to develop the SRBs) were normally distributed, then approximately 5% of cases should fall below the 90% confidence interval and 5% of cases should exceed it. The observed percentages in the final column of Table 6 seem far from ideal. Although some measures generally met expectations (e.g., same-test List Learning = 7% below and 7% above; same-test TMT-B = 6% below and 6% above), most seemed to fall much further from the 5% below and 5% above mark. For example, the same-test HVLT-R Total score had

15% of cases falling below the $SEM \times 1.64$ cutoff, and 11% of cases falling above that cutoff. Given our methods of calculation, falling below the cutoff means that the predicted score was higher than the observed score, and falling above the cutoff means that the observed score was higher than the predicted score. In the case of the same-test SRB for the HVLT-R Total score, three times as many predicted scores exceeded their observed scores as expected, and more than twice as many observed scores exceeded their predicted scores. In an example with a different-test SRB, the predicted Picture Naming score was always below the observed score, and the observed score was always higher than the predicted score (i.e., 19% below and 0% above for all three different-test SRBs for Picture Naming). As noted earlier, the near ceiling scores for this particular subtest of the RBANS probably contribute to both less accurate SRBs and lower prediction scores (e.g., there was no room for the predicted score to exceed the observed score). Despite these less-than-optimal confidence interval numbers, it should be reiterated that the preliminary validation data was mixed, both for the same-test and different-test SRBs. True validation, however, requires an independent sample, and future studies are encouraged to follow this path.

An unexpected finding in the current study was the role of MCI status in the SRBs. In the same-test SRBs, MCI status only contributed to 1 of these 14 models. In the different-test SRBs, however, MCI status contributed to 9 of these 36 models. It is not clear why MCI status contributed to more different- vs. same-test SRBs. Some of the difference must be due to base rates (i.e., only 14 same-test SRBs vs. 36 different-test SRBs), but this is unlikely to be the only reason. Another possibility might be due to accounted-for variance. In the same-test SRBs, there is more variance already accounted for (45%) compared to the different-test SRBs (31%). Since the same-test SRBs capture more variance, there is less remaining variance for other variables (e.g., MCI status) to be linked to. Additionally, it is possible that the higher variance that is already accounted for in the same-test SRBs is redundant with MCI status. That is, some variance that is shared between baseline and one-year scores on the same test is related to MCI status. However, the relationship between baseline and one-year scores is stronger than the relationship between one-year scores and MCI status, so baseline scores enters the equation first. For the same-test SRBs, MCI status no longer adds anything unique to the equation, so it is not included as an additional variable (except for RBANS Story Memory). For the different-test SRBs, MCI might contribute because the relationship between baseline and one-year scores (on different tests) is weaker than in the same-test SRBs. It is interesting, but not overly surprising, that this “effect” primarily occurs on the memory tests (i.e., 1 of 1 same-test SRBs, 7 of 9 different-test SRBs), as our MCI sample represents the amnesic subtype of this condition. This “effect” also occurs twice within the language domain (i.e., 2 of 9 different-test SRBs), but only when COWAT and RBANS Semantic Fluency are involved. It is possible that at least half our sample (i.e., amnesic MCI) is experiencing some semantic network disruption, which might lead to a similar state of affairs as those that might occur with the memory tasks (i.e., base rates, accounted-for variance, redundancy between scores and MCI status). It is important to note that these discussion points are speculative, and they need to be further evaluated.

Several limitations of the current study should be mentioned. First, the sample size is modest, although large sample sizes do not always dramatically improve SRBs (Crawford & Howell, 1998). Second, individuals with intact and mildly impaired cognition were purposely included in the same SRBs. The inclusion of individuals with a wider range of cognitive functioning may expand the applicability of these SRBs. For example, Heaton et al. (2001) observed that SRBs and other change formula developed on healthy samples might be less applicable in clinical samples. In their work, the authors developed change formulae on healthy adults, but then examined their validity in patients with schizophrenia (who were presumed to be relatively stable). Fewer than expected numbers of these patients with schizophrenia were identified as “not changing” across time. Heaton et al. suggested that samples used to develop SRBs might

include individuals that are neurological stable, but not necessarily cognitively normal, so that a wide range of baseline and follow-up scores are represented. We also examined the potential impact of MCI status on SRBs by including it as a potential variable in the regression models. Third, participants were all Caucasian, well-educated, and community-dwelling, and the applicability of these findings to dissimilar individuals is unknown. Fourth, the cognitive tasks examined in the current study were limited, and a wider range of tasks may have produced different results. Future studies might not only examine additional cognitive measures, but also alternate forms of the same measure (e.g., RAVLT Forms A and B, Attix et al., 2009) to see how this additional variable affects SRB accuracy. Given these limitations, the SRB formulas should be used cautiously. However, the intent of these analyses was only partially related to developing SRBs that could be used across different tests. More broadly, we wish to examine the feasibility of expanding SRB methodology to different tests within the same cognitive domain. If taken further, it might be possible to predict follow-up performance in a cognitive domain (e.g., memory) from baseline performance in that same cognitive domain, regardless of the specific tests that were used at each evaluation. For example, baseline scores on the Wechsler Memory Scale – III Logical Memory might be used to predict a baseline memory domain score, which could be compared to a follow-up memory domain score based on follow-up scores from the HVLTL-R. If SRBs become less specifically tied to a given test and more tied to a cognitive domain, then they become much more flexible as a tool for neuropsychologists. More flexible domain-based SRBs might also allow for more conceptually-based diagnostic impressions (e.g., “memory declines across time” as opposed to “declines in performances on memory tests”). Given the preliminary support for within-domain SRBs observed in the current study, future research might also extend these findings to other salient cognitive domains (e.g., attention, executive functioning), as well as further validate them in patient samples.

Acknowledgments

The project described was supported research grants (R03 AG025850-01; K23 AG028417-01A2) from the National Institutes On Aging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute On Aging or the National Institutes of Health.

References

- Attix DK, Story TJ, Chelune GJ, Ball JD, Stutts ML, Hart RP. The prediction of Change: Normative neuropsychological trajectories. *The Clinical Neuropsychologist* 2009;23(1):21–38. [PubMed: 18720272]
- Barr WB. Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectrum* 2002;7(4):300–302. 304–306.
- Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. *Journal of the International Neuropsychological Society* 2001;7(6):693–702. [PubMed: 11575591]
- Bruggemans EF, Van de Vijver FJ, Huysmans HA. Assessment of cognitive deterioration in individual patients following cardiac surgery: correcting for measurement error and practice effects. *Journal of Clinical and Experimental Neuropsychology* 1997;19(4):543–559. [PubMed: 9342689]
- Chelune GJ, Naugle RI, Luders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology* 1993;7(1):41–52.
- Crawford JR, Howell DC. Regression equations in clinical neuropsychology: an evaluation of statistical methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology* 1998;20(5):755–762. [PubMed: 10079050]
- Duff K, Beglinger LJ, Van Der Heiden S, Moser DJ, Arndt S, Schultz SK, et al. Short-term practice effects in amnesic mild cognitive impairment: implications for diagnosis and treatment. *International Psychogeriatrics* 2008;20(5):986–999. [PubMed: 18405398]

- Duff K, Leber WR, Patton DE, Schoenberg MR, Mold JW, Scott JG, et al. Modified Scoring Criteria for the RBANS Figures. *Applied Neuropsychology* 2007;14(2):73–83. [PubMed: 17523881]
- Duff K, Schoenberg MR, Patton D, Paulsen JS, Bayless JD, Mold J, et al. Regression-based formulas for predicting change in RBANS subtests with older adults. *Archives of Clinical Neuropsychology* 2005;20(3):281–290. [PubMed: 15797165]
- Duff K, Schoenberg MR, Patton DE, Mold J, Scott JG, Adams RA. Predicting change with the RBANS in a community dwelling elderly sample. *Journal of the International Neuropsychological Society* 2004;10:828–834. in press. [PubMed: 15637773]
- Duff K, Schoenberg MR, Patton DE, Mold JW, Scott JG, Adams RL. Predicting cognitive change across 3 years in community-dwelling elders. *The Clinical Neuropsychologist* 2008;22(4):651–661. [PubMed: 17853149]
- Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD, et al. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology* 2001;16(1):75–91. [PubMed: 14590193]
- Hermann BP, Seidenberg M, Schoenfeld J, Peterson J, Leveroni C, Wyler AR. Empirical techniques for determining the reliability, magnitude, and pattern of neuropsychological change after epilepsy surgery. *Epilepsia* 1996;37(10):942–950. [PubMed: 8822692]
- Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting & Clinical Psychology* 1991;59(1):12–19. [PubMed: 2002127]
- McSweeney AJ, Naugle RI, Chelune GJ, Luders H. T scores for change: An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist* 1993;7:300–312.
- Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Archives of Neurology* 1999;56(3):303–308. [PubMed: 10190820]
- Sawrie SM, Chelune GJ, Naugle RI, Luders HO. Empirical methods for assessing meaningful neuropsychological change following epilepsy surgery. *Journal of the International Neuropsychological Society* 1996;2(6):556–564. [PubMed: 9375160]
- Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: a comparison of four models. *Journal of the International Neuropsychological Society* 1999;5(4):357–369. [PubMed: 10349298]
- Zegers, FE.; Hafkenscheid, A. *The Ultimate Reliable Change Index: An Alternative to the Hageman & Arrindell Approach*. Groningen, The Netherlands: University of Groningen; 1994.

Table 1

Cognitive scores divided by domains

Cognitive domain/test	Test description	Maximum raw score
Learning and memory		
HVLT-R Total	Words recalled across three learning trials	36
HVLT-R Delayed Recall	Word recalled after a 20 – 25 minute delay	12
RBANS List Learning	Words recalled across four learning trials	40
RBANS List Recall	Words recalled after a brief delay	10
RBANS Story Memory	Details recalled across two learning trials	24
RBANS Story Recall	Details recalled after a brief delay	12
Processing Speed		
TMT-A	Time to completion connecting circles in ascending numerical order	Discontinued at 300"
TMT-B	Time to completion connecting circles in alternating letter and numerical order	Discontinued at 300"
SDMT	Correct digit-symbol pairs in 90"	110
RBANS Coding	Correct digit-symbol pairs in 90"	89
Language		
COWAT	Correct number of words generated across three 60" trials	n/a
Animals	Correct number of words generated in 60"	n/a
RBANS Picture Naming	Number of correctly named pictures	10
RBANS Semantic Fluency	Correct number of words generated in 60"	n/a

Note. HVLT-R = Hopkins Verbal Learning Test – Revised, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, TMT-A = Trail Making Test Part A, TMT-B = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test, COWAT = Controlled Oral Word Association Test.

Table 2

Cognitive scores at baseline and one-year for entire sample

Scores	Baseline	One-year
HVLT-R Total	23.6 (5.5)	24.5 (6.1)
HVLT-R Delayed Recall	6.9 (3.4)	7.5 (3.6)
RBANS List Learning	25.1 (5.5)	26.0 (6.0)
RBANS List Recall	4.8 (2.3)	5.1 (2.6)
RBANS Story Memory	16.3 (3.8)	16.7 (4.2)
RBANS Story Recall	8.6 (2.2)	8.7 (2.6)
TMT-A	43.6 (15.6)	44.0 (19.2)
TMT-B	113.3 (52.9)	112.3 (64.4)
SDMT	39.8 (9.4)	40.1 (11.7)
RBANS Coding	39.8 (9.7)	39.4 (11.1)
COWAT	38.9 (10.8)	39.7 (12.4)
Animals	17.5 (5.2)	18.0 (6.0)
RBANS Picture Naming	9.7 (0.6)	9.6 (0.7)
RBANS Semantic Fluency	18.8 (4.7)	19.1 (4.9)

Note. HVLT-R = Hopkins Verbal Learning Test – Revised, RBANS = Repeatable Battery for the Assessment of Neuropsychological Status, TMT-A = Trail Making Test Part A, TMT-B = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test, COWAT = Controlled Oral Word Association Test. Means and standard deviations in parentheses.

Table 3

Regression equations for predicting one year learning and memory scores

Cognitive scores	F(df)	R ²	SE _{est} ^a	C ^b	Variables in equation ^c
Same test					
HVLT-R Total	89.12 (1,125)	0.42***	4.69	7.56	+ (baseline HVLT-R Total *0.72)
HVLT-R DR	59.08 (2,123)	0.49***	2.57	11.26	+ (baseline HVLT-R DR *0.59) - (age *0.10)
List Learning	51.59 (2,124)	0.45***	4.45	29.98	+ (baseline List Learning *0.52) - (age *0.22)
List Recall	58.35 (2,124)	0.49***	1.88	7.49	+ (baseline List Recall *0.67) - (age *0.07)
Story Memory	31.68 (3,123)	0.44***	3.23	20.47	+ (baseline Story Memory *0.48) - (age *0.14) - (MCI *1.69)
Story Recall	37.74 (2,124)	0.38***	2.08	8.56	+ (baseline Story Recall *0.61) - (age *0.07)
Different test					
HVLT-R Total	29.11 (2,124)	0.32***	5.09	15.13	+ (baseline List Learning *0.42) - (MCI *3.13)
HVLT-R Total	19.45 (3,123)	0.32***	5.10	32.42	+ (baseline Story Memory *0.41) - (age *0.17) - (MCI *3.64)
HVLT-R DR	30.14 (3,122)	0.43***	2.73	15.78	+ (baseline List Recall *0.54) - (age *0.13) - (MCI *1.33)
HVLT-R DR	26.34 (3,122)	0.39***	2.81	14.82	+ (baseline Story Recall *0.42) - (age *0.13) - (MCI *1.88)
List Learning	55.11 (2,124)	0.47***	4.38	29.33	+ (baseline HVLT-R Total *0.55) - (age *0.21)
List Learning	38.95 (2,124)	0.39***	4.72	46.08	+ (baseline Story Memory *0.49) - (age *0.36)
List Recall	51.45 (2,124)	0.45***	1.94	5.93	+ (baseline HVLT-R DR *0.45) - (age *0.05)
List Recall	28.77 (3,123)	0.41***	2.02	6.62	+ (baseline Story Recall *0.44) - (age *0.06) - (MCI *1.33)
Story Memory	32.32 (2,124)	0.34***	3.48	15.75	+ (baseline HVLT-R Total *0.36) - (age *0.10)
Story Memory	18.93 (3,123)	0.32***	3.56	20.70	+ (baseline List Learning *0.21) - (age *0.11) - (MCI *1.84)
Story Recall	32.74 (2,124)	0.35***	2.13	10.40	+ (baseline HVLT-R DR *0.38) - (age *0.05)
Story Recall	16.17 (3,123)	0.28***	2.24	13.74	+ (baseline List Recall *0.29) - (age *0.08) - (MCI *1.00)

Note. HVLT-R = Hopkins Verbal Learning Test – Revised, DR = Delayed Recall. For the R², the following key denotes the significance of the final step within the regression model:

p<0.001,

**
p<0.01,

*
p<0.05.

^aStandard error of the estimate,

^bConstant,

^cUnstandardized beta weights for other variables in the equation. Age is years old at baseline visit. MCI status is coded as 0 = intact, 1 = MCI. To calculate the Predicted One Year score, use the following formula: (Constant value for the cognitive variable) + (Other variables in equation as noted in the Table).

Table 4

Regression equations for predicting one year processing speed scores

Cognitive scores	F(df)	R ²	SE _{est} ^a	C ^b	Variables in equation ^c
Same test					
TMT-A	37.81 (2,124)	0.38***	15.28	-18.80	+ (baseline TMT-A *0.62) + (age *0.45)
TMT-B	65.67 (2,124)	0.51***	45.27	-105.54	+ (baseline TMT-B *0.71) + (age *1.75)
SDMT	132.35 (1,123)	0.52***	8.15	4.10	+ (baseline SDMT *0.90)
Coding	130.33 (2,124)	0.68***	6.34	22.29	+ (baseline Coding *0.84) - (age *0.21)
Different test					
TMT-A	31.44 (2,124)	0.34***	15.79	-11.56	+ (baseline TMT-B *0.17) + (age *0.47)
TMT-A	67.71 (1,125)	0.35***	15.55	92.38	- (baseline SDMT *1.21)
TMT-A	57.72 (1,125)	0.32***	15.97	88.56	- (baseline Coding *1.12)
TMT-B	39.36 (2,124)	0.39***	50.81	-172.37	+ (baseline TMT-A *1.61) + (age *2.72)
TMT-B	44.29 (2,124)	0.42***	49.62	76.19	- (baseline SDMT *3.20) + (age *2.08)
TMT-B	40.88 (2,124)	0.40***	50.43	51.00	- (baseline Coding *2.89) + (age *2.24)
SDMT	26.66 (2,122)	0.30***	9.84	90.85	- (baseline TMT-A *0.21) - (age *0.53)
SDMT	31.61 (2,122)	0.34***	9.57	84.04	- (baseline TMT-B *0.08) - (age *0.44)
SDMT	124.35 (1,123)	0.50***	8.29	5.95	+ (baseline Coding *0.86)
Coding	48.48 (2,124)	0.44***	8.37	93.73	- (baseline TMT-A *0.28) - (age *0.54)
Coding	57.29 (2,124)	0.48***	8.05	86.36	- (baseline TMT-B *0.10) - (age *0.46)
Coding	103.29 (2,124)	0.63***	6.84	27.55	+ (baseline SDMT *0.80) - (age *0.25)

Note. TMT-A = Trail Making Test Part A, TMT-B = Trail Making Test Part B, SDMT = Symbol Digit Modalities Test. For the R², the following key denotes the significance of the final step within the regression model:

*** p<0.001,

** p<0.01,

* p<0.05.

^aStandard error of the estimate,

^b Constant.

^c Unstandardized beta weights for other variables in the equation. Age is years old at baseline visit. To calculate the Predicted One Year score, use the following formula: (Constant value for the cognitive variable) + (Other variables in equation as noted in the Table).

Table 5

Regression equations for predicting one year language scores

Cognitive scores	F(df)	R ²	SE _{est} ^a	C ^b	Variables in equation ^c
Same test					
COWAT	62.13 (2,122)	0.51 ***	8.86	30.42	+ (baseline COWAT [*] 0.79) – (age [*] 0.27)
Animals	41.01 (1,120)	0.26 ***	5.23	7.96	+ (baseline Animals [*] 0.58)
Picture Naming	27.16 (2,124)	0.31 ***	0.57	5.71	+ (baseline Picture Naming [*] 0.53) – (age [*] 0.02)
Semantic Fluency	57.71 (2,124)	0.48 ***	3.56	1.21	+ (baseline Semantic Fluency [*] 0.72) + (education [*] 0.29)
Different test					
COWAT	11.74 (1,121)	0.09 **	12.05	27.22	+ (baseline Animals [*] 0.71)
COWAT	5.70 (1,125)	0.04 *	12.19	0.40	+ (baseline Picture Naming [*] 4.06)
COWAT	7.75 (3,123)	0.16 ***	11.53	5.09	+ (baseline Semantic Fluency [*] 1.02) + (education [*] 0.87) + (MCI [*] 4.59)
Animals	14.25 (2,121)	0.19 ***	5.44	26.38	+ (baseline COWAT [*] 0.19) – (age [*] 0.20)
Animals	9.82 (1,124)	0.07 **	5.76	34.43	– (age [*] 0.21) [Picture Naming does not contribute]
Animals	6.69 (3,122)	0.14 ***	5.60	25.16	+ (baseline Semantic Fluency [*] 0.34) – (age [*] 0.14) – (gender [*] 3.18)
Picture Naming	8.14 (1,124)	0.06 **	0.69	11.34	– (age [*] 0.02) [COWAT does not contribute]
Picture Naming	7.70 (1,121)	0.06 **	0.66	11.32	– (age [*] 0.02) [Animals does not contribute]
Picture Naming	8.38 (1,125)	0.06 **	0.66	11.35	– (age [*] 0.02) [Semantic Fluency does not contribute]
Semantic Fluency	11.26 (4,120)	0.27 ***	4.27	28.46	+ (baseline COWAT [*] 0.11) – (age [*] 0.19) + (gender [*] 2.64) – (MCI [*] 1.86)
Semantic Fluency	21.02 (3,119)	0.35 ***	4.05	18.02	+ (baseline Animals [*] 0.41) – (age [*] 0.11) + (gender [*] 3.01)
Semantic Fluency	15.10 (2,124)	0.20 ***	4.44	35.39	– (age [*] 0.24) + (gender [*] 2.92) [Picture Naming does not contribute]

Note. COWAT = Controlled Oral Word Association Test. For the R², the following key denotes the significance of the final step within the regression model:

p<0.001,

**
p<0.01,

*
p<0.05.

^a Standard error of the estimate,

^b Constant,

^c Unstandardized beta weights for other variables in the equation. Age is years old at baseline visit. Education is years. Gender is coded as 0 = male, 1 = female. MCI status is coded as 0 = intact, 1 = MCI. To calculate the Predicted One Year score, use the following formula: (Constant value for the cognitive variable) + (Other variables in equation as noted in the Table).

Table 6

Validation of regression formulas in entire sample

Measure	Predicted by	Observed One-Year	Predicted One-Year	Difference	t	r	90% CI	% +/-
Same Test								
HVLT-R Total	HVLT-R Total	24.5 (6.1)	24.6 (4.0)	-0.1 (4.7)	-0.2	0.64	5.4	15/11
HVLT-R DR	HVLT-R DR	7.5 (3.6)	7.5 (2.5)	0.0 (2.5)	0.1	0.70	3.2	13/6
List Learning	List Learning	26.0 (6.0)	25.7 (4.1)	0.3 (4.4)	0.6	0.67	5.5	7/7
List Recall	List Recall	5.1 (2.6)	5.2 (1.8)	-0.1 (1.9)	-0.6	0.70	2.2	12/7
Story Memory	Story Memory	16.7 (4.2)	16.6 (2.8)	0.1 (3.2)	0.4	0.66	4.2	8/8
Story Recall	Story Recall	8.7 (2.6)	8.3 (1.6)	0.4 (2.1)	2.2*	0.62	1.9	12/22
Different Test								
HVLT-R Total	List Learning	24.5 (6.1)	24.4 (3.4)	0.1 (5.0)	0.3	0.57	5.4	13/11
HVLT-R Total	Story Memory	24.5 (6.1)	24.2 (3.5)	0.3 (5.0)	0.6	0.57	5.4	15/11
HVLT-R DR	List Recall	7.5 (3.6)	7.6 (2.3)	-0.1 (2.7)	-0.2	0.65	3.2	13/7
HVLT-R DR	Story Recall	7.5 (3.6)	7.4 (2.3)	0.1 (2.8)	0.4	0.63	3.2	13/11
List Learning	HVLT-R Total	26.0 (6.0)	25.8 (4.1)	0.2 (4.3)	0.4	0.69	5.5	9/9
List Learning	Story Memory	26.0 (6.0)	25.7 (3.7)	0.2 (4.7)	0.5	0.62	5.5	9/10
List Recall	HVLT-R DR	5.1 (2.6)	5.1 (1.7)	-0.1 (1.9)	-0.3	0.67	2.2	11/7
List Recall	Story Recall	5.1 (2.6)	5.1 (1.7)	0.0 (2.0)	-0.3	0.64	2.2	13/12
Story Memory	HVLT-R Total	16.7 (4.2)	16.4 (4.2)	0.3 (3.4)	1.0	0.59	4.2	12/14
Story Memory	List Learning	16.7 (4.2)	16.5 (2.4)	0.2 (3.5)	0.5	0.56	4.2	10/11
Story Recall	HVLT-R DR	8.7 (2.6)	9.1 (1.5)	-0.4 (2.1)	-2.3*	0.59	1.9	25/10
Story Recall	List Recall	8.7 (2.6)	8.4 (1.4)	0.3 (2.2)	1.4	0.53	1.9	18/21
Same Test								
TMT-A	TMT-A	44.0 (19.2)	43.6 (11.8)	0.4 (15.2)	0.3	0.62	16.3	4/7
TMT-B	TMT-B	112.3 (64.4)	112.6 (46.3)	-0.3 (44.9)	-0.1	0.72	47.9	6/6
SDMT	SDMT	40.1 (11.7)	40.0 (8.4)	0.1 (8.1)	0.2	0.72	8.1	8/9
Coding	Coding	39.4 (11.1)	39.2 (9.2)	0.2 (6.3)	0.4	0.82	6.8	8/9
Different Test								
TMT-A	TMT-B	44.0 (19.2)	44.7 (11.4)	-0.6 (15.7)	-0.5	0.58	16.3	6/9

Measure	Predicted by	Observed One-Year	Predicted One-Year	Difference	t	r	90% CI	% +/-
TMT-A	SDMT	44.0 (19.2)	44.2 (11.4)	-0.2 (15.5)	-0.1	0.59	16.3	7/8
TMT-A	Coding	44.0 (19.2)	44.0 (10.8)	0.0 (15.9)	0.0	0.56	16.3	7/6
TMT-B	TMT-A	112.3 (64.4)	111.9 (40.1)	0.3 (50.4)	0.1	0.62	47.9	10/8
TMT-B	SDMT	112.3 (64.4)	112.5 (41.6)	-0.2 (49.2)	0.0	0.65	47.9	9/9
TMT-B	Coding	112.3 (64.4)	112.3 (40.6)	0.0 (50.0)	0.0	0.63	47.9	8/9
SDMT	TMT-A	40.1 (11.7)	40.0 (6.4)	0.1 (9.8)	0.1	0.55	8.1	5/15
SDMT	TMT-B	40.1 (11.7)	40.4 (6.7)	-0.3 (9.5)	-0.3	0.58	8.1	14/13
SDMT	Coding	40.1 (11.7)	40.2 (8.3)	0.0 (8.3)	0.0	0.71	8.1	10/6
Coding	TMT-A	39.4 (11.1)	39.0 (7.4)	0.4 (8.3)	0.6	0.66	6.8	19/22
Coding	TMT-B	39.4 (11.1)	38.8 (7.8)	0.6 (8.0)	0.8	0.69	6.8	17/11
Coding	SDMT	39.4 (11.1)	39.7 (8.8)	-0.3 (6.8)	-0.5	0.79	6.8	15/11
Same Test								
COWAT	COWAT	39.7 (12.4)	40.0 (8.9)	-0.2 (8.8)	-0.3	0.71	9.9	9/10
Animals	Animals	18.0 (6.0)	18.1 (3.0)	0.0 (5.2)	0.1	0.51	6.0	8/7
Picture Naming	Picture Naming	9.6 (0.7)	9.3 (0.4)	0.4 (0.6)	7.4*	0.55	0.7	5/21
Semantic Fluency	Semantic Fluency	19.1 (4.9)	19.2 (3.4)	-0.1 (3.5)	-0.3	0.69	4.4	10/13
Different Test								
COWAT	Animals	39.7 (12.4)	39.6 (3.7)	0.0 (12.0)	0.0	0.30	9.9	19/18
COWAT	Picture Naming	39.7 (12.4)	39.7 (2.6)	0.0 (12.1)	0.0	0.21	9.9	21/19
COWAT	Semantic Fluency	39.7 (12.4)	39.7 (4.9)	0.0 (11.4)	0.0	0.40	9.9	19/19
Animals	COWAT	18.0 (6.0)	18.0 (2.6)	0.1 (5.4)	0.1	0.44	6.0	12/10
Animals	Picture Naming	18.0 (6.0)	17.9 (1.6)	0.1 (5.7)	0.3	0.27	6.0	12/12
Animals	Semantic Fluency	18.0 (6.0)	17.9 (2.3)	0.1 (5.5)	0.3	0.38	6.0	12/8
Picture Naming	COWAT	9.6 (0.7)	9.8 (0.2)	-0.1 (0.6)	-2.2*	0.25	0.7	19/0
Picture Naming	Animals	9.6 (0.7)	9.7 (0.2)	-0.1 (0.6)	-1.9	0.25	0.7	19/0
Picture Naming	Semantic Fluency	9.6 (0.7)	9.8 (0.2)	-0.1 (0.6)	-2.4*	0.25	0.7	19/0
Semantic Fluency	COWAT	19.1 (4.9)	19.2 (2.6)	0.0 (4.2)	-0.1	0.52	4.4	16/15
Semantic Fluency	Animals	19.1 (4.9)	18.9 (2.9)	0.2 (4.0)	0.6	0.59	4.4	11/13
Semantic Fluency	Picture Naming	19.1 (4.9)	18.9 (2.2)	0.3 (4.4)	0.7	0.44	4.4	15/20

Note. Observed one-year score is actual score obtained at one-year follow-up visit. Predicted one-year score is the predicted score expected at one-year follow-up visit based on equations in Tables 3 - 5. Difference is Observed minus Predicted one-year scores. $t = t\text{-value in the dependent t-test comparing Observed and Predicted one-year scores}$. $r = \text{Pearson correlation between Observed and Predicted one-year scores}$. $90\% \text{ CI} = 90\% \text{ confidence interval of the one-year scores}$. $\% -/+ = \% \text{ of participants that fall below the confidence interval and \% of participants that fall above the confidence interval}$.