A Method for Assessing Clinically Relevant Individual Cognitive Change in Older Adult Populations

Stephen M. Sawrie, Daniel C. Marson, Amy L. Boothe, and Lindy E. Harrell

The evaluation of individual cognitive change has relied heavily upon the raw change score, defined simply as the difference between follow-up and baseline scores. However, raw change scores are susceptible to the confounding effects of both regression-to-the-mean and practice effect. The clinical relevance of raw change scores for the older adult is also obscured by normal, age-related cognitive change. The present study illustrates the use of a standardized regression-based (SRB) methodology to generate an alternative to the raw change score; the SRB change score. SRB change scores provide a standardized alternative to the raw change score, allowing the clinician to evaluate the magnitude of change on one or more variables along a common metric that controls for practice effect, regression-to-the-mean, and normal cognitive decline. Case data illustrate how SRB change scores can identify clinically relevant cognitive change in the individual older adult patient.

Identification of clinically relevant cognitive change through serial assessment is becoming increasingly prevalent in neuropsychology and related disciplines (Lezak, 1995), and has particular importance in older adult populations. The accurate diagnosis of dementias and other progressive neurologic conditions is often better characterized by a pattern of change over time than a pattern of cognitive performance at any one point in time (Cronin-Golomb, Corkin, & Rosen, 1993; Flicker, Ferris, & Reisberg, 1993). In addition, clinical research trials for the treatment of Alzheimer’s disease and related dementias rely heavily on serial cognitive testing and accurate detection of meaningful cognitive change over time (Antuono, 1995; Bodick et al., 1997; Bryson & Benfield, 1997; Raskind, Sadowsky, Sigurd, Beitr, & Auster, 1997; Rockwood, Beattie, Eastwood, Feldman, Mohr, Prise-Phillips, & Gauthier, 1997; Rogers & Friedhoff, 1996; Talwalker, 1996; Zemlan, 1996; Zemlan, Keys, Richter, & Strub, 1996). Cognitive rehabilitation programs targeting normal older adults (Bodenburg & Technow, 1992), older adult stroke patients (Gonzalez-Torrecillas, Mendlewicz, & Lobo, 1995; Wilson, 1982), and patients with Alzheimer’s disease (Heiss, Kessler, Mielke, Szelies, & Herholz, 1994) also rely on serial testing to accurately detect relevant cognitive change following treatment.

All of the previously referenced serial assessment research has examined cognitive change primarily at the group level. Although group-level longitudinal data demonstrate average population changes, they obscure the individual variability that is of greatest interest to the clinician. Group-level data are therefore of limited benefit to the clinician assessing change in the individual patient. Consider, for instance, the hypothetical example of an older adult patient who initially obtains a standard score of 100 on the Auditory Immediate Memory index of the Wechsler Memory Scale—III (WMS—III: Weschler, 1997) and a score of 100 one month later. Has this hypothetical patient remained perfectly stable on this index over one month? Data presented in the Technical Manual for the WMS—III and WAIS—III (Psychological Corporation, 1997) indicate that normal older adults ranging from 55 to 89 years of age actually average close to a 10-point increase on the Auditory Immediate Memory index upon retesting one month later. Thus, our hypothetical patient may actually be declining somewhat in immediate auditory memory. A more ambiguous case involves the patient who initially obtains a score of 100 on the WMS—III Auditory Immediate Memory index and then obtains a score of 105 at follow-up evaluation one month later. The group-level data presented in the Technical Manual do little to assist the clinician in interpreting this type of change at the individual level.

For the clinician, the datum of most interest is not the group mean but rather the raw change score (i.e., the difference between a baseline and retest score for a patient) (Matarazzo, Cartm, & Jacobs, 1980). Clinical relevance of the raw change score is based largely on its prevalence in the normal population (Chelune, 1998). A statistically common change score is thought to be drawn from a “normal” population, whereas a change score that is statistically rare in the normal population is thought to be drawn from an “abnormal” population (Jacobson & Truax, 1991). However, studies have suggested that raw change scores can be affected to varying degrees by psychometric confounds such as practice effect and regression-to-the-mean (Chelune, Naugle, Lüders, Sedlak, & Awad, 1993; Cronbach & Furby, 1970) that obscure their clinical relevance. Age-associated cognitive decline introduces yet another challenge to the detection of clinically relevant change scores in older adult populations (Chelune, 1998). Thus, the clinician examining a raw change score for the older adult patient faces the difficult task of determining when that change score represents a clinically relevant change versus normal test-retest variability introduced by practice-effect, regression-to-the-mean, and normal cognitive aging. (In this article, we do not equate “clinically relevant change” with “clinically meaningful change.” The former refers to change that is relevant to the clinician, whereas the latter refers to change that is meaningful to the patient.) These variables exist as potential confounds to the accurate detection of the clinically relevant change score in the older adult, and each will be discussed in more detail.

Recent literature has demonstrated the potentially confounding effect that practice exerts on the raw change score. McCaffrey,
Ortega, Orsillo, Nelles, and Haase (1992) demonstrated significant improvements in serial cognitive assessments across two different medical samples in the absence of any intervention. McCaffrey, Ortega, and Haase (1993) replicated and extended these findings by demonstrating steady improvements across multiple cognitive assessments of chronic cigarette smokers, again in the absence of any intervention. Reliable improvements on memory tests have been reported in longitudinal studies of older adults with test-retest intervals as great as four years (Hultsch, Hertzog, Small, McDonald-Miszcak, & Dixon, 1992; McCarty, Siegler, & Logue, 1982; McDonald-Miszcak, Hertzog, & Hultsch, 1995; Zelinski, Gilewski, & Schaie, 1993). In a review of the literature, Zelinski and Burnight (1997) suggested that prior exposure to testing may influence future performance for up to six years.

The effects of regression-to-the-mean are less well understood in the serial assessment of the older adult. Regression-to-the-mean exists when a measure’s test-retest reliability is less than perfect (Nesselroade, Stigler, & Baltes, 1980), a condition that is very common among cognitive tests. Regression-to-the-mean is manifested by the tendency for baseline scores at either extreme of the distribution to move towards the mean upon retesting. The net effect is to amplify measured change (negative or positive) for individuals initially performing at the upper or lower end of the distribution (Furby, 1973; Laird, 1983; Nesselroade et al., 1980). The extent of regression-to-the-mean is largely dependent on two factors: (1) the test-retest reliability of the measure from which the raw change score was obtained, and (2) initial level of performance (Nesselroade et al., 1980; Speer, 1992). In general, a raw change score’s susceptibility to regression-to-the-mean is inversely related to the measure’s test-retest reliability. A raw change score is also more susceptible to regression-to-the-mean when initial level of performance is particularly deviant. However, research has suggested that the phenomenon of regression-to-the-mean is not universal, and can exist in varying degrees depending on the reliability of the measure and the number of repeated testings (Nesselroade et al., 1980). Despite its potential influence on the magnitude of a raw change score, regression-to-the-mean has typically not been addressed in longitudinal studies of normal cognitive aging.

Normal cognitive decline represents an additional challenge to the clinical interpretation of the older adult change score. Review of the literature on normal, age-associated cognitive decline is beyond the scope of this article, and interested readers are referred to secondary sources for a more detailed discussion of this research (e.g., Birren & Schaie, 1996; Rybash, Roodin, & Hoyer, 1995). However, the broad conclusions from this body of literature have implications for the assessment of clinically relevant cognitive change in the older adult. Although recent cognitive aging literature has demonstrated the stability of intelligence and other cognitive abilities over the adult life span (see Schaie, 1994), longitudinal studies continue to identify discrete cognitive abilities that decline in older adulthood (Arenberg, 1983; Colsher & Wallace, 1991; Giambra, Arenberg, Zonderman, Kawas, & Costa, 1995; Schaie, 1994). In a very thorough review of the literature on cognitive aging, Schaie (1994) concluded that abilities such as perceptual speed and numerical ability show decline by the middle of the fifth decade. By the sixth decade, inductive reasoning and spatial orientation begin to show decline, and by the middle of the seventh decade, verbal ability and verbal memory decline. Within the specific domain of memory, research has suggested that the abilities to learn and retain novel information decrease with age whereas remote memory and overall attention generally remain stable (Cullum, Butters, Troster, & Salmon, 1990; Perlmutter & Nyquist, 1990; Salthouse, 1982). There has also been a distinction drawn between crystallized and fluid cognitive abilities, with the latter demonstrating the greater degree of decline with age (Albert & Heaton, 1988; Botwinick, 1977; Cornelius, 1984; Mack & Carlson, 1978). Thus, many cognitive abilities do decline normally with increasing age. The challenge for the clinician is in determining when a raw change score represents normal cognitive decline and when it represents something more pathological. This challenge is amplified as a function of increasing test-retest duration, where normal cognitive decline can become increasingly pronounced.

The combined effects of normal cognitive aging, practice, and regression-to-the-mean on a raw change score are even less well understood. Theoretically, individuals initially performing at the lower end of the distribution for a particular measure should benefit from both practice effect and regression-to-the-mean. Conversely, these effects should oppose each other in individuals initially performing at the upper end of the distribution, thereby diminishing the measured amount of overall change (Chelune, 1998). For the older adult, the benefits received through practice may be diminished as a result of normal cognitive decline (Chelune, 1998). Indeed, Zelinski and Burnight (1997) have suggested that practice effect may diminish the psychometric measurement of normal cognitive decline, resulting in an overall depiction of cognitive stability over time. Considered separately or in combination, the effects of practice effect, regression-to-the-mean, and normal cognitive decline can obscure the detection of clinically relevant cognitive change in the older adult. What is needed are systematic and empirically grounded approaches for controlling these potential confounds in studies of cognitive change in older adults.

A standardized regression-based (SRB) approach to the evaluation of cognitive change was introduced recently by McSweeney, Naugle, Chelune, and Lüders (1993) in their study of intelligence and memory change following temporal lobe epilepsy surgery. The SRB approach uses test and retest performances from a control sample to develop regression equations that predict retest scores from observed baseline scores. These equations can then be used in a clinical setting to generate predicted follow-up scores relative to the referenced normal control population. These predicted follow-up scores can then be contrasted to observed follow-up scores as a means of determining how far a clinical patient’s score has deviated from normal control scores on follow-up testing. Because SRB regression equations model change in a normal population, the predicted follow-up score will reflect the extent of practice, regression-to-the-mean, and any other test-retest confounds observed in the normal population for a particular measure. Thus, SRB methodology controls for any test-retest confound that may be present for a particular measure in the normal population. Dividing the difference between the predicted and observed retest scores by the standard error of the estimate from the regression equation provides a standardized change score (i.e., an SRB change score). This SRB change score represents an improvement upon the raw change score in that the SRB score reflects both the direction and the magnitude of change for an individual patient or a particular clinical group while controlling for test-retest confounds. The SRB change score is also standardized to a common metric, allowing comparisons of change to be made.
after including demographic variables in the initial SRB regression equations, potential moderating variables such as gender, race, and education can also be controlled.

SRB methodology has already been employed in epilepsy patient populations (Chelune et al., 1993; Hermann et al., 1996; McSweeney et al., 1993; Sawrie, Chelune, Naugle, & Lüders, 1996), but as yet has not been utilized in studies of cognitive change in older adult populations. The present study extends SRB methodology to the study of individual cognitive change in older adult populations. Specifically, this study illustrates the application of SRB methodology using test-retest data from a small sample of normal older adults. Test-retest data from three patients illustrate the clinical application and advantages of the SRB approach.

METHOD

Participants

Participants consisted of 23 neurologically intact, community-dwelling older adults recruited from the Clinical Core of an ongoing Program Project on AD (AD: A Multidisciplinary Approach: NIA 5 PO1 AG06569-05), and from the Clinical Core of a prior Alzheimer Disease Center Core grant (ADC) (NIA, 1 P30 AG101163-01). Diagnosis of normalcy was based on extensive medical, neuroradiological, psychiatric, and neuropsychological testing and on consensus of a neurologist, neuropsychologist, and psychiatrist. Neuropsychological test-retest results were reviewed by a neuropsychologist (S.M.S, author), and no evidence of an emerging dementing process was identified in any of the 23 normal subjects.

Table 1 presents demographic characteristics of the study sample, including gender, age, race, and education. The average test-retest duration was 369 days (SD = 53). The racial composition of the study sample was comparable to that of the Birmingham metropolitan and surrounding areas (35% African American, 65% Caucasian).

Assessment Measures

Each participant was administered a comprehensive neuropsychological baseline assessment followed by retesting approximately one year later. The neurocognitive domains and measures used in the present study are commonly used in the assessment of the older adult (Butters, Salmon, & Butters, 1994).

1. Global cognitive status.—The Dementia Rating Scale total score (DRS; Mattis, 1976, 1988) was used as a measure of global cognitive status. The DRS total score ranges from 0 to 144, and measures cognition across the domains of attention, perseveration, construction, conceptualization, and memory.

2. Verbal memory.—The Logical Memory I raw score (LM I), Logical Memory II raw score (LM II), and Logical Memory percent retention (LM %) \((LM I - LM II) \times 100\) from the Wechsler Memory Scale–Revised (WMS–R: Wechsler, 1987) were used to assess verbal memory. LM I requires immediate recall of 50 bits of information from two paragraphs presented to the patient orally. LM II requires free recall of this information after a 30-minute delay. LM % represents the percentage of material held over the 30-minute delay.

3. Attention.—Attention was assessed using the Digit Span (DS) raw score from the Wechsler Adult Intelligence Scale–Revised (WAIS–R: Wechsler, 1981), and Trails A time in seconds (Reitan & Wolfson, 1993). The WAIS–R DS subtest is comprised of two related but separate tasks. The first task requires subjects to immediately recall a series of orally presented numeric strings of increasing length. The second task requires subjects to recall and verbalize numeric strings backwards. Trails A is a timed visuomotor task that requires participants to draw a line connecting consecutively numbered circles on a worksheet.

4. Language.—Language was assessed using the 60-item Boston Naming Test (BNT: Kaplan, Goodglass, & Weintraub, 1983), a measure of phonemic word fluency (C-F-L: Benton & Hamsher, 1989), and a measure of semantic word fluency (Animals: Spreen & Strauss, 1991). The BNT is a measure of confrontation naming that requires subjects to generate the name of a visually presented line drawing of an object. Semantic and phonemic cues are included as prompts when the subject cannot freely generate the correct name. The word fluency tasks require the participant to generate as many words as possible in one minute to a specific phonemic or semantic category.

5. Executive function.—The cognitive domain of executive function was assessed using the WAIS–R Similarities (Sim) raw score and Trails B time in seconds (Reitan & Wolfson, 1993). The WAIS–R Sim subtest is a verbal conceptualization task that requires the use of abstract abilities to generate similarities between orally presented word pairs (e.g., boat and automobile). Trails B is a measure of mental flexibility that requires the subject to draw a line connecting an alternating series of numbers and letters.

The means and standard deviations of each test at baseline and retesting are presented in Table 2.

Data Analysis

Statistical analysis proceeded in three phases. The first phase examined the extent of normal decline, practice effect, and regression-to-the-mean present in each of our neurocognitive measures. To examine cognitive decline and practice effect, we calculated mean change scores for each test. Paired-sample \(t\) tests were conducted to determine significant movement from baseline to follow-up. Bonferroni correction was employed to reduce the possibility of experimentwise error. Two quantitative indicators were employed to examine regression-to-the-mean. Test-retest reliability coefficients were calculated for each measure. (Recall that regression-to-the-mean occurs only when the correlation between test and retest scores is less than perfect; Nesselroade et al., 1980). We then correlated baseline scores with corresponding new change scores for each neurocognitive
measures. According to Speer (1992), "regression-to-the-mean occurs only when the correlation between the amount of change and the initial score is negative" (p. 403).

The second phase of analysis involved the development of the SRB regression equations. As discussed previously, these equations predict follow-up scores from baseline scores and other demographic moderating variables. Their development in the present study utilized two separate blocks of predictor variables to predict one-year follow-up scores in the normal control sample. The first predictor block included only the baseline score, which was forced into the equation. The second predictor block included race, gender, and education, and was entered into the model in a stepwise fashion using \( p = .05 \) as the criterion for entrance and \( p = .10 \) as criterion for removal. Using this methodology, a regression equation was developed for each of the 11 measures used in the present study.

The third phase of analysis involved generating SRB change scores using test-retest data from three patients with differing dementia types: Alzheimer's disease (AD), vascular dementia (VaD), and Pick's disease. Predicted follow-up scores were generated on each measure for each patient using the regression equations derived in this study in the previous phase of data analysis. To generate SRB change scores, the differences between the predicted and observed follow-up scores were transformed into standardized z-scores (SRB change scores) using the following equation:

\[
SRB \ change \ score = \frac{(Y_o - Y_p)}{SE_{est}}
\]

where \( Y_o \) is the observed retest score, \( Y_p \) is the predicted score, and \( SE_{est} \) is the standard error of the estimate from the regression analysis. In the present study, a 90% confidence interval (±1.64 SRB change scores) was employed such that any SRB change score exceeding the boundary at either end of the distribution represented statistically rare and clinically relevant change. The SRB change scores were depicted graphically for each of the three case examples.

Table 2. Baseline and Retest Scores Across Neurocognitive Domains

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline M</th>
<th>SD</th>
<th>Retest M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Cognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRS Total</td>
<td>139.74</td>
<td>4.01</td>
<td>139.35</td>
<td>4.23</td>
</tr>
<tr>
<td>Verbal Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS-R LM I</td>
<td>29.57</td>
<td>8.48</td>
<td>28.57</td>
<td>8.21</td>
</tr>
<tr>
<td>WMS-R LM II</td>
<td>26.04</td>
<td>10.29</td>
<td>24.96</td>
<td>8.69</td>
</tr>
<tr>
<td>WMS-R LM (% retained)</td>
<td>85%</td>
<td>18%</td>
<td>86%</td>
<td>16%</td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS-R Digit Span</td>
<td>14.70</td>
<td>4.32</td>
<td>15.35</td>
<td>4.61</td>
</tr>
<tr>
<td>Trails A</td>
<td>34.22</td>
<td>13.78</td>
<td>39.30</td>
<td>17.94</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>54.26</td>
<td>6.94</td>
<td>54.78</td>
<td>6.10</td>
</tr>
<tr>
<td>C-F-L</td>
<td>39.52</td>
<td>7.81</td>
<td>39.43</td>
<td>10.84</td>
</tr>
<tr>
<td>Animals</td>
<td>18.05</td>
<td>4.90</td>
<td>16.09</td>
<td>4.28</td>
</tr>
<tr>
<td>Executive Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS-R Similarities</td>
<td>19.48</td>
<td>5.34</td>
<td>21.18</td>
<td>4.49</td>
</tr>
<tr>
<td>Trails B</td>
<td>99.14</td>
<td>77.86</td>
<td>99.43</td>
<td>60.80</td>
</tr>
</tbody>
</table>

RESULTS

Test-Retest Psychometric Confounds

Table 3 sets forth the three different measurements of potential test-retest psychometric confounds: mean raw change scores, test-retest reliabilities, and the correlations between baseline and raw change scores. None of the baseline to retest mean differences exceeded the corrected (Bonferroni) \( p \) value of .005. Test-retest reliability coefficients ranged from .67 (Trails B) to .97 (BNT). The WMS-R and WAIS-R subtests, BNT, and the DRS ranged from an acceptable .81 (LM %) to an extremely robust .97 (BNT). The other measures demonstrated marginal test-retest reliabilities, ranging from .67 (Trails B) to .76 (C-F-L). The correlations between baseline performance and change scores revealed that LM %, BNT, Animals, WAIS-R Similarities, and Trails B were significantly susceptible to regression-to-the-mean.

SRB Regression Equations

Table 4 sets forth the results of the regression analyses used to derive the SRB change scores. Demographic variables entered into only 4 of the 11 analyses. Race significantly enhanced the prediction of the DRS, whereas gender significantly enhanced the prediction of LM I, Trails A, and C-F-L. Education did not enter into any of the equations.

Each of the regression equations generated a predicted 1-year follow-up score for each person. The standardized difference between the predicted and observed retest score provided a quantitative indicator of change using a common metric (i.e., the SRB change score). The clinical application of these SRB change scores are illustrated below through individual case examples.

Individual Clinical Case Examples

The application of SRB change scores and their utility in understanding clinical change is demonstrated using cognitive
Table 4. Regression Equations for Cognitive Test Measures by Domain

<table>
<thead>
<tr>
<th>Variable</th>
<th>R*</th>
<th>SE of R*</th>
<th>C*</th>
<th>β1</th>
<th>β2</th>
<th>β3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Cognition</td>
<td>.89</td>
<td>1.98</td>
<td>59.97</td>
<td>.60</td>
<td>-3.50</td>
<td></td>
</tr>
<tr>
<td>DRS Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS-R LM I</td>
<td>.92</td>
<td>3.39</td>
<td>-1.26</td>
<td>.85</td>
<td>3.47</td>
<td></td>
</tr>
<tr>
<td>WMS-R LM II</td>
<td>.89</td>
<td>4.05</td>
<td>5.92</td>
<td>.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS-R LM (% retained)</td>
<td>.81</td>
<td>0.10</td>
<td>0.22</td>
<td>.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS-R Digit Span</td>
<td>.85</td>
<td>2.51</td>
<td>2.52</td>
<td>.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trails A</td>
<td>.81</td>
<td>11.11</td>
<td>16.94</td>
<td>1.13</td>
<td>-10.38</td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>.97</td>
<td>1.54</td>
<td>8.39</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-F-L</td>
<td>.86</td>
<td>5.69</td>
<td>-12.58</td>
<td>1.14</td>
<td>5.24</td>
<td></td>
</tr>
<tr>
<td>Animals</td>
<td>.69</td>
<td>3.37</td>
<td>4.81</td>
<td>.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Executive Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAIS-R Similarities</td>
<td>.82</td>
<td>2.72</td>
<td>7.67</td>
<td>.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trails B</td>
<td>.67</td>
<td>46.22</td>
<td>44.86</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Multivariate correlation coefficient and standard error of the estimate from regression analysis; C*constant; β1unstandardized beta for baseline measure; β2unstandardized beta for race (1 = White, 2 = Black); β3unstandardized beta for gender (1 = female, 2 = male).

baseline and follow-up test data from three actual patients with different clinical diagnoses: Alzheimer’s disease (AD), vascular dementia (VaD), and Pick’s disease. Each case is described in three parts. The first part outlines the nature of the referral and summarizes the results of the initial cognitive evaluation. The second part summarizes the results of the follow-up evaluation. The third part summarizes a retrospective SRB change score analysis for the 11 measures used in the present study. The figures associated with each case include tabulated baseline, follow-up, and change score data for each of the 11 measures, as well as a graphical representation of the SRB change score “profile.” As outlined above, any change score exceeding the ± 1.64 SRB change score interval was considered statistically rare (p < .05) and therefore clinically relevant. This boundary is denoted by reference lines in each figure.

Case 1: Alzheimer’s Disease

Referral and initial evaluation.—The patient in Figure 1 was a 66-year-old, right-handed, retired engineer with 15 years of education who presented to the neurologist for evaluation of mild, progressive short-term memory complaints. Mental status examination revealed difficulties in verbal recall and visuospatial construction. A cranial CT and short term scalp EEG were normal. The results of his initial neuropsychological evaluation were consistent with possible mild dementia (DRS = 136/144, CDR = 1.0). However, his only impaired test performances were on measures of short-term and delayed visual memory (mild), high load verbal learning and recall (mild), mental flexibility (severe), propositional auditory comprehension (mild), semantic word fluency (mild), and block design (moderate). Short-term and delayed narrative verbal memory and overall delayed recall for verbal and visual information were diminished relative to premorbid estimates but still within normal limits. On the basis of interview and test data, the patient appeared to be in the early stage of AD.

Follow-up evaluation.—At retesting approximately one year later, the patient’s overall cognitive status had deteriorated significantly (DRS = 115, CDR = 1.0). The patient now demonstrated mildly impaired short-term verbal memory and severely impaired high load verbal learning and recall. However, delayed narrative verbal memory was still in the low average range, although diminished relative to baseline testing and to estimated premorbid status. The diagnostic assignment remained mild AD.

SRB change score analysis.—Retrospective application of SRB normative equations to the patient’s baseline and follow-up testing revealed significant declines over a one-year period in the areas of short-term and delayed verbal memory (LM I, LM II), retention of verbal material (LM %), auditory attention (WAIS–R DS), confrontation naming (BNT), semantic word fluency (Animals), and mental flexibility (Trails B). The patient did not demonstrate significant change in areas such as visual-motor tracking/attention (Trails A), phonemic word fluency (C-F-L), and verbal concept formation (WAIS–R Sim). This pattern of change appeared consistent with the decline in cognitive function found in mild AD.

Case 2: Vascular Dementia

Referral and initial evaluation.—The patient in Figure 2 was a 69-year-old, right-handed, retired anesthesiologist who presented to the neurologist for evaluation of possible peripheral neuropathy. Mental status examination revealed difficulties
with concentration, calculation, and memory suggestive of a possible dementia. A cranial MRI revealed possible old ischemic changes in the areas of the thalamus and pons. There was no clear evidence of recent stroke. A comprehensive neuropsychological evaluation revealed evidence of a mild dementia of unclear etiology (DRS = 139/144, CDR = 0.5), although the patient’s only clearly impaired performances were on measures of visuospatial construction, mental flexibility, and high load auditory attention. She demonstrated a mild impairment in fine motor dexterity for her dominant, right hand. Her performance on measures of executive function and delayed visual memory were in the low average range, and diminished relative to estimated premorbid status. Her overall cognitive profile and MRI findings raised the possibility of ischemic VaD.

Follow-up evaluation.—Upon retesting one year later, the patient’s only impaired performances were on measures of mental flexibility and fine motor dexterity for the dominant hand. Executive function was still in the low average range for age and occupational background. However, visuospatial construction was clearly intact, and delayed visual memory had improved to the high average range. The patient’s follow-up cognitive profile again was viewed to be potentially consistent with ischemic VaD.

**SRB change score analysis.**—Retrospective application of SRB norms to this patient’s baseline and follow-up test scores revealed significant improvements over a one-year period in the areas of verbal memory (LM I, LM II) and confrontation naming (BNT), with no areas of significant decline. As discussed below, this pattern of change appeared consistent with the spontaneous cognitive recovery that follows a vascular insult or event.

**Case 3: Pick’s Disease**

**Referral and initial evaluation.**—The patient in Figure 3 was a 63-year-old, right-handed pastoral counselor with two master’s degrees and 18 years of combined formal education. She presented to the neurologist at the request of her husband for evaluation of mild cognitive difficulties and depression. Mental status examination revealed mild evidence of anomia and decreased abstraction with relatively preserved short-term memory. Cranial CT revealed very mild microangiopathic involvement with no acute abnormalities. Subsequent neuropsychological evaluation was consistent with a possible mild dementia (DRS = 130/144, CDR = 0.5). Short-term and delayed verbal and visual memory were in the average range, delayed recall was in the low average range, and high load verbal learning and recall were in the low average to very mildly impaired range. Etiology of the patient’s possible mild dementia was unclear.

**Follow-up evaluation.**—Upon retesting approximately one year later, the patient now demonstrated impairment on virtually every cognitive measure. In interview, the patient’s husband reported drastic personality changes characterized by extreme disinhibition,
impulsivity, and repetitive, purposeless behavior. On the basis of these findings, her overall neurocognitive profile and behavioral changes were viewed to be frontally mediated. The patient was eventually diagnosed with rapidly progressive Pick's disease.

**SRB change score analysis.**—Retrospective use of the SRB norms illustrated the severe and global cognitive decline from baseline to follow-up one year later. The pattern of cognitive decline depicted by the SRB norms revealed a dramatic decline in overall cognition (DRS), short-term and delayed verbal memory (LM I, LM II, LM %), confrontation naming (BNT), phonemic and semantic verbal fluency (C-F-L, Animals), verbal concept formation (WAIS-R Sim), and mental flexibility (Trails B). The patient did not decline dramatically in auditory attention (WAIS-R DS). As discussed below, this pattern of change appeared consistent with a dramatic loss of frontally mediated executive abilities.

**DISCUSSION**

Consistent with previous literature (Hermann et al, 1991; McCaffrey et al., 1992), our findings suggested that raw change scores are influenced by varying combinations of practice effect and regression-to-the-mean. Test-retest reliability coefficients revealed marginal reliability for Trails A, Animals, and C-F-L. Correlations between baseline and change scores revealed statistically significant susceptibility to regression-to-the-mean for WMS–R LM %, BNT, Animals, and WAIS-R Sim. The present study also suggested that normal cognitive decline presents an additional challenge to the accurate detection of clinically relevant change in older adults. In the present study, means for each of the measures remained generally stable over the one-year test-retest duration. However, it is interesting to note that in Sawrie and colleagues (1996), 15 of 32 measures administered to a younger adult sample of epilepsy surgery candidates demonstrated statistically significant improvements, with no significant declines. Of the five measures used in both studies (LM I, LM II, LM %, BNT, & C-F-L), four of the five demonstrated statistically significant improvements over an approximately eight-month retest duration in Sawrie and colleagues (1996), whereas none of the five in the present study demonstrated significant improvement or decline over approximately 12 months. Thus, it may be that for some measures the effects of normal cognitive decline may actually suppress the artificial improvement introduced by practice, resulting in a greater degree of overall test stability in older adult populations. This is consistent with findings reported by Chelune (1998) suggesting that the effects of normal cognitive decline may begin to outweigh the effects of practice in older adult populations.

We found that SRB methodology clarified the interpretation of the raw change score. For instance, the patient depicted in Figure 2 (VaD) completed Trails B in 135 seconds during her initial evaluation. Upon retesting one year later, she completed the task in 164 seconds, for a raw change score of +29. This kind of decline may intuitively appear to be significant clinically. However, the relatively low test-retest reliability coefficient for Trails B (see Table 3) suggests that this measure may not be reliable over one year in a control sample of neurologically intact older adults. When the difference between the patient's observed and predicted follow-up scores was standardized using the appropriate SRB equation, her resulting SRB change score was only –1.3, which was still within the boundary of our cutoff values of ±1.64. Thus, the patient's increase in performance by 29 seconds on Trails B was considered statistically common relative to our control group, and therefore not clinically relevant. Conversely, consider this same patient's test-retest performance on BNT. She obtained a 55/60 during her initial evaluation, and a 58/60 one year later. This raw change score of only 3 points seems insignificant at first glance. However, the SRB change score was +1.9, which exceeds the upper boundary of our SRB change score cutoff values. Thus, the patient's seemingly modest improvement on BNT was actually statistically rare and therefore clinically relevant relative to our control sample of neurologically intact older adults. Consider also the patient with mild AD who declined by only three raw score points on LM I (Figure 1). This raw change score of only –3 points may seem small intuitively. However, application of SRB methodology generated a statistically rare and clinically relevant SRB change score of –1.9. Thus, SRB methodology clarified the clinical significance of each of these potentially misleading raw change scores by providing a standardized measure of change relative to the test-retest performance of our older adult control sample.

SRB methodology places each SRB change score on a common metric, which allows for the generation of an overall change score profile. The common metric allows for statements to be made about the relative magnitude of change across the different measures in the SRB change score profile. The patient depicted in Figure 1 (AD) obtained an SRB change score profile that was generally consistent with the cognitive decline seen in mild AD. For instance, he demonstrated significant decline on all measures of verbal memory (LM I, LM II, LM %), consistent with the decline in short-term and delayed memory seen in AD (Butters et al., 1994). He also demonstrated a characteristic decline on measures of semantic knowledge (BNT; Animals) (Butters et al., 1994; Smith, Murdoch, & Chenery, 1989). The patient in Figure 2 (VaD) obtained an overall SRB change score profile that revealed significant improvement in verbal memory and expressive language function, with no significant cognitive declines. This profile was consistent with spontaneous cognitive recovery following a vascular insult or event. The SRB change score profile for the patient depicted in Figure 3 (Pick's disease) revealed a significant and pervasive cognitive decline. This patient's profile probably reflects the extent to which executive function may mediate performance on neurocognitive measures to varying degrees. However, the profile suggests further that although the patient with Pick's disease experienced a global cognitive decline, this decline was most prominent in the areas of short-term verbal memory (LM I), retention of verbal material (LM %), confrontation naming (BNT), and verbal concept formation (WAIS-R Sim).

SRB change scores may also be viewed as dynamic, changing as a function of initial level of performance. Consider, for instance, the patient with mild AD who had a raw change score of –3. This patient's baseline score of 14 on LM I fell to 11 in follow-up testing one year later. Using the appropriate regression equation (see Table 4), this patient's predicted follow-up score on LM I was 18. The standardized difference between the predicted follow-up score of 18 and the observed follow-up score of 11 yielded an SRB change score of –1.9. Now consider a hypothetical patient who obtains a score of 30 on LM I at baseline and a score of 27 one year later. This hypothetical
patient has exactly the same raw change score as the patient with mild AD. However, the predicted follow-up score for the hypothetical patient is only 31. The resulting SRB change score in this instance is only -1.2, which is statistically common relative to our control group, and therefore not clinically relevant. Thus, identical raw change scores can take on much different meanings depending on initial level of performance. For this reason, SRB change scores operate dynamically depending on the level of baseline performance.

For all of these reasons, SRB methodology serves as a potentially powerful technique in assessing clinically relevant change in the individual older adult. Not only does it offer the clinician an empirically grounded means of detecting change while accounting for test-retest confounds, it also depicts the magnitude of that change. Furthermore, standardization allows for the detection of a pattern of change across a number of neuropsychological variables. SRB methodology may also be used in research settings. SRB change score “profiles” may be generated for any disease entity, allowing more accurate comparisons and distinctions to be made on the basis of their neurocognitive sequelae. This approach may be of particular importance in studies evaluating the efficacy of change agents such as medications or treatment programs. SRB methodology can be used to depict clinically relevant cognitive change in any longitudinal design (clinical, scientific, or otherwise) that generates baseline and retest scores in an older adult population over specified periods of time.

An important distinction made in this article concerns the difference between clinical relevance and clinical meaningfulness. The SRB approach offers an objective, empirical means of identifying cognitive change scores that have relevance to the clinician. The meaning of these change scores to the individual patient has yet to be determined. Accordingly, future studies should investigate the clinical meaning of statistically derived and clinically relevant SRB change scores.

The present study has several limitations. The generation of SRB normative equations is limited by the “ceiling” of each cognitive test. SRB regression equations cannot be developed for tests with low ceilings in which virtually every normal older adult obtains a perfect score. Tests such as these do not reflect sufficient variability in normal populations. Such limited variability in the normal population precludes the development of a predictive regression equation, which is the analytic underpinning of SRB methodology. Furthermore, an SRB change score is limited by the “floor” of each test. Once a patient has reached the floor of a particular test, he or she no longer demonstrates decline on that measure. This does not necessarily mean that the patient has stopped declining in that particular cognitive area. It simply means that the test is no longer able to capture the patient’s true performance in that cognitive domain.

We wish to emphasize that the present study is intended simply to illustrate a new approach to the assessment of clinically relevant individual cognitive change in older adults. It should be noted that our sample size was small and therefore probably not fully representative of the general older adult population. For instance, education did not enter into any of the 11 regression equations in the present study. This probably reflects the restricted range of education in the present sample, and suggests that the sample may not represent a broad segment of the older adult population. Readers should therefore be cautioned against using the regression equations from the present study to examine relevant cognitive change in their own clinical patients or samples.

We feel that SRB methodology may be useful in generating actual change score normative data when applied to test-retest data from larger, population-based samples of normal older adults. Although several recent studies have provided normative data for older adults on a variety of cognitive measures (see Malec, Ivnik, & Smith, 1993, for a concise summary), no normative data for cognitive change scores currently exists for older adults. Such normative data would provide the clinician with an empirically grounded basis for distinguishing clinically relevant change from both psychometric confounds and normal cognitive decline in the serial assessment of the older adult. These studies might also develop SRB equations for cognitive measures other than the 11 used in the present study. Finally, studies should apply SRB methodology to measures administered over differing test-retest durations. This would allow the clinician to identify clinically significant cognitive change scores derived from test-retest time intervals other than the one-year interval used in the present study.

Although this study specifically targeted change in older adult populations, it also addressed problematic issues of evaluating cognitive change in any population or setting. It is our hope that these and other change score methodologies will be used in other populations and settings in order to expand our ability to assess clinically relevant individual cognitive change objectively.

ACKNOWLEDGMENTS

Address correspondence to Dr. Stephen M. Sawrie, UAB Department of Neurology, Jefferson Tower (1216), 625 19th Street South, Birmingham, AL 35233. E-mail: ssawrie@uabmc.edu

REFERENCES
