Assessment of cognitive deterioration in individual patients following cardiac surgery

Bruggemans, E.F.; van de Vijver, F.J.R.; Huysmans, H.A.

Published in:
Journal of Clinical and Experimental Neuropsychology

Publication date:
1997

Citation for published version (APA):
Bruggemans, E. F., van de Vijver, F. J. R., & Huysmans, H. A. (1997). Assessment of cognitive deterioration in individual patients following cardiac surgery: Correcting for measurement error and practice effects. *Journal of Clinical and Experimental Neuropsychology*, (19), 543-559.

Journal of Clinical and Experimental Neuropsychology
1997, Vol. 19, No. 4, pp. 543-559

1380-3395/97/1904-543$12.00

Assessment of Cognitive Deterioration in Individual Patients Following Cardiac Surgery: Correcting for Measurement Error and Practice Effects*

Eline F. Bruggemans[1], Fons J. R. Van de Vijver[2], and Hans A. Huysmans[1]
[1]Department of Cardio-Thoracic Surgery, University Hospital, Leiden, The Netherlands, and [2]Tilburg University, The Netherlands

ABSTRACT

Assessment of cognitive change in individual patients may be confounded by unreliability of test scores and effects of repeated testing. An index correcting for both problems is proposed and compared with change indices that do not or do not adequately deal with measurement error and practice effects. These indices were used to examine cognitive deterioration in a sample of 63 patients undergoing cardiac surgery. It was demonstrated that for test measures with a low reliability, failure to correct for measurement error resulted in overestimation of deterioration rates. For test measures with a high reliability, but showing substantial practice effects, failure to correct for practice effects resulted in underestimation of deterioration rates. With the proposed index, cognitive deterioration shortly after cardiac surgery was most frequently observed for attention and psychomotor speed, less frequently for verbal fluency, and only occasionally for learning and memory.

Medical intervention may be accompanied by changes in cognitive functioning. Negative cognitive effects have been reported, for example, following epilepsy surgery (Chelune, 1992; Naugle, 1992), cardiac surgery involving extracorporeal circulation (Benedict, 1994; Newman, 1993), intrathecal chemotherapy or central nervous system radiation therapy (Brown & Madan-Swain, 1993; Fletcher & Copeland, 1988; Roman & Sperduto, 1995), and pharmacotherapy of Parkinson's disease (Karayanidis, 1989; Saint-Cyr, Taylor, & Lang, 1993). Positive effects have sometimes been reported after pharmacotherapy of hypertension (Waldstein, Manuck, Ryan, & Muldoon, 1991) and of AIDS or AIDS-related complex (Everall, 1995; Maj, 1990). Cognitive effects of medical intervention are most frequently studied using a pretest-posttest design. Pretest to posttest changes in performance on neuropsychological tests are usually evaluated by means of a statistical test of the changes in group means. This method, however, is often not satisfactory for clinical practice because overall changes in group performance do not disclose changes in individual patients. Assessment of cognitive change at the individual level is important for evaluating individual risks or benefits of medical treatment and informing patients about anticipated outcomes.

Different indices for identifying individual change can be found in the literature. Most of these indices define an individual's test performance as significantly changed if the difference in test scores obtained before and after treatment (i.e., the observed change score) is larger than some criterion measure. For instance, in recent studies of the negative cognitive effects of temporal lobectomy or cardiac surgery, a decline in

---

* This study was supported by Grant 89.249 from the Netherlands Heart Foundation.
Address correspondence to: Eline F. Bruggemans, Department of Cardio-Thoracic Surgery, University Hospital, P.O. Box 9600, 2300 RC Leiden, The Netherlands.
Accepted for publication: January 29, 1997.

test performance was defined significant if a patient's pre- to post-operative test score decrement was at least one standard deviation (*SD*), the *SD* being generally determined on the distribution of the preoperative scores in the patient sample (Loring et al., 1995; Phillips & McGlone, 1995; Sellman, Holm, Ivert, & Semb, 1993; Shaw et al., 1986, 1987; Treasure et al., 1989). Such an SD criterion, however, can be criticized because it does not adequately deal with the problem of measurement error (i.e., imperfect reliability of test scores). When an unreliable test is applied, score fluctuations of more than one *SD* that are entirely due to measurement error will frequently be observed. Moreover, random score fluctuations may cause change scores of less than one *SD* when intervention effects do exist. Hence, application of the SD criterion may lead to both overestimation and underestimation of incidence rates.

Several suggestions have been made as to how measurement error can be taken into account in the assessment of individual change. Some authors have proposed an index in which the observed change score is divided by the standard error of measurement (*SEM*) of the pretest scores instead of the *SD* (e.g., Edwards, Yarvis, Mueller, Zingale, & Wagman, 1978; Jacobson, Follette, & Revenstorf, 1984; Shatz, 1981). More specifically, these authors suggested that a change score larger than 1.96 *SEM*, or, with rounding, 2 *SEM*, could be considered as statistically significant ($p < .05$). At first sight, this SEM criterion seems less arbitrary than the SD criterion because it takes into account test reliability: the less reliable the test, the greater the amount of score change required to be significant. Yet, it does not compensate for all effects of measurement error. A first criticism concerns the fact that the SEM statistic does not make allowance for the unreliability of the posttest scores. Score changes should be corrected for unreliability of the instrument at both the pretest and the posttest. Accordingly, alternative standard error measures have been put forward as a more appropriate denominator in the above change index, such as the standard error of the difference between the observed pre- and post-test scores (e.g., Christensen & Mendoza, 1986; Jacobson & Truax, 1991) and the standard error of prediction (e.g., Hsu, 1989, 1995; Knight, 1983).

A second criticism concerns the use of the observed change score in the numerator of the above change index. Unfortunately, all indices using this numerator can be criticized for having ambiguous statistical properties, irrespective of the type of standard error in the denominator. The reason is that the standard errors make, in effect, reference to estimated true change scores rather than observed change scores. Sampling distributions of statistics that are based on mixtures of observed-score and estimated-true-score characteristics are usually unknown. Consequently, the confidence intervals that are required to judge whether the individual has significantly changed cannot properly be determined. Thus, in order to correct for all effects of measurement error, the numerator of the index has also to be adapted. It is important to note that the recommendation to employ the individual's estimated true pretest score as a substitute for the observed pretest score (e.g., Knight, 1983; Speer, 1992, 1993) does not entirely solve the problem. The numerator should comprise both the estimated true pretest and the estimated true posttest score (e.g., Hageman & Arrindell, 1993). Recently, Zegers and Hafkenscheid (1994) suggested an index that adequately deals with the above criticisms with regard to both numerator and denominator. Their Reliable Change (RC) index is defined as the ratio of the estimated true change score to the estimated standard error of this score (details are given below).

All change indices described thus far have still one problem in common: They do not address the practice effects that frequently arise from repeated neuropsychological testing. In healthy subjects, improved test scores at retesting have been reported for many cognitive performance tests (e.g., Feinstein, Brown, & Ron, 1994; Macciocchi, 1990; Wing, 1980). Practice effects and measurement error should be carefully distinguished. Both reliable and unreliable tests may show practice effects. Some test measures with a high test-retest reliability, such as the Verbal IQ, Performance IQ, and Full Scale

IQ of the Wechsler Adult Intelligence Scale (WAIS) or WAIS-Revised, showed a low test-retest stability in absolute test scores (Matarazzo, Carmody, & Jacobs, 1980; Matarazzo & Herman, 1984). Thus, also with a reliable test, practice effects should be dealt with in the assessment of intervention effects in individual patients. Otherwise, negative cognitive effects of intervention could be underestimated and positive effects overestimated.

A method that considers both measurement error and practice effects was recently proposed by Chelune, Naugle, Lüders, Sedlak, and Awad (1993) in a study on cognitive change following temporal lobectomy. These authors subtracted the mean practice effect observed in a control group from the patient's pre- to post-operative change in test scores before applying the RC index as defined by Jacobson and Truax (1991). There are, however, two difficulties associated with this method. First, measurement error is not adequately dealt with because the numerator of the RC index of Jacobson and Truax includes observed change scores rather than estimated true change scores. Second, it is counterintuitive to correct individual change scores by using the mean change score of controls because there is ample evidence of marked individual differences in the magnitude of practice effects (Matarazzo et al., 1980; Matarazzo & Herman, 1984).

A method that makes allowance for individual differences in practice effects and that is more familiar in the literature was employed by the same group of authors in analyzing the same data set (McSweeny, Naugle, Chelune, & Lüders, 1993). In a linear regression model, the posttest scores of the control group were predicted on the basis of their pretest scores. The regression coefficient and intercept of the control group's regression line were then applied to obtain predicted posttest scores for individual patients. A patient was said to be significantly changed at the posttest if the difference between the observed and the predicted posttest score divided by the standard error of prediction was larger than a criterion value. Yet, this method can again be criticized for not adequately dealing with measurement error. The regression model makes the incorrect assumption that the predictor (i.e., the pretest score) has been measured free of error. As a consequence, the unreliability of the pretest scores is not adequately taken into account (cf., Alder, Adam, & Arenberg, 1990; Blomqvist, 1977; Geenen & Van de Vijver, 1993; Jin, 1992; Myrtek & Foerster, 1986). To some extent, this criticism to the regression-based index adopted by McSweeny et al. mirrors that raised to the SEM index described above because both indices consider the measurement error of one testing occasion only.

In the current study, an alternative technique is proposed that aims to deal with the above criticisms to existing change indices. Because it is an elaboration of the RC index of Zegers and Hafkenscheid (1994), it corrects for measurement error by using estimated true change scores. In addition, it corrects for individual variations in practice effects by contrasting a patient's estimated true change score with the estimated true change scores in a matched control group (i.e., a group of controls with similar pretest scores as the patient). To illustrate the results of this technique, an empirical comparison is made among six change indices, falling into three categories: (a) the SD index, which considers neither measurement error nor practice effects; (b) the RC indices of Jacobson and Truax (1991) and Zegers and Hafkenscheid, correcting for measurement error; and (c) the indices of Chelune et al. (1993) and McSweeny et al. (1993), and the currently proposed index, correcting for both measurement error and practice effects (hereafter called reliability-stability or RST indices). The neuropsychological data analyzed for this purpose were collected in a study performed on cardiac surgery patients. An analysis of the current data in terms of group means has been published elsewhere (Bruggemans, Van Dijk, & Huysmans, 1995).

METHOD

The data used for illustrative analysis were obtained as part of a longitudinal study of the cognitive effects of coronary artery bypass graft (CABG) surgery. A battery of neuropsychological tests was administered to 63 CABG patients 2 weeks preoperatively (T1), and 1 week (T2), 1 month (T3), and 6 months (T4) postoperatively. The effects of practice and relieved distress after surgery were controlled for by including the spouses of these patients, exposed to the same stressors associated with the operation and daily life, as a control group. A more detailed description of the methodology of this study can be found in Bruggemans et al. (1995).

Subjects

The 63 CABG patients (55 men, 8 women) and their 63 spouses were enrolled in the study at the University Hospital of Leiden, The Netherlands. Exclusion criteria for patients and spouses were previous cardiac surgery with extracorporeal circulation; a history of neurological disorders (including cerebrovascular diseases), psychiatric illness, or alcohol or drug abuse; a preoperative Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) score lower than 24; and major visual or hearing deficits. Unpaired t tests (two-tailed) revealed no significant differences between patients and spouses for age (M = 58.98 years, SD = 9.16 vs. M = 56.22 years, SD = 11.00), t (124) = 1.53, $p > .05$; years of education (M = 10.46, SD = 3.34 vs. M = 9.63, SD = 2.18), t (124) = 1.64, $p > .05$; and MMSE score at entry (M = 27.50, SD = 2.27 vs. M = 27.33, SD = 2.47), t (121) = 0.40, $p > .05$.

Efficacy of controlling for the effects of changes in distress was evaluated by assessing self-reported mood states in patients and spouses concurrent with the neuropsychological examinations. Changes in mood state scores were not significantly different between the two groups (Bruggemans et al., 1995).

Neuropsychological Test Measures

The following neuropsychological test measures, pertaining to three categories of cognitive functioning, were used in the current analysis: *Learning and memory.* (a) The total score (i.e., Forward plus Backward) of the Digit Span and Visual Memory Span subtests of the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987); and (b) the number of words correctly recalled on the first, the fifth, and a 30-min delayed recall trial of the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964), using a Dutch version with alternate forms (Kingma & Van den Burg, 1985). No interference list was used. *Verbal fluency.* The sum of all admissible words for the letters N, K, and A used in the Controlled Oral Word Association Test (Benton & Hamsher, 1978). *Attention and psychomotor speed.* (a) The average time per line in the Bourdon-Vos Test (Vos, 1988), a cancellation task; (b) the time required for Parts A and B of the Trail Making Test of the Halstead-Reitan Neuropsychological Battery (Reitan, 1958). This test was not included in the test battery for the first nine couples participating in this study; (c) the interference score (i.e., the time required for color-naming of color names printed in nonmatching colored inks minus the time required for color-naming of colored dots) of the Stroop Color-Word Test (Stroop, 1935). The interference score could not be obtained in 2 patients suffering from color blindness; and (d) the number of correctly placed digits in the Symbol Digit Modalities Test (Smith, 1982).

Surgical Parameters

The following surgical parameters were registered: duration of extracorporeal circulation, aortic cross-clamp time, minimum nasopharyngeal temperature, and mean flow and mean arterial pressure during extracorporeal circulation and during aortic cross-clamping (details can be found in Bruggemans et al., 1995).

Statistical Analysis

For the assessment of negative cognitive effects of CABG surgery in individual patients, the raw test scores obtained at times T1 and T2 of the original study were used. Cognitive deterioration in individual patients was defined according to six different change indices:

The Standard Deviation Index

According to this method, cognitive deterioration is expressed by the following formula:

$$C = \frac{X_2 - X_1}{s_1} \tag{1}$$

where X_1 and X_2 are the patient's observed pre- and post-operative test scores, and s_1 is the standard deviation of the distribution of the preoperative scores in the patient sample. As in previous studies on the cognitive effects of cardiac surgery, cognitive deterioration in a patient was said to be significant if $|C| > 1$, the sign of the index being dependent on the denotation of the test score (e.g., Sellman et al., 1993; Shaw et al., 1986, 1987; Treasure et al., 1989).

The Reliable Change Index of Jacobson and Truax (1991)

This index is described by the authors by the following formula:

$$RC_{\mathrm{JT}} = \frac{X_2 - X_1}{SE_{\mathrm{diff}}} \qquad (2)$$

in which X_1 and X_2 are defined as in Equation 1, and SE_{diff} denotes the standard error of the difference between the observed pre- and post-test scores. The SE_{diff} term is computed from $SE_{\mathrm{diff}} = [2(SEM)^2]^{1/2}$, where SEM is the standard error of measurement, computed from $SEM = s_1(1-r_{12})^{1/2}$. The change in a patient's test score from pre- to post-testing is usually taken to be significant if $|RC_{\mathrm{JT}}| > z_{\alpha/2}$ (two-tailed), where α is the Type I error rate. In the present study, it was decided to work with a one-tailed significance test ($\alpha = .05$, $z = 1.645$) because the interest was in cognitive deterioration.

The Reliable Change Index of Zegers and Hafkenscheid (1994)

This RC index is given by the following formula:

$$RC_{\mathrm{ZH}} = \frac{(X_2 - X_1)r_{\mathrm{DD}} + (M_2 - M_1)(1 - r_{\mathrm{DD}})}{\sqrt{r_{\mathrm{DD}}}\sqrt{(1 - r_{\mathrm{DD}})}s_{\mathrm{D}}} \qquad (3)$$

in which X_1 and X_2 are defined as in Equation 1, M_1 and M_2 are the observed pre- and post-operative means of the patient sample, r_{DD} denotes the reliability of the pre- to post-test change scores, and s_{D} denotes the standard deviation of these change scores. The r_{DD} term is computed by the formula:

$$r_{\mathrm{DD}} = \frac{r_{11}s_1^2 + r_{22}s_2^2 - 2r_{12}s_1s_2}{s_1^2 + s_2^2 - 2r_{12}s_1s_2} \qquad (4)$$

where r_{11} and r_{22} are the reliabilities of the pre- and post-test scores, r_{12} is the correlation between the pre- and post-test scores (test-retest reliability), and s_1 and s_2 are the standard deviations of the pre- and post-test scores. The s_{D} term is computed by $s_{\mathrm{D}} = (s_1{}^2 + s_2{}^2 - 2r_{12}s_1s_2)^{1/2}$. Because the numerator of Equation 3 expresses the patient's estimated true change score and the denominator represents he error of this estimate (i.e., error of prediction; see also Hsu, 1995), RC_{ZH} is a normally distributed z-score. As for RC_{JT}, the absolute value of RC_{ZH} ndicated significant cognitive deterioration when t was larger than 1.645.

In the present study, we used the data of the control group to compute the reliability (r_{DD}) and standard deviation (s_{D}) of the change scores because large variances in patients' posttest scores nay reflect individual differences in surgery effects rather than test score properties. Variability in posttest scores has generally been shown to be substantially larger among cardiac surgery patients than among controls (Benedict, 1994). In order to estimate the reliabilities of the pretest (r_{11}) and posttest (r_{22}) scores in the control group, a factor analysis was carried out on the controls' test scores at times T1 to T4. The estimated communalities of the scores at times T1 and T2 were taken as a measure of r_{11} and r_{22}, respectively (see Crocker & Algina, 1986, p. 295).

The Reliability-Stability Index of Chelune et al. (1993)

Chelune et al. used the above RC index of Jacobson and Truax (1991), with a modification to accommodate for practice effects. Their index can be described by the following formula:

$$RST_{\mathrm{Chel}} = \frac{(X_2 - X_1) - (M_2 - M_1)}{SE_{\mathrm{diff}}} \qquad (5)$$

in which X_1, X_2, and SE_{diff} are defined as in Equation 2, and M_1 and M_2 are the observed pre- and post-operative means of the control group. The absolute value of RST_{Chel} had to be larger than 1.645 in order to indicate significant cognitive deterioration.

The Reliability-Stability Index of McSweeny et al. (1993)

This index, based on the regression model, deals with individual differences in practice effects. A regression line is fit to the data of the control group, the posttest scores being predicted on the basis of the pretest scores. A patient's predicted posttest score is then calculated on the basis of his or her pretest score, using the regression parameters (i.e., regression coefficient and intercept) for the regression line of the control group. The regression formula for predicting a patient's posttest score, X'_2, from his or her pretest score is given by:

$$X'_2 = bX_1 + c \qquad (6)$$

where X_1 is defined as in Equation 1, and b and c are the regression coefficient and intercept, respectively, for the regression line to the control group data. The index of McSweeny et al. is given by:

$$RST_{\mathrm{McS}} = \frac{X_2 - X'_2}{SE_{\mathrm{pred}}} \qquad (7)$$

in which X_2 and X'_2 are defined as in Equation 1 and Equation 6, respectively, and SE_{pred} is the standard error of prediction, computed by $SE_{\text{pred}} = s_2(1-r_{12}^2)^{1/2}$. For significant deterioration, the absolute value of RST_{McS} had to be larger than 1.645.

The proposed Reliability-Stability Index

The currently proposed index is an elaboration of the above RC index of Zegers and Hafkenscheid (1994), purporting to correct for individual differences in practice effects. Each patient is matched with a group of control subjects with similar pretest scores (cf. the above regression-based index of McSweeny et al., 1993). Matching with multiple controls rather than with a single subject is carried out to minimize the effects of irrelevant idiosyncratic variations in the change scores of individual control subjects. Matched control group sizes may vary among patients, depending on the opportunities to find appropriate matches. In the present study, on average 10 matched controls were gathered for each patient. The mean change score of the matched control group is considered to represent the change score of a single ideal control subject. Analogous to the change scores of patients, change scores of ideal controls will be subject to measurement error. Consequently, the RC_{ZH} index (Equation 3) is applied to the observed mean pre- and post-test scores of the matched control (mc) subjects, that is:

$$RC_{\text{mc}} = \frac{(M_{\text{mc}_2} - M_{\text{mc}_1})r_{\text{DD}} + (M_2 - M_1)(1 - r_{\text{DD}})}{\sqrt{r_{\text{DD}}}\sqrt{(1 - r_{\text{DD}})}s_{\text{D}}} \qquad (8)$$

where M_{mc_1} and M_{mc_2} are the observed mean pre- and post-test scores of the matched controls, M_1 and M_2 are the observed pre- and post-test means for the entire control group, s_{D} denotes the standard deviation of pre- to post-test change scores of all control subjects, and r_{DD} denotes the reliability of these change scores. The proposed new index (RST_{new}) is then a composite index based on two RC_{ZH} indices:

$$RST_{\text{new}} = RC_{\text{pat}} - RC_{\text{mc}} \qquad (9)$$

where RC_{pat} is the RC_{ZH} index of the patient as described in Equation 3 and RC_{mc} is the RC_{ZH} index of his or her matched controls as given in Equation 8. Because RST_{new} is the difference between two z-scores, like for the other z-score indices, its absolute value had to be larger than 1.645 in order to indicate significant cognitive deterioration.

In addition to comparing the incidence of cognitive deterioration for the six different indices, the question was addressed as to how the indices differ in their relationships with predictor variables, that is patients' age and surgical parameters. In order to answer this question, Pearson correlations were computed between, on the one hand, individual change scores as generated by the indices and, on the other hand, age and surgical parameters. The correlations were subjected to one-tailed t tests on the basis of expectations about the effects of predictor variables on cognitive deterioration.

RESULTS

For each of the neuropsychological test measures, reliabilities and practice effects observed in the control group are reported first as they form the basis of the RC (RC_{JT} and RC_{ZH}) and RST (RST_{Chel}, RST_{McS}, and RST_{new}) indices. Then, the results for the six indices will be presented.

Test Score Reliabilities in the Control Group

For each test measure, the three reliability coefficients (r_{11}, r_{22}, and r_{12}) are presented in Table 1. Only for the immediate memory measure (i.e., Trial I) of the RAVLT, reliabilities were low with coefficients ranging from .32 to .49. Coefficients for the other measures ranged from .61 to .95. Reliability coefficients for the measures of learning and memory (WMS-R and RAVLT) were smaller than those for the measures of verbal fluency (Controlled Oral Word Association) and attention and psychomotor speed (Bourdon-Vos, Trail Making Test, Stroop Interference, and Symbol Digit Modalities). Similar differences in reliability coefficients have been reported in the literature (Spreen & Strauss, 1991).

Practice Effects in the Control Group

In order to evaluate practice effects, paired t tests (one-tailed) were carried out on the pre- to post-test change scores of the entire control group. For change scores (i.e., observed posttest minus observed pretest scores) on each test measure, the mean value, standard deviation, and associated t value are presented in Table 2. Significant improvements in performance were

Table 1. Reliability Coefficients of the Neuropsychological Test Measures in the Control Group.

Test measure	Reliability at T1 (r_{11})	Reliability at T2 (r_{22})	Test-retest reliability (r_{12})
Learning and memory			
WMS-R Digit Span	.79	.82	.71
WMS-R Visual Memory Span	.68	.72	.61
RAVLT Trial I	.45	.49	.32
RAVLT Trial V	.62	.78	.62
RAVLT Delayed Recall	.70	.76	.63
Verbal fluency			
Controlled Oral Word Association	.89	.89	.85
Attention and psychomotor speed			
Bourdon-Vos	.89	.95	.89
Trail Making Test, Part A	.76	.80	.73
Trail Making Test, Part B	.80	.89	.79
Stroop Interference	.82	.84	.77
Symbol Digit Modalities	.92	.93	.90

Note. WMS-R = Wechsler Memory Scale-Revised; RAVLT = Rey Auditory Verbal Learning Test.

found for the verbal fluency measure and for all measures of attention and psychomotor speed, whereas no significant changes were obtained for the learning and memory measures.

The above mean values are the basis of the correction for practice effects as applied in the RST_{Chel} index. The other two RST indices, RST_{McS} and RST_{new}, correct for practice effects using change scores of control subjects with similar pretest scores as the patient. Table 2 presents the change scores for controls matched to the patient with the lowest and the patient with the highest pretest score on each test measure. For the RST_{McS} index, the change scores represent differences between predicted posttest and observed pretest scores. For RST_{new}, the values represent the mean estimated true change scores for the on average 10 matched controls gathered for each patient. As expected, the change scores used in both indices showed considerable variability among pretest score levels, underlining the necessity to apply level-dependent corrections of patients' scores.

Comparison of Change Indices

Critical values and deterioration rates

Table 3 presents critical values for significant test score deterioration (i.e., observed posttest minus observed pretest scores) for each of the six change indices. For the SD, RC_{JT}, RC_{ZH}, and RST_{Chel} indices, the values apply to the entire patient group. For the RST_{McS} and RST_{new} indices, critical values depend on a patient's pretest score. In Table 3, the values are presented only for the patient with the lowest and the patient with the highest pretest score on each test measure.

The effects of score corrections for measurement error can be examined by comparing the results for the SD and RC indices. For the majority of test measures, both RC indices produced larger absolute critical values than the SD index, the RC_{ZH} index tending to be more conservative than the RC_{JT} index. The discrepancies between the SD and RC indices are mainly attributable to the reliability of the test scores: For test measures associated with relatively low reliability coefficients (see Table 1), the RC indices tended to be more conservative than the SD index. This behavior of the RC indices is intu-

Table 2. Change Scores in the Control Group.

Test measure	Entire control group			Matched controls			
				RST_{McS}		RST_{new}	
	M	*SD*	*t*	Lowest pretest score	Highest pretest score	Lowest pretest score	Highest pretest score
Learning and memory							
WMS-R Digit Span[a]	0.06	2.02	0.25	1.57	–1.46	0.40	–0.22
WMS-R Visual Memory Span[a]	0.21	2.28	0.72	1.63	–1.90	0.34	–0.09
RAVLT Trial I[a]	–0.40	1.96	–1.60	2.56	–1.79	0.01	–0.68
RAVLT Trial V[a]	0.17	1.74	0.80	2.31	–1.03	0.52	0.09
RAVLT Delayed Recall[a]	–0.38	2.31	–1.31	2.17	–1.55	0.18	–0.71
Verbal fluency							
Controlled Oral Word Association[a]	4.38	6.09	5.71*	7.84	2.04	4.75	3.72
Attention and psychomotor speed							
Bourdon-Vos[b]	–0.69	0.94	–5.79*	–0.66	–0.76	–0.58	–0.63
Trail Making Test, Part A[b]	–4.91	9.10	–3.96*	2.10	–11.14	–4.46	–5.48
Trail Making Test, Part B[b]	–10.06	20.13	–3.67*	–0.24	–40.79	–8.13	–14.16
Stroop Interference[b]	–8.13	14.30	–4.51*	4.06	–30.65	–5.75	–10.00
Symbol Digit Modalities[a]	3.79	4.56	6.60*	4.89	2.56	4.24	4.05

Note. Change scores represent observed posttest minus observed pretest scores for the entire control group, predicted posttest minus observed pretest scores for matched controls in the Reliability-Stability index of McSweeny et al. (RST_{McS}; 1993), and estimated true posttest minus estimated true pretest scores for the matched controls in the currently proposed Reliability-Stability index (RST_{new}). WMS-R = Wechsler Memory Scale-Revised; RAVLT = Rey Auditory Verbal Learning Test.

[a] A positive change score indicates improvement in performance. [b] A negative change score indicates improvement in performance.

* $p < .001$, one-tailed.

Table 3. Critical Values for Significant Test Score Deterioration.

Test measure	SD index	RC indices (Correcting for measurement error)		RST indices (Correcting for measurement error and practice effects)				
		RC_{JT}	RC_{ZH}	RST_{Chel}	RST_{McS}		RST_{new}	
					Lowest pretest score	Highest pretest score	Lowest pretest score	Highest pretest score
Learning and memory								
WMS-R Digit Span[a]	-2.97	-3.18	–4.78	-3.12	-1.63	–4.65	-3.58	-5.45
WMS-R Visual Memory Span[a]	-2.38	-3.64	-5.69	-3.44	-1.85	-5.38	–4.24	-6.07
RAVLT Trial I[a]	-1.52	-3.46	-7.19	-3.86	0.14	–4.21	-7.15	-10.30
RAVLT Trial V[a]	-2.25	-2.74	–4.93	-2.56	–0.38	-3.71	-2.64	–4.54
RAVLT Delayed Recall[a]	-2.74	-3.60	-3.87	-3.98	-1.43	-5.15	-3.22	-6.36
Verbal fluency								
Controlled Oral Word Association[a]	-11.10	-9.77	-20.48	-5.39	-2.06	-7.86	-2.16	-6.13
Attention and psychomotor speed								
Bourdon-Vos[b]	2.38	1.43	1.39	0.75	0.90	0.80	-0.26	–0.43
Trail Making Test, Part A[b]	10.83	14.99	26.31	10.08	16.11	2.87	4.09	–0.96
Trail Making Test, Part B[b]	37.09	33.01	30.17	22.95	31.43	-9.12	1.19	-20.31
Stroop Interference[b]	17.07	24.87	29.13	16.75	20.56	-14.15	13.84	2.54
Symbol Digit Modalities[a]	–9.30	-7.31	-3.74	-3.51	-2.62	–4.94	14.73	13.88

Note. Critical values represent observed posttest minus observed pretest scores. RC_{JT} = Reliable Change index of Jacobson and Truax (1991); RC_{ZH} = Reliable Change index of Zegers and Hafkenscheid (1994); RST_{Chel} = Reliability-Stability index of Chelune et al. (1993); RST_{McS} = Reliability-Stability index of McSweeny et al. (1993); RST_{new} = currently proposed Reliability-Stability index; WMS-R = Wechsler Memory Scale-Revised; RAVLT = Rey Auditory Verbal Learning Test. [a]A positive change score indicates improvement in performance. [b]A negative change score indicates improvement in performance.

itively appealing in that they correct for the increased uncertainty by enlarging their critical value. In general, for test measures having a low reliability, critical values for the RC indices can become much larger than for the SD index.

Comparison of the results for the RC and RST indices reflects the influence of correcting for practice effects. As could be expected, all three RST indices produced in general much less conservative critical values than the RC indices for test measures showing significant practice effects (i.e., the measures of verbal fluency and attention and psychomotor speed; see Table 2), the RST_{new} index being less conservative than the RST_{Chel} and RST_{McS} indices. Correction for practice effects may lead to seemingly counterintuitive behavior of the RST indices. According to the RST_{McS} and RST_{new} indices, for example, a score improvement up to 9 or 20 points, respectively, on the Trail Making Test, Part B, could mean a net worsening. In general, for test measures showing substantial practice effects, even a score improvement might imply a significant cognitive deterioration.

Table 4 presents the deterioration rates for the six indices, based on the critical values in Table 3. In Table 4, once again, the impact of score corrections for measurement error and practice effects is reflected. For test measures having a relatively low reliability and not showing significant score improvements due to practice (i.e., the learning and memory measures), the SD index generated larger deterioration rates than the RC and RST indices. On the other hand, for test measures with a relatively high reliability but showing significant practice effects (i.e., the measures of verbal fluency and attention and psychomotor speed), the RST indices revealed in general much larger deterioration rates than the SD and RC indices, with the RST_{new} index by far showing the largest rates. The most salient example of this concerns the Symbol Digit Modalities measure. According to the RST_{new} index, all patients were deteriorated on this test measure.

It can be concluded from Table 4 that the six indices showed different results regarding the negative effects of CABG surgery on cognitive functioning. When using the SD index, moderate deterioration rates were found for the measures of learning and memory and those of attention and psychomotor speed (8–29%), whereas deterioration on verbal fluency occurred in only 3% of the patients. The two RC indices produced generally low deterioration rates for the learning and memory measures and for verbal fluency (0–18%). They showed moderate rates for the measures of attention and psychomotor speed, the RC_{ZH} index showing a larger variation in rates (2–48%) than the RC_{JT} index (5–19%). Like the RC indices, the RST indices generally resulted in low deterioration rates for the learning and memory measures (0–21%). However, moderate to high rates were obtained for the attention and psychomotor speed measures, with moderate rates (12–48%) for the RST_{Chel} and RST_{McS} indices, and high rates (25–100%) for RST_{new}. Verbal fluency was affected in a moderate number of patients (16–24%) for the RST indices.

Correlations with age and surgical parameters

Table 5 presents the correlations between individual change scores as generated by the six indices and relevant predictor variables. Because intercorrelations among the SD index, the two RC indices, and the RST_{Chel} index by definition are equal to one (since they all imply a linear transformation of the observed change score), their correlations with predictor variables are identical.

The data show that the relationships with age and surgical parameters were essentially similar for all indices. Only a limited number of significant correlations were found. Low mean flow during extracorporeal circulation as well as during aortic cross-clamping and low minimum nasopharyngeal temperature appeared to be the most pertinent predictors of cognitive deterioration. The neuropsychological test measures most affected by these surgical parameters were the delayed recall measure of the RAVLT and the Symbol Digit Modalities measure. These two test measures have in common that they make a demand on memory retrieval.

Table 4. Deterioration Rates (as percentages).

Test measure	SD index	RC indices (Correcting for measurement error)		RST indices (Correcting for measurement error and practice effects)		
		RC_{JT}	RC_{ZH}	RST_{Chel}	RST_{McS}	RST_{new}
Learning and memory						
WMS-R Digit Span	10	0	0	0	3	0
WMS-R Visual Memory Span	16	7	2	7	5	5
RAVLT Trial I	18	2	0	2	0	0
RAVLT Trial V	14	14	3	14	11	3
RAVLT Delayed Recall	29	18	18	18	21	11
Verbal fluency						
Controlled Oral Word Association	3	6	0	18	16	24
Attention and psychomotor speed						
Bourdon-Vos	11	19	21	35	32	73
Trail Making Test, Part A	15	11	2	15	15	33
Trail Making Test, Part B	15	19	19	28	33	67
Stroop Interference	8	5	5	12	28	25
Symbol Digit Modalities	10	13	48	48	46	100

Note. RC_{JT} = Reliable Change index of Jacobson and Truax (1991); RC_{ZH} = Reliable Change index of Zegers and Hafkenscheid (1994); RST_{Chel} = Reliability-Stability index of Chelune et al. (1993); RST_{McS} = Reliability-Stability index of McSweeny et al. (1993); RST_{new} = currently proposed Reliability-Stability index; WMS-R = Wechsler Memory Scale-Revised; RAVLT = Rey Auditory Verbal Learning Test.

Table 5. Correlations between Change Indices and Predictor Variables.

	Age			ECC			ACC			Temperature		
Test measure	SD^{a}	RST_{McS}	RST_{new}	SD^{a}	RST_{McS}	RST_{new}	SD^{a}	RST_{McS}	RST_{new}	SD^{a}	RST_{McS}	RST_{new}
Learning and memory												
WMS-R Digit Span	.04	.03	.05	–.07	.03	.00	–.04	.03	.00	.13	.04	.06
WMS-R Visual Memory Span	–.09	–.15	–.17	.03	.03	.03	–.02	–.05	–.04	.01	.04	.06
RAVLT Trial I	.06	.08	.09	–.02	–.03	–.05	–.06	–.09	–.10	.22*	.18	.17
RAVLT Trial V	–.18	–.18	–.18	–.03	–.06	–.03	.06	.00	.03	.07	.14	.12
RAVLT Delayed Recall	–.05	–.06	–.14	–.15	–.18	–.22*	–.21*	–.25*	–.29*	.27*	.34**	.42***
Verbal fluency												
Controlled Oral Word Association	–.02	–.03	–.03	–.23*	–.21*	–.23*	–.25*	–.25*	–.26*	.08	.06	.07
Attention and psychomotor speed												
Bourdon-Vos	.18	.19	.17	–.03	–.03	–.01	–.03	–.03	–.01	–.03	–.03	–.03
Trail Making Test, Part A	.15	.21	.17	–.02	–.06	–.03	.00	–.01	.00	–.27*	–.22	–.22
Trail Making Test, Part B	.20	.24*	.24*	.11	.06	.08	.04	.01	.03	–.27*	–.22	–.24*
Stroop Interference	–.12	.09	–.02	–.04	–.04	–.06	–.07	–.04	–.08	.01	–.01	–.01
Symbol Digit Modalities	–.07	–.11	–.07	–.15	–.13	–.13	–.15	–.15	–.14	.29*	.29*	.26*

	Flow 1			Flow 2			Pressure 1			Pressure 2		
Test measure	SD^{a}	RST_{McS}	RST_{new}	SD^{a}	RST_{McS}	RST_{new}	SD^{a}	RST_{McS}	RST_{new}	SD^{a}	RST_{McS}	RST_{new}
Learning and memory												
WMS-R Digit Span	.05	.07	.07	.12	.10	.10	.10	.11	.12	–.07	–.05	–.04
WMS-R Visual Memory Span	.07	.17	.13	.02	.07	.04	–.09	–.13	–.11	–.24	–.20	–.19
RAVLT Trial I	.15	.26*	.23*	.20	.24*	.24*	.03	.10	.11	.12	.11	.15
RAVLT Trial V	.17	.23*	.20	.32**	.36**	.35**	.11	.11	.10	.04	.02	.00
RAVLT Delayed Recall	.21*	.26*	.27*	.22*	.25*	.28*	–.01	.01	.01	–.15	–.14	–.13
Verbal fluency												
Controlled Oral Word Association	.10	.13	.12	.02	.04	.04	.00	.01	.01	.06	.07	.07
Attention and psychomotor speed												
Bourdon-Vos	–.09	–.10	–.08	–.10	–.11	–.09	.25	.24	.23	.12	.11	.10
Trail Making Test, Part A	–.18	–.22	–.23*	–.18	–.18	–.20	.11	.09	.11	.03	.01	.00
Trail Making Test, Part B	–.10	–.11	–.13	–.19	–.18	–.21	.07	.07	.04	.00	–.03	–.05
Stroop Interference	–.13	–.23*	–.19	–.18	–.24*	–.21	.00	–.08	–.04	–.13	–.16	–.15
Symbol Digit Modalities	.36**	.38**	.35**	.37**	.37**	.34**	–.05	–.05	–.09	–.01	–.02	–.05

Note. ECC = duration of extracorporeal circulation; ACC = aortic cross-clamp time; Temperature = minimum nasopharyngeal temperature; Flow 1 and Flow 2 = mean flow during extracorporeal circulation and during aortic cross-clamping, respectively; Pressure 1 and Pressure 2 = mean arterial pressure during extracorporeal circulation and during aortic cross-clamping, respectively; RST_{McS} = Reliability-Stability index of McSweeny et al. (1993); RST_{new} = currently proposed Reliability-Stability index; WMS-R = Wechsler Memory Scale-Revised; RAVLT = Rey Auditory Verbal Learning Test.

[a] Correlations for the SD, RC_{JT}, RC_{ZH}, and RST_{Chel} indices are identical.

* $p < .05$, ** $p < .01$, *** $p < .001$, one-tailed.

DISCUSSION

The interpretation of individual changes in neuropsychological test scores is hampered by problems of measurement error and practice effects. In this study, six different indices of individual change were applied to assess cognitive deterioration in patients following CABG surgery: the SD index, which considers neither measurement error nor practice effects; the RC indices of Jacobson and Truax (1991) and Zegers and Hafkenscheid (1994), correcting for measurement error; and the RST indices of Chelune et al. (1993) and McSweeny et al. (1993), and the currently proposed RST index, correcting for both measurement error and practice effects. Results for these indices were compared on various neuropsychological test measures. The purpose was to examine the impact of score corrections for measurement error and practice effects in the assessment of individual change and to evaluate the consequences for the incidence of cognitive deterioration in CABG patients.

The indices showed marked differences in deterioration rates on the various test measures due to measurement error and practice effects, as may be expected. For test measures with a relatively low reliability and showing no significant practice effects (the learning and memory measures), the SD index produced higher deterioration rates than did the RC and RST indices. In contrast, for test measures with a relatively high reliability but showing significant score improvements due to practice (the verbal fluency and attention and psychomotor speed measures), the RST indices produced higher deterioration rates than did both the SD index and the RC indices. These findings are consistent with the mathematical differences among the indices that imply that the discordance of the SD index versus the RC and RST indices will increase when tests become less reliable, and that the discordance of both the SD index and the RC indices versus the RST indices will increase when tests show larger practice effects.

Although the RC_{ZH} index more adequately corrects for measurement error than does the RC_{JT} index, systematic differences in deterioration rates between these two indices were not found in the present study. It should, however, be emphasized that with other data sets the indices may yield a larger discrepancy. This can be explained by the fact that the mathematical difference between the indices is not merely a function of test reliability. In comparison with the RC_{JT} index, the RC_{ZH} index involves additional sample statistics, such as the mean of the observed pretest scores, and the mean and standard deviation of the observed posttest scores. Therefore, the discordance between the two RC indices will depend on the characteristics of the study sample.

Analogous reasoning holds for the behavior of the three RST indices. In the present study, the statistically more adequate RST_{new} index revealed considerably higher deterioration rates than the RST_{Chel} and RST_{McS} indices on the measures of attention and psychomotor speed, the latter two indices showing almost equal rates. As for the RC indices, however, application of the RST indices in other study samples may result in other incongruities. Because the three RST indices employ partially different sample statistics, their discordance may vary with the characteristics of the study sample.

The current results clearly show that conclusions about the incidence of cognitive change after medical intervention strongly depend on the index utilized. Because of their psychometric shortcomings, application of the SD index, the two RC indices, and the RST_{Chel} and RST_{McS} indices may lead to fallacious conclusions. Thus, with the SD index, presumably too many CABG patients in the present sample were considered as significantly deteriorated on learning and memory (i.e., false positives), whereas too few patients were considered as significantly deteriorated on verbal fluency and attention and psychomotor speed (i.e., false negatives). The major problem of the two RC indices concerned false negatives on verbal fluency and attention and psychomotor speed. Finally, false negatives on attention and psychomotor speed were also observed for the RST_{Chel} and RST_{McS} indices.

Utilization of the RST_{new} index in the analysis of the current data indicated differential effects of CABG surgery on cognitive functioning, with deterioration frequently occurring on attention and psychomotor speed (25–100%), to a smaller degree on verbal fluency (24%), and only occasionally on learning and memory (0–11%). This pattern of early cognitive dysfunctioning (i.e., deterioration detectable 1 week following CABG surgery) is consistent with that quantified in terms of group means for this patient sample (Bruggemans et al., 1995). At 1 week postoperatively, patients' mean scores on attention and psychomotor speed as well as on verbal fluency were significantly deteriorated compared with the mean scores of spouses, whereas mean change scores for memory and learning showed no differences between both groups. This pattern is also consistent with the early pattern revealed by other controlled studies reporting group means (e.g., Hammeke & Hastings, 1988; Townes et al., 1989). On the other hand, there is little agreement with studies that analyzed group means but did not include a control group (e.g., Fish et al., 1987; Zeitlhofer et al., 1993). These studies in general did not show important differences in deterioration among different cognitive functions and sometimes even suggested improvement in functioning, probably due to practice effects.

As mentioned before, the SD index has been commonly used to report on the incidence of cognitive deterioration following CABG surgery. In two studies applying this index, the deterioration rates on the various test measures were inconsistent and also differed considerably from those of the $\mathrm{RST}_{\mathrm{new}}$ index in the current study. In one study, learning and memory were slightly more often affected than attention and psychomotor speed (Shaw et al., 1986, 1987), whereas this pattern was not replicated in the other study (Harrison et al., 1989). The other studies applying the SD index (see introduction), however, can in general be criticized for underrating the influence of the composition of the test battery. In these studies, cognitive deterioration in a patient was generally said to occur in case of a drop in test performance of at least one *SD* on two or more of the administered tests (see also Mahanna et al., 1996). Even though using this additional criterion may well reduce the influence of random score fluctuations on the diagnosis of cognitive deterioration, the differential pattern of deterioration is neglected in this definition. Hence, the likelihood of assessing cognitive deterioration will depend on the homogeneity of the test battery.

In the present study, relationships between relevant predictor variables (i.e., patients' age and surgical parameters) and individual change scores showed much correspondence for the six indices, indicating high intercorrelations among the indices. As stated before, the SD index, the two RC indices, and the $\mathrm{RST}_{\mathrm{Chel}}$ index by definition show a perfect intercorrelation. The corrections for practice effects as applied in the $\mathrm{RST}_{\mathrm{McS}}$ and $\mathrm{RST}_{\mathrm{new}}$ indices, however, may entail a nonlinear transformation of the observed change score because these corrections depend on a patient's pretest score. Nevertheless, Pearson correlations between observed change scores and the $\mathrm{RST}_{\mathrm{McS}}$ and $\mathrm{RST}_{\mathrm{new}}$ indices were high, ranging from .80 to 1.00 and from .91 to .99, respectively, and thus suggest almost linear corrections for practice effects. Empirical studies have frequently shown that estimated true change scores (RST_{new}) are linearly related to initial scores (e.g., Alder et al., 1990; Fahrenberg, Foerster, & Wilmers, 1995; Myrtek & Foerster, 1986). The same holds probably for residualized change scores (RST_{McS}). It may, therefore, be assumed that individual score corrections for practice effects tend to amount to an almost linear transformation of observed change scores and that, in general, correlations for the $\mathrm{RST}_{\mathrm{McS}}$ and $\mathrm{RST}_{\mathrm{new}}$ indices with external variables will only slightly deviate from those for the SD index, the two RC indices, and the $\mathrm{RST}_{\mathrm{Chel}}$ index.

The above findings have implications for choosing the most appropriate index in the assessment of individual change. If one is interested in the incidence of cognitive change rather than in its prediction, the use of the $\mathrm{RST}_{\mathrm{new}}$ index is to be preferred over the use of the other indices. Application of this new index will reduce the likelihood of incorrect decisions about the significance of change scores because of the corrections for both measurement error and practice effects. If one is primarily interested in predicting cognitive change on the basis of external variables, then the different indices will frequently yield similar results and, therefore, there is no need to select among the indices.

Three practical problems may be encountered in applying the RST_{new} index. First, the different reliability coefficients (i.e., r_{11}, r_{22}, and r_{12}) may not always be available. When using standardized neuropsychological tests, useful reliability estimates may be found in the test manual or reports in the literature. Here the question may arise as to what types of estimates are appropriate for inclusion in the index. For both r_{11} and r_{22}, one could use one of the various estimates of internal consistency, such as the alpha coefficient, the split-half coefficient, or the communality derived from factor-analysis as employed in the present study. The only meaningful estimate for r_{12} is the test-retest correlation coefficient. It is important to note that the intraclass correlation coefficient as a measure of agreement (cf., McGraw & Wong, 1996; Shrout & Fleiss, 1979), which in other circumstances may be considered a very useful reliability estimate (cf., Brown, Rourke, & Cicchetti, 1989), cannot be employed here. A high value of this coefficient indicates both a reliable and a stable (in absolute test scores) measurement. A low value reflects poor reliability, instability, or both. The RST_{new} index, however, assumes reliability estimates that only reflect the influence of measurement error.

Second, in some cases estimates of internal consistency such as the split-half coefficient will seriously underestimate the true reliability of the instrument. In such cases, the value of the test-retest correlation can be almost as high as the internal consistency estimates and, as a consequence, the reliability of the change score can become close to or even equal to zero. The value of the RST_{new} index will then be mainly determined by the difference in group means at both test occasions, thereby underrating the relevance of the change in the individual. In such a case, however, the use of the other indices may not lead to more appropriate results.

The third problem in applying the RST_{new} index may involve the lack of information about practice effects. As for test reliability, test manuals or reports in the literature may provide useful data on test-retest stability. It is highly unlikely, however, that these will contain data on groups with a pretest score similar to the patient's. Applying corrections for mean practice effects such as those applied in the RST_{Chel} index will then be the only viable option.

In summary, six different indices for assessing individual change, differing in the extent that corrections were made for measurement error and practice effects, have been examined. The present findings suggest that when one is interested in incidence rates, the RST_{new} index will be very useful. This index enables a statistically adequate test of individual change both when measurement error and practice effects jeopardize the validity of classical approaches. The index was applied here to assess the negative cognitive effects of cardiac surgery but the domain of application is broader and includes all pretest-posttest designs in which individual change is examined.

REFERENCES

Alder, A. G., Adam, J., & Arenberg, D. (1990). Individual-differences assessment of the relationship between change in and initial level of adult cognitive functioning. *Psychology and Aging, 5,* 560-568.

Benedict, R. H. B. (1994). Cognitive function after open-heart surgery: Are postoperative neuropsychological deficits caused by cardiopulmonary bypass? *Neuropsychology Review, 4,* 223-255.

Benton, A. L., & Hamsher, K. (1978). *Multilingual Aphasia Examination Manual-Revised.* Iowa City, IA: University of Iowa.

Blomqvist, N. (1977). On the relation between change and initial value. *Journal of the American Statistical Association, 72,* 746-749.

Brown, R. T., & Madan-Swain, A. (1993). Cognitive, neuropsychological, and academic sequelae in children with leukemia. *Journal of Learning Disabilities, 26,* 74-90.

Brown, S. J., Rourke, B. P., & Cicchetti, D. V. (1989). Reliability of tests and measures used in the neuropsychological assessment of children. *The Clinical Neuropsychologist, 3,* 353-368.

Bruggemans, E. F., Van Dijk, J. G., & Huysmans, H. A. (1995). Residual cognitive dysfunctioning at 6 months following coronary artery bypass graft surgery. *European Journal of Cardio-Thoracic Surgery, 9,* 636-643.

Chelune, G. J. (1992). Using neuropsychological data to forecast postsurgical cognitive outcome. In H.

O. Lüders (Ed.), *Epilepsy surgery* (pp. 477-485). New York: Raven Press.

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7,* 41-52.

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC Index. *Behavior Therapy, 17,* 305-308.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Orlando, FL: Holt, Rinehart, and Winston.

Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test-taking and the stability of adjustment scales. Can we assess patient deterioration? *Evaluation Quarterly, 2,* 275-291.

Everall, I. P. (1995). Neuropsychiatric aspects of HIV infection. *Journal of Neurology, Neurosurgery, and Psychiatry, 58,* 399-402.

Fahrenberg, J., Foerster, F., & Wilmers, F. (1995). Is elevated blood pressure level associated with higher cardiovascular responsiveness in laboratory tasks and with response specificity? *Psychophysiology, 32,* 81-91.

Feinstein, A., Brown, R., & Ron, M. (1994). Effects of practice of serial tests of attention in healthy subjects. *Journal of Clinical and Experimental Neuropsychology, 16,* 436-447.

Fish, K. J., Helms, K. N., Sarnquist, F. H., Van Steennis, C., Linet, O. I., Hilberman, M., Mitchell, R. S., Jamieson, S. W., Miller, D. C., & Tinklenberg, J. S. (1987). A prospective, randomized study of the effects of prostacyclin on neuropsychologic dysfunction after coronary artery operation. *Journal of Thoracic and Cardiovascular Surgery, 93,* 609-615.

Fletcher, J. M., & Copeland, D. R. (1988). Neurobehavioral effects of central nervous system prophylactic treatment of cancer in children. *Journal of Clinical and Experimental Neuropsychology, 10,* 495-538.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189-198.

Geenen, R., & Van de Vijver, F. J. R. (1993). A simple test of the Law of Initial Values. *Psychophysiology, 30,* 525-530.

Hageman, W. J. J. M., & Arrindell, W. A. (1993). A further refinement of the Reliable Change (RC) Index by improving the pre-post difference score: Introducing RC_{ID}. *Behaviour Research and Therapy, 31,* 693-700.

Hammeke, T. A., & Hastings, J. E. (1988). Neuropsychologic alterations after cardiac operation. *Journal of Thoracic and Cardiovascular Surgery, 96,* 326-331.

Harrison, M. J. G., Schneidau, A., Ho, R., Smith, P. L. C., Newman, S., & Treasure, T. (1989). Cerebrovascular disease and functional outcome after coronary artery bypass surgery. *Stroke, 20,* 235-237.

Hsu, L. M. (1989). Reliable changes in psychotherapy: Taking into account regression toward the mean. *Behavioral Assessment, 11,* 459-467.

Hsu, L. M. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 141-144.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336-352.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19.

Jin, P. (1992). Toward a reconceptualization of the Law of Initial Value. *Psychological Bulletin, 111,* 176-184.

Karayanidis, F. (1989). Parkinson's disease: A conceptualization of neuropsychological deficits within an information-processing framework. *Biological Psychology, 29,* 149-179.

Kingma, A., & Van den Burg, W. (1985). *Vijftien Woorden Test* [Fifteen Words Test]. Groningen, The Netherlands: University Hospital.

Knight, R. G. (1983). On interpreting the several standard errors of the WAIS-R: Some further tables. *Journal of Consulting and Clinical Psychology, 51,* 671-673.

Loring, D. W., Meador, K. J., Lee, G. P., King, D. W., Nichols, M. E., Park, Y. D., Murro, A. M., Gallagher, B. B., & Smith, J. R. (1995). Wada memory asymmetries predict verbal memory decline after anterior temporal lobectomy. *Neurology, 45,* 1329-1333.

Macciocchi, S. N. (1990). "Practice makes perfect": Retest effects in college athletes. *Journal of Clinical Psychology, 46,* 628-631.

Mahanna, E. P., Blumenthal, J. A., White, W. D., Croughwell, N. D., Clancy, C. P., Smith, L. R., & Newman, M. F. (1996). Defining neuropsychological dysfunction after coronary artery bypass grafting. *Annals of Thoracic Surgery, 61,* 1342-1347.

Maj, M. (1990). Psychiatric aspects of HIV-1 infection and AIDS. *Psychological Medicine, 20,* 547-563.

Matarazzo, J. D., Carmody, T. P., & Jacobs, L. D. (1980). Test-retest reliability and stability of the WAIS: A literature review with implications for clinical practice. *Journal of Clinical Neuropsychology, 2,* 89-105.

Matarazzo, J. D., & Herman, D. O. (1984). Base rate data for the WAIS-R: Test-retest stability and VIQ-PIQ differences. *Journal of Clinical Neuropsychology, 6,* 351-366.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

McSweeny, A. J., Naugle, R. I., Chelune, G. J., & Lüders, H. (1993). "*T* Scores for change": An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist, 7,* 300-312.

Myrtek, M., & Foerster, F. (1986). The Law of Initial Value: A rare exception. *Biological Psychology, 22,* 227-237.

Naugle, R. I. (1992). Neuropsychological effects of surgery of epilepsy. In H. O. Lüders (Ed.), *Epilepsy surgery* (pp. 637-645). New York: Raven Press.

Newman, S. (1993). Neuropsychological and psychological changes. In P. L. Smith & K. M. Taylor (Eds.), *Cardiac surgery and the brain* (pp. 34-54). London: Hodder & Stoughton.

Phillips, N. A., & McGlone, J. (1995). Grouped data do not tell the whole story: Individual analysis of cognitive change after temporal lobectomy. *Journal of Clinical and Experimental Neuropsychology, 17,* 713-724.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8,* 271-276.

Rey, A. (1964). *L'examen clinique en psychologie* [Clinical psychological assessment]. Paris: Presses Universitaires de France.

Roman, D. D., & Sperduto, P. W. (1995). Neuropsychological effects of cranial radiation: Current knowledge and future directions. *International Journal of Radiation Oncology, Biology, Physics, 31,* 983-998.

Saint-Cyr, J. A., Taylor, A. E., & Lang, A. E. (1993). Neuropsychological and psychiatric side effects in the treatment of Parkinson's disease. *Neurology, 43* (Suppl. 6), S47-S52.

Sellman, M., Holm, L., Ivert, T., & Semb, B. K. H. (1993). A randomized study of neuropsychological function in patients undergoing coronary bypass surgery. *Thoracic Cardiovascular Surgeon, 41,* 349-354.

Shatz, M. W. (1981). WAIS practice effects in clinical neuropsychology. *Journal of Clinical Neuropsychology, 3,* 171-179.

Shaw, P. J., Bates, D., Cartlidge, N. E. F., French, J. M., Heaviside, D., Julian, D. G., & Shaw, D. A. (1986). Early intellectual dysfunction following coronary bypass surgery. *Quarterly Journal of Medicine, 225,* 59-68.

Shaw, P. J., Bates, D., Cartlidge, N. E. F., French, J. M., Heaviside, D., Julian, D. G., & Shaw, D. A. (1987). Long-term intellectual dysfunction following coronary artery bypass graft surgery: A six month follow-up study. *Quarterly Journal of Medicine, 239,* 259-268.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Smith, A. (1982). *Symbol Digit Modalities Test. Manual-Revised 1982.* Los Angeles: Western Psychological Services.

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60,* 402-408.

Speer, D. C. (1993). Correction to Speer. *Journal of Consulting and Clinical Psychology, 61,* 27.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary.* New York: Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-662.

Townes, B. D., Bashein, G., Hornbein, T. F., Coppel, D. B., Goldstein, D. E., Davis, K. B., Nessly, M. L., Bledsoe, S. W., Veith, R. C., Ivey, T. D., & Cohen, M. A. (1989). Neurobehavioral outcomes in cardiac operations: A prospective controlled study. *Journal of Thoracic and Cardiovascular Surgery, 98,* 774-782.

Treasure, T., Smith, P. L. C., Newman, S., Schneidau, A., Joseph, Ph., Ell, P., & Harrison, M. J. G. (1989). Impairment of cerebral function following cardiac and other major surgery. *European Journal of Cardio-Thoracic Surgery, 3,* 216-221.

Vos, P. G. (1988). *Bourdon-Vos Test. Manual.* Lisse, The Netherlands: Swets & Zeitlinger.

Waldstein, S. R., Manuck, S. B., Ryan, C. M., & Muldoon, M. F. (1991). Neuropsychological correlates of hypertension: Review and methodologic considerations. *Psychological Bulletin, 110,* 451-468.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised. Manual.* New York: Psychological Corporation.

Wing, H. (1980). Practice effects with traditional mental test items. *Applied Psychological Measurement, 4,* 141-155.

Zegers, F. E., & Hafkenscheid, A. (1994). *The Ultimate Reliable Change Index: An alternative to the Hageman & Arrindell approach.* (Heymans Bulletins HB-94-1154-EX). Groningen, The Netherlands: University of Groningen.

Zeitlhofer, J., Asenbaum, S., Spiss, C., Wimmer, A., Mayr, N., Wolner, E., & Deecke, L. (1993). Central nervous system function after cardiopulmonary bypass. *European Heart Journal, 14,* 885-890.