EUROPEAN JOURNAL OF
CARDIO-THORACIC
SURGERY

Assessment of statistical change criteria used to define significant change in neuropsychological test performance following cardiac surgery

Paul D. Raymond, Anton D. Hinton-Bayre, Michael Radel, Michael J. Ray and Neville A. Marsh

Eur J Cardiothorac Surg 2006;29:82-88
DOI: 10.1016/j.ejcts.2005.10.016

European Journal of Cardio-thoracic Surgery 29 (2006) 82—88

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Assessment of statistical change criteria used to define significant change in neuropsychological test performance following cardiac surgery

Paul D. Raymond [a,e,*], Anton D. Hinton-Bayre [b], Michael Radel [c], Michael J. Ray [d], Neville A. Marsh [a]

[a] *Research Concentration in Biological and Medical Sciences, School of Life Sciences, Queensland University of Technology, Brisbane, Australia*
[b] *Discipline of General Practice, School of Medicine, University of Queensland, Brisbane, Australia*
[c] *Department of Psychology and Neuropsychology, The Prince Charles Hospital, Brisbane, Australia*
[d] *Department of Haematology, The Prince Charles Hospital, Brisbane, Australia*
[e] *Department of Perfusion Services, Main Operating Theatres, The Prince Charles Hospital, Rode Road, Brisbane Q4032, Australia*

Received 29 March 2005; received in revised form 31 August 2005; accepted 7 October 2005; Available online 6 December 2005

Abstract

Objective: This paper compares four techniques used to assess change in neuropsychological test scores before and after coronary artery bypass graft surgery (CABG), and includes a rationale for the classification of a patient as overall impaired. **Methods:** A total of 55 patients were tested before and after surgery on the MicroCog neuropsychological test battery. A matched control group underwent the same testing regime to generate test—retest reliabilities and practice effects. Two techniques designed to assess statistical change were used: the Reliable Change Index (RCI), modified for practice, and the Standardised Regression-based (SRB) technique. These were compared against two fixed cutoff techniques (standard deviation and 20% change methods). **Results:** The incidence of decline across test scores varied markedly depending on which technique was used to describe change. The SRB method identified more patients as declined on most measures. In comparison, the two fixed cutoff techniques displayed relatively reduced sensitivity in the detection of change. **Conclusions:** Overall change in an individual can be described provided the investigators choose a rational cutoff based on likely spread of scores due to chance. A cutoff value of $\geq$20% of test scores used provided acceptable probability based on the number of tests commonly encountered. Investigators must also choose a test battery that minimises shared variance among test scores.

Keywords: Cardiopulmonary bypass; Neurocognitive deficits; Brain; Cerebral complications

1. Introduction

The incidence of neuropsychological impairment following cardiac surgery is difficult to assess, as there is no standardisation on appropriate statistical techniques used to measure change. Individual change from baseline score in neuropsychological tests may be favoured over group analysis, yet the definition of what constitutes a significant change score remains elusive. By far, the most commonly used technique in the literature has been the standard deviation (SD) method [1—5], which describes significant change as a decline from preoperative test score by at least 1SD of the mean baseline score of the population sample. Another technique used defines significant score change as a decline of at least 20% from baseline [6]. While these methods may be simple to implement, they do not adequately evaluate the psychometric and statistical issues surrounding change scores. Because most tests have less than perfect reliability, scores may vary by 1SD or more entirely due to test error. Further, even with perfect reliability a test may be susceptible to practice effects. In order to evaluate 'true' change, a method should determine whether the observed change exceeds that expected from measurement error and improvement over time due to practice or regression to the mean [7]. Test—retest data from matched control subjects are used to generate a range in which variation in test performance is considered 'normal'. Any deviation from this range is considered a significant change. Two methods of analysis that deal with statistically significant change in individuals are presented in this paper: the Reliable Change Index (RCI), modified for practice [7], and the Standardised Regression-based (SRB) technique [8].

* Corresponding author. Tel.: +61 7 3350 8705; fax: +61 7 3350 8659.
E-mail address: p.raymond@optusnet.com.au (P.D. Raymond).

1010-7940/$ — see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejcts.2005.10.016

The RCI with adjustment for practice (RCIp) has previously been employed by Kneebone et al. [9] for analysis of cardiac surgical patients. This method provides criteria for meaningful change based on the calculated measurement error for each score. A patient's predicted retest score equals their Time 1 score plus mean practice effect detected in a matched control group. If the difference between the actual and expected retest score exceeds the likely variation based on matched controls, a significant change is considered to have occurred. A disadvantage of this method is the use of mean practice effect rather than individualised practice, as it may not be reasonable to expect all people to show similar effects. Similarly, this method does not consider regression toward the mean—the tendency for people with outlying performance at initial test to drift toward the mean at follow-up. The SRB method generates a regression equation to predict a patient's score change on the basis of Time 1 performance, plus any demographic variables that contribute significantly to the prediction model in matched controls. The expected change is therefore dependent on initial performance rather than mean practice effect.

The aim of this study was to compare the SRB model with the RCIp, and to use the SD and 20% methods as an illustration of the importance of statistical change criteria. In addition, a discussion is presented on the methodology of selecting criteria for the classification of an individual as overall impaired when using a battery of test scores.

2. Methods

2.1. Subjects

Seventy-four patients undergoing first-time elective coronary artery bypass grafting (CABG) surgery were enrolled. Exclusion criteria included age less than 50 years, prior cardiac surgery, or a history of stroke or any progressive neurological disease. Seven patients were excluded because the surgery was performed without cardiopulmonary bypass; a further 12 patients who did not complete both the preoperative and postoperative assessments were also excluded from analysis. The remaining sample consisted of 55 patients, with demographics shown in Table 1. The study was approved by the ethics committee of both The Prince Charles Hospital and Queensland University of Technology, with informed consent given by all subjects.

Table 1
Demographic variables

Variable	Controls	Subjects
n	40	55
Age (years)		
Mean	64.7	64.4
SD	8.3	9.0
Range	50–83	50–78
Education (years)		
Mean	11.4	10.4
SD	3.1	3.6
Range	6–18	5–22
Sex		
Males	19	48*
Females	21	7*

* $p < 0.001$.

2.2. Procedure

Patients underwent a complete neuropsychological testing package in the week prior to surgery (mean 9.7 ± 4.5 days prior; range 2–26 days). The follow-up assessment was performed immediately before discharge from hospital (mean 5.3 ± 1.2 days postoperation; range 3–10 days). The mean time between tests was 15 ± 4.7 days (range 8–31 days). Neuropsychological assessment was performed using the MicroCog: Assessment of Cognitive Functioning [10]. This is a computer administered and scored test, assessing a wide range of cognitive functions. Results are presented for five cognitive domain scores (attention/mental control, memory, reasoning/calculation, spatial processing, and reaction time) and four higher-order scores (information processing speed, information processing accuracy, general cognitive functioning, and general cognitive proficiency) generated from speed and accuracy scores of the subtests comprising the domain scores. A more detailed description of the MicroCog has been supplied elsewhere [11]. All patients completed the standard form MicroCog at baseline, taking approximately 1 h to complete, with the short form being administered at follow-up (approximately 30 min). A control group of adult participants was obtained from local community groups [11]. Control subjects were matched for exclusion criteria, and completed the same testing protocol over a 2-week retest interval. Demographics for the control group are also shown in Table 1. Control subjects differ from surgical patients only in gender ratio. Gender differences in performance were not seen in the control group [11].

2.3. Methods of defining change in neuropsychological scores

Four different analysis techniques were used to define significant change in neuropsychological scores.

2.3.1. Standardised Regression-based technique

Regression equations were generated for each neuropsychological score using retest data from the control group. Using standard multiple linear regression analysis, age, sex, education level, and score at Time 1 were evaluated as potential predictors of score at Time 2. For all measures in this series, only performance at Time 1 was shown to be a significant predictor. Results for the regression analysis are presented as regression coefficient and intercept, and standard error of the estimate (SE_{est}), calculated as $SE_{est} = SD_2\sqrt{[1-(r_{xx})^2]}$, where SD_2 is the SD of post-test scores and r_{xx} is the test–retest reliability. A patient's predicted post-test score (X_2') can be calculated on the basis of their initial score using the formula $X_2' = bX_1 + c$, where X_1 is the patient's Time 1 score, and b and c are the regression coefficient and intercept, respectively. This data can be used to assess significant change from pre- to post-test using the SRB formula:

$$\text{SRB } Z \text{ score} = \frac{X_2' - X_2}{SE_{est}}$$

where X_2 is the patient's observed post-test score. When the SRB Z score is greater than ± 1.645 (90% confidence), a significant change from expected performance is considered to have occurred. Examples of more detailed regression formulas incorporating the effects of demographic variables in the prediction of Time 2 performance are presented by Temkin et al. [12].

Ninety percent confidence intervals for SRB may be generated using the following approach.

$$\text{SRB confidence interval} = \mathrm{SE}_{\mathrm{est}}(\pm 1.645) + \text{predicted post-test score}$$

2.3.2. Reliable Change Index modified for practice

RCI Z scores can be generated from test—retest data from the normative control group, using the formula:

$$\text{RCI } Z\text{ score} = \frac{(X_2 - X_1) - \text{Practice effect}}{\mathrm{SE}_{\mathrm{diff}}}$$

where $\mathrm{SE}_{\mathrm{diff}} = \sqrt{[2(\mathrm{SE}_{\mathrm{m}})^2]}$, and $\mathrm{SE}_{\mathrm{m}} = \mathrm{SD}_1[\sqrt{(1 - r_{xx})}]$, where SD_1 is the standard deviation of the baseline score. Practice effect was calculated by change in mean score over the test—retest interval, and was analysed for significance using repeated measures t-tests for each measure ($p < 0.05$). For each patient, the postoperative minus preoperative score was calculated ($X_2 - X_1$). When this value was greater than ± 1.645, a significant change was considered to have occurred. The RCI method can also be used to give 90% confidence intervals, using the following formula:

$$\text{RCI confidence interval} = \mathrm{SE}_{\mathrm{diff}}(\pm 1.645) + \text{mean practice effect}$$

Both the SRB and RCIp techniques provide a confidence interval for the detection of significant change, for example a 90% level of confidence indicates 5% of cases may be expected to fall above and 5% of cases below the cutoff due to chance rather than real change.

2.3.3. Standard deviation method

According to the SD method, a change in score on any measure is considered significant if it is greater than 1SD of the baseline score of the surgical sample. This may be represented as:

$$\text{SD score} = \frac{X_2 - X_1}{\mathrm{SD}_1}$$

2.3.4. 20% method

Using the 20% change method, a measurement score must change by at least 20% from baseline to be considered significant.

$$20\%\text{ method} = \frac{X_2 - X_1}{X_1}$$

2.4. Method of defining significant change in an individual

Patients were classified as impaired if they demonstrated significant deterioration on ≥ 2 of the nine test scores used. This decision was based on the probable distribution of false changes when using the prediction models (SRB and RCIp). A more detailed discussion of the process used to define overall change in individuals is included.

3. Results

3.1. Normative data

Test—retest data from the control group used to generate change intervals and regression equations are shown in Table 2. Reliability coefficients ranged from 0.49 to 0.84. All measures demonstrated a statistically significant practice effect at retest. Spatial processing was most vulnerable to practice, with an improvement of 24.1 (32% improvement from baseline). Attention/mental control was least affected, with an improvement of 7.3 (8% improvement). The full analysis of retest performance for this control group has previously been reported [11].

3.2. Surgical patient results

Table 3 lists all nine MicroCog scores, showing the percentage of patients with postoperative change detected by each of the four analytic models. There is a wide variation in the incidence of decline across scores, and across analysis techniques. The SRB and RCIp techniques classified more patients as declined than the SD or 20% methods on all scores

Table 2
Control group data used to generate reliable change and regression-based change statistics for MicroCog retested at 2 weeks

Domain	Test—retest reliability	$\mathrm{SE}_{\mathrm{diff}}$	Practice effect	RC−	RC+	Slope	Intercept	$\mathrm{SE}_{\mathrm{est}}$
Attention/mental control	0.68	13.66	7.3*	−15.2	29.8	0.58	43.67	11.56
Reasoning/calculation	0.67	14.38	10.5**	−13.2	34.2	0.62	41.32	12.71
Memory	0.65	14.08	9.3**	−13.9	32.5	0.68	36.00	13.18
Spatial processing	0.49	15.19	24.1**	−0.9	49.1	0.42	66.93	12.21
Reaction time	0.51	18.59	13.0**	−17.6	43.6	0.50	48.74	16.17
Information processing speed	0.84	10.30	11.7**	−5.2	28.6	0.64	40.93	7.84
Information processing accuracy	0.75	11.86	10.9**	−8.6	30.4	0.74	31.86	11.18
General cognitive functioning	0.77	10.98	13.7**	−4.4	31.8	0.70	36.20	9.92
General cognitive proficiency	0.81	7.87	10.4**	−2.5	23.3	0.78	27.13	7.44

$\mathrm{SE}_{\mathrm{diff}}$: standard error of difference scores, RC−: lower limit of 90% confidence interval, RC+: upper limit of confidence interval, $\mathrm{SE}_{\mathrm{est}}$: standard error of estimate.
* $p = 0.002$.
** $p < 0.001$.

Table 3
Percentage of surgical patients demonstrating postoperative changes in neuropsychological test performance detected by the four analytic models

Domain	SRB		RCIp		SD		20%	
	Decline	Improve	Decline	Improve	Decline	Improve	Decline	Improve
Attention/mental control	27.9	0	16.4	0	14.8	9.8	16.4	11.5
Memory	8.2	3.3	4.9	4.9	3.3	23.0	4.9	23.0
Reasoning/calculation	4.9	3.3	6.6	1.6	3.3	26.2	4.9	29.5
Spatial processing	6.6	0	3.3	3.3	1.6	62.3	1.6	52.5
Reaction time	6.6	3.3	8.2	3.3	8.2	24.6	6.6	26.2
Information processing speed	21.3	3.3	6.6	1.6	0	23.0	0	24.6
Information processing accuracy	14.8	1.6	8.2	3.3	1.6	21.3	1.6	23.0
General cognitive functioning	19.7	0	9.8	0	0	31.1	1.6	31.1
General cognitive proficiency	9.8	3.3	4.9	1.6	0	29.5	0	21.3

except reaction time. The average incidence of decline across scores was 3.6% (SD method) and 4.2% (20% method), compared to 7.7% (RCIp method) and 13.3% (SRB method). Overall, the SRB method tends to classify more patients as impaired than all other analysis techniques. Both the SRB and RCIp also classified a small number of patients as improved; this number was not greater than the prediction of 5% change due to chance. In comparison, both the SD and 20% methods classified a large proportion of patients as significantly improved. This highlights the importance of accounting for the effects of practice, which can particularly be seen with the spatial processing domain, on which the control subjects demonstrated considerable practice (Table 3). Across methods, attention/mental control consistently demonstrated the greatest decline for any measure. The SRB technique also detected considerable decline for the higher order scores such as information processing speed and general cognitive functioning. For reference, the mean and standard deviation performance at both time points for the control and surgical groups is shown in Table 4.

The distribution of score changes that may be expected using the SRB or RCIp was calculated using the binomial distribution. Assuming a 5% rate of decline due to chance in each score, the estimated probability of detecting $\geq$2 declines due to chance from the nine scores used was found to be 0.07. Table 5 shows the estimated probability of change scores due to chance that may be expected in either direction when using a cutoff of $\pm$1.645. Using the criteria of change in $\geq$2 test scores, the incidence of postoperative impairment and improvement indicated by each of the four criteria is shown in Fig. 1. Considerable differences can be seen between methods. The SRB technique yields the greatest number of patients classified as impaired (32.7%). In comparison, the RCIp classified only 16.4%, while the SD and 20% methods detected just 3.6% and 5.5%, respectively. Again, both the SRB and RCIp techniques revealed a small number of patients classified as improved, which was not greater than the 7% predicted by the binomial distribution of change. In comparison, the SD and 20%

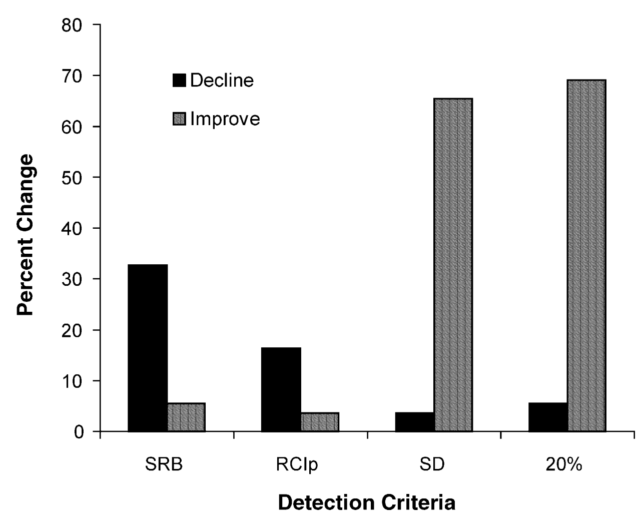

Fig. 1. Incidence of postoperative decline and improvement across detection criteria. SRB: Standardised Regression-based technique, RCIp: Reliable Change Index modified for practice, SD: standard deviation method, 20%: 20% change method.

Table 4
Mean and SD performance of control and surgical cohorts at baseline (Time 1) and repeat testing (Time 2)

Domain	Controls				Surgical			
	Time 1		Time 2		Time 1		Time 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Attention/mental control	87.3	18	94.6	16	83.3	15	80.1	17
Memory	80.4	18	90.8	17	79.2	16	88.2	17
Reasoning/calculation	82.0	17	91.3	17	74.6	15	82.3	16
Spatial processing	74.1	16	98.2	14	74.4	15	95.5	15
Reaction time	71.3	19	84.3	19	77.5	18	86.0	20
Information processing speed	79.3	19	90.9	14	75.2	14	83.4	16
Information processing accuracy	80.3	17	91.3	17	75.5	14	81.4	17
General cognitive functioning	75.5	17	89.2	15	69.7	14	78.8	16
General cognitive proficiency	76.2	13	86.6	12	72.2	11	80.9	13

Table 5
Estimated probability of decline in change scores due to chance when increasing the number of tests

Number of tests	1 or more		2 or more		3 or more	
	p	% of tests	p	% of tests	p	% of tests
3	**0.143**	**≥33.3**	0.007	≥66.7		
4	**0.185**	≥25.0	0.014	≥50.0		
5	0.226	≥20.0	**0.023**	**≥40.0**	0.001	≥60.0
6	0.265	≥16.7	**0.033**	**≥33.3**	0.002	≥50.0
7	0.302	≥14.3	**0.044**	**≥28.6**	0.004	≥42.9
8	0.337	≥12.5	**0.057**	**≥25.0**	0.006	≥37.5
9	0.370	≥11.1	**0.071**	**≥22.2**	0.008	≥33.3
10	0.401	≥10.0	**0.086**	**≥20.0**	0.012	≥30.0
11	0.431	≥9.1	0.102	≥18.2	**0.015**	**≥27.3**
12	0.460	≥8.3	0.118	≥16.7	**0.020**	**≥25.0**
13	0.487	≥7.7	0.135	≥15.4	**0.025**	**≥23.1**
14	0.512	≥7.1	0.153	≥14.3	**0.030**	**≥21.4**
15	0.537	≥6.7	0.171	≥13.3	**0.036**	**≥20.0**

p: probability of decline due to chance. % of tests: percentage of test scores used when set as a cutoff for overall decline. Bold values represent optimal cutoff when using a given number of tests. A target value $p < 0.10$ was used where possible.

methods both classified large numbers of patients as significantly improved (65.5% and 69.1%, respectively).

4. Discussion

The aim of this study was to discuss the use of more statistically sound techniques when evaluating change in cognitive performance after cardiac surgery. Obviously, factors such as time since surgery impact on observed changes; however, this report highlights the effect different analytical techniques have on reported outcomes using the same data set. Earlier reports [6,9,13] have shown different analysis techniques to strongly influence the reported incidence of decline, a finding supported in this comparison of four methodologies. Because the incidence of change is highly dependent on the method of analysis, it is essential that rational criteria be used. While traditional fixed cutoff techniques such as the SD method are simple to use, they do not employ sound statistical criteria for change, and as such are highly dependent on the reliability and stability of the test. In other words, these methods do not anticipate the likely changes that may be expected in test performance over time, even though the deviation from 'normal' is exactly what is being measured in the clinical setting. Because the expected changes are not accounted for, a fixed cutoff may seriously miscalculate true change in the patient population. Accordingly, it has been suggested that techniques such as SD and 20% change be abandoned for change criteria using an appropriately matched control group [9,13—15]. This study builds on earlier methodological publications by directly comparing the application of two statistical approaches widely accepted in mainstream clinical neuropsychological research, and by the presentation of a novel approach to the statistical classification of overall impaired on the basis of a number of test scores.

Two statistical criteria for change that account for measurement error and practice are presented in this study: the Reliable Change Index, modified for practice, and the Standardised Regression-based technique. Both techniques are simple to apply, and are based on similar principles. They do, however, differ in a number of ways that may lead to different conclusions. The RCIp accounts for measurement error and mean practice effects detected in controls for each measurement score. This technique tends to identify more patients as declined on the nine scores than either the SD or 20% methods. However, it does not account for individual practice or regression toward the mean. The potential advantages of the SRB technique lie not only with the influence of these factors in the prediction of retest performance, but also with the inclusion of demographic variables such as age and education. The equation may also include the influence of tests of mood such as anxiety and depression, which may have an effect on retest performance. In this series, however, only Time 1 performance was found to be significant in the prediction of retest performance in controls. Previous work with the SRB has shown only small influence of demographics [12], although when the influence is significant the effect should be entered into the equation to obtain optimal results. With a suitably sized, well-matched control group, the SRB prediction model will be better suited to the full range of patients being studied.

Because it predicts retest performance on the basis of initial performance, the SRB appears to give narrower detection intervals than the RCIp. This is due mainly to accounting for regression to the mean and is consistent with previous studies [12]. As a result, the SRB tended to give a higher sensitivity for most scores used: the SRB identified 27.9% of patients as declined on the attention/mental control domain, nearly double the rate given by the RCIp or fixed cutoff techniques. Further, the SRB was the only technique that classified a large number of patients as impaired on the higher-order scores of cognitive functioning. These scores, such as general cognitive functioning, are more likely to reflect impairment that may affect day-to-day functioning in an individual. Despite the narrower detection intervals for the SRB compared to the RCIp, there was little variation in the detection of improvement. Both techniques accounted well for practice as there were few improvements detected, and this was not above the 5% detection due to chance. The SRB technique, therefore, has the advantage of greater sensitivity in the detection of decline without increased classification of improvement. This does not necessarily reflect an ideal accuracy of detection; indeed more complex formula have been proposed as both the SRB and RCIp techniques have been criticised for not adequately dealing with measurement error [13]. However, both methods are simple to use and conceptualise, and are vastly superior to the traditional SD or 20% methods. While the debate over the best method for defining significant change remains open, the use of either statistical change criteria discussed here deals well with test imperfections, and makes considerable advancements over traditional cutoff techniques.

Two domain scores in this series, reaction time and spatial processing, highlight the importance of accounting for both reliability and stability. The reaction time score had poor retest reliability, and when using the fixed cutoff techniques there was an incidence of decline in this score similar to the SRB and RCIp techniques. This incidence of decline is high, relative to the much lower incidences detected using the fixed cutoff techniques across other domain scores. This may be because the large 'normal' fluctuations that occur in retest scores due to poor reliability more easily exceed the fixed cutoff used by the SD or 20% methods. In contrast, spatial processing while having poor reliability also had considerable practice effect. The two fixed cutoff techniques recorded low declines, but recorded considerable improvements, as the effect of practice for many patients would have exceeded the fixed cutoff.

One further question remains unanswered when defining individual change in cognitive performance: on how many tests must a patient demonstrate decline before they are considered to have shown significant overall change? The commonly used criteria include $\geq$1 test score, $\geq$2 test scores, and $\geq$20% of test scores. Again, these definitions are based on arbitrary decisions rather than on a theoretical underpinning. Using either the SRB or RCIp to assess decline on any one measurement score, the probability of detecting a decline entirely due to chance equals 0.05. Obviously, the more measurements used in a test battery, the greater the chance of recording a decline in any one or more scores due to chance. The choice of optimal criteria for the detection of overall change was analysed using the binomial distribution of false changes across a range of scores. The aim was to provide rational criteria for a cutoff based on describing the probability of decline detected entirely by chance. It must be stressed that the binomial distribution of score changes used here provides only an *estimate* of false changes as it assumes independence of tests—rarely the case in psychometric assessment. It does, however, provide some theoretical underpinning in the choice of cutoff rather than relying on an arbitrary number. In this study the estimated probability of false decline in $\geq$2 scores from the nine used was calculated to be 0.07. Compare this to the probability of false decline in $\geq$1 score ($p = 0.37$), which is unacceptably high. As a guide, the criteria of $\geq$20% of test scores used was mostly found to provide acceptable probability on the range of score numbers commonly encountered (Table 5).

The other factor that must be considered when describing change is shared variance among tests used. This occurs when two or more tests used in the overall analysis measure the same or similar cognitive functions. When this occurs, a decline in one score indicates likely decline on other scores that share variance. This will in turn inappropriately increase the probability of reaching any chosen cutoff in the detection of change. Shared variance needs to be controlled for by choosing a broad range of neuropsychological tests assessing independent constructs, and through presentation of sum scores for a domain where possible rather than relying on the presentation of a large number of related subtests. Where changes in overall function are not the focus of investigation, for instance therapies targeting specific brain functions, then more detailed presentations of similar, targeted neuropsychological assessments would be the desirable approach.

In conclusion, it has been demonstrated that the choice of statistical models used to assess post-event cognitive decline has a strong influence on reportable outcomes. Two methods employing statistical change criteria, namely the SRB and RCIp, demonstrated greater sensitivity in the detection of decline compared to fixed cutoff techniques. These methods are more likely to reflect 'true' change in the performance of an individual, as they detect significant variation from the spread of score changes that may be reasonably expected over time, based on retest data from matched controls. The SRB in particular was shown to be a useful prediction model as it provides an estimate of retest performance based on initial score for an individual, and as such considers individual practice effects and regression toward the mean. This technique also has the advantage of accounting for the effects of demographic variables such as age and education, should they influence the prediction model. When these models were used to assess whether a person could be classified as significantly impaired through the use of a battery of sub-tests, it could be seen that the number of tests used to define change has a strong influence on reported outcomes. Investigators should minimise shared variance by avoiding the presentation of similar sub-tests in the analysis. From a suitable selection of tests, the definition of overall change needs also to be based on sound statistical criteria. When using either RCIp or SRB, the cutoff of $\geq$20% of test scores used was found to provide acceptable probability on the range of score numbers commonly encountered.

References

[1] Shaw PJ, Bates D, Cartlidge NE, French JM, Heaviside D, Julian DG, Shaw DA. Neurologic and neuropsychological morbidity following major surgery: comparison of coronary artery bypass and peripheral vascular surgery. Stroke 1987;18:700–7.

[2] Aris A, Solanes H, Cámara ML, Junqué C, Escartin A, Caralps JM. Arterial line filtration during cardiopulmonary bypass. Neurologic, neuropsychologic, and hematologic studies. J Thorac Cardiovasc Surg 1986;91:526–33.

[3] Hall RA, Fordyce DJ, Lee ME, Eisenberg B, Lee RF, Holmes JH, Campbell WG. Brain SPECT imaging and neuropsychological testing in coronary artery bypass patients. Ann Thorac Surg 1999;68:2082–8.

[4] Croughwell ND, Newman MF, Blumenthal JA, White WD, Lewis JB, Frasco PE, Smith LR, Thyrum EA, Hurwitz BJ, Leone BJ, Schell RM, Reves JG. Jugular bulb saturation and cognitive dysfunction after cardiopulmonary bypass. Ann Thorac Surg 1994;58:1702–8.

[5] McKhann GM, Goldsborough MA, Borowicz LM, Selnes OA, Mellits D, Enger C, Quaskey SA, Baumgartner WA, Cameron DE, Stuart S, Gardner TJ. Cognitive outcome after coronary artery bypass: a one-year prospective study. Ann Thorac Surg 1997;63:510–5.

[6] Mahanna EP, Blumenthal JA, White WD, Croughwell ND, Clancy CP, Smith R, Newman MF. Defining neuropsychological dysfunction after coronary artery bypass grafting. Ann Thorac Surg 1996;61:1342–7.

[7] Chelune GJ, Naugle RI, Lüders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: practice effects and base-rate information. Neuropsychology 1993;7:41–52.

[8] McSweeny AJ, Naugle RI, Chelune GJ, Luders H. T scores for change: an illustration of a regression approach to depicting change in clinical neuropsychology. Clin Neuropsychol 1993;7:300—12.
[9] Kneebone AC, Andrew MJ, Baker RA, Knight JL. Neuropsychologic changes after coronary artery bypass grafting: use of reliable change indices. Ann Thorac Surg 1998;65:1320—5.
[10] Powell DH, Kaplan EF, Whitla D, Weintraub S, Catlin R, Funkestein HH. MicroCog assessment of cognitive functioning. San Antonio, TX: The Psychological Corporation, 1993.
[11] Raymond PD, Hinton-Bayre AH, Radel M, Ray MJ, Marsh NA. Test—retest norms and reliable change indices for the MicroCog battery in a healthy community population over 50 years of age. Clin Neuropsychol, in press.
[12] Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: a comparison of four models. J Int Neuropsychol Soc 1999;5:357—69.
[13] Bruggemans EF, Vandevijver FJR, Huysmans HA. Assessment of cognitive deterioration in individual patients following cardiac surgery—correcting for measurement error and practice effects. J Clin Exp Neuropsychol 1997;19:543—59.
[14] Murkin JM, Newman SP, Stump DA, Blumenthal JA. Statement of consensus on assessment of neurobehavioral outcomes after cardiac surgery. Ann Thorac Surg 1995;59:1289—95.
[15] Collie AC, Darby DG, Falleti MG, Silbert BS, Maruff P. Determining the extent of cognitive change after coronary surgery: a review of statistical procedures. Ann Thorac Surg 2002;73:2005—11.