Detection of cognitive decline after coronary surgery: a comparison of computerized and conventional tests

B. S. Silbert1*, P. Maruff1 2*, L. A. Evered1, D. A. Scott1, M. Kalpokas1, K. J. Martin1, M. S. Lewis1 2 and P. S. Myles3

1Centre for Anaesthesia and Cognitive Function, Department of Anaesthesia, St Vincent’s Hospital, Melbourne, Australia. 2School of Psychological Science, La Trobe University, Melbourne, Australia. 3Department of Anaesthesia and Pain Management, Alfred Hospital, Melbourne, Australia

*Corresponding author: Department of Anaesthesia, St Vincent’s Hospital, Victoria Parade, Melbourne, Victoria 3065, Australia. E-mail: silberbs@svhm.org.au

Background. Postoperative cognitive decline is a common complication after coronary artery bypass graft (CABG) surgery. Postoperative cognitive decline is defined on the basis of change in cognitive function detected with repeated assessments using neuropsychological tests. Therefore improvement in neuropsychological testing instruments may increase our understanding of postoperative cognitive decline.

Methods. Fifty patients undergoing CABG surgery completed both a conventional and a computerized battery of tests before and 6 days after CABG surgery. Fifty age-and education-matched controls completed the same test batteries 6 days apart. The reliability and the sensitivity to postoperative cognitive decline were computed for each battery.

Results. Both test batteries detected postoperative cognitive decline 6 days after CABG surgery. For the computerized battery, the reliability of the reaction times (intraclass correlation 0.89–0.92) was greater than for any test from the conventional battery (intraclass correlation 0.56–0.71), although accuracy measures were less reliable (intraclass correlation 0.61–0.89). The computerized battery detected all the cases of POCD identified by the conventional test battery and also five cases that were classified as normal by the conventional tests.

Conclusion. Computerized tests are suitable for measuring cognitive change after CABG surgery and may detect change in a greater proportion of patients 6 days after CABG surgery than conventional neuropsychological tests.


Keywords: complications, cognitive deficit; monitoring, neuropsychological testing; surgery, cardiovascular

Accepted for publication: February 1, 2004

Postoperative cognitive decline is a common complication of coronary artery graft (CABG) surgery and has been described extensively.1–4 Postoperative cognitive decline is identified on the basis of neuropsychological assessment in which performance on one or more tests (a test battery) is classified as having decreased from a preoperative level.1 2

A variety of neuropsychological tests assessing different cognitive functions have been used to define the disorder.5 6 To introduce uniformity into the assessment and definition of postoperative cognitive decline, specific sets of conventional neuropsychological tests have been recommended.7

The inclusion of tests such as the Rey Auditory Verbal Learning Test (RAVLT), the trail making tests (parts A and B) and the grooved pegboard test in these recommendations is based on findings that they consistently show sensitivity to postoperative cognitive decline (i.e. they must be able to detect change). The ability of neuropsychological tests to detect postoperative cognitive decline depends heavily on their sensitivity to detect change in cognitive function over

1Declaration of interest. Paul Maruff is a principal at Cogstate Ltd, Melbourne, Australia (www. cogstate.com).
time. The properties of the tests recommended for the detection of postoperative cognitive decline reveal common features that also make them useful for repeated administration and therefore for detection of change. The recommended tests all have relatively high test–retest reliability, which provides an estimate of the stability of the test over time. They have short administration times, parallel versions, small or non-existent practice effects, they assess simple cognitive functions and are relatively portable. Each test also yields data that have a broad range of possible scores, rarely suffer from ceiling or floor effects and are at least of an ordinal level of measurement.5 6 8

It is appealing intuitively to propose that the computerization of cognitive tests can increase their reliability through the standardization of test administration and recording of responses. However, computerization per se is not sufficient for increased sensitivity to change. Rather, sensitivity to change occurs if the features of the tests (computerized or conventional) required for detecting change can be enhanced. In addition to standardization of administration, computerization allows the rapid generation of infinite numbers of stimulus sets, and allows the accurate measurement of reaction time and accuracy while keeping the time for administration to a minimum.9–12 As with conventional tests, practice effects can be minimized if the actual tests are kept simple.9 A test battery designed explicitly to detect change has been shown to be more sensitive to cognitive change than conventional neuropsychological tests after sports injury, concussion and fatigue.12–14 The salient features include minimal practice effects, suitability for repeated administration (e.g. multiple alternative forms, standardized stimulus sets and response requirements and short administration times) and the generation of normally distributed performance scores.

The aim of the present study was to determine whether this computerized test battery could detect postoperative cognitive decline after CABG surgery in the immediate postoperative period and to compare the sensitivity with that of conventional neuropsychological tests.

Methods

After institutional ethics committee approval and informed consent, patients scheduled to undergo CABG surgery and healthy control patients were entered into the study. The CABG group were all aged 55 yr or above, presenting for first time for elective CABG surgery. Exclusion criteria were poor ventricular function (ejection fraction <30%) associated major systemic illness, pre-existing neurological disease, or anticipated difficulty with neuropsychological assessment (e.g. difficulty with eyesight or hearing, English not the prime language, hemiparesis). Anaesthesia consisted of oral temazepam premedication (10–20 mg), midazolam (up to 0.1 mg kg−1) for insertion of monitoring lines, fentanyl (intermediate or high dose: 10 or 50 µg kg−1), rocuronium (0.1 mg kg−1) before tracheal intubation and a propofol infusion as required. The lungs were ventilated with oxygen and air and volatile anaesthetic agents were not used. Proximal anastomoses were performed under aortic cross-clamping. For cardiopulmonary bypass, the circuit was primed with 2 litres of heparinized crystalloid solution (Plasma-lyte® 148; Baxter Healthcare, Toongabbie, NSW, Australia). Standard haemodynamic management involved maintaining a mean arterial pressure of 60–80 mm Hg (phenylephrine or isoflurane as required) with moderate hypothermia (nasopharyngeal temperature 30–34°C) and alpha-stat pH management. Myocardial preservation involved tepid blood cardioplegia (~25°C) given antegrade to induce asystole with subsequent doses administered retrograde at ~15 min intervals. Perfusion flow rates were 2.0–2.4 litres min−1 m−2 using non-pulsatile flow with either a centrifugal pump head (Medtronic Bio-Pump® Model 550 with a BP-80 head (Medtronic Australasia, Sydney, Australia)) or a roller pump head (using 0.5 inch PVC tubing; Cobe Cardiovascular, Arvada, CO, USA). All circuits used a Cobe® Optima membrane oxygenator (Cobe Cardiovascular) and a 40 µm arterial line filter. Rewarming was commenced at the completion of the final distal anastomosis (before proximal aortic grafting) and adjusted to achieve a nasopharyngeal temperature of not greater than 37°C at any time before weaning. Postoperative pain relief consisted of a morphine infusion for the first 48 h, paracetamol (with or without codeine) orally, and indomethacin orally or rectally.

The control group consisted of 50 healthy people with normal medical and psychiatric function, selected to match the CABG group for age, years of education and baseline IQ (Table 1). The controls were a subset of a larger group (n=254) recruited through advertisements for volunteers to study cognitive changes associated with normal ageing. All were deemed to have normal cognitive function after completing the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) neuropsychological battery.15

Parametric data were compared using t-tests, nominal data using Fisher’s exact test and ranked data using the Mann–Whitney U test.

| Table 1 Clinical characteristics of the control and CABG groups |
|-------------------|-------------------|-------------------|---|
|                  | Control (n=50)    | CABG (n=50)       | P  |
| Age (yr): mean (range) | 68.7 (54–87)   | 66.3 (55–82)   |    |
| Sex ratio (male/female) | 27/13         | 41/9            |    |
| Education (yr): mean (SD) | 8.9 (6.7)     | 7.9 (5.4)       |    |
| IQ estimate: mean (SD) | 112.5 (11.1)   | 110 (13.2)      |    |
| Smoker             | 0               | 34              | <0.001 |
| Diabetes           | 0               | 35              | <0.001 |
| Hypertension       | 12              | 30              | <0.001 |
| History of stroke  | 0               | 0               |    |
| Pulmonary disease  | 0               | 9               | <0.01 |
| Peripheral vascular disease | 0 | 1 | 0.34 |
| Obesity            | 0               | 7               | <0.01 |
Both the control and CABG groups completed a battery of conventional tests followed by the computerized tests. In the CABG group, neuropsychological assessment was undertaken before surgery (pre-admission clinic or day before) and before discharge (median 6 days, range 5–7 days). The control group was tested on the two test batteries approximately 6 days apart (median 6 days, range 5–9 days) to correspond with the CABG group.

The conventional tests were selected because they had been commonly used and recommended in the consensus statements. This battery (see Appendix in Supplementary data online) consisted of the CERAD Word Learning test, Symbol Digit Modalities test, Trail Making test parts A and B, Semantic Fluency test, and the Grooved Pegboard test. The conventional tests were scored according to standard protocols. Baseline intelligence (IQ) was estimated with the National Adult Reading Test.

The computerized test battery (see Appendix in Supplementary data online), a subset of tests from the CogState™ (Melbourne, Australia) battery, was selected to be brief and deliver the same number of performance measures as the conventional tests. For each test, the stimuli consisted of playing cards. Six test results were derived from the reaction time and accuracy of the three tests administered (detection, identification and matching). For each of the three tasks, 15 trials were presented.

For each computerized test, the number of trials on which an incorrect answer was given and the remaining correct trials were calculated and expressed as a percentage of the total trials. Trials on which incorrect responses occurred were then excluded from further analysis. Reaction times <100 ms and >1000 ms were classified as abnormally fast or slow responses and excluded. The reaction time distributions for each test showed significant negative skew (i.e. skew/SE skew >1.96); therefore, these were normalized using a logarithmic base 10 transformation. The mean of the log10 reaction time was calculated for each participant for each task.

The test–retest reliability of all tests was calculated from the control sample tested at baseline and at day 6 using the intraclass correlation. The intraclass correlation is interpreted in the same way as Pearson’s product moment correlation; however, unlike Pearson’s r, it is robust to order effects (e.g. the practice effect). The test–retest reliability is an estimate of the amount of error of measurement that is contained in a specific test. Therefore, test–retest reliability in controls provides an estimate of the extent to which a test provides a stable measure of performance over time. This is important for tests used to measure change, because the more performance can vary by chance, the harder it is to detect true change.

For both test batteries, the presence of postoperative cognitive decline for each test was determined at the group level by submitting each participant’s score on each test to a series of 2 (group: control, CABG) × 2 (time: baseline, day 6) repeated measures analyses of variance. For all of the analyses conducted at a group level, statistical significance was set at P<0.01.

The hypothesis that cognitive change would be greater in the CABG group than in controls was tested by the presence of a significant group × time interaction. Significant interactions were investigated with matched pairs t-tests. The magnitude of change detected in both the control and CABG group on all tests was computed using Cohen’s d statistic. This statistic expresses the differences between means as a function of their pooled standard deviation. It therefore provides a common metric for evaluating the magnitude of change on tests with different units and can be compared with published conventions that specify whether an effect is small (0.2), medium (0.5) or large (0.8).

In addition to group change, the incidence of cognitive impairment in individuals for both test batteries was determined by calculating the reliable change index (RCI). To calculate the RCI, the difference in performance from the baseline to the day-6 assessment was computed for each test in each of the control subjects. The mean and standard deviation (SD) of the distributions of difference scores for each test were then established. The mean of the difference scores from the control group reflects the expected effect of practice for each test, while the SD of the difference scores provides an estimate of the expected variability in normal performance on the cognitive tests over the study period. For each subject, assessment at the baseline and day 6 for each test was used to calculate the RCI as follows: RCI=(day 6 score−baseline score)−(practice effect estimated from controls)/(SD of difference scores estimated from control group).

An RCI can be interpreted like a standard (Z) score. For each test, a deterioration in RCI ≥1.65 was classified as abnormal decline in cognitive performance. The cut-off of RCI=1.65 is the point beyond which 5% of the values from the normal sample population will fall (i.e. P<0.05, one-tailed). To further protect against Type I error, postoperative cognitive decline was defined only if an individual showed cognitive decline on 2 or more performance measures derived from the tests in each battery.

Results
Table 1 shows the characteristics of the two groups. Participants were similar in age, years of education and estimated IQ. Details of the surgery for patients in the CABG are shown in Table 2.

In the computerized battery, there was no difference between groups for the percentage of trials excluded for being abnormally fast (control median 2.1%, range 0–4%; CABG median 2.0%, range 0–3.4%; Mann–Whitney U, P=0.2) or abnormally slow (control median 0.6%, range 0–1.5%; CABG median 0.5%, range 0–2.1%; Mann–Whitney U, P=0.2).

The test–retest reliability for conventional and computerized tests in the control group is shown by the intraclass
correlations in Table 3. The computerized tests were more reliable than the conventional tests. Within the computerized battery, measures of the reaction time were more reliable than measures of accuracy. For the conventional tests the grooved pegboard had the highest reliability.

Table 4 shows the group mean performance and SD on each test at baseline and day 6 in both the control and CABG groups, and the magnitude of change across time (Cohen’s $d$). Significant group × time interactions were found for all measures of the reaction time and for performance accuracy on the matching task in the computerized battery. Of the conventional tests, only the grooved pegboard test showed the interaction. Post hoc t-tests indicated that each of the changes with time arose because there was significant decline in cognitive performance at day 6 in the CABG group but not the control group. Importantly, no meaningful practice effects (e.g. Cohen’s $d<0.2$) were detected in the control group for any of the conventional or computerized tests. The greatest change detected at 6 days after CABG surgery was observed for the matching accuracy test.

Table 5 shows the mean (SD) of the change scores for each test in both batteries in the control group used to calculate the RCI and the number with abnormal performance for each test. Of the computerized tests, the measures of reaction time detected the greatest number of cases in which there was an abnormal decline in performance. Of conventional tests, the grooved pegboard showed the most cases. When the criterion of decline on two or more tests was applied to each battery, the conventional tests detected a rate of postoperative cognitive decline at day 6 of 32% (16 cases), while the computerized battery detected 42% (21 cases).
The rate of agreement between the two batteries was high (kappa=0.79; \(P=0.001\)).

## Discussion

Cognitive decline is an important and debilitating complication of CABG surgery. The measurement of postoperative cognitive decline has proved problematic. Selection of tests should be based on proven sensitivity to cognitive dysfunction, test–retest reliability, construct validity, low error rates, cultural insensitivity and ease of administration. Computerized cognitive tests have been used after non-cardiac anaesthesia and there is one report of such use after CABG surgery, which was able to detect decline 1 week after surgery.

The properties of the computerized test battery satisfy most of the requirements for detecting cognitive change. Indeed, the computerized battery has been designed specifically for detecting change in contrast to the conventional tests, many of which were originally designed for single administration. Thus, the computerized tests are sensitive to change, avoid practice effects, and are not prone to floor and ceiling effects. In addition, the computerized battery has other suitable characteristics for testing after surgery. It may overcome cultural and language difficulties and is easy to administer, standardized and quick. The results can be downloaded and analysed automatically, eliminating human error in both assessing and transcribing. In view of both the neuropsychological properties and utility of the computerized battery, it seems opportune to compare its use with that of the standard conventional testing.

The present results are consistent with previous studies showing cognitive decline in the immediate postoperative period after CABG surgery. Both the conventional and computer test batteries detected postoperative cognitive decline 6 days after CABG surgery.

Analysis of the intraclass correlations indicated that, for the computerized battery, the reliability of all three reaction times was higher than for any test from the conventional battery. There were minimal practice effects for reaction time or accuracy on the computer measures. The accuracy measures were less reliable than the reaction times, although the reliability of identification and matching accuracy were equivalent to those of the Trail Making and Symbol Digit Modalities tests. The low reliability of the accuracy measure from the identification task occurred because participants in both groups gave performance scores close to the maximum (100%) at both the preoperative and postoperative assessments (Table 3).

At the group level, the computerized battery showed cognitive decline for detection, identification and matching, and matching accuracy. The only conventional test to show a significant effect of CABG surgery was the grooved pegboard test. This pattern of results is related to the test–retest reliability, as only measures with reliabilities >0.7 gave results that indicated a significant effect of the surgery at the group level. The importance of choosing an appropriate test battery is best demonstrated by referring to a previous study by Mullges et al. that used conventional tests to detect postoperative cognitive decline. Deterioration in cognitive performance at a group level was noted 3 days after CABG surgery but performance returned to the baseline level by day 6. This was interpreted as showing that postoperative cognitive decline resolved relatively quickly after CABG surgery. However, our results suggest that the Trail Making A and B tests may not be sensitive enough to detect the cognitive decline that is present 6 days after CABG surgery. In the present study no decline was detected on the Trail Making A and B Tests 6 days after CABG surgery. However, a significant decline in performance was observed for the reaction time tests from the computerized battery and the grooved pegboard test.

### Table 5 Control group difference score (mean, SD) and number of participants with test deficits

<table>
<thead>
<tr>
<th>Test battery</th>
<th>Control group change (day 6 minus baseline)</th>
<th>Number abnormal RCI &gt;1.65</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean change score (SD)</td>
<td>Control</td>
</tr>
<tr>
<td>Conventional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory recall (n words)</td>
<td>0.40 (5.70)</td>
<td>1</td>
</tr>
<tr>
<td>Symbol digit modalities (n boxes)</td>
<td>3.77 (8.20)</td>
<td>0</td>
</tr>
<tr>
<td>Trail Making A (s)</td>
<td>-7.60 (13.20)</td>
<td>1</td>
</tr>
<tr>
<td>Trail Making B (s)</td>
<td>-6.60 (15.20)</td>
<td>0</td>
</tr>
<tr>
<td>Semantic fluency (n words)</td>
<td>1.6 (7.9)</td>
<td>0</td>
</tr>
<tr>
<td>Grooved pegboard (s)</td>
<td>-10.0 (13.2)</td>
<td>1</td>
</tr>
<tr>
<td>Computerized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detection reaction time (log10 ms)</td>
<td>-0.003 (0.050)</td>
<td>0</td>
</tr>
<tr>
<td>Detection accuracy (% correct)</td>
<td>0.01 (1.20)</td>
<td>1</td>
</tr>
<tr>
<td>Identification reaction time (log10 ms)</td>
<td>-0.01 (0.07)</td>
<td>0</td>
</tr>
<tr>
<td>Identification accuracy (% correct)</td>
<td>1.00 (1.50)</td>
<td>1</td>
</tr>
<tr>
<td>Matching reaction time (log10 ms)</td>
<td>-0.03 (0.07)</td>
<td>0</td>
</tr>
<tr>
<td>Matching accuracy (% correct)</td>
<td>1.20 (2.10)</td>
<td>0</td>
</tr>
</tbody>
</table>
The ability of any test to detect true change in cognitive functions depends on its ability to resist practice effects and the extent to which the variability within individuals, over time, is kept to a minimum. The use of a healthy control group to derive estimates of a practice effect, variability in change scores and the reliability of the tests is consistent with other studies of postoperative cognitive decline after CABG surgery and anaesthesia. RCI in which the denominators are estimated from the analysis of change in a non-operated control group provide the most reliable and valid method for determining the significance of any cognitive change within individuals. Minimal practice effects were found for the computerized tests in the control group. This is consistent with a previous study in which no practice effects were observed when the same tests were given four times to a group of healthy older people. Importantly, the tests in the CogState battery were designed explicitly to minimize practice effects by restricting the assessment to very simple attention, memory and executive functions.

In this study, the conventional battery identified postoperative cognitive decline in 32% of participants at day 6. The rate of postoperative cognitive decline was 42% of subjects when calculated from performance on the computerized test battery. Both estimates of postoperative cognitive decline are consistent with previously reported studies in CABG surgery populations. If the conventional tests were considered as the gold standard, then the sensitivity of the computer tests would be 100% and the specificity 85%. However, as noted by Arrowsmith et al. because of the difficulties in psychometric testing, there is no gold standard against which to compare new tests and the comparison against the conventional tests must be interpreted in this light.

The computerized cognitive test battery represents a viable alternative to conventional testing. All of the cases of postoperative cognitive decline classified by the conventional battery were also classified by performance on the computerized battery, although five additional cases were identified by their performance on the computerized battery. The question arises as to whether these five additional cases of postoperative cognitive decline were true cases or false positives. We believe that the additional cases of postoperative cognitive decline identified on the computerized battery may be true cases. This is because the greater sensitivity of the computerized tests to postoperative cognitive decline is consistent with their high test–retest reliability, absence of practice effects and relatively small variability found in the difference scores generated in the control group. Secondly, the same computerized tests have been shown to possess greater sensitivity to subtle cognitive decline than conventional tests in other circumstances. Finally, the five additional cases of postoperative cognitive decline identified by the computerized test battery all showed some impairment on conventional testing (three patients deteriorated >1.65 SD on one conventional test and >1.0 SD on at least one other; two deteriorated >1.1 SD on two or more of the tests).

The concept of specific neuropsychological domains being affected and impaired during surgery is appealing. However, most tests require multiple cognitive processes for successful performance. Therefore, the disruption of different cognitive systems may present as the same cognitive deficit. Thus, the current measurement tools reliably and sensitively measure change and make no inference as to which neuronal pathways or resultant cognitive domains may be disturbed.

In conclusion, we have compared the use of a computerized test battery with conventional tests to detect postoperative cognitive decline after CABG surgery. The computerized tests provided suitable test properties, were easy to administer and analyse, and proved to be sensitive to postoperative cognitive decline after CABG surgery. The use of the computerized battery warrants further investigation in the detection of postoperative cognitive decline after CABG surgery and has the potential to overcome many of the methodological difficulties that have hampered research in this area.

Supplementary data
The appendix can be found as supplementary data in British Journal of Anaesthesia online.

Acknowledgements
This study was supported by the National Health and Medical Research Council, Australia.

References
17 Darby D, Maruff P, Collie A, McStephen M. Mild cognitive impairment can be detected by multiple assessments in a single day. Neurology 2002; 59: 1042–6
20 Zakzanis KK. Statistics to tell the truth, the whole truth, and nothing but the truth: Formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. Arch Clin Neuropsychol 2001; 16: 653–67