The influence of different error estimates in the detection of postoperative cognitive dysfunction using reliable change indices with correction for practice effects

Matthew S. Lewis\textsuperscript{a,b,*}, Paul Maruff\textsuperscript{a,b}, Brendan S. Silbert\textsuperscript{a}, Lis A. Evered\textsuperscript{a}, David A. Scott\textsuperscript{a}

\textsuperscript{a} Centre for Anaesthesia and Cognitive Function, Department of Anaesthesia, St. Vincent’s Hospital, Victoria Parade, Melbourne 3065, Australia

\textsuperscript{b} School of Psychological Science, La Trobe University, Melbourne 3086, Australia

Abstract

The reliable change index (RCI) expresses change relative to its associated error, and is useful in the identification of postoperative cognitive dysfunction (POCD). This paper examines four common RCIs that each account for error in different ways. Three rules incorporate a constant correction for practice effects and are contrasted with the standard RCI that had no correction for practice. These rules are applied to 160 patients undergoing coronary artery bypass graft (CABG) surgery who completed neuropsychological assessments preoperatively and 1 week postoperatively using error and reliability data from a comparable healthy nonsurgical control group. The rules all identify POCD in a similar proportion of patients, but the use of the within-subject standard deviation (WSD), expressing the effects of random error, as an error estimate is a theoretically appropriate denominator when a constant error correction, removing the effects of systematic error, is deducted from the numerator in a RCI.

Keywords: Coronary artery bypass graft surgery; Neuropsychological assessment; Postoperative cognitive dysfunction; Reliable change index; Statistics

1. Introduction

In neuropsychology, there is increasing interest in the methodological and statistical processes by which change in cognitive function can be identified in individual patients. This issue has been raised in the context of the assessment of cognitive dysfunction after surgery for epilepsy (Chelune, Naugle, Luders, Sedlak, & Awad, 1993), and we have begun to investigate the use of different individual change calculations when using cognitive tests to decide whether athletes should return to play after head injury resulting in concussion (Collie et al., 2004), or determining the response to stimulant medication in individual children with attention deficit disorder (Mollica, Maruff, & Vance, 2004). However, this issue has been extremely important in the identification of postoperative cognitive impairment after surgery, in particular coronary artery bypass graft (CABG) surgery (Collie, Darby, Falleti, Silbert, & Maruff, 2002; Lewis, Maruff, \textsuperscript{1}DOIs of the original articles:10.1016/j.acn.2006.05.004, 10.1016/j.acn.2007.01.002.

* Corresponding author. Tel.: +61 3 9288 2251; fax: +61 3 9288 4255.

E-mail address: matt.lewis2@gmail.com (M.S. Lewis).

0887-6177/ – see front matter © 2007 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.acn.2007.01.019
Silbert, Evered, & Scott, 2006), and there is currently a high level of investigation and debate regarding the best methods to reliably identify cognitive change in individual patients.

It is accepted that some patients will show a decline in cognitive function after CABG surgery that was not present prior to surgery and is believed to have arisen peri-operatively. This decline is identified by comparing the preoperative to postoperative performances on objective tests of cognition in individual patients, has no objective clinical presentation or external markers, and is termed postoperative cognitive dysfunction (POCD). Thus, the identification of POCD depends entirely on the application of statistical rules to neuropsychological performance data of individual patients. To date, the statistical POCD rules used most commonly to classify abnormal change in performance require the individual to have declined by 1 standard deviation (computed from the group baseline performance) on a nominated number of tests (Mullges, Babin-Ebell, Reents, & Toyka, 2002; Shaw et al., 1986; Zamvar et al., 2002), or by 20% of their baseline score on 20% of the tests in the battery (Dowd et al., 2001; Dumas, Dupuis, Searle, & Cartier, 1999; Grichnik et al., 1999; Khatri et al., 1999; Stroobant, van Nooten, Belleghem, & Vingerhoets, 2002; Van Dijk et al., 2002). The application of these methods in different studies has indicated that POCD may be present in up to 80% of patients at 1 week following CABG (Shaw et al., 1986) and that some performance detriment may be observed in 40% of patients 5 years after surgery (Newman et al., 2001). These statistical rules for classifying POCD have been criticized however because they do not address the impact of practice effects, do not properly account for the impact of error in the statistical determination of change, and do not include a control group to gauge the expected false-positive classification rates (Collie et al., 2004; Rasmussen et al., 2001). Furthermore, the application of these rules to healthy nonsurgical control groups has shown that they are also associated with large false-positive classification rates (Lewis et al., 2006; Rasmussen et al., 2001).

Critical reviews of the methods used to determine change in the cognitive function in individuals now suggest that the application of reliable change indices (RCIs) to neuropsychological data may provide a better basis for decisions about cognitive change generally (Collie et al., 2004) and POCD specifically (Collie et al., 2002; Rasmussen et al., 2001). This is because RCIs can control the sources of error that are associated with repeated neuropsychological assessment (Collie et al., 2002; Kneebone et al., 1998; Lewis et al., 2006; Rasmussen et al., 2001). However, the term “reliable change index” has been applied to numerous formulae that express change in terms of the error associated with its assessment. All of these differ slightly in their computation and therefore their application to the same data set will be likely to yield different results. There has been some discussion in the literature regarding the best computation of the RCI, but no consensus has been reached (Maassen, 2004). Researchers may not be aware that different calculations of the RCI exist, and that they may classify different rates of change among the same population. Therefore, in the present study, estimates of the incidence of POCD after CABG, based on commonly used neuropsychological tests, is determined using four calculations of the RCI.

In its most basic form, the RCI [RCIJ&T (Jacobson & Truax, 1991)] expresses the difference in performance between two assessments (e.g., postoperative performance minus preoperative performance) as a function of the standard error of the difference (SEdifference) score for the test (see Table 1)( Jacobson & Truax). Expressing the change in terms of its error allows the change in the individual to be viewed in context of the spread of the distribution (Jacobson & Truax). This provides an indication of the magnitude of the change. Variations to this RCI have been developed that may also be relevant to the classification of POCD. Each modification to the RCI estimates the error associated with repeated neuropsychological assessment differently. Some groups have employed a correction for systematic error in order to control the practice effects that occur commonly when standard neuropsychological tests are given repeatedly over relatively short re-test intervals [e.g., RCIChelune (Chelune et al., 1993); RCISP OC D: ISPOCD—international study of postoperative cognitive dysfunction (Rasmussen et al., 2001)] (see Table 1). In the context of POCD, Kneebone et al. (1998) first demonstrated that this rule was useful in cardiac research, as it more adequately controlled the impact of practice and expected variability in neuropsychological performance than other methods for classification of POCD.

In both RCIChelune and RCISP OC D, the expected practice effects (estimated from some control group assessed at the same intervals) are included in the numerator and this results in a term in which the “average individual change of a control group” is subtracted from the “change in the individual.” However, these rules differ in their estimate of variability applied in the denominator. For example, the RCIChelune uses Jacobson and Truax’s (1991) original estimate of the standard error of the difference, whereas the RCISP OC D uses the standard deviation of the change scores estimated from the control group used to derive the group mean practice effect (Rasmussen et al., 2001). More recently, the RCI has been calculated using the within-subject standard deviation (WSD), drawn from the estimate of residual error in a repeated measures analysis of variance that has compared perfor-
Table 1  
Formula for the RCI,J&T, the RCI,SPOCD, the RCI,Chelune, and the RCI,WS D, the systematic change observed in the control group, and the error estimates used in the RCI calculations for each neuropsychological test

<table>
<thead>
<tr>
<th>RCI</th>
<th>Formula</th>
<th>Error and group change estimates used in the RCI equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔXc</td>
<td>ΔX/SE(difference)</td>
<td>WLT TMTA TMTB COWAT DSST GPD GPND</td>
</tr>
<tr>
<td>RCI,J&amp;T</td>
<td>0.8 (3.8)</td>
<td>−0.1 (15.6) 8.5 (24.7) 4.5 (10.5) −1.5 (8.3) −7.6 (14.5) −0.1 (15.6)</td>
</tr>
<tr>
<td>RCI,SPOCD</td>
<td>(ΔX−ΔXc)/SD(ΔXc)</td>
<td>3.80 8.28 15.59 24.70 10.47 14.48 15.58</td>
</tr>
<tr>
<td>RCI,Chelune</td>
<td>ΔX−ΔXc/SD(difference)</td>
<td>3.54 8.90 12.85 26.49 11.02 12.64 12.10</td>
</tr>
<tr>
<td>RCI,WS D</td>
<td>ΔX−ΔXc/WSD(ΔXc)</td>
<td>2.73 5.92 10.96 18.37 8.02 11.53 10.96</td>
</tr>
</tbody>
</table>

Note. ΔX = individual time 2 performance – time 1 performance; ΔXc = mean control group time 2 performance – time 1 performance; SE(difference) = √(2×(SD(baseline control)×(1−rxx)))², where rxx = test–retest reliability of the measure; SD(ΔXc) = standard deviation of the control group change; WSD(ΔXc) = within subject standard deviation of the control group.

Tasks: WLT, CERAD word learning task; TMTA, Trail Making Task part A; TMTB, Trail Making Task part B; COWAT, Controlled Oral Word Association Task; DSST, Digit Symbol Substitution Task; GPD, Grooved Pegboard Dominant; GPND, Grooved Pegboard Nondominant.

Table 1: Formula for the RCI,J&T, the RCI,SPOCD, the RCI,Chelune, and the RCI,WS D, the systematic change observed in the control group, and the error estimates used in the RCI calculations for each neuropsychological test.

mance over multiple time-points time in a control group (see Table 1; RCI,WS D) (Falleti, Maruff, Collie, & Darby, in press; Mollica et al., 2004). Bland and Altman (1996) have argued that the WSD can be used to reflect the error that has come from variation within the individual, or variability in the measurement process or from both sources.

To demonstrate the relative efficacy of the different RCIs to POCD, we applied all four to the baseline and Day 7 postsurgery neuropsychological data of a large (n = 204) cohort of patients drawn from the Australian Trial Investigating Postoperative Deficit, Early Extubation and Survival (ANTIPODES) study with the estimates of systematic and random error for the RCIs derived from 90 healthy, nonsurgical controls assessed over at the same times as the CABG group. As higher rates of POCD have been consistently detected at the Day 7 assessment than at other times (e.g., Newman et al., 2001; Stroobant et al., 2002; Toner, Taylor, Newman, & Smith, 1998; Treasure et al., 1989), cognitive data from this time-point provides the best opportunity to compare the ability of the different RCI calculations to classify POCD following CABG surgery.

2. Method

2.1. Participants

2.1.1. CABG patients

Two hundred and four patients scheduled to undergo first time elective CABG surgery were drawn from two sites involved in the ANTIPODES, a prospective randomized controlled trial investigating the cognitive impact of anesthetic technique (mean age = 68.8, SD = 7.0; mean IQ = 108.9, SD = 11.0; 152 males, 52 females). Patients aged 55 years or older were included. Exclusion criteria were poor left 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, poor English comprehension, hemiparesis).

Anaesthesia consisted of temazepam premedication, midazolam, fentanyl (either 10 or 50 μg/kg), propofol and rocuronium. Inhalational anaesthetics were not used. Arrest of the heart was managed using a combination of antegrade and retrograde tepid blood cardioplegia without active cooling. Proximal anastomoses were performed under aortic cross-clamping, CPB was undertaken with a membrane oxygenator and roller or centrifugal pump with continuous flow of 2.0–2.4 L/min/m², moderate hypothermia (32–34 °C) and mean systemic pressures maintained at 60–80 mmHg. Surgical details are presented in Table 2.
Table 2
Surgical data for CABG group (N=204)

<table>
<thead>
<tr>
<th></th>
<th>Median (IQR)</th>
<th>Range (min:max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N grafts</td>
<td>3 (1)</td>
<td>1:6</td>
</tr>
<tr>
<td>Cross-clamp time</td>
<td>78 (27.3)</td>
<td>32:132</td>
</tr>
<tr>
<td>CBP time</td>
<td>98 (34)</td>
<td>43:180</td>
</tr>
<tr>
<td>OPCAB (n)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Stroke/mortality</td>
<td>1 CVA—Day 7 post op (no residual symptoms)</td>
<td>1—Day 2 post op 2—Within 3 months post op</td>
</tr>
</tbody>
</table>

Note. CBP, cardio-pulmonary bypass; OPCAB, off-pump coronary artery bypass; CVA, cerebro-vascular accident; post op, postoperative.

2.1.2. Controls

Ninety controls were recruited from friends and family members of patients undergoing CABG surgery at a major metropolitan hospital (mean age = 67.8, \( SD = 7.9 \); mean IQ = 106.4, \( SD = 18.4 \); 65 males, 25 females). All individuals underwent psychiatric and neurological assessment at entry to the study and a full medical history was taken to assess each person’s health status. The exclusion criteria included a history of respiratory, circulatory or endocrine disease, personal or family history of psychiatric illness, head injury or substance abuse, mini mental status examination <27, and English as not the primary language.

2.2. Procedure

Following institutional ethics approval (March 2001) and informed consent, each participant completed a neuropsychological assessment battery at two time-points: at baseline (preoperative assessment), and at 1 week (postoperative Day 6—day of discharge in CABG group).

The National Adult Reading Test (NART) (Nelson, 1992) was completed at the baseline assessment. At the baseline and 1 week assessments participants completed a series of neuropsychological tests yielding seven outcome variables: the CERAD word learning task (WLT: total number of words recalled on immediate trials—maximum score 30), the Trail Making Task Part A (TMTA: number of seconds required to complete), the Trail Making Task Part B (TMTB: number of seconds required to complete), the Digit Symbol Substitution Task (DSST: number of symbols correctly transcribed in 90 s), the Controlled Oral Word Association task (COWAT: total number of words for the three, 60 s letter presentations), the Grooved Pegboard Task, dominant hand (GPD: number of seconds to completion), and the Grooved Pegboard Task, nondominant hand (GPND: number of seconds to completion). Each of the tasks was administered according to protocol by trained investigators. More comprehensive descriptions of the tasks can be found elsewhere (Lezak, 1995; Silbert et al., 2004).

2.3. Missing data

As each rule may be affected by the number of outcome variables applied against the rule, missing data for one or more outcome variable resulted in an individual being excluded from analysis at that time-point, rather than being

Table 3
Predominant reasons for excluding patients from analysis

<table>
<thead>
<tr>
<th>Missing all data at follow-up</th>
<th>n</th>
<th>Missing some data at follow-up</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient refused</td>
<td>8</td>
<td>Fatigue</td>
<td>4</td>
</tr>
<tr>
<td>Patient unwell/ICU/complications</td>
<td>9</td>
<td>Pain (graft site)</td>
<td>6</td>
</tr>
<tr>
<td>Depression/fatigue/other</td>
<td>3</td>
<td>Patient stopped task</td>
<td>5</td>
</tr>
<tr>
<td>Death</td>
<td>2</td>
<td>Other</td>
<td>3</td>
</tr>
<tr>
<td>Unexpected early discharge</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
estimated or being replaced with the mean value of the population. This resulted in 160 of the 204 CABG patients being available for analysis at the Day 7 follow-up. All controls completed all the tasks at both time-points. The reasons for the missing data are shown in Table 3 and include physical ailment (i.e., arthritis), impaired mobility or sensation as a result of the surgery (i.e., nerve damage arising from radial graft site), and time constraints (i.e., discharge procedures).

2.4. Data analysis

For all analyses, the direction of data was corrected so that positive changes indicated improvement, whereas negative changes indicated deterioration. All percentage values are presented as percentage of valid cases.

2.4.1. Calculation of RCIs and estimates of error relative to each

The algebraic formulae for each of these RCI calculations and their associated error estimates are shown in Table 2.

a. RCI_J&T: Change was calculated for each neuropsychological task in the individual by subtracting the preoperative performance from the postoperative performance. This value was divided by the SE_difference, which is calculated from the standard deviation of baseline performance in that group (s_1) and the test–retest reliability (r_xx) expressed as \( \sqrt{2s_1^2(1 - r_{xx})^2} \).

b. RCI_Chelune: Change was calculated for each neuropsychological task in the individual by subtracting the preoperative performance from the postoperative performance. The same calculation was conducted for each person in the control group to estimate the impact of any systematic error. The averaged control group change was then subtracted from the change in the individual. This value was divided by the SE_difference, calculated using the standard deviation of baseline performance in that group (s_1) and the test–retest reliability (r_xx) expressed as \( \sqrt{2s_1^2(1 - r_{xx})^2} \).

c. RCI_ISPOCD: Change was calculated for each neuropsychological task in the individual by subtracting the preoperative performance from the postoperative performance. The same calculation was conducted for each person in the control group to estimate the impact of any systematic error. The averaged control group change was then subtracted from the change in the individual. This value was then divided by the standard deviation of the change observed in the control group.

d. RCI_WSD: Change was calculated for each neuropsychological task in the individual by subtracting the preoperative performance from the postoperative performance. The same calculation was conducted for each person in the control group to estimate the impact of any systematic error. The averaged control group change was then subtracted from the change in the individual. This value was then divided by the WSD drawn from the change data in the control group and calculated using the methods described by Bland and Altman (1996).

2.4.2. Abnormal cognitive decline and the calculation of POCD

For each of the RCI calculations, the procedure described by Rasmussen et al. (2001) was used to classify POCD. Change relative to error was determined for each neuropsychological task providing an indication of change in the individual expressed as a Z-score. These scores were then used to create a compound test score (Z_combined) using the sum of Z-scores for each test (\( \Sigma Z_{a,b,c,d,...} \)) divided by the standard deviation of this summation in the control group using the same RCI calculation [SD(\( \Sigma Z_{control} \))]. On individual neuropsychological tests, performance met the criteria for abnormal cognitive decline if the Z-score on that task was \( \leq -2 \). POCD was then classified if the patient had two or more Z-scores for single tests \( \leq -2 \), or a Z_combined score \( \leq -2 \). This approach classified POCD on the basis of a substantial failure on two or more tests, or a more pervasive subtle decline across the neuropsychological test battery.

2.4.3. Statistical analyses

The incidence data on individual tasks and over the battery was compared between the different RCI calculations using Cochran’s Q for related dichotomous data (Cochran, 1950). An initial comparison of the outcome data from each of the four RCIs was conducted. If a significant difference was indicated between the four RCIs, subsequent analyses were conducted by ranking the incidences of dysfunction detected by the four RCIs and conducting a second series of Cochran’s Q-tests on the two RCIs that detected the greatest incidence of dysfunction for each test. Group analysis was conducted on the POCD classifications using Cochran’s Q.
### Table 4
Incidence of abnormal cognitive dysfunction (percentages) on individual neuropsychological tasks and POCD calculation detected by each RCI classification in CABG patients

<table>
<thead>
<tr>
<th>RCIs</th>
<th>WLT (n = 178)</th>
<th>TMTA (n = 178)</th>
<th>TMTB (n = 172)</th>
<th>COWAT (n = 178)</th>
<th>DSST (n = 172)</th>
<th>GPD (n = 173)</th>
<th>GPND (n = 164)</th>
<th>POCD (n = 160)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCI&lt;sub&gt;J&amp;T&lt;/sub&gt;</td>
<td>7.87</td>
<td>15.17</td>
<td>23.84</td>
<td>2.25</td>
<td>4.07</td>
<td>26.01</td>
<td>28.66</td>
<td>33.1</td>
</tr>
<tr>
<td>RCI&lt;sub&gt;ISPOCD&lt;/sub&gt;</td>
<td>7.87</td>
<td>12.92</td>
<td>30.23</td>
<td>6.74</td>
<td>4.07</td>
<td>15.61</td>
<td>21.95</td>
<td>33.1</td>
</tr>
<tr>
<td>RCI&lt;sub&gt;Chelune&lt;/sub&gt;</td>
<td>7.87</td>
<td>15.17</td>
<td>27.91</td>
<td>5.06</td>
<td>3.49</td>
<td>19.08</td>
<td>28.66</td>
<td>33.8</td>
</tr>
<tr>
<td>RCI&lt;sub&gt;WSD&lt;/sub&gt;</td>
<td>11.8</td>
<td>19.66</td>
<td>37.79</td>
<td>14.04</td>
<td>9.88</td>
<td>20.23</td>
<td>29.88</td>
<td>36.9</td>
</tr>
<tr>
<td>RCI&lt;sub&gt;WSD&lt;/sub&gt; vs. next</td>
<td>Q(3) = 21.0</td>
<td>Q(2) = 16.0</td>
<td>Q(1) = 13.0</td>
<td>Q(2) = 20.2</td>
<td>Q(1) = 10.0</td>
<td>Q(2) = 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most sensitive</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .001</td>
<td>p &lt; .01&lt;sup&gt;a&lt;/sup&gt;</td>
<td>p = .135</td>
</tr>
</tbody>
</table>

*Note.* Abnormal cognitive decline classified on one task if performance ≤−1.96, corrected for direction; POCD classified if performance on more than two tasks ≤−1.96, or combined Z-score ([Σindividual scores]/[SD<sub>individual scores control group</sub>]) was ≤−1.96.

*Tasks:* WLT, CERAD word learning task; TMTA, Trail Making Task part A; TMTB, Trail Making Task part B; COWAT, Controlled Oral Word Association Task; DSST, Digit Symbol Substitution Task; GPD, Grooved Pegboard Dominant; GPND, Grooved Pegboard Nondominant.

<sup>a</sup>RCIWSD significantly lower than RCI<sub>J&T</sub>.

### 3. Results

#### 3.1. Calculation of RCIs and estimates of error relative to each

Table 1 presents the four RCIs used in the current analysis and the estimates of error used in the denominator of each equation. The estimated SE<sub>difference</sub> was similar to the standard deviation of the change that was observed in the control group, but for different tests it both underestimated and overestimated the error associated with the assessment of change at each time. The WSD provided a smaller estimate of error for each task, and this was most readily observed for the COWAT data.

#### 3.2. Abnormal cognitive decline and the calculation of POCD

The proportion of individuals classified by each rule as having abnormal cognitive decline on each neuropsychological task, and the incidence rate of POCD are shown in Table 4. The total number of cases used in the analysis is also shown for each outcome measure and varies slightly between tests due to missing data. It can be seen for all of the neuropsychological tests, excepting the GPD task, that the RCI<sub>WSD</sub> detected a higher incidence of abnormal cognitive decline than the other RCI classifications.

For all neuropsychological tasks, there were significant differences in the incidence of dysfunction between the RCIs (p < .001 for all tasks). On individual neuropsychological tasks, the RCI<sub>WSD</sub> was generally the most sensitive rule (see Table 1). This is significant for all tasks except for the GPD and GPND. For the GPD the RCI<sub>J&T</sub> classified impairment in a significantly greater proportion of patients and indicates that rather than improve, the control group has unexpectedly deteriorated on this task. Applying the criteria for classifying POCD used by ISPOCD (taken from Moller et al., 1998) using each of the RCIs, the RCI<sub>WSD</sub> classified a greater proportion of POCD (see Table 1) relative to the other rules with significant differences evident between groups [Q(3) = 10.24, p < .001].

### 4. Discussion

The four RCI calculations all classified a similar proportion of CABG patients as having abnormal cognitive decline. In this context, such classifications indicate that these individuals have POCD and many researchers consider POCD to reflect that disruption to brain function has arisen from the cardiac surgery (see Lewis, Maruff, & Silbert, 2004). Thus, as POCD is considered to reflect brain injury, its classification should not be taken lightly. When considered on a test-by-test basis, the RCI<sub>WSD</sub> classified significantly more abnormality than the next most sensitive rule on
all but the Grooved Pegboard Dominant hand (GPD) measure. The results were anomalous for the GPD task as the RCIJ&T classified a higher incidence of abnormal cognitive decline on this than the rules with correction for learning. This suggests that the control group deteriorated on this task rather than improved. The RCIWSD classified a greater incidence of POCD than the other three rules, and this was also seen to be significant. Interestingly, the incidence of decline on each test was similar to that detected using the RCIJ&T and suggests that the impact of systematic error (i.e., practice effects) on these results were minor.

The RCIJ&T was introduced to determine true change in psychotherapy, a field where the assessment of change is difficult given the absence of external markers and gold standards (Jacobson & Truax, 1991), which are also problems associated with the study of POCD (Lewis et al., 2004). Although improvements to RCIJ&T have been proposed (Maassen, 2004), no universally accepted standard has emerged (Hinton-Bayre, 2004; Maassen, 2004; Temkin, 2004). However, the RCIJ&T has two important limitations. The first is that it assumes that there is no systematic variability arising from the repeated administration of neuropsychological tests (Maassen, 2004), which is an unlikely condition in studies of POCD. Practice effects are commonly observed when neuropsychological tests are re-administered, particularly over short re-test intervals (Chelune et al., 1993; Selnes et al., 2001). The second is that the RCIJ&T uses an estimate of error that is calculated from the baseline variance and the test retest reliability. In studies that have access to a control group, a more accurate assessment of error could be drawn from observed change in the control group (Collie et al., 2004).

Improvements to the RCI have been made that address both of these shortcomings. Both the RCISPOCD and the RCICheChelune deduct the average change in a control group from the individual change calculation to minimize the effect of learning or other systematic error. The RCICheChelune uses the same estimate of error as the RCIJ&T, and the RCISPOCD uses the standard deviation of the change, an error estimate drawn directly from the control group. On this basis, the RCISPOCD is preferable to the RCICheChelune; however, both equations use error estimates in the denominator that do not reflect the removal of the practice effect in the change calculation of the numerator. Although an estimate of systematic error calculated in the control group has been deducted from the change in the individual, the error estimate in the denominator still reflects the influence of both systematic and random error. As this is controlling systematic error twice, the true incidence of POCD will be underestimated as the numerator is reduced (individual change – expected change), and the denominator is magnified (error estimate = systematic + random error).

By its nature, systematic error is likely to affect the whole sample in a uniform way, and deducting a constant error correction can be seen to reduce its impact. Random error is more difficult to control, as it is highly individualistic, reflecting a confluence of factors that may not affect the sample equally. It is therefore justified to use an error estimate that reflects random error, rather than one that reflects both random and systematic error. We have suggested previously that it is more appropriate to calculate the RCI by expressing the change in terms of just the random error associated with the assessment (Mollica et al., 2004), as this will more closely represent the change against its unaccounted variance. Such an estimate of random variance is readily available using the WSD, which is calculated from the residual error term in an analysis of variance comparing performance on a neuropsychological test over time (Bland & Altman, 1996). Thus, in the current study, the WSD improved the ability of the rule to detect change; however, it is possible that this may reflect an increased Type I error rate. This can be controlled however by calibrating the rule against a control group to minimize error. To date, the theoretical appropriateness of the error estimate in these rules with error corrections has not undergone the same degree of scrutiny as has been applied to error estimates used in the RCIJ&T (Maassen, 2004), and therefore it is felt that such a dialogue is necessary to improve the accuracy of individual change calculations.

The WSD improves upon commonly used RCI calculations that account for systematic variability by utilizing a more appropriate estimate of error readily derived using the methodology outlined by Bland and Altman (1996). The RCIWSD draws its error estimate from the same source as the standard deviation of the change and provides a compelling alternative to existing RCIs that had already demonstrated distinct advantages over other POCD rules (Kneebone et al., 1998; Lewis et al., 2006; Rasmussen et al., 2001). Although it is possible that the increased ability to detect change in the current study reflects increased error, it is believed that the RCIWSD has greater theoretical justifications than other RCI equations in the classification of POCD.

The derivation and calculation of the RCIWSD in the current study may have application beyond the current use and be applicable to the assessment of individual neuropsychological change more widely. It is felt that greater effort is needed to improve the statistical techniques used to calculate cognitive change at an individual level, and we encourage
greater dialogue in this regard. As neuropsychological assessments are often associated with less than optimal metric properties, the interpretation of individual change needs to be cautiously approached.

Acknowledgements

Funding was provided by the National Health and Medical Research Council, Australia (Project Grant #140510).

References


