## ORIGINAL ARTICLE

# Statistical procedures for determining the extent of cognitive change following concussion

## A Collie, P Maruff, M Makdissi, M McStephen, D G Darby, P McCrory

See end of article for authors' affiliations

Correspondence to:
Dr Collie, 68 Molesworth Street, North Melbourne 3051, Victoria, Australia; acollie@cogstate.com

Accepted 1 May 2003

Neuropsychological (NP) testing is now often used to help to determine if the cognitive function of a concussed athlete has declined. The NP test score after concussion is compared with the baseline test score. Many clinicians simply subtract one from the other and make a clinical decision about the significance or otherwise of the resulting "difference score". Such techniques are inadequate, as they fail to account for the many factors that may confound interpretation of serially acquired cognitive test scores. This is a review of a number of alternative approaches used in other areas of medicine for differentiating "true" changes from changes caused by these confounding factors. A case example is used to illustrate the effect that the statistical approach may have on clinical decision making.

In the United States alone, an estimated 300 000 sports related concussions are reported each year. This figure may underestimate the incidence of concussion because of non-reporting or lack of awareness of concussive symptoms.[1][2] Other reports suggest that the incidence of concussion in junior and community based sports may also be higher than previous estimates.[3] The cognitive function of athletes after concussion is now commonly used to determine suitability to return to play and rehabilitation strategies. Investigations of cognitive function after concussion follow a common experimental protocol, whereby the athlete is assessed on a short battery of neuropsychological (NP) tests before the season (baseline assessment) and again after the concussion. Any change in the athlete's cognitive status is then determined by comparing the two scores.[4][5] Considerable attention has been paid to methodological problems associated with the assessment of cognitive function before and after concussion, including selection of NP tests, the setting in which testing occurs, and the potential effects of other athlete related factors[6]—for example, age, learning difficulty. Much less consideration has been given to the statistical techniques used to guide decisions about the presence or absence of cognitive impairment following concussion.

Judgments on the presence and magnitude of cognitive impairment following concussion are often made by a medical practitioner to whom the athlete has been referred. When baseline data are not available, the clinician must make a judgment about the athlete's performance relative to normative data.[6][7] This approach is mirrored in many published research studies, in which groups of concussed athletes have been compared with control groups on NP tests.[5][8][9] In contrast, when baseline data are available, a clinician can compare the score after concussion with the baseline score to determine if cognitive change has occurred. This approach has been adopted in more recent research studies,[10–12] and has been advocated by neuropsychologists and neurologists involved in sports medicine.[6][7] With the increasing uptake of baseline NP testing in clinical sports medicine settings, a review of the statistical techniques that can be used to compare baseline and post-concussion data is appropriate.

Although some of these techniques have begun to be adopted in sports medicine,[12][13] there remains little understanding of how these techniques may aid (or inhibit) accurate clinical decision making. Before entering this discussion, it is worth briefly describing the problems that these techniques are designed to overcome.

## PROBLEMS IN SERIAL ASSESSMENT

Most conventional NP tests are designed for the investigation of brain-behaviour relations in cognitively impaired subjects.[14] However, many of these tests have psychometric properties that restrict their use in serial investigations. For example, many "paper and pencil" and some computerised tests have limited or non-equivalent alternative forms, which may result in performance changes resulting from practice effects.[15][16] These tests may also have poor reliability, which results in increased measurement error and regression to the mean (see below).[17] When administered to healthy young people, many tests display floor or ceiling effects and have a restricted range within which a healthy subject usually scores. Combined, these confounding factors ensure that large changes in cognition are required for small changes in NP test score to be observed, and may mean that mild but "true" changes in cognition may not be reflected as a change in test score. Other factors that may affect test score on serial assessment include age, education, intelligence, sex, and severity of concussion. Furthermore, assessment related factors such as anxiety, fatigue, and stress may also affect the magnitude of change in test score on serial assessment.[17]

One factor that can confound interpretation of serial data is the practice effect or learning effect. On many NP tests, practice effects produce a slight improvement in performance,[18][19] which may be sufficient to mask any decline in performance occurring after a concussion. Further, the magnitude of practice effects may be modulated by the length of the test-retest interval, as longer intervals result in reduced practice effects and vice versa.[18][19] It is therefore important to minimise the effects of practice to help ensure accurate interpretation of changes in test data. The use of alternative forms of a test may help to achieve this aim. However, practice effects still occur in studies in which alternative forms are used.[15] Another strategy for reducing practice effects is to administer the test twice at baseline, and take either the second or the "optimal"[11] test as the "true" baseline against which subsequent comparisons are made.

**Abbreviations:** DSST, digit symbol substitution test; NP, neuropsychological; RCI, reliable change index; TMT, trail making test

**Table 1** Performance of athlete X on computerised and paper and pencil neuropsychological tasks

| | Baseline | | Day 1 | | Day 4 | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Psychomotor speed | 272.8 | 87.1 | 372.1 | 143.2 | 319.8 | 143.0 |
| Decision making speed | 489.0 | 111.4 | 734.2 | 190.5 | 722.2 | 201.4 |
| Problem solving speed | 598.3 | 215.5 | 851.2 | 272.7 | 609.5 | 210.2 |
| Memory speed | 1116.5 | 444.0 | 1404.4 | 439.4 | 1143.2 | 572.2 |
| Digit symbol (number) | 68 | – | 56 | – | 63 | – |
| Trail making test (s) | 43 | – | 80 | – | 35 | – |

One limitation of such dual baseline strategies is that assessment becomes time consuming.

Regression to the mean is a statistical phenomenon whereby an extreme test score from a individual at one assessment tends to revert toward the mean of the group to which that individual belongs at a follow up assessment.[17] Thus, an athlete who scores poorly at one assessment is likely to improve at a subsequent assessment, whereas an athlete who scores highly at one assessment is likely to decline at a subsequent assessment. This occurs without any interim evidence of injury. The effects of regression to the mean were evident in a recent study by Erlanger and colleagues,[20] who reported that athletes with fast baseline scores on tests of simple and choice reaction time became slower at a second test administration, while athletes with slow baseline scores became faster at follow up. The magnitude of regression to the mean is exacerbated when the test used to rate cognitive status has poor reliability, as greater amounts of measurement error result in greater regression to the mean.

This brief review suggests that there are significant methodological challenges for reducing error in serial assessment. This has led to the development of statistical techniques that attempt to differentiate true changes in cognitive test score caused by an independent variable—for example, concussion—from artificial or test related changes and measurement error. A discussion of these techniques in the context of a clinical case example follows.

## CASE EXAMPLE

Athlete X was a 19 year old professional footballer concussed during the second quarter of a game played early in the 2002 Australian football season. The athlete describes being hit twice during the general course of play, each blow separated by minutes. At the first contact, he experienced a visual disturbance lasting only seconds. This resolved quickly and he continued to play. The second contact was more major, and the athlete was removed from the ground and took no further part in the game. Headache, blurred vision, dizziness/nausea, and confusion were reported initially. There was no loss of consciousness or post-traumatic amnesia. Concussion was diagnosed by the team sports physician.

Cognitive testing was performed at a baseline assessment exactly three months before the injury, and then again one and four days after the concussion. The CogSport test battery was administered on all occasions,[11 21] as were the digit symbol substitution test (DSST) and the trail making test (TMT). At the baseline assessment, the athlete performed CogSport twice to help minimise practice effects, as described above. On day 1 after the concussion, the athlete was still symptomatic, experiencing a persistent headache, intermittent dizziness, and fatigue. Performance on CogSport tasks, the DSST, and TMT was considered to be impaired relative to baseline (table 1), and the athlete was instructed to return for a further assessment. On day 4 after the concussion, all clinical symptoms had resolved and performance on the TMT had returned to baseline. Performance on the DSST remained below baseline but had improved from day 1. Performance on the CogSport psychomotor and decision making tasks remained worse than at baseline. The athlete was withheld from playing the following week.

Figure 1 shows performance on the CogSport psychomotor speed task. Data from athlete X will be used to illustrate the advantages and disadvantages of the statistical techniques described below. Specifically, data generated at baseline and day 1 after concussion will be used. Test-retest reliability and normative data are required for some calculations. These data were taken from our previous work.[22] While reading the section below, it is important to remember that the treating doctor deemed the decline observed between baseline and day 1 in athlete X sufficient for him to be prevented from returning to play and training.

This review is concerned primarily with describing those techniques that may be applied with relative ease to individual level data. More specifically, it is concerned with describing techniques that have been used in the acute stages after concussion to assist clinical decision making. These include simple change scores and reliable change indices (RCIs). A brief summary of statistics that have been used in other areas of medicine is also provided. Table 2 lists most of these techniques, defines them statistically, and provides a reference in which they have been applied to NP test data.

## TECHNIQUES USED COMMONLY IN SPORTS CONCUSSION

''Simple change scores'' are perhaps the most commonly used methods of measuring the degree of change in a cognitive test score. These methods may be applied to both individual and group level data.[23] Simple change scores are calculated by subtracting baseline score from follow up score, ensuring that changes are reported in test appropriate units of measurement. For athlete X, this provides a value of
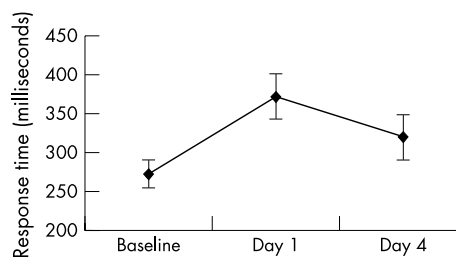


**Figure 1** Mean (SE) reaction time of athlete X on the CogSport psychomotor task. Performance is shown at baseline and one and four days after concussion. A decline in performance on this task is shown as an increase in response time, as occurs between baseline and day 1. Improvement in performance is shown as a decrease in response time, as occurs between day 1 and day 4.

**Table 2** Common statistical techniques for determining change in cognitive test score

| Statistical technique | Reference | Accounts for | Calculation | Extra data required |
|---|---|---|---|---|
| Simple change score<br>"True change" core | Mitrushina & Satz[29]<br>Maassen[30] | –<br>Measurement error<br>Regression to mean | $X_2-X_1$<br>$(X_2-X_1)\times r^2$ | Test reliability |
| Standard deviation index<br>Standard error of measurement index | Kneebone et al[31]<br>Jacobson et al[32] | –<br>Measurement error<br>Regression to mean | $(X_2-X_1)/\mu SD_1$<br>$(X_2-X_1)/\mu SEM_1$ | Control group SD<br>Control group SEM |
| Cohen's d<br>Simple regression | Benedict & Zgaljardic[15]<br>McSweeney et al[14] | –<br>Regression to mean | $(X_2-X_1)/SD_1$<br>$Y=BX+C$. Y is then compared with actual post-injury score | Control group baseline mean<br>Control group follow up mean |
| Multiple regression | Temkin et al[27] | Regression to mean<br>Variables such as age, education, socioeconomic, and history of concussion | $Y=BX+C+fAge+fEducation+....$ Y is then compared with actual post-injury score | Control group baseline mean<br>Control group follow up mean<br>Variables such as age, education, socioeconomic status, history of concussion |
| RCI | Jacobson & Traux[33] | Measurement error<br>Regression to mean* | $(X_2-X_1)/SE_{diff}$ | Test reliability<br>Control group SD |
| Modified RCI 1 | Chelune et al[34] | Measurement error<br>Regression to mean*<br>Practice effects | $(X_2-X_1)-(\mu_2-\mu_1)/SE_{diff}$ | Control group baseline mean<br>Control group follow up mean<br>Test reliability<br>Control group SD |
| Modified RCI 2 | Zegers & Hafkenschield[24] | Measurement error<br>Regression to mean*<br>Practice effects | $((X_2-X_1)r)+((\mu_2-\mu_1)]1-r)/(\iota\mu r_\iota(1-\mu r)SD)$ | Control group baseline mean<br>Control group follow up mean<br>Test reliability<br>Control group SD |

*Although RCIs allow for regression to the mean, they do not account for it statistically, as regression techniques do.
$X_1$ = baseline score; $X_2$ = retest score; SD = standard deviation; $SE_{diff}$ = standard error of the difference; $r$ = test-retest reliability coefficient; Y = predicted post-concussion score; X = slope of the regression equation; C = intercept of the regression equation; B = assessment number; $\mu$ = group mean; M = score of matched control subject; SEM = standard error of measurement; $f$ = value of multiple regression predictor; RCI, reliable change index.

99.25 milliseconds for the CogSport psychomotor task. The ''true change'' score represents the proportion of the simple change score that is reliable or not due to measurement error. Calculation of the ''true change'' score for athlete X provides a value of 75.43 milliseconds. The obvious problem here is determining what indicates a ''significant change''. This is left to the clinician's judgment and experience, as this method provides no definitive criterion above or below which significant change can be said to have occurred. Accurate interpretation of simple and true change scores can be difficult even for experienced clinicians. These change methods are also severely limited by lack of consideration of test-retest reliability and both within and between test variability. Further, no statistical adjustment is made for practice effects or regression to the mean.

RCIs, and modified RCIs, provide more guidance to the clinician in the decision making process. This is because, unlike simple change techniques, RCIs provide a criterion value above which an observed change can be said to be meaningful. Specifically, an RCI greater than 1.96 is likely to occur randomly in only 5% of cases (p<0.05), and is thus considered a significant change. Athlete X recorded an RCI of 2.01, indicating that the observed baseline to day 1 change depicted in fig 1 was significant. RCIs include an estimate of reliability. Operationally, this means that a less reliable test will require a greater test-retest difference score for that difference score to be rated as significant. RCIs do not, however, provide any direct statistical adjustments to minimise the effects of regression to the mean.

The standard RCI does not correct for the effects of measurement error caused by practice or other confounding variables. This requires manipulation of the numerator and has led to the development of modified RCIs. For example, Hinton-Bayre and colleagues[12] describe an RCI corrected for practice effects (see table 2 for calculation). With this formula, athlete X records a value of 2.23, higher than the 2.01 recorded with the uncorrected RCI described above. Again, this value indicates that athlete X's performance has changed significantly. Another modified RCI described by Zegers and Hafkenscheid[24] provides correction for measurement error. This method requires appropriate control data and also knowledge of test reliability. Using this formula, athlete X records a value of 2.20, again above the 1.96 cut off defining significant change.

These modified RCIs may be limited by their use of control group data to correct for individual practice effects, as prior research suggests that the magnitude of practice effects may vary considerably between individuals.[25] These modified RCIs also require that data be available for an appropriate control group assessed over a test-retest interval similar to that of the concussed athlete. Such data are rarely available in clinical settings. Despite these limitations, the outcome from RCI calculations is directly interpretable by clinicians, including those with limited experience administering cognitive tests.

Despite the ease with which RCIs may be interpreted, clinicians should always exercise their judgment when making return to play decisions. For example, a modified RCI calculated for athlete X between the baseline and day 4 assessments produces a result of 1.18. Although this result is below the criterion value of 1.95 and is therefore not statistically significant compared with baseline, it was sufficient for him to be withheld from playing for a further week. In this case, the clinician considered the RCI of 1.18 to indicate cognitive function that was recovering from the larger impairment observed on day 1 after concussion, but had not yet returned to baseline. Consistent with this interpretation, previous authors have suggested that an RCI>1.03 be considered borderline and worthy of further investigation.[13]

## OTHER AVAILABLE TECHNIQUES

The following section seeks to briefly describe techniques that have been applied in other areas of medicine to determine whether a individual's cognitive function has changed. It is expected that these will be investigated in future sports concussion research studies. Table 2 describes the mathematical derivation of these techniques.

The standard deviation index expresses the individual athlete's change score as a proportion of a control group standard deviation. For athlete X, application of this technique provides a value of 1.39, indicating that his mean performance has changed by greater than 1 standard deviation of the performance of a matched control group. This statistic is often applied in medical research as the outcome is easy to understand and interpret.[26] However, the clinical significance of a 1 standard deviation change following sports related concussion is yet to be established. Further, this statistic will be affected by the size, homogeneity, and appropriateness of the control group, and may only be calculated by those with access to control group data.

The standard error of measurement index technique replicates the standard deviation index described above, but the standard error of measurement (SEM) replaces the SD in the denominator. The advantage is that the SEM incorporates some sources of measurement error—for example, sample size. However, this technique is still subject to the same limitations as the standard deviation index. Calculation of the standard error of measurement index for athlete X provides a value of 5.58. Again, the clinical significance of this value is unknown.

Cohen's d is a common method that is easy to apply to both individual and group level data, requiring knowledge only of the mean and standard deviation of test performance at baseline and follow up. For athlete X, this technique reveals a value of 1.14, indicating that his performance has changed by more than 1 standard deviation of his own baseline standard. Cohen's d can be calculated for the data from individuals only if the test used provides estimates of the mean and standard deviation of performance for each testing session. For example, this technique could not be applied to the DSST or TMT, as a standard deviation cannot be derived from any individual performance. Cohen's d is subject to the same limitations as the standard deviation index, and, in addition, is inappropriate for serial analysis within individuals as it was designed specifically for comparison of two independent groups.

A very promising approach yet to be applied clinically in sport concussion is regression techniques. There exists a modest body of evidence to suggest that regression techniques are very accurate at determining if cognitive change has occurred.[14 27 28] Simple and multiple regression methods may be used to predict the subject's score after concussion from their baseline score. In the case of multiple regression techniques, the equation used to predict the score after concussion may include estimates of the effects of variables such as age, level of education, socioeconomic status, and history of concussion. A significant change is said to have occurred when the difference between the predicted and observed score is greater than a certain criterion. These techniques require access to serially collected normal control data and, in the case of the multiple regression technique, data from a normal control group—for example, age, education, sex, number of prior concussions.[14 27 28] One of the advantages is that these techniques directly account for regression to the mean. However, they incorrectly assume that the baseline score is perfectly reliable—that is, free from measurement error.

## EXAMPLES FROM THE LITERATURE

In the sports medicine literature, most published studies of cognitive function at baseline and after concussion investigate

## Take home message

The use of appropriate statistical techniques to determine both the clinical and statistical significance of change in neuropsychological test score after concussion is advocated. Such techniques include reliable change calculations and regression methods, which are designed to minimise sources of measurement error. Use of inappropriate techniques may contribute to erroneous clinical decisions, endangering the health of the concussed athlete.

test performance in groups of athletes.[5][9][10][23] The statistical techniques used in most of these investigations, although appropriate at the group level—for example, analysis of variance—may not be applicable to NP test data collected in individuals and are therefore not clinically useful. Two exceptions to this rule include papers by Hinton-Bayre and colleagues[12] and Erlanger and colleagues.[13]

Hinton-Bayre et al[12] used an RCI to assess individual variations in test performance following concussion in rugby players. Significant decline was observed in 80% (16 of 20) of concussed players assessed in the first three days after injury. Of the 16 players with significant decline, three were impaired on all three tests administered, six on two tests, and seven on one test. Similarly, Erlanger et al[13] assessed 26 concussed athletes using the "concussion resolution index" and used RCIs to determine the proportion of this group with significant post-concussive cognitive deterioration. Significant cognitive impairment was observed in 58% (15 of 26) of concussed athletes at the first evaluation after concussion. A further 12% (3 of 26) of athletes were considered to have borderline cognitive function. Although the validity of this classification has not been established, it is in keeping with the clinical interpretation in the case study discussed above.

Recent cognitive investigations in medical specialties other than sports medicine—for example, psychiatry, cardiology, and neurology—have aimed to determine the ability of the different statistical techniques described in this review to differentiate between "true" changes in cognition and changes attributable to sources of measurement error. These studies typically aim to determine the practical ability of these techniques to predict a follow up score from a baseline score. For example, Temkin and colleagues[27] compared standard and practice effect corrected RCIs with linear and multiple regression techniques as predictors of follow up performance in a group of 384 healthy adults. The corrected RCI linear and multiple regression methods were equally accurate at predicting follow up score, and the standard RCI was least accurate. The accuracy of these prediction models was further investigated in a later study by the same group[28] by applying them to serial cognitive data collected from a smaller, non-clinical sample (n = 124), a group of patients with schizophrenia (n = 69), a group of subjects recovering from traumatic brain injury (n = 23), and a group of subjects in whom a traumatic brain injury occurred between baseline and follow up assessments (n = 10). All statistical techniques performed best in predicting the follow up score of the non-clinical group, and poorly in predicting the follow up score of the schizophrenia and brain injured groups. This indicates that prediction models developed in non-clinical samples may not be transferable to patients with head injury and other patient groups. The accuracy of each model did not differ substantially, and therefore Heaton and colleagues[28] recommended use of the simpler RCI with practice effect correction over the more

complex regression models for clinically ill patients. Although not conducted in concussed athletes, studies such as these provide valuable information on the most appropriate statistical technique to use with both individual and group level data. Similar studies in sports related concussion are expected.

## SUMMARY AND CONCLUSIONS

Many of the statistical techniques used currently in the NP literature to differentiate "true" change in test score from change caused by measurement error and practice effects are summarised here. Whereas some techniques perform quite well and facilitate accurate clinical decisions, others fail to adequately account for possible confounding factors. RCIs provide output in a fashion that may be interpreted meaningfully by a clinician. Further, RCIs have been found to be statistically valid in other medical settings. Research in these areas also suggests that regression techniques are highly accurate; however, they are difficult to apply to individual athletes. A comparative study investigating the different statistical techniques in sports concussion is expected. When establishing cognitive testing as part of concussion management practices, clinicians should also incorporate methodological approaches to reducing the effects of confounding factors—for example, having multiple baseline assessments to reduce practice effects. If applied appropriately, these measures will ensure accurate assessment of cognitive function in sports related concussion research and clinical practice.

. . . . . . . . . . . . . . . . . . . .

## Authors' affiliations
**A Collie, D G Darby,** Center for Neuroscience, University of Melbourne, Parkville, Victoria, Australia
**A Collie, P Maruff, M McStephen, D G Darby,** CogState Ltd, Carlton South, Victoria, Australia
**A Collie, M Makdissi, P McCrory,** Centre for Health, Exercise and Sports Medicine, University of Melbourne
**P Maruff,** School of Psychological Science, La Trobe University, Bundoora, Victoria, Australia
**P McCrory,** Brain Research Institute, University of Melbourne

## REFERENCES
1 **Thurman DJ**, Branche CM, Sneizek JE. The epidemiology of sports related concussion in the United States: recent developments. *J Head Trauma Rehabil* 1998;**13**:1–8.
2 **Johnston K**, McCrory P, Mohtadi NG, et al. Evidence-based review of sports-related concussion: clinical Science. *Clin J Sport Med* 2001;**11**:150–9.
3 **Orchard J**, Wood T, Seward H, et al. Comparison of injuries in elite senior and junior Australian Football. *J Sci Med Sport* 1998;**1**:82–8.
4 **Barth J**, Alves W, Ryan T, et al. Mild head injury in sports: neuropsychological sequelae and recovery of function. In: Levin H, Eisenberg H, Benton A, eds. *Mild head injury*. New York: Oxford University Press, 1989.
5 **Maddocks D**, Saling M. Neuropsychological deficits following concussion. *Brain Injury* 1996;**10**:99–103.
6 **Lovell MR**, Collins MW. Neuropsychological assessment of the college football player. *J Head Trauma Rehabil* 1998;**13**:9–26.
7 **Collie A**, Darby DG, Maruff P. Computerised cognitive assessment of athletes with sports related brain injury. *Br J Sports Med* 2001;**35**:297–302.
8 **Matser JT**, Kessels AG, Jordan BD, et al. Chronic traumatic brain injury in professional soccer players. *Neurology* 1998;**51**:791–6.
9 **Butler R**. Neuropsychological investigation of amateur boxers. *Br J Sports Med* 1994;**28**:187–90.
10 **Macciocchi SN**, Barth JT, Alves W, et al. Neuropsychological functioning and recovery after mild head injury in collegiate athletes. *Neurosurgery* 1996;**39**:510–14.
11 **Makdissi M**, Collie A, Maruff P, et al. Computerized cognitive assessment of concussed Australian rules footballers. *Br J Sports Med* 2001;**35**:354–60.

12 **Hinton-Bayre AD**, Geffen GM, Geffen LB, *et al.* Concussion in contact sports: reliable change indices of impairment and recovery. *J Clin Exp Neuropsychol* 1999;**21**:70–86.

13 **Erlanger D**, Saliba E, Barth JT, *et al.* Monitoring resolution of post-concussion symptoms in athletes: preliminary results of a Web-based neuopsychological test protocol. *J Athl Train* 2001;**36**:280–7.

14 **McSweeney AJ**, Naugle RI, Chelune GJ, *et al.* ''T scores for change'': an illustration of a regression approach to depicting change in clinical neuropsychology. *Clin Neuropsychol* 1993;**7**:300–12.

15 **Benedict RHB**, Zgaljardic DJ. Practice effects during repeated administrations of memory test with and without alternate forms. *J Clin Exp Neuropsychol* 1998;**20**:339–52.

16 **McCaffrey RJ**, Ortega A, Orsillo SM, *et al.* Practice effects in repeated neuropsychological assessments. *Clin Neuropsychol* 1992;**6**:32–42.

17 **McCaffrey RJ**, Duff K, Westervelt HJ. *Practitioner's guide to evaluating change with neuropsychological assessment instruments.* New York: Kluwer Academic/Plenum Publishers, 2000.

18 **Catron DW**. Immediate test-retest changes in WAIS scores among college males. *Psychol Rep* 1978;**43**:279–90.

19 **Catron DW**, Thompson CC. Test-retest gains in WAIS scores after four test-retest intervals. *J Clin Psychol* 1979;**35**:352–7.

20 **Erlanger D**, Feldman D, Barth JT. Statistical techniques for interpreting post-concussion neuropsychological tests. *Br J Sports Med* 2001;**35**:370–1.

21 **Collie A**, Maruff P, Darby DG. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc* 2003;**9**:419–28.

22 **Collie A**, Maruff P, Makdissi, *et al.* CogSport: reliability and correlation with conventional cognitive tests used in post-concussion medical evaluations. *Clin J Sport Med* 2003;**13**:28–32.

23 **Collins MW**, Grindel SH, Lovell MR, *et al.* Relationship between concussion and neuropsychological performance in college football players. *JAMA* 1999;**282**:964–70.

24 **Zegers FE**, Hafkenscheid A. *The ultimate reliable change index: an alternative to the Hageman & Arrindell approach.* Groningen, the Netherlands: University of Groningen, 1994.

25 **Matarazzo JD**, Herman DO. Base rate data for the WAIS-R: t-retest stability and VIQ-PIQ differences. *J Clin Neuropsychol* 1984;**6**:351–66.

26 **Bruggemans EF**, Van de Vijver 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.

27 **Temkin NR**, Heaton RK, Grant I, *et al.* Detecting significant change in neuropsychological test performance: a comparison of four models. *J Int Neuropsychol Soc* 1999;**5**:357–69.

28 **Heaton RK**, Temkin N, Dikmen SS, *et al.* Detecting change: a comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychol* 2001;**16**:75–91.

29 **Mitrushina M**, Satz P. Effect of repeated administration of a neuropsychological battery in the elderly. *J Clin Psychol* 1991;**47**:790–800.

30 **Maassen GH**. Principles of defining reliable change indices. *J Clin Exp Neuropsychol* 2000;**22**:622–32.

31 **Kneebone AC**, Andrew MJ, Baker RA, *et al.* Neuropsychologic changes after coronary artery bypass grafting: use of reliable change indices. *Ann Thorac Surg* 1998;**65**:1320–5.

32 **Jacobson NS**, Follette WC, Revenstorf D. Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behav Ther* 1984;**15**:336–52.

33 **Jacobson NS**, Traux P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;**59**:12–19.

34 **Chelune GJ**, Naugle RI, Luders H, *et al.* Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology* 1993;**7**:41–52.