REVIEW

See end of article for

authors' affiliations

Correspondence to:

Dr Maassen, Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht

University, PO Box 80140, 3508 TC Utrecht, the

Accepted 7 August 2006 **Published Online First**

Netherlands; g.h.maassen@fss.uu.nl

22 August 2006

Reliable change assessment with practice effects in sport concussion research: a comment on Hinton-Bayre

G H Maassen, E R Bossema, N Brand



Br J Sports Med 2006;40:829-833. doi: 10.1136/bjsm.2005.023713

In his comments on our previous article, Hinton-Bayre advocates the use of the regression based approach in most cases of determining reliable change. This article comments on Hinton-Bayre's argument, discusses cases where the regression method might not be the preferred method, and presents adjustments that make the method more generally preferable.

.....

his review may be regarded as a sequel to our comment¹ on the proposals and reviews of Collie et al,2 and is elicited by Hinton-Bayre's commentary thereupon,3 published earlier in this journal. Hinton-Bayre seemed surprised that our article did not provide more extensive coverage of the regression approach to the assessment of reliable change, as this approach can also be encountered in the sports concussion literature.45 In our first article, it was indeed noted that this approach is preferable under certain circumstances, without further elaboration. The main reason for this reservation was that the regression approach was not the central subject of the papers of Collie et al, which made it expedient to avoid going into more detail. In our view, weighing the advantages of the regression approach is a complicated matter requiring a separate address.

In his comments, Hinton-Bayre expresses his preference for the regression approach in most circumstances. We understand this preference, but have some reservations. Firstly, in our opinion, the circumstances in which the regression approach is not preferred do not rarely occur. Moreover, we have reservations about the reasoning behind Hinton-Bayre's position. To clarify our views, we have chosen to structure this article around comments on some salient pronouncements made by Hinton-Bayre to underpin his position. This will amount to an explanation of the disadvantages of the regression method, and culminate in a proposal to adjust the regression method to make it generally preferable.

Before starting the discussion, we repeat Hinton-Bayre's outline of the methods for assessment of reliable change that are involved. A reliable change index (RCI) assessed for person i can be represented by the following generic formula:

 $RCI = \frac{Y_i - Y_i'}{s.e.}$

where Y is the observed posttest score, Y' is the posttest score estimated according to the method chosen, Y - Y' is the estimated true change, and s.e. the corresponding standard error. Originally, the standard error is a criterion for what is regarded an exceptional difference within the conceivable distribution of possible change scores of a participant under the null hypothesis of zero true change. The s.e. also depends on the method chosen. In the classic approach of Jacobson and Truax, which does not account for practice effects, Y' is estimated by the pretest score X. Jacobson and Truax proposed the use of:

$$S_{E_n} = S_{E_X} \sqrt{2} = S_X \sqrt{2(1 - r_{XY})}$$
 (2)

as the estimated standard error of measurement of the difference, with S_X being the standard deviation of the pretest, and rxy the test-retest reliability within the control group. Chelune et al⁶ proposed a method that accounts for the effects of retesting—for example, practice effects-often referred to as RCIp. Within this method, Y' = X - M, where M is the average change observed in the control group. Chelune et al adopted s.e.(2) as the accompanying standard error, which has usually been followed in sport concussion research.4 5 This standard error assumes equality of initial and final variance of the outcome measure. However, when differential effects of retesting are expected, this assumption is not justified. Commenting on the standard error recently presented by Collie et al in this journal,2 and dealing with methods that do not account for effects of testing, our previous article advocated the use of:

$$S_{E_n} = \sqrt{(S_X^2 + S_Y^2)(1 - r_{XY})}$$
 (3)

with S_Y^2 being the variance of the posttest. (This standard error has already been applied in sport concussion research by Iverson and colleagues.7) The test-retest reliability is only an appropriate estimator of the reliability of the outcome measure, if the true scores of pretest and posttest correlate perfectly-that is, if no differential effects of testing occur. However, this article deals with differential effects of testing, implying that the value of r_{XY} is probably too low as an estimation of ρ_{XX} and ρ_{YY} (respectively the

Abbreviations: RCI, reliable change index; SEP, standard error of prediction

830 Maassen, Bossema, Brand

reliability coefficient of pretest and posttest within the population from which the control group is sampled). As a consequence, s.e.(2) or (3) is enlarged, making RCIp a more conservative criterion. S.e.(2) bears the additional disadvantage of not accounting for change of score variance as a consequence of differential practice effects, which can make the standard error much too large, as well as much too small. Therefore, in the case of differential effects of testing also, s.e.(3) seems to be preferable to s.e.(2). In the regression approach, hereafter referred to as RCIsrb, $Y' = b_c X + c_c$, which is a regression estimate of the posttest score (b_c is the regression coefficient of posttest on pretest and c_c is a constant); index c refers to values of variables in the control group. McSweeny et al,8 who first presented RCIsrb, proposed use of the standard error of prediction (SEP) as the corresponding standard error:

$$S_{Y,X} = S_Y \sqrt{1 - r_{XY}^2} (4)$$

We will now discuss several statements and opinions on these methods, expressed by Hinton-Bayre in his commentary on our first article.

Statement 1: Comparison of false positive rates based on each of the theoretical approaches would be far more convincing to the clinician.

Like many authors, Hinton-Bayre9 compares the performance of various methods of determining reliable change by establishing the numbers of false positives resulting in a normative sample. This seems natural, because these authors contend that an RCI statistic should follow a standard normal distribution in the normative sample. Consequently, for instance, when using a 90% confidence interval, they verify whether indeed about 10% of all the members of a control group are wrongly assessed as being reliably changed. However, the claim of standard normally distributed RCI values can be challenged for several reasons. To begin with, both RCIp and RCIsrb should be regarded as a t statistic rather than a standard normally distributed statistic, because the estimates involved are based on sample statistics rather than population parameters. Secondly, as for RCIp, the use of the standard error of measurement is based on the assumption that the measurement errors in a given person are normally distributed and that the same normal distribution applies to any given person belonging to the population from which the control group is sampled. This is, of course, a rather stringent assumption. Nevertheless, if assumptions are not dramatically violated, and if the numerator values of the RCIs are not notably different, then a difference between chosen standard errors is only effective in the tails of the standard normal distribution and will only affect a few cases. As a consequence, differences between the numbers of false positives established in comparative studies usually prove to be small. Therefore comparisons of false positives in the control group rarely enable a researcher to conclude that one method is preferable. Unsurprisingly, Hinton-Bayre came to the following pronouncement.

Statement 2: Some comfort can be taken that the practical difference between the approaches appears to be minor.

The finding that differences between numbers of false positives, as established by RCIp or RCIsrb, are often small suggests that the normative sample is not particularly adequate for a comparison of approaches. The sport concussion studies that compared outcomes within a sample of concussed athletes, ^{4 5} however, also did not show strikingly different reliable change assessments. This raises the question of whether significantly different outcomes are actually found in practice.

Addressing this question, we present results of our own research, 10 which admittedly is not an example of sport

concussion research. It does, however, show that the results of the various approaches may differ considerably and gives an impression of which parameter values may induce spectacular differences. Our research was conducted to examine whether cognitive functioning of severely atherosclerotic patients was improved after carotid endarterectomy. It included a group of 59 patients, who were assessed before and three months after surgery, using a neuropsychological test battery and a mood questionnaire. RCI values were determined for 16 of these test outcomes. In addition, the study comprised a control group of 46 healthy people, comparable to the patient group with regard to demographic characteristics. This group was also assessed twice, with an interval of three months.

Most of the tests did not show spectacular differences between results of the various approaches, although some differences were notable. For our argument, we have selected the three outcome variables that showed the largest differences, including two neuropsychological tests (motor planning test/planning times of motor behaviour in milliseconds, and verbal fluency/number of words with a specific letter within one minute) and the mood state scale vigor. Table 1 displays the percentages of false positives according to several approaches in the control group, as well as percentages of patients who were submitted to surgery and who were assessed as reliably changed. The left hand columns show striking differences between the results of applying RCIp (with s.e.(3)) and RCIsrb in the experimental sample. As could be expected from our argument above, the differences in the normative sample were much smaller. Obviously, these results do not endorse Hinton-Bayre's comforting statement. In later sections we will discuss possible causes of the divergent results and recommend an approach of how to deal with those causes.

Sport concussion studies comparing false positives within the normative group, as well as studies comparing reliable change assignments within the injured group, show little difference between the results yielded by RCIp or RCIsrb. Therefore, as yet, the following two pronouncements expressed in Hinton-Bayre's comments are based on little evidence.

Statement 3: The regression based approach is preferable because it accounts for regression to the mean.

This opinion seems to have become established among sport concussion researchers.4 5 In our view, it is a misunderstanding to contend that the classic approach (when no practice effects are present) or RCIp (when practice effects are expected) do not account for regression to the mean as a consequence of unreliability. When no sources other than unreliability (such as, for instance, practice effects) are effective, regression to the mean is a direct manifestation of measurement errors, and the standard errors (2) and (3) afford tests of which amount of change can be expected as a consequence of unreliability. On the other hand, the use of a uniform standard error of measurement with regard to any given pretest score is debatable. For instance, related to the phenomenon of regression to the mean, a larger standard error seems to be required when the given initial score is extremely high or low. As for RCIsrb, Crawford and Howell¹¹ noted that s.e.(4) is not generally the correct formula to be used as the denominator with regard to a new person. A larger standard error is appropriate for a given person with a relatively high or low initial score, which is incorporated into the formula cited by Crawford and Howell.11

Statement 4: The regression approach is preferable because its standard error is usually smaller than the standard error of RCIp.

A sport concussion researcher or clinician indeed may prefer a method with a standard error that is possibly too small. Such a method induces a confidence interval that is

Table 1 Percentages of the normative and experimental sample in the study by Bossema et al° classified as reliably changed on the basis of a 90% confidence interval according to various procedures accounting for effects of testing

		(a) RCIp s.e.(3)	(a) RCIp with s.e.(3)		(b) RCIsrb with s.e.(4)		(c) Adj RCIsrb with s.e.(3)		(d) Adj RCIsrb with associated SEP		
Test	N	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos		
Normative sample											
MPT	39 '	7.7	0.0	7.7	5.1	5.1	0.0	7.7	0.0		
VF	46	6.5	4.3	2.2	6.5	2.2	2.2	2.2	4.3		
Vigor	46	4.3	4.3	2.2	8.7	2.2	4.3	6.5	4.3		
Experimental sample											
MPT	59	6.8	18.6	3.4	47.5	3.4	30.5	3.4	33.9		
VF	59	16.9	8.5	25.4	8.5	18.6	8.5	20.3	8.5		
Vigor	58	12.1	12.1	43.1	5.2	15.5	8.6	20.7	12.1		

Notable differences between corresponding outcomes for the same test appear in bold. s.e., Standard error; SEP, standard error of prediction; Neg/Pos, percentage of those classified as reliably changed on the basis of a negative/positive change score; MPT, Motor planning test/planning times of motor behaviour; VF, verbal fluency: number of words.

possibly too narrow, thus reducing the probability of not detecting significant deterioration of cognitive function of an athlete who has sustained a concussion. A too narrow interval also increases the probability of detecting subtle changes of cognitive function still present after a period of recovery, thus hampering the conclusion that the cognitive function of this athlete has returned to baseline.

Firstly, we will examine in what cases the standard error of the regression approach (SEP) is indeed smaller than that of RCIp. The difference between the squares of s.e.(2) and (4) equals $(1 - r_{XY})(2S_X^2 - S_Y^2(1 + r_{XY}))$, which reveals that SEP is smaller than the standard error usually used in sport concussion research, if $S_X^2/S_Y^2 > (1 + r_{XY})/2$. S.e.(2) exceeds SEP to the extent that the initial variance exceeds the final variance, and rxy differs from 1. Hinton-Bayre9 provided an example where this condition is not fulfilled, where SEP is larger than s.e.(2) (namely speed of comprehension, of which the initial variance is considerably smaller than the final variance). The difference between the squares of s.e.(3) and (4) equals $(1 - r_{XY})(S_X^2 - r_{XY}S_Y^2)$, which reveals that SEP is smaller than s.e.(3), if $S_X^2/S_Y^2 > r_{XY}$. s.e.(3) also exceeds SEP to the extent that the initial variance exceeds the final variance, and rxy differs from 1. Hinton-Bayre also pointed out a situation in which this condition is not fulfilled (namely the postconcussion symptoms scale in the study of Iverson *et al*⁷). Thus SEP occasionally proves to be larger than the standard error of measurement of the difference, but Hinton-Bayre is right in concluding that the SEP is smaller in most of the cases, particularly if the final variance is smaller than the initial variance, or if the test-retest reliability is low, or both.

Is the usually smaller SEP indeed a decisive argument for preferring RCIsrb? A complication is the bias of the true change estimation, preventing RCIsrb from being standard normally distributed. The regression approach seems preferable to RCIp because it more thoroughly uses information provided by the control group. The estimation of the true change in a person according to RCIsrb may indeed be more precise than according to RCIp, but this is just half of the story. The estimation is only unbiased under a restrictive assumption.12 Thus the estimation will be generally biased, implying that the conditional probability distribution of the RCI (given the true initial and change score of the person to be assessed) is not centred at 0. This means that the probability statements derived from the standard normal distribution usually associated with an RCI outcome are not justified. The estimation of the true change according to RCIp is also usually biased, but the bias using RCIp is expected to be smaller than the bias using RCIsrb under rather prevalent conditions. These conditions are 12:

$$b_c > \rho_{XX}$$
 and $b_c < 1$ (5a)

OI

$$b_c < \rho_{\chi\chi}$$
 and $\frac{\rho_{\chi\chi} - \rho_{\chi\chi}b_c}{\rho_{\chi\chi} - b_c} > 2$ (5b)

(Later we will discuss examples where these conditions are verified.) Searching for the smallest standard error may magnify the undesirable effect of the estimation bias. This effect will probably be masked within the control group, but may become spectacular within the experimental sample. In the following section, we will introduce an RCI that may be of interest to sport concussion researchers, because it is less biased and usually still associated with a smaller standard error than RCIp.

PROPOSAL TO ADJUST THE REGRESSION BASED METHOD

Maassen¹² has shown how the estimation of the posttest score can be modified to preclude occurrence of situations where the bias of true change estimation using RCIsrb is expected to exceed the bias when RCIp is used—that is, the conditions 5a and 5b:

$$Y_i' = \frac{b_c}{\rho_{XX}} X_i + c_c$$
, where $c_c = \overline{Y}_c - \frac{b_c}{\rho_{XX}} \overline{X}_c$

If Y - X is denoted by D, the numerator of RCIsrb becomes:

$$D_i - \overline{D_c} + \left(1 - \frac{b_c}{\rho_{xx}}\right) (X_i - \overline{X_c}) \tag{6}$$

Estimating ρ_{XX} , the reliability coefficient of the pretest by the test-retest reliability yields:

$$D_i - \overline{D_c} + \left(1 - \frac{S_{\gamma}}{S_{\chi}}\right) (X_i - \overline{X_c}) \tag{7}$$

The bias of the true change estimation using expression 6 is expected to be smaller than when RCIp is used. As for the standard error to be used in this adjusted RCIsrb, the researcher has several options. One option would be the SEP associated with expression 6 or 7, which can be easily calculated as the standard deviation of expression 6 or 7

832 Maassen, Bossema, Brand

Table 2 Statistics of the two assessments in the normative sample of the study by Bossema et allo

									Standard error			
									Measurement		Prediction	
Test	N	M_{X}	$M_Y - M_X$	SD_X	SD_{Υ}	r _{XY}	$b_{Y,X}$? _{Y.X}	s.e.(2)	s.e.(3)	s.e.(4)	Adjust
MPT VF Vigor	39 46 46	407.15 14.03 14.76	-10.93 0.70 -0.24	64.29 4.93 3.05	48.76 4.17 2.42	0.78 0.87 0.38	0.59 0.74 0.30	0.76 0.85 0.79	42.45 2.51 3.41	37.67 2.33 3.08	30.80 2.08 2.27	32.20 2.12 2.70

Disattenuated regression coefficient $\beta_{Y,X}$ is estimated by $b_{Y,X}/r_{XY}$. X, Pretest; Y, posttest; Adjust, standard error associated with adjusted RCIsrb; MPT, motor planning test/planning times of motor behaviour; VF, verbal fluency: number of words

within the control group (hereafter referred to as the SEP associated with the adjusted RCIsrb). This may be an attractive option in the field of sports medicine as we will discuss later, but, generally, for several reasons we prefer a second option: the standard error of measurement of the difference score as an estimate. This leads to the following RCI:

$$RC = \frac{D_i - \overline{D_c} + \left(1 - \frac{b_e}{\rho_{XX}}\right)(X_i - \overline{X_c})}{\sigma_{E_c}}$$
(8)

which, if ρ_{XX} is estimated by r_{XY} , becomes:

$$RC = \frac{D_i - \overline{D_c} + \left(1 - \frac{S_{\gamma}}{S_{\chi}}\right) (X_i - \overline{X_c})}{\sigma_{\kappa}}$$
(9)

where s.e.(3) is again advocated as the preferred estimation of the standard error. The first reason to prefer this option is that the standard error of measurement of the difference score is more closely related to the original concept of reliable change assessment, whereas the standard error associated with expression 6 or 7 is just a criterion for what size of prediction error is regarded as exceptional within the normative sample. Secondly, when practice effects do not occur (that is, $\overline{D_a} = 0$, $S_X = S_Y$), the classic approach, RCIp, and the regression based approach should coincide.

When differential practice effects do not occur, then $S_X =$ Sy, and RCIp and the regression based method should approximately coincide. This would be the case if expression 9 were applied. Finally, the results of applying several approaches discussed in the present text in our own research appear to be in favour of the second option. Table 1 displays the results of applying the following approaches: (a) RCIp with s.e.(3), (b) the original RCIsrb, (c) the adjusted RCIsrb with s.e.(3), and (d) adjusted RCIsrb with the associated SEP. We have already pointed out the spectacular differences between the outcomes of the approaches (a) and (b) within the experimental sample. Now, we will consider the outcomes of all four approaches applied to the mood scale vigor and the motor planning test. (The outcomes of the verbal fluency test showed the same pattern as those of the motor planning test, albeit less dramatically.) We will also discuss reasons for the differences between the outcomes in relation to the values of the variables central to the various approaches (mean difference score, initial and final variance, test-retest reliability, and standard error), which are displayed in table 2.

This variable showed the largest difference (26%) between s.e.(3) and the SEP of the original RCIsrb. This difference is predominantly caused by the low value of r_{XY} (0.38). The diminution of the variance is notable, but fails to be significant (t(44) = 1.67, p>0.10, two tailed). Note, however, that a low value of rxy may prevent notably different variances from being significantly different. It can be verified that inequality 5b holds for this scale, indicating the use of an adjusted RCIsrb (approach (c) or (d)). This should lead to a reduction in the number of false positives in the control group as compared with RCIp. Table 1 shows that this only holds for variant (c). As the mean score did not change significantly in the control group (paired samples ttest: t(45) = 0.52, p = 0.60, two tailed), we conclude that practice effects did not occur. In that case, methods that do and do not account for practice effects should not yield markedly different results. Table 1 shows that this holds for RCIp versus approach (c). The other variants produced considerably more designations of reliable change, which can be explained by a more biased estimation of the true change, a standard error that is possibly too small, or both (the original RCIsrb). We therefore put most confidence in the results of RCIp or variant (c) of the adjusted RCIsrb, which were similar.

Motor planning test

Table 2 shows very different standard errors with regard to this variable. This was caused by a significant diminution of the variance (t(37) = 2.73, p<0.05, one tailed) while the value of rxy was satisfactory. As the mean score had changed significantly in the control group (t(38) = 1.70, p<0.005, one tailed), we conclude that practice effects did occur. Methods that account differently for practice effects that occur may yield different results, but as the inequalities 5a and 5b did not hold for the motor planning test, in principle, approaches (b)–(d) should produce no more false positives in the control group than RCIp. Table 1 shows that, at both tails of the RCI distribution, this only clearly holds for variant (c) and (d). In the experimental group, the largest number of reliable change designations was produced by the original RCIsrb, considerably more than the two variants of adjusted RCIsrb, which are expected to be less biased.

DISCUSSION

Attempting to resolve the confusion about the standard error that should be used in reliable change calculations, our previous article showed that s.e.(3) should be used rather than the commonly used s.e.(2). The present article shows that the outcomes of RCIp s.e.(3) incorporated and of RCIsrb within the experimental sample may be spectacularly different when the final variance in the control group is notably smaller than the initial variance, or when the testretest reliability of the outcome measure is low, or both. (The difference would be even larger if s.e.(2) were incorporated into RCIp, because the standard error is then solely based on the larger initial variance.) In cases of low test-retest reliability or large diminution of variance, the performance

What is already known on this topic

 As demonstrated by Hinton-Bayre's commentary, the regression approach to assessing reliable change appears to be preferred by sport concussion researchers

of an adjusted RCIsrb—that is, expression 8 or 9—in our own research was most consistent with theory. It designated more people as reliably changed than RCIp, which can be explained by a more refined estimation of the true change in the numerator. It produced considerably less reliable change than the original RCIsrb, which can be explained by a standard error that is possibly too small or a more biased estimation of the true change used in the latter method. The differences in results produced by the two options for the adjusted RCIsrb (variants (c) and (d) introduced above) were not spectacular, except that variant (c) was a little more conservative in designating people as being reliably changed. $\sigma_{\rm ED}$ has been shown¹² to be a theoretically safe upper limit for the standard error associated with the adjusted RCIsrb, provided that the control group is not too small. In practice, the latter standard error is usually smaller than s.e.(3), which is also shown in table 2. Therefore, ultimately, sport concussion researchers may find variant (d)—that is, expression 6 or 7 combined with their associated SEP calculated as the standard deviation of expression 6 or 7 in the control grouppreferable. However, the caution provided by variant (d) can also be realised by raising the alpha value (thus narrowing the confidence interval) in variant (c).

In sports medicine, multiple regression estimation is sometimes applied^{4 5} rather than regression involving only the initial assessment of the outcome measure as predictor. We note that the impact of the outcome measure always overshadows that of other predictors. However, our recommendations can easily be extended to multiple predictor situations. Further research should examine to what extent our findings would apply to such designs.

From the above discussion, the effect of the test-reliability of the outcome measure is obvious. Fortunately, authors of sport concussion studies are aware of the problems of low test-retest reliability. Some⁵ have advocated construction and use of alternative test forms, as this is expected to improve the estimation of ρ_{XX} , for instance, by reducing the occurrence of differential effects of testing. Of further interest is the suggestion¹³ of assessing a population of healthy athletes twice at the beginning of the playing season, with an interval comparable to the period between an actual concussion and the assessment of cognitive function after concussion. The values derived from such preliminary assessments may be helpful in reliably determining deterioration of cognitive function after concussion and of eventual recovery.

For the moment, we conclude that sport concussion research is at risk, as circumstances that may result in differences often occur in that field. Test-retest reliabilities of about 0.50 for tests commonly used in sports medicine (TMT-B, VIGIL-1, Digit Span), and even as low as 0.43 (TMT-A), have been cited,⁵ but we do not know of any report of

What this study adds

- The reasoning behind Hinton-Bayre's point of view is considered
- Situations in which the regression approach is not the best approach are shown to occur often
- Adjustments to the regression approach to make it more generally preferable are proposed

eventual differences in outcomes of RCIp and RCIsrb for injured athletes. Barr and McCrea⁴ proposed the standardised assessment of concussion (SAC), which has a test-retest reliability of 0.55, but its effect is masked by an increase in variance. The symbol digit modalities test⁹ showed notable diminution of variance, but, again, eventual differences in outcomes of RCIp and RCIsrb applied to an experimental sample were not reported.

Authors' affiliations

G H Maassen, Department of Methodology and Statistics, Utrecht University, Utrecht, the Netherlands

E R Bossema, Leiden University Medical Center, Leiden, the Netherlands **N Brand**, Department of Health Psychology, Utrecht University

Competing interests: none declared

REFERENCES

- 1 Maassen GH. Reliable change assessment in sport concussion research: a comment on the proposal and reviews of Collie et al. Br J Sports Med 2005;39:483–8.
- 2 Collie A, Maruff P, McStephen M, et al. Are Reliable Change (RC) calculations appropriate for determining the extent of cognitive change in concussed athletes? Br J Sports Med 2003;37:370-6.
- 3 Hinton-Bayre A. Commentary on Maassen's "Reliable change assessment in sport concussion research". Br J Sports Med 2005;39:488-9.
- 4 Barr WB, McCrea M. Sensitivity and specificity of standardized neurocognitive testing immediately following sports concussion. J Int Neuropsychol Soc 2001;7:693–702.
- 5 Erlanger D, Feldman D, Kutner K, et al. Development and validation of a web-based neuropsychological test protocol for sports-related return-to-play decision-making. Arch Clin Neuropsychol 2003;18:293–316.
- 6 Chelune GJ, Naugle RI, Lüders H, et al. Individual change after epilepsy surgery: practice effects and base-rate information. Neuropsychology 1993;7:41–52.
- Iverson GL, Lovell MR, Collins MW. Interpreting change on ImPACT following sport concussion. *Clin Neuropsychol* 2003;17:460–7.
 McSweeny AJ, Naugle RI, Chelune GJ, et al. "T Scores for change": an
- 8 McSweeny AJ, Naugle RI, Chelune GJ, et al. "T Scores for change": ar illustration of a regression approach to depicting change in clinical neuropsychology. Clin Neuropsychol 1993;7:300–12.
- Hinton-Bayre A. Holding out for a reliable change from confusion to a solution: a comment on Maassen's "The standard error in the Jacobson and Truax Reliable Change Index." J Int Neuropsychol Soc 2004;10:894–8.
 Bossema ER, Brand N, Moll FL, et al. Does carotid endarterectomy improve
- Bossema ER, Brand N, Moll FL, et al. Does carotid endarterectomy improve cognitive functioning? J Vasc Surg 2005;41:775–81.
 Crawford JR, Howell DC. Regression equations in neuropsychology: an
- 11 Crawford JR, Howell DC. Regression equations in neuropsychology: an evaluation of statistical methods for comparing predicted and obtained scores. J Clin Exp Neuropsychol 1998;20:755–62.
- 12 Maassen GH. Principes voor de definitie van reliable change (2): reliable change indices en practice effects [Principles of defining reliable change (2): reliable change indices and practice effects. (English version is submitted and can be obtained from the first author)]. Ned Tijdschr Psychol 2003;58:69–79.
- 13 Hinton-Bayre AD, Geffen G, McFarland K. Mild head injury and speed of information processing: a prospective study of professional rugby players. J Clin Exp Neuropsychol 1997;19:275–89.