



Reproducibility of computer based neuropsychological testing among Norwegian elite football players

T M Straume-Naesheim, T E Andersen and R Bahr

Br. J. Sports Med. 2005;39:64-69
doi:10.1136/bjsem.2005.019620

Updated information and services can be found at:
http://bjsm.bmjournals.com/cgi/content/full/39/suppl_1/i64

These include:

- | | |
|-------------------------------|---|
| References | This article cites 31 articles, 13 of which can be accessed free at:
http://bjsm.bmjournals.com/cgi/content/full/39/suppl_1/i64#BIBL |
| | 1 online articles that cite this article can be accessed at:
http://bjsm.bmjournals.com/cgi/content/full/39/suppl_1/i64#otherarticles |
| Rapid responses | You can respond to this article at:
http://bjsm.bmjournals.com/cgi/eletter-submit/39/suppl_1/i64 |
| Email alerting service | Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article |
-
- Topic collections** Articles on similar topics can be found in the following collections
[Sports Medicine](#) (1079 articles)

Notes

To order reprints of this article go to:
<http://www.bmjournals.com/cgi/reprintform>

To subscribe to *British Journal of Sports Medicine* go to:
<http://www.bmjournals.com/subscriptions/>

SUPPLEMENT

Reproducibility of computer based neuropsychological testing among Norwegian elite football players

T M Straume-Naesheim, T E Andersen, R Bahr

Br J Sports Med 2005;39(Suppl 1):i64-i69. doi: 10.1136/bjsm.2005.019620

Background: Head injuries account for 4–22% of all football injuries. The rate of brain injuries is difficult to assess, due to the problem of defining and grading concussion. Thus computerised testing programs for cognitive function have been developed.

Objective: To assess the reliability of a computerised neuropsychological test battery (CogSport) among Norwegian professional football players.

Methods: Norwegian professional football league players (90.3% participation) performed two consecutive baseline CogSport tests before the 2004 season. CogSport consists of seven different subtasks: simple reaction time (SRT), choice reaction time (ChRT), congruent reaction time (CgRT), monitoring (MON), one-back (OBK), matching (Match) and learning (Learn).

Results: There was a small but significant improvement from repeated testing for the reaction time measurements of all seven subtasks (SRT: 0.7%, ChRT: 0.4%, CgRT: 1.2%, MON: 1.3%, OBK: 2.7%, Match: 2.0%, Learn: 1.1%). The coefficient of variation (CV) ranged from 1.0% to 2.7%; corresponding intraclass correlation coefficients ranged from 0.45 (0.34 to 0.55) to 0.79 (0.74 to 0.84). The standard deviation data showed higher CVs, ranging from 3.7% (Learn) to 14.2% (SRT). Thus, the variance decreased with increasing complexity of the task. The accuracy data displayed uniformly high CV (10.4–12.2) and corresponding low intraclass correlation coefficient (0.14 (0.01 to 0.26) to 0.31 (0.19 to 0.42)).

Conclusion: The reproducibility for the mean reaction time measures was excellent, but less good for measures of accuracy and consistency. Consecutive testing revealed a slight learning effect from test 1 to test 2, and double baseline testing is recommended to minimise this effect.

See end of article for authors' affiliations

Correspondence to:
T M Straume-Naesheim,
Oslo Sports Trauma and
Research Center, Oslo,
Norway;
truls.straume-nesheim@
nih.no

Football is the only contact sport that exposes a large number of participants to purposeful use of the head for controlling and advancing the ball.¹ Based on a series of cross-sectional studies using neurological examinations, neuropsychological tests, computed tomography scans and electroencephalographic examinations on active and older retired Norwegian football players, Tysvaer² postulated that heading the ball could lead to chronic brain injuries as seen in boxing. Since then, several cross-sectional studies have indicated that football can cause sustained measurable brain impairment,^{3–6} although not all studies have reported such a relation.^{7–8}

Head injuries account for 4–22% of all football injuries² with a reported incidence during matches of 1.7 injuries per 1000 player hours.⁹ However, this figure incorporates all types of head injury, including facial fractures, concussions, lacerations, and eye injuries. The incidence of concussion has been estimated to be 0.5 injuries per 1000 match hours⁹ but is difficult to assess, due to the problem of defining and grading concussions.^{1–10} When using the traditional diagnostic criteria for concussions, which require loss of consciousness or amnesia, only a fraction of these are recognised as concussions. Trauma to the neck and/or head that is sufficient to cause facial fractures or lacerations, will potentially also cause damage to the brain, although this is easily overlooked because of the more visible injuries. Although most athletes with head injuries recover uneventfully following a single episode of concussion, repetitive mild head trauma may be implicated in the development of cumulative cognitive deterioration.¹ Accurate monitoring of symptom resolution and cognitive recovery is therefore important to ensure the athlete's safety and indicate whether the player should return to play or not.

The change of paradigm in the diagnosis and management of concussion has evoked the need for new diagnostic instruments within sports related head injuries. One item in such tests is deterioration in cognitive test performance.¹¹ In the sports arena, changes in cognition following a concussion injury are conventionally determined by administering a battery of neuropsychological tests during the pre-season to establish a baseline for comparison after an injury. In studies using such a design, any changes from baseline are considered to be a consequence of the concussion injury.

In the past decade, computerised cognitive function testing programs have been developed—for example, CogSport (CogState Ltd, Melbourne, Australia), ImPACT (ImPACT Inc., Pittsburgh, PA), ANAM (Automated Neuropsychological Assessment Metrics; developed by the US Department of Defense), CRI by HEADMINDER (concussion resolution index; Headminder Inc., New York). The conventional paper and pencil tests were designed primarily for assessment of cognitive dysfunction caused by neuronal or psychiatric disorders and not for the assessment of mild changes in cognitive function over time.¹² Therefore, these tests often have poor psychometric properties for serial studies including a limited range of possible scores, floor and ceiling effect(s), learning effects, and poor test–retest reliability.^{13–14} Computerised testing using infinitely variable test paradigms may overcome these concerns.¹⁵

Makdissi *et al*¹⁶ compared the sensitivity of the CogSport test and conventional paper and pencil tests to detect cognitive changes following mild concussion in a cohort of elite players from the Australian Football League by comparing baseline tests with post-injury tests. Their data suggested that computerised tests may be particularly sensitive to the cognitive consequences of sports related concussions, and

Table 1 Description of the seven CogSport subtasks and their assumed corresponding cognitive function

Test	Description	Cognitive function
Simple reaction time	A single card was presented face down in the centre of the screen. The was player instructed to press "yes" whenever the card turned face-up. Fifteen trials were presented and the test was repeated three times; at the beginning, in the middle, and at the end of the battery. All other tests were presented just once	Motor function
Choice reaction time	This test used the same stimuli as above, but the player was now instructed to indicate whether the card was red by pressing "yes" or black "no"	Decision making
Congruent reaction time	Two cards were presented and the player had to indicate if they were same colour or not by pressing "yes" or "no"	Simple attention
Monitoring	Five cards moved simultaneously on the screen and the player was instructed to press "yes" as soon as one card moved outside a predefined area	Divided attention
One-back	The player was instructed to indicate whether a new card was identical to or different from the last by pressing "yes" or "no"	Working memory
Matching	Six card pairs were presented at the top of the screen and the player had to decide whether a pair presented at the bottom of the screen matched any of the above	Complex attention
Learning	<i>Incidental learning:</i> this followed immediately after the matching task, and was identical to that task except that the six pairs were turned face down <i>Associate learning:</i> similar to the matching task, however, the pairs were turned face down when the player correctly indicated a matching pair presented at the bottom of the screen	Learning and memory

also that conventional neuropsychological tests do not show this sensitivity in athletes with mild concussion. Similar findings have recently been reported in studies on high school athletes with head injuries using ImPACT.¹⁷⁻¹⁹ Computer based cognitive tests have many advantages over paper and pencil tests that may allow them to detect subtle impairments such as those expected to occur in mildly concussed athletes.²⁰ In general, repeated tests of healthy adults in different age groups have shown that computer based tests are reliable²⁰⁻²¹; although there is a learning effect between test 1 and 2, this effect seems to decrease after the first two tests.²²

The test properties of the CogSport test, a computer based neuropsychological evaluation tool widely used in football concussion management, have not been assessed by independent researchers, nor has it been examined among elite athletes. Therefore, the objective of this study was to evaluate the CogSport test by investigating the reproducibility of two consecutive baseline tests in a cohort of elite football players.

METHODS

We invited all the 14 clubs of the Norwegian professional male football league (Tippeligaen) with their A-squad contract players (about 300) to participate in the study; 289 players (90%) agreed to take part. Written informed consent was obtained from all participants and the project was approved by the Regional Ethics Committee for Southern Norway.

Neuropsychological testing

The teams meet for the La Manga Cup and pre-season training camp in February/March every year at the Norwegian Football Association training centre in La Manga, Spain. We conducted testing among 13 of the 14 teams of Tippeligaen at La Manga prior to the 2004 season in a test lab set up in the residential complex, Los Lomas II. Trained personnel administered and supervised the neuropsychological testing, and the tests were completed by the players in groups of three in the same quiet room to allow efficient data collection. The last team was tested at its home field in Norway two weeks later under similar standardised conditions. There is no time difference between Spain and Norway and the testing was performed at the same time of day with the same person instructing and supervising the test for each team.

We used the computer based neuropsychological test CogSport (versions 2.2.0 and 2.2.1). Norwegian speaking players were tested with the Norwegian language version of the test, where instructions for each subtask were in Norwegian, and the rest of the players used the English language version. The test has been described in detail elsewhere.¹³⁻²³⁻²⁴ The stimulus for all tasks consists of playing cards and responses are given using the keyboard. The **d** key indicates "no" and the **k** key "yes" (vice versa for left handed players). These are the only keys used throughout the whole test.

The CogSport test battery includes seven subtasks testing different cognitive brain functions (table 1). All subtasks

Table 2 Comparison between the results from test 1 and test 2 for the main CogSport outcome measures

Subtask	Mean difference (95% CI)	Improvement (%)	p
Simple reaction time	-0.016 (-0.020 to -0.012)	-0.7	<0.001
Choice reaction time	-0.009 (-0.0158 to -0.003)	-0.4	0.004
Congruent reaction time	-0.034 (-0.040 to -0.027)	-1.2	<0.001
Monitoring, reaction time	-0.031 (-0.043 to -0.0194)	-1.3	<0.001
One-back, reaction time	-0.074 (-0.083 to -0.066)	-2.7	<0.001
Matching, reaction time	-0.062 (-0.072 to -0.0519)	-2.0	<0.001
Learning, reaction time	-0.032 (-0.039 to -0.025)	-1.1	<0.001
One-back, accuracy	0.066 (0.036 to 0.096)	5.0	<0.001
Matching, accuracy	-0.014 (-0.044 to 0.016)	-1.1	0.300
Learning, accuracy	0.039 (0.019 to 0.058)	3.8	0.001

Data are the mean difference of the \log_{10} of the reaction times (ms) and arcsine of the per cent correct responses for the accuracy data for test 1 and 2 with the corresponding 95% confidence intervals (CI) in parenthesis. Test results were compared using paired Student's *t* test.

include between 15 and 40 trials, and the data are reported by the CogSport program as mean reaction times with corresponding standard deviations for all subtasks, accompanied by accuracy data for all tasks except simple reaction time and monitoring. Anticipatory responses (reaction times <100 ms) and abnormally slow responses (reaction times >3500 ms) are recorded as errors and excluded from the analyses. Accuracy data are calculated as the number of true positive responses divided by the number of trials.

The computer program sends a test report by e-mail to the test supervisor with basic analyses of the result. In addition, the test report includes an estimate of whether the player's performance meets the minimum requirements of the test with regard to alertness throughout the test and the plausibility of whether they understood the instructions or not. This built-in decision is based on the variability of performance on simple reaction time and a threshold value for accuracy on the three final tasks. If a player has more than 40 incorrect responses on one task, the test is stopped.

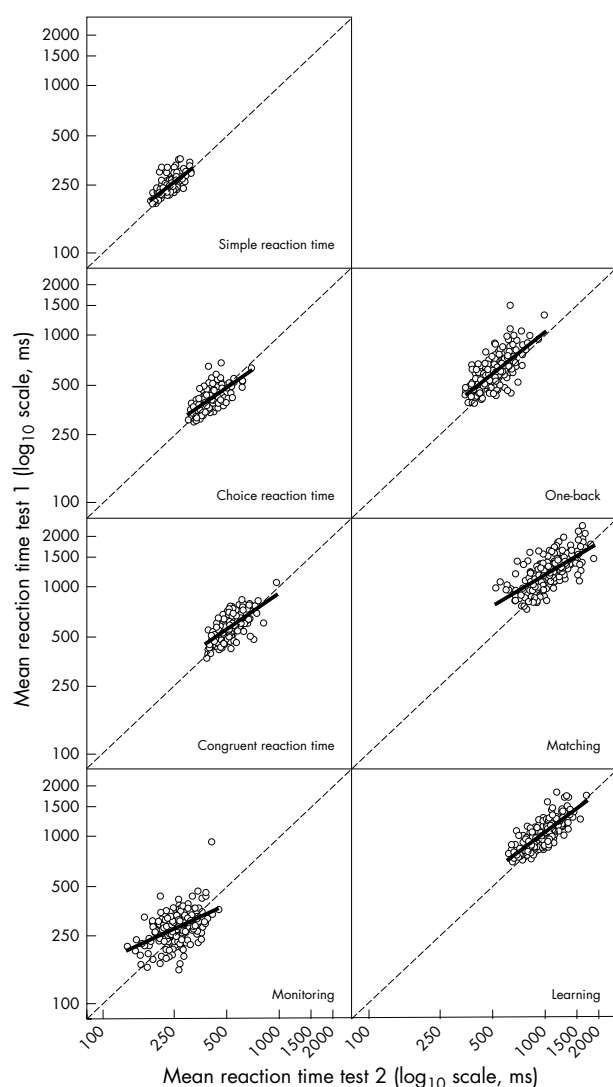


Figure 1 Reproducibility of mean reaction time (\log_{10} , ms) for five CogSport subtasks; test 1 plotted against test 2 ($n=232$). The hatched line is the identity line ($x=y$). Regression lines (dotted) have been added to illustrate whether there were systematic differences between test 1 and test 2. The subtasks are arranged vertically and from left to right according to their complexity from top left (easiest) to bottom right (most difficult).

Data analysis

Reliability and correlation studies of CogSport on young adults recommend the mean reaction time for all seven subtasks and accuracy data from the three final tasks (one-back, matching, learning) as the main outcome measures.²⁰ Our data analysis therefore focused on these 10 measures. Before all calculations, the mean reaction times and standard deviation data were \log_{10} transformed and the accuracy data were arcsine transformed to obtain a more normal distribution.²⁰

Reproducibility analyses were performed using the method error (ME), calculated as the standard deviation (SD) of the mean difference between test 1 and 2 divided by the square root of the number of tests performed: $ME = SD_{\text{mean diff}}/\sqrt{2}$.²⁵ From the ME we calculated the coefficient of variation (CV), which quantifies the variation between each measurement as a percentage of the joined mean: $CV = ME / [(X_1\text{mean} + X_2\text{mean})/2]$. These calculations were done for all outcome measures supplied by the test. We also calculated the intraclass correlation coefficient for the same measures. The intraclass correlation coefficient is defined as the ratio of the "true" variance, or the variance between subjects (S^2_b), relative to the total variance given by the variance between subjects adding the variance within subjects (S^2_w).²⁶ The intraclass correlation coefficient ranges from 0 to 1, and from the equation in its simplest form ($S^2_b / (S^2_b + S^2_w)$), we see that when the variation within the subjects (that is, a player's test score on two consecutive tests) moves towards 0, the intraclass correlation coefficient approaches 1 indicating good reproducibility.

We used SPSS version 11 for the statistical analyses and its two way random single measure model for calculating the intraclass correlation coefficients. Paired Student's *t* tests were used to investigate significant differences and any directional trends between the groups.

RESULTS

Demographics

Of the 289 players (96.3%) who agreed to participate in the study, 18 did not report for testing, leaving us with 271 (90.3%) players who underwent two consecutive neuropsychological tests. However, due to technical problems with some tests (unrelated to test performance), the number of players with dual tests decreased to 247. In addition, 15 tests did not fulfil the minimum requirements set by the computer program and therefore could not be included in the analyses. Thus, a total of 232 players (83% Norwegians, 8% Scandinavian (with no problems in understanding Norwegian), and 9% from other countries (mainly European)) were included in the study. The mean (SD) age of the investigated group was 25.7 (4.6) years (range 17–35); 87.5% were right and 12.5% left handed; 62.9% had completed secondary education (that is, high school), and 36.6% had a tertiary level of education (that is, college or beyond). The demographic characteristics of excluded group did not differ significantly in any way from the included group.

Reproducibility

There was a significant improvement in the CogSport subtasks from test 1 to test 2, ranging between 0.4% and 2.7% for the \log_{10} transformed reaction time measures (table 2, fig 1). The improvement in reaction time was slightly higher for the more complex tasks compared with the simpler ones (table 2). The accuracy data for the more complex subtasks (one-back and learning) also indicated a better performance (higher percentage of correct responses) in test 2, except for matching (table 2, fig 2).

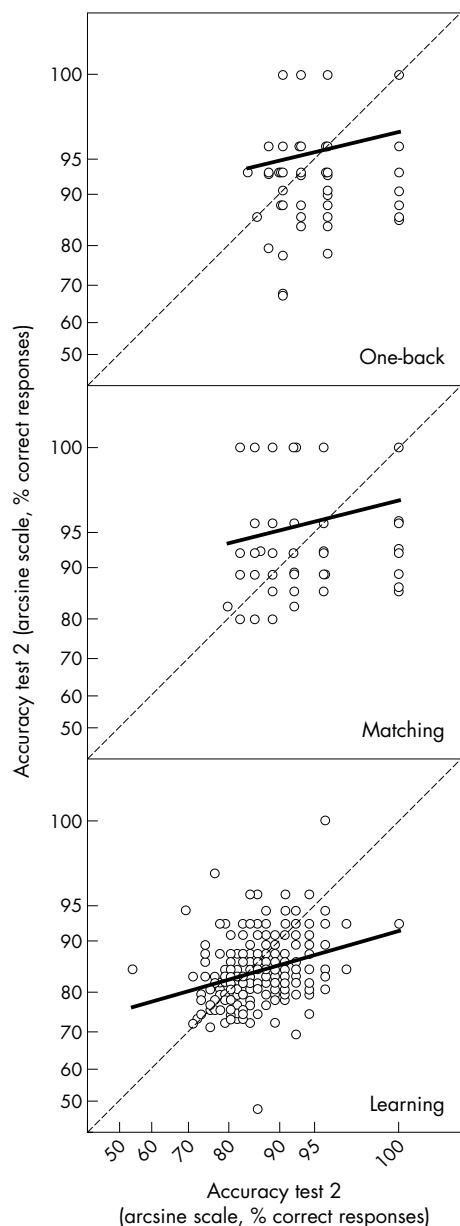


Figure 2 Reproducibility of accuracy (arcsine of % correct responses) for three CogSport subtasks; test 1 plotted against test 2 ($n=232$). See fig 1 for further details.

The reproducibility tests resulted in a CV ranging from 1.0% to 2.7% for the reaction time measures (table 3). A closer look at fig 1 reveals higher variability for subjects with slower reaction times and Bland-Altman plots (not shown) were used to examine this phenomenon more closely. They uniformly indicated a somewhat increasing difference in favour of test 2 with increasing reaction time. Thus, a poor performance on test 1 indicated a larger improvement on test 2. The intraclass correlation coefficients were also generally high for the reaction time measurements. All but one task, monitoring (0.45 (0.34 to 0.55)), resulted in intraclass correlation coefficients above 0.65 (up to 0.79 for the most complex task, learning, thus indicating good reproducibility; table 3).

The accuracy data for the three more complex tasks, one-back, matching, and learning, showed poorer reproducibility. The CV ranged from 10.4% to 12.4% and the intraclass correlation coefficient from 0.31 to 0.14 (table 3).

Table 3 Reproducibility reported as the coefficient of variation and the intraclass correlation coefficient between test 1 and test 2 for the seven CogSport subtasks

Subtask	Coefficient of variation (%)	Intraclass correlation coefficient
Mean reaction time		
Simple reaction time	1.0	0.73 (0.67 to 0.79)
Choice reaction time	1.4	0.65 (0.57 to 0.72)
Congruent reaction time	1.4	0.69 (0.61 to 0.75)
Monitoring	2.7	0.45 (0.34 to 0.55)
One-back	1.8	0.71 (0.64 to 0.77)
Matching	1.8	0.69 (0.61 to 0.75)
Learning	1.3	0.79 (0.74 to 0.84)
Standard deviation		
Simple reaction time	14.2	0.12 (-0.01 to 0.24)
Choice reaction time	9.2	0.39 (0.28 to 0.49)
Congruent reaction time	7.0	0.35 (0.23 to 0.45)
Monitoring	6.5	0.32 (0.20 to 0.43)
One-back	8.8	0.37 (0.25 to 0.47)
Matching	4.7	0.38 (0.27 to 0.49)
Learning	3.7	0.61 (0.52 to 0.69)
Accuracy		
Choice reaction time	11.4	0.14 (0.01 to 0.26)
Congruent reaction time	11.1	0.23 (0.11 to 0.35)
One-back	12.2	0.21 (0.09 to 0.33)
Matching	12.4	0.24 (0.12 to 0.36)
Learning	10.4	0.31 (0.19 to 0.42)

For each test, reproducibility results are shown for the mean and the corresponding standard deviation. Accuracy data are shown for the tasks requiring a 'yes' or 'no' response.

Additionally, as indicated in fig 2, one-back and matching tasks suffered from a ceiling effect with many participants managing 100% correct responses.

Measures of consistency, as given by the standard deviations of the mean reaction times for each subtask, were subject to greater variability than the mean result and inversely related to the complexity of the task. The CV for the standard deviation ranged from 14.2% for simple reaction time to 3.7% for learning, the most complex task (table 3). In the same way, the corresponding intraclass correlation coefficients increased with increasing task complexity, ranging from 0.12 for simple reaction time to 0.61 for learning (table 3).

DISCUSSION

This is the first study to examine the test properties of a computer based neuropsychological test battery performed by an independent research group. The main finding was that the day to day reproducibility for the mean reaction time measures was excellent in a large cohort of professional football players, but that the accuracy and consistency measures were less reliable. We also observed a slight learning effect from the first to the second test. Thus our results are in accordance with those of recent studies examining the reliability of computerised neuropsychological tests among healthy young adults and elderly people.²⁰ Collie *et al* assessed the reliability of CogSport by serial testing at a one hour and a one week interval 60 young volunteers recruited through advertisements around university campuses in Melbourne, Australia.²⁰ Elite athletes are select individuals, who may differ from this group in many different ways, including background characteristics such as education level and socioeconomic status. However, even more important is that superior neurocognitive skills may be one of the selection criteria to become an elite footballer. In fact, a closer look at the reaction time data of Collie *et al*'s 60 volunteers reveals that they were considerably slower than the footballers on all subtasks. The reproducibility of the CogSport test on elite athletes has not been thoroughly investigated before. The apparent difference between regular

What is already known on this topic

- Computerised neuropsychological testing programs have been proved to be sensitive and reliable in the evaluation of cognitive function after concussions in sport
- Dual baseline testing is recommended to minimise learning effects

controls and elite athletes illustrates the need to develop appropriate reference data in populations of elite athletes, and supports the practice of individual baseline testing in the elite as a basis for the management of concussion.

The CV ranged from 1.0% to 2.7% for the mean reaction time measures and all values under 5% must be considered as good. Collie *et al.*²⁰ in their study on 60 healthy non-athletic young volunteers, reported intraclass correlation coefficients for the reaction time measurement higher than 0.69 for all of the four tested subtasks. Except for simple reaction time, the results were similar when comparing the test–retest results with both the one hour and the one week interval between the tests. The intraclass correlation coefficient for the mean reaction time measures from our material were within the same range. Reaction time measures have been shown to provide the most sensitive index of cognitive changes following a head injury,²⁷ which in part is due to the fact that they are highly reproducible, as indicated in both our study and previous studies on other study populations.^{16 20 22 28} In contrast, of the consistency measures only the standard deviations for the most complex tasks (matching and learning) were within this limit. Although there was a uniform trend of less variation on test 2, the reproducibility data imply that these measures are unlikely to be helpful for follow up evaluations. The simpler tasks were the least consistent and one may speculate that the lack of complexity in these tasks causes the player to lose focus during the task. In the CogSport testing program, simple reaction time testing is repeated three times during the session, which may exaggerate this effect.

In our study, the accuracy data showed inadequate reproducibility and the highest improvement from test 1 to test 2. The ceiling effect found on both one-back and matching may also make these less suitable as outcome measures, even with a dual baseline setting. Previous analyses using this computerised battery have shown ceiling effects for all accuracy data except matching and learning,²⁹ but our results indicate that this is also the case for matching.

It should be noted that intraclass correlation coefficients must be interpreted with caution. From the simplified equation for the intraclass correlation coefficient ($S^2_b / (S^2_b + S^2_w)$) it is evident that data of a homogeneous group (that is, where the between-subjects variability (S^2_b) is small compared with the within-subjects variability (S^2_w)) will produce a poorer intraclass correlation coefficient than data of a heterogeneous group (that is, with high between-subject variability with respect to the within-subject variability), even if within-subject variability is exactly the same for the two groups. It is therefore recommended not to compare directly the intraclass correlation coefficients from different study populations without knowing the variance within the tested groups.³⁰ We have therefore, as recommended,²⁵ also presented the test–retest coefficients of variation, which are independent of test result range and therefore can be compared directly between studies. It should be noted that, compared with the performance data reported by Collie *et al.*,²⁰

What this study adds

- The computerised test battery (CogSport) showed excellent reproducibility in a large cohort of professional Norwegian football players using a translated version of the test
- The reaction time measures proved to be the most reliable for all subtasks tested, and these are therefore recommended as primary outcome measures

our footballers displayed both faster mean reaction times and a more homogeneous performance. When this is taken into consideration, a comparison of the test–retest intraclass correlation coefficients with Collie *et al.* indicates that the reproducibility of the mean reaction time measures may be even better among elite footballers than non-athletic controls. In a one year follow up of 84 elite Australian Rules footballers, the test–retest coefficients of variation were not reported.³¹

In our group of professional football players, there was a significant improvement from test 1 to test 2 for the mean reaction time measures on all subtasks of CogSport. Collie *et al.* found a similar practise effect when a group of elderly volunteers (mean age 64 (8) years) performed four consecutive CogSportTM tests in three hours.²² Whereas our professional football players tended to display a more pronounced practise effect when the tasks became more complicated, Collie *et al.*'s elderly volunteers showed an opposite trend. More relevant is a comparison with elite Australian Rules footballers, and, as mentioned above, 84 of these were tested after an injury-free season (the exact timeframe was not stated) without displaying any significant differences in performance since baseline for any of the subtasks of CogSport (for the final two tasks, matching and learning, accuracy data were presented instead of mean reaction times). A practise test was conducted before the baseline test, but it is not clear if this was done for the follow-up as well (either in full or shortened).

Since we performed only two tests, we are not in a position to say whether the practise effect will decrease with further testing. However, Falleti *et al.* followed 26 young volunteers who performed four different baseline CogSport tests on three different days, where the first two tests were performed on the same day with a two hour break in between, and the first was discarded. In the three remaining baseline tests (time intervals not stated) there were no differences in performance on reaction time or accuracy measures.²³

Another aspect of the mean reaction time measurements, which became evident on Bland–Altman plots, was that the improvement from test 1 to 2 was not evenly distributed. The players with the slowest mean reaction times improved the most, and on some subtasks those with the fastest mean reaction times were actually slower on test 2. Such regression towards the mean has also been described by Erlanger on simple and choice reaction time measures from a similar computerised neuropsychological test package from HEADMINDER.³²

Due to the practise effect, we agree with previous studies conducted on other populations that the test requires a dual baseline, where the first test is discarded.^{22 23} Whether this procedure should be used in follow up testing if more than a couple of weeks have passed since the baseline testing, needs further investigation. One problem with a dual baseline tests is that the test becomes more time consuming and there is a risk is that the player will lose their focus. A large study of patient with head injuries found that effort explained 53% of

the variance in neuropsychological test performance (in comparison, educational level accounted for only 11% and age only 4%).³³ It has to be noted that the group used the old paper and pencil test, which implies a different testing setting. The results can therefore not be transferred directly to computerised testing. Nevertheless, the issue of including some kind of effort measure when conducting a neuropsychological test was recently stressed at the Second International Symposium on Concussion in Sport in Prague.³⁴

In conclusion, the reproducibility for the mean reaction time measures was excellent in the cohort on professional footballers included in the present study. However, the accuracy and consistency measures were less reliable, and may therefore be less sensitive as outcome measures in post-concussion management. Consecutive testing revealed a slight learning effect from test 1 to test 2, and dual baseline testing with rejection of the first test is recommended to minimise this effect.

ACKNOWLEDGEMENTS

This study was paid for by a grant from FIFA. In addition, financial support came from the Oslo Sports Trauma Research Center, which has been established at the Norwegian University of Sport and Physical Education through generous grants from the Eastern Norway Regional Health Authority, the Royal Norwegian Ministry of Culture and Church Affairs, the Norwegian Olympic Committee and Confederation of Sports, Norsk Tipping AS, and Pfizer AS. CogState Ltd. provided the necessary software and technical support free of charge. A special thanks to Jiri Dvorak and Astrid Junge from the FIFA-Medical Assessment and Research Centre (F-MARC) for their collaboration on developing the study protocol and Alex Collie for technical support. The authors thank Jostein and Grete Steene-Johannessen for test supervision, Ingar Holme and Lars Bo Anderson for statistical assistance, and the players, team physicians, physiotherapists, and coaches for their cooperation.

Authors' affiliations

T M Straume-Naesheim, T E Andersen, Oslo Sports Trauma and Research Center, Oslo, Norway

R Bahr, Norwegian University of Sport and Physical Education, Oslo, Norway

Competing interests: none declared

REFERENCES

- Kirkendall DT, Jordan SE, Garrett WE. Heading and head injuries in soccer. *Sports Med* 2001;**31**:369–86.
- Tysvaer AT. Head and neck injuries in soccer. Impact of minor trauma. *Sports Med* 1992;**14**:200–13.
- Downs DS, Abwender D. Neuropsychological impairment in soccer athletes. *J Sports Med Phys Fitness* 2002;**42**:103–7.
- Matser JT, Kessels AG, Jordan BD, et al. Chronic traumatic brain injury in professional soccer players. *Neurology* 1998;**51**:791–6.
- Matser EJ, Kessels AG, Lezak MD, et al. Neuropsychological impairment in amateur soccer players. *JAMA* 1999;**282**:971–3.
- Matser JT, Kessels AG, Lezak MD, et al. A dose-response relation of headers and concussions with cognitive impairment in professional soccer players. *J Clin Exp Neuropsychol* 2001;**23**:770–4.
- Guskiewicz KM, Marshall SW, Broglio SP, et al. No evidence of impaired neurocognitive performance in collegiate soccer players. *Am J Sports Med* 2002;**30**:157–62.

- Rutherford A, Stephens R, Potter D. The neuropsychology of heading and head trauma in Association Football (soccer): a review. *Neuropsychol Rev* 2003;**13**:153–79.
- Andersen TE, Arnason A, Engebretsen L, et al. Mechanisms of head injuries in elite football. *Br J Sports Med* 2004;**38**:690–6.
- Delaney JS, Lacroix VJ, Leclerc S, et al. Concussions among university football and soccer players. *Clin J Sport Med* 2002;**12**:331–8.
- Echemendia RJ, Putukian M, Mackin RS, et al. Neuropsychological test performance prior to and following sports-related mild traumatic brain injury. *Clin J Sport Med* 2001;**11**:23–31.
- McSweeney A, Naugle R, Chelune G. "T Scores for change": an illustration of a regression approach to depicting in clinical neuropsychology. *Clin Neuropsychol* 1993;**7**:300–12.
- Collie A, Darby D, Maruff P. Computerised cognitive assessment of athletes with sports related head injury. *Br J Sports Med* 2001;**35**:297–302.
- Macciocchi SN. "Practice makes perfect": retest effects in college athletes. *J Clin Psychol* 1990;**46**:628–31.
- Aubry M, Cantu R, Dvorak J, et al. Summary and agreement statement of the First International Conference on Concussion in Sport, Vienna 2001. Recommendations for the improvement of safety and health of athletes who may suffer concussive injuries. *Br J Sports Med* 2002;**36**:6–10.
- Makdissi M, Collie A, Maruff P, et al. Computerised cognitive assessment of concussed Australian Rules footballers. *Br J Sports Med* 2001;**35**:354–60.
- Collins MW, Field M, Lovell MR, et al. Relationship between postconcussion headache and neuropsychological test performance in high school athletes. *Am J Sports Med* 2003;**31**:168–73.
- Lovell MR, Collins MW, Iverson GL, et al. Grade 1 or "ding" concussions in high school athletes. *Am J Sports Med* 2004;**32**:47–54.
- Lovell MR, Collins MW, Iverson GL, et al. Recovery from mild concussion in high school athletes. *J Neurosurg* 2003;**98**:296–301.
- Collie A, Maruff P, Makdissi M, et al. CogSport: reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clin J Sport Med* 2003;**13**:28–32.
- Iverson GL, Lovell MR, Collins MW. Interpreting change on IMPACT following sport concussion. *Clin Neuropsychol* 2003;**17**:460–7.
- Collie A, Maruff P, Darby DG, et al. 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.
- Falletti MG, Maruff P, Collie A, et al. Qualitative similarities in cognitive impairment associated with 24 h of sustained wakefulness and a blood alcohol concentration of 0.05%. *J Sleep Res* 2003;**12**:265–74.
- Westerman R, Darby D, Maruff P, et al. Computerised cognitive function testing of pilots. *Aust Defense Forces Health Services J* 2001;**2**:29–36.
- Sale DG. Testing strength and power. In: MacDougall JD, Wenger HA, Green HJ, eds. *Physiological testing of the high-performance athlete*. Champaign, IL: Human Kinetics Books, 1990:21–106.
- Benestad HB, Laake P. Metode og Planlegging. In: Benestad HB, Laake P, eds. *Forskningsmetode i medisin og biotag*. Norway: Gyldendal Norsk Forlag AS, 2004:83–114.
- Stuss DT FAU, Stethem LL FAU, Hugenholtz HF, et al. Reaction time after head injury: fatigue, divided and focused attention, and consistency of performance. *J Neurol Neurosurg Psychiatry* 1989;**52**:742–8.
- Collie A, Maruff P, McStephen M, et al. Psychometric issues associated with computerised neuropsychological assessment of concussed athletes. *Br J Sports Med* 2003;**37**:556–9.
- Collie A, Maruff P, McStephen M, et al. CogSport. In: Echemendia RJ, eds. *Sports Neuropsychology: A clinical primer*. New York: Guilford Publications, 2005.
- Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994;**13**:2465–76.
- Collie A, Makdissi M, Moriarity, et al. Post-concussion cognitive function in symptomatic and asymptomatic athletes. In: Second international symposium on concussion in sport. *Br J Sports Med* 2004;**38**:254–64, Abstract no 2.
- Erlanger DM. Statistical techniques for interpreting post-concussion neuropsychological test. In: International symposium on concussion in sport. *Br J Sports Med* 2001;**35**:367, Abstract no 11.
- Green PF, Rohling ML, Lees-Haley PR, et al. Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Inj* 2001;**15**:1045–60.
- McCrory P, K Johnston, W Meeuwisse, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *Br J Sports Med* 2005;**39**(suppl 1):i78–i86.