

ORIGINAL ARTICLE

Reliability of the revised wheelchair rugby trunk impairment classification system

VC Altmann¹, BE Groen², J van Limbeek^{2,3}, YC Vanlandewijck⁴ and NLW Keijsers²

Study design: Observational, cross-sectional.

Objectives: A new classification system for trunk impairment in wheelchair rugby was introduced in 2010. It consists of 10 tests, arranged in an algorithm, to assign four different trunk scores (0, 0.5, 1.0 or 1.5) to athletes. The purpose of this study was to assess the inter-rater reliability of this classification system.

Setting: National competition for wheelchair rugby and wheelchair basketball in the Netherlands and Belgium.

Methods: Three experienced wheelchair rugby classifiers independently assigned trunk scores to wheelchair rugby and wheelchair basketball athletes in two sessions. After each session, test descriptions were adjusted. The inter-rater reliability was evaluated by determining the agreement and Fleiss Kappa.

Results: In the first session, all classifiers agreed on the trunk score in 13 out of 16 athletes; the overall Kappa was 0.76 ($P < 0.001$). The Kappa per trunk score ranged from 0.29 to 1. Four test descriptions were adjusted after the first session. In the second session, there was an agreement in trunk score between the classifiers in 15 out of 21 athletes. The overall Kappa was 0.75 ($P < 0.0001$), and the Kappa per trunk scores ranged from 0.58 to 0.92. After the second session, two test descriptions were improved.

Conclusion: The revised classification system for trunk impairment in wheelchair rugby showed a adequate inter-rater reliability for the allocation of trunk scores.

Spinal Cord (2013) 51, 913–918; doi:10.1038/sc.2013.109; published online 17 September 2013

Keywords: paralympic sport; wheelchair athletes; athletes with spinal cord injury; trunk impairment; classification

INTRODUCTION

Wheelchair rugby is a team sport, invented by athletes with tetraplegia due to spinal cord injury (SCI). They perceived an inability to compete in wheelchair basketball due to impaired arm and hand muscle power.¹ Because the International Paralympic Committee states that in paralympic sports, winning or losing the game should depend on training, talent, motivation and skills, rather than on being the more or less impaired athlete, athletes are evaluated using a sport-specific classification system.^{2,3}

Initially, almost all wheelchair rugby athletes had complete SCI, American Spinal Injury Association Impairment Scale A (AIS A) at level C5–8.⁴ As a result, they had complete paralysis of all muscles of the legs and the trunk, and partial paralysis of the arm muscles. The influence of impairment on performance was largely based on impairment in arm muscle power, which was reflected by the emphasis on arm impairment in classification.⁵ Studies about performance in sport-specific activities⁶ and performance during the game^{7–9} in relation to wheelchair rugby sport class have used the overall sport class, without any specification of the score for arm and for trunk and leg impairment. As the popularity of the sport increased, athletes with other types of eligible impairment started to compete, for example, athletes with amputations of all four limbs. In addition, athletes with SCI tended to have incomplete lesions (AIS B–D).⁴ In contrast to the first wheelchair rugby athletes, many

new athletes had limited-to-no trunk and leg impairment. These new athletes were perceived as dominant over athletes with severe trunk and leg impairment in the same sport class. Therefore, the classification system no longer seemed adequate for the changed athlete population.¹⁰

In 2009, an expert group of classifiers and athlete representatives was formed to systematically assess the influence of trunk impairment on performance in wheelchair rugby. The following sources were used: the classification database,¹¹ a review of the literature,¹² interviews with athletes and coaches, and expert opinion of the most experienced classifiers.¹³ In 2010, the expert group proposed a revised trunk impairment classification system with a maximum of 10 tests (t1, t2, t2a, t2b, t3–t8), arranged in an algorithm leading to allocation of a trunk score (0, 0.5, 1.0 or 1.5) after failure of one or more tests. Each test was described concisely so the classifiers would be able to apply them after reading the test descriptions. This proposed revision was adopted by the International Wheelchair Rugby Federation later that year. (For description of the trunk impairment classification system, see Supplementary Information.)

Although face validity for the assessment of trunk impairment was established, content validity, construct validity and reliability have to be established to guarantee an effective and evidence-based classification system. As a first step towards evidence-based classification, the purpose of this study is to assess the inter-rater reliability of the

¹Department of Rehabilitation, Sint Maartenskliniek, Nijmegen, The Netherlands; ²Department of Research, Sint Maartenskliniek, Nijmegen, The Netherlands; ³De Viersprong, Postbus 7, Halsteren, The Netherlands and ⁴Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Leuven, Belgium
Correspondence: VC Altmann, Department of Rehabilitation, Sint Maartenskliniek, PO Box 9011, 6500 GM Nijmegen, The Netherlands.
E-mail: v.altmann@maartenskliniek.nl

Received 1 April 2013; revised 25 July 2013; accepted 20 August 2013; published online 17 September 2013

revised trunk impairment classification system in an observational, cross-sectional study in two sessions. After both sessions, test descriptions were improved if necessary.

MATERIALS AND METHODS

Participants

For each of the two sessions, three internationally certified classifiers with at least 7 years of experience at an international level participated. They were recruited by sending an invitation to all 33 registered active international classifiers for wheelchair rugby fulfilling these criteria. Three classifiers with the least travel distance and costs were selected. Two of the three classifiers in the second session had not participated in session one. There was a 6 months time period between the sessions.

In the first session, athletes from wheelchair rugby from the Netherlands and Belgium with at least 1 year experience in their sport and at least 18 years of age were invited to participate. Sixteen athletes participated in the first session. Four athletes also played wheelchair basketball and 12 athletes played only wheelchair rugby. For the second session, a limited additional number of wheelchair rugby athletes from the Netherlands and Belgium with limited-to-no trunk impairment were available (only 17 active athletes with limited-to-no trunk impairment were registered in the Netherlands and Belgium). We permitted athletes with limited-to-no trunk impairment who had already participated in the first session to participate again in the second session. Because the athletes did not know the allocated trunk score of the first session and because only one out of three classifiers in the second session could probably recognise the athletes from the first session, there would be no positive influence on the results. Also, invitations were sent to wheelchair basketball athletes to increase the number of athletes without full trunk impairment. Because there is a huge overlap in sport-specific activities between wheelchair rugby and wheelchair basketball, the influence on trunk impairment score would be minimal. The second session consisted of 22 athletes: 11 who played wheelchair basketball, 8 who played wheelchair rugby and 3 who played both sports. Three athletes participated in both studies, they all had limited-to-no trunk impairment. Two of them did not have agreement on their trunk score in the first session. Nineteen athletes did not participate in the first session. Table 1 shows the health conditions of the participating athletes.

All procedures were conducted according to the Helsinki Declaration.¹⁴ All athletes signed informed consent. To secure the athletes' safety during the assessment, all classifiers had support from an assistant who was either a nationally certified classifier or a physiotherapist.

Experimental design and procedures

The participating classifiers received the algorithm and the descriptions of the tests (t1–t8) four weeks before the testing day (See Supplementary Information). The classifier who participated twice did not get any information on which changes were made and why, since the first session. The classifiers were instructed to prepare by reading and practising all 10 tests of the classification algorithm. If they had any questions, the classifiers had the

opportunity to ask the researchers (VCA, Anne Leota Hart) for additional information the weeks before testing. On the testing day, the classifiers had to make an independent decision on passing or failing every test in the algorithm without consulting other classifiers or assistants. They were instructed to mark their test decisions and final trunk score using a score sheet showing the algorithm.

All classifier–athlete interactions were filmed while testing. At the end of the testing day, the classifiers and researchers discussed the classifiers' questions and remarks about all tests. Using the videos, one of the researchers (VCA) confirmed whether the athlete was in the correct testing position, and whether the tests were performed according to the descriptions. Any inconsistencies with the descriptions were noted and summarised. Inconsistencies between the classifiers in any of the tests were evaluated, either by self-report or observed in the video analysis. If inconsistencies were based on an insufficient description of a test, descriptions were adjusted using the advice of the participating classifiers.

Statistics

The percentage agreement and Fleiss Kappa for agreement were calculated in both sessions using SPSS (version 17.0; SPSS Inc., Chicago, IL, USA). Full agreement was defined as the percentage of athletes that received the same trunk score by all three classifiers. Because in wheelchair rugby, the majority of the athletes have full trunk impairment, higher numbers of athletes with trunk score 0 were expected compared with the other trunk scores. This could potentially influence the value for Fleiss Kappa, with an overall value for all classes that would not be representative for the individual trunk scores. To make sure that the reliability was adequate for each trunk score, Fleiss Kappa's were calculated for each individual trunk score in addition to the overall Fleiss Kappa (all trunk scores together). All tests were arranged in an algorithm in which failing in one test leads to a trunk score. Therefore, differences in interpretation of one test will lead to inconsistency in trunk score. To identify tests with inconsistent interpretation by the classifiers, the number of athletes that received full agreement between the classifiers per test was evaluated.

Significance level was set at $P < 0.05$. For interpretation of the Fleiss Kappa, the guidelines of Koch and Landis¹⁵ were used. We decided that for the overall value of Fleiss Kappa, the minimum value should be 0.61 (substantial agreement), and the minimum value of Fleiss Kappa for individual trunk scores should be 0.41 (moderate agreement).

Statement of ethics

The authors certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during the course of this research (Medical Ethical Committee of the Netherlands, region Arnhem and Nijmegen registration number 2011/378).

RESULTS

First session

Table 2 shows the trunk scores per classifier for all athletes. There was full agreement between the three classifiers for the trunk scores for 13 out of 16 athletes (81%).

The overall Fleiss Kappa was 0.76 ($P < 0.001$). Fleiss Kappa for the 0 trunk score was 1, whereas Fleiss Kappa for the other trunk scores ranged between 0.29 and 0.68 (Table 3), based on a total of seven athletes with trunk score 0.5–1.5.

The agreement per test is shown in Table 4. Note that as a result of the used decision algorithm, many of the tests were performed in a limited number of athletes. The discussion with the classifiers and the analysis of video revealed that the test descriptions for tests t3 and t5 about fixation of the athlete to prevent losing balance were missing. In addition, the quality of movement in lateral flexion in test t5 was not clearly described. Furthermore, the outcome for tests t3, t4, t5 and t6 was not described in case of an asymmetry in test performance. Therefore, adjustments to the descriptions of test t3 to t6 were made after the first session for these specific issues (see Supplementary Information).

Table 1 Health conditions of participating athletes

	First session (N)	Second session (N)
Spinal cord injury (AIS A–B) ^{a,b}	9	4
Spinal cord injury (AIS C–D) ^{a,c}	3 ^d	8 ^d
Spina bifida	0	1
Neuromuscular disease	2 ^d	4 ^d
Multiple amputations	1 ^d	1 ^d
Cerebral palsy	1	0
Achondroplasia	0	1
Orthopaedic conditions	0	3

^aAIS, American Spinal Injury Association Impairment Scale.

^bA–B, motor complete.

^cC–D, motor incomplete.

^dThree athletes participated in both sessions, one with spinal cord injury AIS C, another with neuromuscular disease and the other with multiple amputations.

Table 2 Trunks scores, highest achieved test of the algorithm and agreement on trunk score in the first session

Athletes	Sport participation		Classifier 1		Classifier 2		Classifier 3 ^a		Agreement trunk score
	Rugby	Basket ball	Trunk score	Highest achieved test	Trunk score	Highest achieved test	Trunk score	Highest achieved test	
	1 ^b	X		0.5	t2	0.5	t4	0.5	
2 ^b	X	X	0.5	t4	1.5	t7	0.5	t4	No
3 ^b	X		1.5	t8	1.0	t5	0.5	t4	No
4	X	X	0.5	t1	0.5	t2b	0.5	t1	Yes
5	X		1.5	t8	1.5	t8	1.5	t8	Yes
6	X	X	0.5	t1	0.5	t4	0.5	t4	Yes
7	X	X	0.5	t1	1.0	t6	1.0	t6	No
8	X		0	None	0	None	0	None	Yes
9	X		0	None	0	None	0	None	Yes
10	X		0	None	0	None	0	None	Yes
11	X		0	t1	0	None	0	None	Yes
12	X		0	t1	0	None	0	t1	Yes
13	X		0	None	0	None	0	None	Yes
14	X		0	None	0	None	0	None	Yes
15	X		0	t1	0	None	0	None	Yes
16	X		0	t1	0	None	0	t1	Yes
Total agreement									13/16 (81%)

All other athletes only participated in session 1.

^bt = test with the test number, that is, t1 = test 1.

^aClassifier 3 also participated in session 2.

^bAthletes 1, 2 and 3 also participated both sessions and have the same number in Tables 2 and 5.

Table 3 Value for Fleiss Kappa per trunk impairment score for both sessions

Trunk impairment score	Fleiss Kappa first session	Fleiss Kappa second session
0	1	0.78
0.5	0.68	0.58
1.0	0.29	0.74
1.5	0.55	0.92

Second session

One athlete was excluded from the analysis because the athlete did not consistently perform the tests in all three classifications, which was mentioned in the discussion with the classifiers and confirmed by the video analysis. The results for the remaining 21 athletes are shown in Table 5.

All three classifiers gave the same trunk score to 15 out of 21 athletes (71%). The overall Fleiss Kappa for all classifiers for all athletes was 0.75 ($P < 0.001$). Fleiss Kappa ranged between 0.58 for trunk score 0.5 and 0.92 for trunk score 1.5. (Table 3).

Table 4 shows the number of athletes that performed a test and the number of athletes all classifiers agreed on passing or failing this test. Test t2, t2a and t2b are the least used tests (respectively 2, 1 and 2 times in the second session). Most tests had the same outcome for each athlete by all classifiers. Test t5 showed less agreement (8 out of 12 athletes).

The discussion with the classifiers after the second session revealed that it was not clear how to proceed with test t2. The classifiers suggested that tests t2a and t2b were appropriate and easy to administer to all athletes, and test t2 did not seem to be of additional value. The classifiers also experienced difficulties with the instruction of test t5 to the athlete and the interpretation of the test result. In the analysis of the videos, it appeared that the athletes' performance of test t5 improved if the classifier gave repeated instructions. However, classifiers did not consistently repeat the instructions in each athlete. Moreover, the videos showed differences in judgement between the classifiers on how strict the lateral movement should be, and their subsequent interpretation of the movement resulted in differences in judgement of passing or failing this test.

Table 4 Agreement amongst all classifiers per test

Test number	First session (N = 16)		Second session (N = 21)	
	Performed in (N) athletes	Agreement of all classifiers in (N) athletes	Performed in (N) athletes	Agreement of all classifiers in (N) athletes
t1	16	16	21	21
t2	0	Not applicable	2	2
t2a	0	Not applicable	1	0
t2b	0	Not applicable	2	2
t3	6	4	15	14
t4	6	3	12	12
t5	6	3	12	8
t6	1	1	9	9
t7	1	1	6	6
t8	1	1	6	6

^bt = test with the test number. that is, t1 = test 1.

Table 5 Trunks scores, highest achieved test of the algorithm and agreement on trunk score in the second session

Athletes	Sport participation		Classifier 4		Classifier 5		Classifier 3 ^a		Agreement trunk score
	Rugby	Basket ball	Trunk score	Highest achieved test	Trunk score	Highest achieved test	Trunk score	Highest achieved test	
1 ^b	X		0.5	t4	0.5	t4	0.5	t4	Yes
2 ^b	X	X	1.5	t7	1.5	t7	0.5	t4	No
3 ^b	X		1.5	t8	1.5	t8	1.5	t7	Yes
4		X	0	None	0	None	0	None	Yes
5	X		0	None	0	None	0	None	Yes
6		X	0.5	t4	1.0	t5	1.0	t5	No
7	X		1.5	t7	1.5	t8	1.5	t8	Yes
8		X	0.5	t1	0.5	t1	0.5	t1	Yes
9	X	X	1.5	t8	1.5	t8	1.5	t8	Yes
10		X	0.5	t1	0.5	t1	0.5	t1	Yes
11	X	X	0.5	t2b	0.5	t2b	0	t1	No
12		X	1.0	t5	1.0	t5	1.0	t5	Yes
13		X	1.5	t8	1.5	t8	1.5	t7	Yes
14		X	0	None	0	None	0	None	Yes
15		X	0.5	t4	1.0	t5	0.5	t4	No
16		X	0	None	0.5	t1	0	t1	No
17		X	1.5	t8	1.5	t8	1.5	t8	Yes
18	X		0	None	0	None	0	None	Yes
19	X		0	None	0	None	0	None	Yes
20	X		1.0	t5	1.0	t5	1.0	t5	Yes
21	X		0.5	t2	0.5	t2	0	None	No
Total agreement									15/21 (71%)

^t = test with the test number, that is, t1 = test 1.

All other athletes only participated in session 2.

^aClassifier 3 also participated in session 1.

^bAthletes 1, 2 and 3 participated in both sessions and have the same number in Tables 2 and 5.

DISCUSSION

In this study, the inter-rater reliability of the revised classification system for trunk impairment for wheelchair rugby was assessed in two sessions. In the first session, the overall Fleiss Kappa for athletes for all trunk scores was substantial. For trunk score 0, Fleiss Kappa was perfect. However, Fleiss Kappa was no more than fair for trunk score 1.0. The value for Kappa may have been negatively influenced by the low numbers of athletes who were allocated trunk class 0.5–1.5, especially the 1.0 trunk score. This prevalence problem can cause Kappa to be unrepresentatively low. Hence it was concluded that classifiers were able to consistently make the distinction between athletes with full trunk impairment and athletes with limited-to-no trunk impairment using the revised classification system. However, the trunk scores for athletes with limited-to-no trunk impairment appeared to be inconsistent due to incomplete test descriptions and a low number of athletes with trunk scores 0.5–1.5.

For the second session, the test descriptions were adjusted and wheelchair basketball athletes participated to increase the number of athletes with trunk scores 0.5–1.5. Nevertheless, the total number of participating athletes was still rather low. The improved test descriptions resulted in a large improvement in the inter-rater reliability of the trunk impairment classification system for the trunk scores 0.5–1.5, whereas the inter-rater reliability for trunk score 0 was still sufficient but decreased compared with the first session. Compared with the first session, there was a better distribution of athletes for all trunk scores.

Also, the allocation of the trunk scores per classifier showed a more balanced distribution. Based on the overall Fleiss Kappa (0.75) and Fleiss Kappa per trunk score (0.58–0.92) in relation to our preset criteria, we concluded after the second session that the new classification system for trunk impairment had adequate reliability with substantial to almost perfect agreement for trunk scores 0, 1.0 and 1.5, although there was still room for improvement. Special attention was needed for trunk score 0.5 (Fleiss Kappa 0.58, indicating moderate agreement), which had a role in all disagreements. The athlete's ability to recruit shoulder muscles enhancing performance during the tests may have had a role in the increase of disagreement between trunk score 0 and 0.5 compared with the first session. In the first session, almost all athletes with trunk score 0 had impairments in the arms. In the second session, the athletes with trunk score 0 had limited-to-no arm impairment. It is known that athletes can compensate for trunk impairment by using their shoulder muscles.^{16–18} This substitution possibly made it more difficult for classifiers to distinguish trunk impairment from compensation, which led to inconsistency in allocating trunk score 0 versus trunk score 0.5. This possible compensation should be emphasised in the description of test number t1.

Another factor in the disagreement involving the 0.5 trunk score, but between trunk score 0.5 and 1.0, were the inconsistencies in test number t5. In the analysis of the individual tests of the algorithm, high agreement was found between all classifiers in the second session in passing or failing each test, except for test t5.

A likely explanation is that lateral flexion is a movement that is not performed as an isolated movement in daily life or in wheelchair sport but always combined with rotation, flexion or extension. Therefore, test t5 was adjusted in order to be less specific for the angle of the spine but rather for lateral flexion of the trunk as a whole, without compensation by weight shifting onto one leg. After final adjustments of this test description, the reliability of the trunk impairment classification system is expected to become even higher than in the present study, especially for trunk scores 0.5 and 1.0.

To the authors' knowledge there are hardly any studies on reliability of classification systems in paralympic sports. One study was performed on reliability of classification for Nordic skiing that is largely based on trunk impairment. In this study, two classifiers allocated one of the five existing Nordic skiing classes to 33 patients of a rehabilitation centre with a variety of health conditions.¹⁹ Although the researchers found a slightly higher inter-rater reliability (overall Kappa of 0.8 based on the raw data) for the classification outcome than we did, the lowest Kappa per Nordic skiing class was much lower (Kappa \approx 0.20). However, disagreement between the raters per test was higher (23 out of 132 tests (18%) versus 6 out of 80 performed tests (7.5%) in the present study). The most important difference between the Nordic skiing study and the current study was the experience of the classifiers. The study on Nordic skiing used the two most experienced international classifiers who had been working together and using the system for a long time. In this study, three classifiers who had sparsely worked together were selected from the entire pool of more than 30 eligible classifiers. Moreover, the trunk assessment for wheelchair rugby was new and had never been used at any tournament before the study. Therefore, it is remarkable that almost the same inter-rater reliability as in the Nordic skiing research was found.

It must be noted that the protocol used in the present study was different in several aspects from the classification procedure in wheelchair rugby.⁵ Firstly, classification is performed by a panel consisting of three classifiers who are allowed to discuss test performance. In case of disagreement between panel members, there will be a vote and the final decision is made by the majority opinion. Secondly, the process of classification is finalised by observation of sport-specific activities both during off court tests and during on court observation. Based on the high inter-rater reliability found in this study, it can be expected that an athlete will receive the same score for trunk impairment, no matter which panel conducts the classification.

In addition to the inter-rater reliability, the test-retest reliability is another important criteria for an effective test for classification, but this was not examined in the present study. In classification practise, the interval between two classifications is at least 11 months.⁵ Because the number of athletes was limited, it can be expected that classifiers would remember their decision on trunk score for each athlete, especially if they performed a test and retest with a short interval between the tests. From the video analysis, it was observed that the athletes performed the same if retested on the same day by different classifiers. There did not seem to be a training or learning effect by performing the tests repeatedly. With an inter-rater reliability as high as found in the second session and a consistency in test performance found in the video analysis, it is likely that the test-retest reliability for revised trunk impairment classification system will be adequate.

Trunk impairment does not only have an important role in activities in wheelchair sport. It also has an important role in activities in daily life. As such, assessment of trunk impairment is important in rehabilitation.²⁰ For patients with SCI, the ASIA score is

the most widely used scale to assess trunk impairment.⁴ However, the ASIA score only uses impairment in sensation at the level of the trunk and does not take impairment in trunk muscle power, trunk range or coordination of trunk movements (biomechanical trunk impairment) into account. In patients with incomplete SCI (AIS B–D), sensation in the trunk is unlikely to adequately reflect trunk muscle weakness. Moreover, Bjerkefors *et al.*²¹ indicated that impairment in sensation does not reflect trunk impairment in patients with complete SCI (AIS A). A small number of scales for biomechanical trunk impairment exist,^{22–25} but their application is limited to a specific medical condition (stroke, cerebral palsy and multiple sclerosis). The current studied wheelchair rugby scale for biomechanical trunk impairment can be applied irrespective of the medical condition. Therefore, it is potentially a useful tool for assessing biomechanical trunk impairment in rehabilitation for patients with SCI, as well as for patients with other medical conditions.

Although this research on reliability is an important step towards evidence-based classification in wheelchair rugby, it should be followed by additional research to establish content validity and construct validity of the trunk classification system with regards to eligible impairment types and severity, and the contribution of this impairment to sport-specific activity limitation.² Only with this additional research, evidence-based classification as defined in the International Paralympic Committee position stand in wheelchair rugby can be achieved.

DATA ARCHIVING

There were no data to deposit.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all classifiers and athletes who participated in this study. We thank Gehandicaptensport Nederland and Hollister Nederland for contributing to the travel costs of the classifiers and the athletes.

- 1 Orr K, Malone LA. Wheelchair rugby. In: Goosey-Tolfrey Vicky *Wheelchair sport. A complete guide for athletes, coaches and teachers. Human Kinetics*, 2010, pp 151–166.
- 2 Tweedy SM, Vanlandewijck YC. International Paralympic Committee position stand- Background and scientific rationale for classification in Paralympic sport. *Br. J. Sports Med* 2011; **45**: 259–269. published online 22 Oct 2009.
- 3 Tweedy S, Howe DP. Introduction to the Paralympic movement. In: Yves C Vanlandewijck, Walther R Thompson (eds) *The Paralympic athlete*. Wiley Blackwell: Oxford, UK, 2011, pp p19–p29.
- 4 American Spinal cord Injury Association Impairment Scale. Available from http://www.asia-spinalinjury.org/elearning/ISNCSCI_Exam_Sheet_r4.pdf.
- 5 International wheelchair rugby classification manual. 3rd revised edition June 2011. Available from [www.iwrf.com/resources/iwrf_docs/IWRF_Classification_Manual_3rd_Edition_rev-2011_\(English\).pdf](http://www.iwrf.com/resources/iwrf_docs/IWRF_Classification_Manual_3rd_Edition_rev-2011_(English).pdf).
- 6 Morgulec-Adamowicz N, Kosmol A, Molik B, Yilla AB, Laskin J. Aerobic, anaerobic and skill performance with regard to classification in wheelchair rugby athletes. *Res Q Exerc Sport* 2011; **82**: 61–69.
- 7 Molik B, Lubelska E, Kosmol A, Bogdan M, Yilla AB, Hyla E. An examination of the international wheelchair rugby federation classification system utilizing parameters of offensive game efficiency. *Adapt Phys Activ Q* 2008; **25**: 335–351.
- 8 Sarro KJ, Misuta MS, Burkett B, Malone LA, Barros RM. Tracking of wheelchair rugby players in the 2008 Demolition Derby final. *J Sports Sci* 2010; **13**: 1–8.
- 9 Sporer ML, Grindle CG, Kelleher A, Teodorski EE, Cooper R, Cooper RA. Quantification of activity during wheelchair basketball and rugby at the National Veterans Wheelchair games: A pilot study. *Prosthet Orthot Int* 2009; **33**: 21–27.
- 10 Altmann VC, Hart AL, van Limbeek J, Vanlandewijck YL. Improvement of the classification system for wheelchair rugby: athlete priorities. 5th IPC Vista conference 2011: A multi-disciplinary approach to Paralympic success. Conference proceedings http://www.paralympic.org/pidas/vista/2011_08_VISTA2011_Booklet.pdf.

- 11 Altmann VC, Hart A, Parkinson A. Wheelchair rugby classification database: aid to the improvement of classification. 4th IPC Vista conference 2006. Classification: solutions for the future. Conference proceedings.
- 12 Washington L, Hart A. Influence of Trunk on Wheelchair Propulsion: Using Evidence to Guide Classification System Review. 5th IPC Vista conference 2011: A multi-disciplinary approach to Paralympic success. Conference proceedings http://www.paralympic.org/pidas/vista/2011_08_VISTA2011_Booklet.pdf.
- 13 Wu SK. Classifiers and social control in wheelchair rugby. *Kaohsiung J Med Sci* 2001; **17**: 90–98.
- 14 World medical association declaration of Helsinki, October 2008. Available from <http://www.wma.net/en/30publications/10policies/b3/17c.pdf>.
- 15 Landis JR, Koch GG. The measurements of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
- 16 Potten YJM, Seelen HAM, Drukker J, Reulen JPH, Drost MR. Postural muscle responses in the spinal cord injured persons during forward reaching. *Ergonomics* 1999; **42**: 1200–1215.
- 17 Janssen-Potten YJ, Seelen HA, Drukker J, Reulen JPH. Chair configuration and balance control in persons with spinal cord injury. *Arch Phys Med Rehabil* 2000; **81**: 401–408.
- 18 Janssen-Potten YJ, Seelen HA, Drukker J, Huson T. The effect of seat tilting on pelvic position, balance control and compensatory postural muscle use in paraplegic subjects. *Arch Phys Med Rehabil* 2001; **82**: 1393–1402.
- 19 Pernot HFM, Lannem AM, Geers RPJ, Ruijters EFG, Bloemendal M, HAM Seelen. Validity of the test-table-test for Nordic skiing for classification of paraplegic sit ski sports participants. *Spinal Cord* 2011; **49**: 935–941.
- 20 Chen CL, Yeung KT, Bih LI, Wang CH, Chien JC. The relationship between sitting stability and functional performance in patients with paraplegia. *Arch Phys Med Rehabil* 2003; **84**: 1276–1281.
- 21 Bjerkefors A, Carpenter MG, Cresswell AG, Thorstensson A. Trunk muscle activation in a person with clinically complete thoracic spinal cord injury. *J Rehabil Med* 2009; **41**: 390–392.
- 22 Fujiwara T, Liu M, Tsuji T, Sonoda S, Mizuno K, Akaboshi K *et al*. Development of a new measure to assess trunk impairment after stroke (Trunk Impairment Scale): its psychometric properties. *Am J Phys Med Rehabil* 2004; **83**: 681–688.
- 23 Heyrman L, Molenaers G, Desloovere K, Verheyden G, De Cat J, Monbaliu E *et al*. A clinical tool to measure trunk control in children with cerebral palsy: the Trunk Control Measurement Scale. *Res Dev Disabil* 2011; **32**: 2624–2635.
- 24 Verheyden G, Nieuwboer A, Mertin J, Preger R, Kiekens C, De Weerd W *et al*. The trunk impairment scale: a new tool to measure motor impairment of the trunk after stroke. *Clin Rehabil* 2004; **18**: 326–433.
- 25 Verheyden G, Nuyens G, Nieuwboer A, Van Asch P, Ketelaer P, De Weerd W *et al*. Reliability and validity of trunk assessment for people with multiple sclerosis. *Phys Ther* 2006; **86**: 66–76.

Supplementary Information accompanies this paper on the Spinal Cord website (<http://www.nature.com/sc>)