Research article

Reliability and validity of the capabilities of upper extremity test (CUE-T) in subjects with chronic spinal cord injury

Ralph J. Marino[1], Stephen B. Kern[2], Benjamin Leiby[3], Mary Schmidt-Read[4], M. J. Mulcahey[2]

[1]Department of Rehabilitation Medicine, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA, USA, [2]Department of Occupational Therapy, College of Health Professions, Thomas Jefferson University, Philadelphia, PA, USA, [3]Department of Pharmacology and Experimental Therapeutics, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA, [4]Magee Rehabilitation Hospital, Philadelphia, PA, USA

Objective: To determine the reliability and validity of the capabilities of upper extremity test (CUE-T), a measure of functional limitations, in patients with chronic tetraplegia.
Design: Repeated measures.
Setting: Outpatient rehabilitation center.
Participants: Fifty subjects (36 male/14 female) with spinal cord injury (SCI) of ≥1-year duration participated. Subjects were 17–81 years old (mean 48.1 ± 18.2); neurological levels ranged from C2 through T6, American Spinal Injury Association Impairment Scale grades A–D.
Interventions: Not applicable.
Outcome measures: Intraclass correlation coefficients (ICC), weighted kappa and repeatability values for CUE-T; Spearman correlations of CUE-T with upper extremity motor scores (UEMS), and self-care and mobility portions of the Spinal Cord Independence Measure, vIII (SCIM III).
Results: Score ranges for UEMS were 8–50, CUE-T 7–135, self-care SCIM 0–20, and mobility SCIM 0–40. The ICC values for total, right, and left side scores were excellent (0.97–0.98; 95% confidence interval 0.96–0.99). Item weighted kappa values were ≥0.60 for all but five items, four of which were right and left pronation and supination. Repeatability of total score was 10.8 points, right and left sides 6.3 and 6.1 points. Spearman correlations of the total CUE-T with the UEMS and SCIM self-care and mobility scores were 0.83, 0.70, and 0.55 respectively.
Conclusions: The CUE-T displays excellent test–retest reliability, and good–excellent correlation with impairment and capacity measures in persons with chronic SCI. After revising pronation and supination test procedures, the sensitivity to change should be determined.

Keywords: Outcomes assessment, Quadriplegia, Reproducibility of results, Spinal cord injuries

Introduction

Recent years have seen a number of interventions to restore lost neurological function after traumatic spinal cord injury (SCI) progress from the preclinical stage to the clinical trial stage.[1] The expectation that improvement will be seen in the spinal cord segments adjacent to the injury level has focused attention on recovery in the upper extremities in persons with cervical SCI.[2,3] One approach has been to determine the amount of neurological recovery typically seen after traumatic tetraplegia, and identify thresholds for recovery that can be used as outcomes in clinical trials.[4] It is acknowledged, however, that there should be functional as well as neurological improvement demonstrated before an intervention can be recommended for clinical use. This

Correspondence to: Ralph J. Marino, Thomas Jefferson University, 132 South 10th Street, 375 Main Building, Philadelphia, PA 19107, USA.
Email: ralph.marino@jefferson.edu

The work should be attributed to the Department of Rehabilitation Medicine, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA.

© The Academy of Spinal Cord Injury Professionals, Inc. 2015
DOI 10.1179/2045772314Y.0000000272

in turn has led to the development of measures to evaluate functional improvement in the upper extremities.[5–7]

In this paper, we will present the reliability and validity of the capabilities of upper extremity test (CUE-T), which assesses functional limitations in the arm and hand.[8] Functional limitations are restrictions performing generic actions that are employed to accomplish many specific activities.[9] An action such as pushing with your index finger, for example, may be used to ring a doorbell, dial a touch-tone phone, or type on a keyboard. Details of the test development and scoring have been presented previously.[7] While the ultimate purpose of the CUE-T is to assess change in functional capabilities, it is first necessary to determine whether it has good levels of reliability and agreement.[10] This is done by testing persons with stable levels of the attribute in question two or more times. If the CUE-T displays high levels of agreement, the next step will be to evaluate its sensitivity to change.

Methods

Subjects with traumatic SCI of at least 1-year duration, with neurological levels from C2-T6, American Spinal Injury Association Impairment Scale (AIS) grades A–D, and upper extremity motor score (UEMS) > 0 were recruited. We attempted to enroll subjects in blocks by level (C2–5, C6, C7, C8–T1, and T2–T6) and severity of injury (motor complete (AIS A\B and motor incomplete (AIS C\D). Target enrollment was six subjects in each of the 10 blocks for a total enrollment of 60. The purpose of the block enrollment was to ensure that subjects spanned the levels and severity of injury seen in cervical SCI. The subjects with high thoracic injuries were included to evaluate the upper range of the test; these subjects do not have upper limb weakness but could have limited trunk control that would make certain items such as "reach down" difficult.

Subjects were tested twice approximately 2 weeks apart. On the first testing session, we performed motor and sensory testing of the upper extremities according to the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI) guidelines,[11] and administered the CUE questionnaire (CUE-Q),[12] Spinal Cord Independence Measure (SCIM) III self-care and mobility subscales,[13] and the CUE-T. At the second testing session, we only administered the CUE-T. The CUE-Q was given before the CUE-T so as not to bias responses based on performance during the CUE-T. All examiners received training, but there was no requirement to keep the same examiners at both testing sessions. Thirty-six subjects had the same and 14 had different testers at the second testing session.

Outcome measures

International Standards for Neurological Classification of Spinal Cord Injury

The ISNCSCI examination is the gold standard for evaluating impairment after SCI.[14] The motor examination consists of manual muscle testing of five muscles in each extremity, each scored on a 6-point scale (0–5). We limited testing to the upper limb muscles for this study: elbow flexors, wrist extensors, elbow extensors, flexor digitorum profundus, and abductor digiti minimi.

Capabilities of upper extremity questionnaire

The CUE-Q is a 32-item questionnaire evaluating perceived difficulty completing actions using the right (15 items), left (15 items), or both (2 items) upper extremities.[12] The original version was found to have high test–retest reliability in persons with chronic tetraplegia, with an intraclass correlation coefficient (ICC) = 0.94. Items were rated on a 7-point scale, which has since been revised to a 5-point scale, from 0 = unable/complete difficulty to 4 = no difficulty, with similar reliability.[15] The CUE-Q was always administered before the CUE-T so that responses would not be influenced by performance on the test.

Capabilities of upper extremity test

The CUE-T consists of 19 tasks, 17 unilateral (tested separately on the right and left sides) and 2 bilateral, for a total of 38 items. Depending on the item, scoring is based on completion of the action, the number of repetitions of the action, or time to complete the action. Raw scores are converted to a 5-point scale (0–4) with 4 being best. Total scores are the sum of item scores; there is no item weighting. Right or left side scores can be obtained by adding the score of the unilateral items on each side.

Spinal Cord Independence Measure – III

The SCIM is a scale developed specifically for people with SCI to evaluate their performance of activities of daily living (ADLs) and to make functional assessments of this population sensitive to change. The most recent version, SCIM III, is composed of 19 items in three subscales: (1) self-care, (2) respiration and sphincter management, and (3) mobility.[16] This study utilized the self-care and mobility subscales of the SCIM III. The self-care subscale consists of six items (feeding, upper body bathing, lower body bathing, upper body dressing, lower body dressing, and grooming), with a maximum total subscale score of 20 points. The mobility subscale

Table 1 Distribution of subjects by level and ASIA impairment scale grade

AIS grade grouping	Motor levels (*n*)				
	C2–C5	C6	C7	C8–T1	T2–T6
Motor complete (AIS A–B)	2	6	4	1	7
Motor incomplete (AIS C–D)	10	5	7	4	4

consists of nine items (bed mobility; four transfer items: bed-wheelchair, wheelchair-tub/toilet, wheelchair-car, and ground-wheelchair; three mobility items: indoors, moderate distances, outdoors; and stairs). The maximum score on the mobility subscale is 40 points. The self-care subscale of the SCIM III has been used by other researchers to evaluate the functional impact of upper extremity motor improvement and to assess validity of the Graded Redefined Assessment of Strength, Sensibility and Prehension.[17,18] The mobility subscale was included to assess discriminant validity of the CUE-T.

The SCIM III is felt to be the most sensitive, reliable, and valid measure of global disability that exists for individuals with SCI.[19] On inpatients the SCIM is typically obtained by observation, but can be obtained by interview with comparable results.[20] We developed a structured questionnaire to obtain self-reported functioning in self-care and mobility.

Statistical analysis

We looked at item score distributions to evaluate ceiling and floor effects, and total score distributions to evaluate the range assessed in this study. We evaluated item agreement using the weighted kappa coefficient, with a target for kappa values of >0.6.[21] We determined test–retest reliability of the total scale and subscales using the ICC, with target values >0.94, sufficient to make a decision about individuals.[22] Bland–Altman plots, the difference in score between testing sessions for each subject against the mean score, were examined for systematic differences.[23] In addition, we calculated the standard error of measurement (SEM) and the repeatability values,[23] also referred to as the smallest real difference.[24] The SEM is defined as the square root of the within-subject variance in a one-way analyses of variance; the repeatability coefficient is $1.96 \times \sqrt{2} \times \text{SEM}$. A difference at least as large as the repeatability coefficient indicates that with 95% confidence there is a real difference between the true scores.

Construct validity was evaluated using Spearman correlation coefficients among the CUE-T, UEMS, and SCIM III self-care and mobility scores. We hypothesized that the CUE-T would be moderately to highly correlated with the UEMS, and self-care SCIM scores, and that the CUE-T would be more highly correlated to self-care SCIM scores than to mobility SCIM scores. Finally, we calculated the mean and range of UEMS and CUE-T scores by enrollment block to determine whether better scores were obtained in groups with lower and less severe injuries.

Results

Subject characteristics

Subjects consisted of 50 persons with chronic stable SCI, and a mean age of 48.1 ± 18.2 years old at testing. Ages ranged from 17 to 81 years. Thirty-six subjects were male. We were more successful with block enrollment for motor incomplete subjects than for motor complete subjects (Table 1).

Item score distribution and agreement

The median and range of scores on the various tests can be found in Table 2. Scores spanned all or most of the range of all the assessments. The distribution of item scores for the CUE-T did not reveal any floor effects, but there was a ceiling effect for push and pull items (Table 3). Scores for most items were distributed over all five values; there were only 11 out of 180 possible item scores that no subject in this sample received. Item agreement was above the weighted kappa target of 0.6 for all items except the pronation and supination actions, and the right wrist up item which just missed the target (Table 4).

Agreement for total scale scores and subscales was excellent, with ICC values ranging from 0.978 to 0.987 (Table 5). The mean difference in total score was only 1.4 points (±5.4 points), and mean differences in subscale scores were all less than 1 point. Bland–Altman plots for the total score and right/left arms show that only a few total score differences were >10 points

Table 2 Range of scores for outcome measures used in the study

Measure (max score)	Median (IQR)	Low	High
UEMS (50)	42.5 (30–47)	8	50
CUE-T (144)	101 (66–119)	7	135
CUE-Q (128)	104 (82–120)	6	128
SCIM SC (20)	16 (10–19)	0	20
SCIM MOB (40)	22 (12–37)	0	40

Table 3 Distribution of item scores for the Capabilities of Upper Extremity test

Item	Right side score					Left side score				
	0	1	2	3	4	0	1	2	3	4
Reach forward	2	9	9	24	6	1	10	11	17	11
Reach up	5	9	13	17	6	4	8	17	16	5
Reach down	9	5	3	13	20	11	7	2	8	22
Pull	1	0	0	1	48	2	0	0	3	45
Push	1	1	0	2	46	2	0	1	3	44
Wrist up	1	11	11	17	10	2	8	12	14	14
Pronation	1	11	7	23	8	1	11	9	15	14
Supination	4	7	8	21	10	1	8	9	18	14
Grasp dynamometer	13	4	12	15	6	13	5	7	16	9
Pinch die	8	2	17	21	2	9	2	16	20	3
Pencil	13	3	5	21	8	13	2	7	23	5
Key force	13	1	4	22	10	14	0	4	21	11
Container	12	2	1	1	34	10	5	0	1	34
Manipulate	18	2	14	6	10	15	3	15	12	5
Push index	6	1	10	21	12	10	1	10	18	11
Push thumb	18	2	3	11	16	14	5	4	15	12
Acquire-release*	9	1	4	1	32	11	0	2	0	35
Lift up weight**	11	0	2	5	32					
Push down**	25	4	2	3	16					

*There were three patients on the right and two on the left with missing values.
**Bilateral items.

(Fig. 1) and only a few unilateral score differences were >5 points (Fig. 2A and B).

Repeatability values for the CUE-T total score and subscales are found in Table 5. For the right or left side, a change in score of at least 7 points would be needed to consider this a true change (with 95% confidence), and a change score of at least 11 points would be needed for the entire scale. A change of this magnitude would require improvement on at least two items for right or left side, and on at least three items for the entire scale.

Table 4 Weighted kappa values for items in the Capabilities of Upper Extremity test

Action	Right	Left
Reach forward	0.69	0.66
Reach up	0.73	0.66
Reach down	0.68	0.67
Pull weight	0.91	0.91
Push weight	0.74	0.77
Wrist up	0.59	0.62
Pronate	*0.39*	*0.41*
Supinate	*0.29*	*0.33*
Grasp dynamometer	0.80	0.85
Pinch die (2 finger)	0.68	0.81
Pencil (3-finger)	0.72	0.82
Key pinch	0.90	0.93
Wide grasp	0.90	0.94
Manipulate	0.90	0.83
Push index finger	0.76	0.79
Push with thumb	0.83	0.79
Acquire/release	0.85	0.91
	Bilateral	
Lift up weight	0.78	
Push down	0.84	

Concurrent and discriminant validity

The CUE-T displayed the expected correlations with the other scales (Table 6). The highest correlation for the CUE-T was with the UEMS and the CUE-Q (range Spearman ρ 0.78–0.83). As hypothesized, the correlation of the CUE-T with the SCIM self-care score ($\rho = 0.70$) was higher than with the SCIM mobility score ($\rho = 0.55$), supporting discriminant validity. Mean and range of CUE-T scores and UEMS by AIS and motor level group are shown in Table 7. CUE-T scores were progressively higher as motor level group descended, and subjects with motor incomplete injuries scored higher than those with motor complete injuries at the same level.

Discussion

The CUE-T has been developed to evaluate changes in functional capabilities/limitations in the upper extremities of persons with tetraplegia. As a result, the focus of test items is on the performance of a specified action, such as pushing numbers on a calculator with your index finger, rather than an activity – using a calculator. It is important how the action is accomplished, not just that the activity is completed. This focus differs from that of many ADL assessments, where the focus is on task completion and assistance needed. The score for just completing a task using an adaptive device may be lower than without a device, for example "Modified Independent" versus "Independent" levels of the FIM, but credit is given for task completion.

Table 5 Reliability and repeatability coefficients for the Capabilities of Upper Extremity test

	Mean difference (points)	SD (points)	ICC	Repeatability coefficient (points)
Total score	1.4	5.4	0.985	10.8
Right side	0.9	2.9	0.980	6.0
Left side	0.6	3.4	0.978	6.6
Right hand	0.5	1.7	0.985	3.5
Left hand	0.3	1.7	0.987	3.3

ICC, intraclass correlation coefficient.

The present study evaluates reliability and agreement of the CUE-T, a prerequisite to the determination of responsiveness. The more variability there is in scores for stable subjects the greater the change in score needed to be considered as a true change. The CUE-T has excellent test–retest reliability and agreement in persons with chronic tetraplegia. Reliability scores (ICC) for the total score and for subscales of right or left side and right or left hand were all greater than the desired value of 0.94. The reliability values of the CUE-T are comparable to measures of Impairment and Activities used to evaluate persons with SCI. Inter-rater reliability values of the sensory and motor scores of the ISNCSCI range from 0.88 to 0.97,[25,26] and values of the SCIM-III total score are between 0.91 and 0.95.[27] The repeatability coefficient of the CUE-T, reflecting the amount of change needed to exceed measurement error, was low – a change in as few as two items on a side or three items on the entire test could result in a valid change score.

For individual item agreement, weighted-kappa values for the pronation and supination items were below acceptable values. This was surprising because the test involves standard measurement of active range of motion. A review of the data sheets found several subjects where the starting point for range of motion at session 1 differed from that on session 2 by 90°. This suggests that the testers did not use standard values to indicate the start and stop range, or did not consistently rotate the wrist passively to the start position. We are revising the test procedure to standardize the recording of angles of rotation by using a protractor oriented with 0° lateral, 90° vertically up, and 180° medially.

Although the push and pull items displayed ceiling effects, we are retaining these items for now. In order for the measure to be able to detect change at the lower levels of ability, there needs to be some items that are easy for most of the intended population. We purposely limited the reliability testing to subjects who had a UEMS >0, and in fact the lowest UEMS was 8

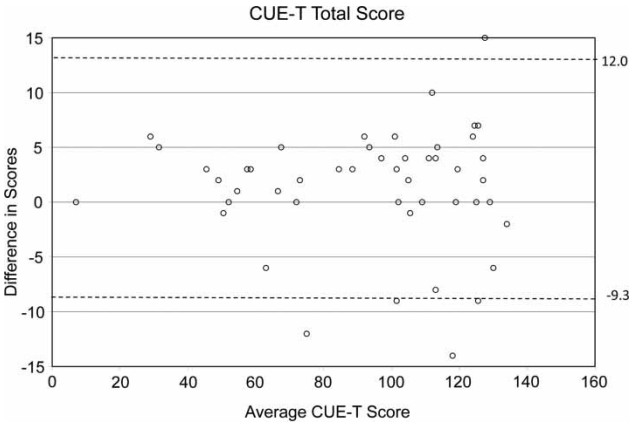

Figure 1 Bland–Altman plot for total CUE-T score, which is the individual mean scores plotted against the difference in scores. The dotted horizontal lines indicate the 95% confidence limits for repeatability.

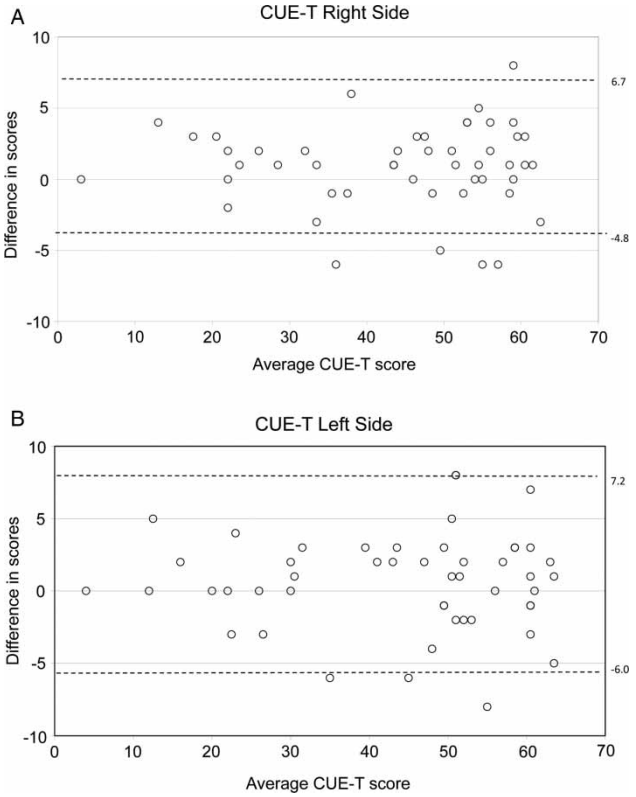

Figure 2 Bland–Altman plots for right (A) and left (B) side CUE-T scores, which are the individual mean scores plotted against the difference in scores. The dotted horizontal lines indicate the 95% confidence limits for repeatability.

Table 6 Spearman correlation coefficients of CUE-T with UEMS, CUE-Q, and SCIM

		UEMS	CUE-Q	SCIM combined	SCIM self-care	SCIM mobility
CUE-T	Total	0.827	0.822	0.617	0.695	0.550
	Right	0.808	0.814			
	Left	0.801	0.781			

UEMS, upper extremity motor score; CUE-T, Capabilities of Upper Extremity test; CUE-Q, Capabilities of Upper Extremity questionnaire; SCIM, Spinal Cord Independence Measure.

points. In the responsiveness testing phase, we will attempt to recruit subjects with UEMS closer to 0 and the potential to improve, such as persons with C4 motor levels at the time of injury and persons with high cervical incomplete injuries.

Validity of the CUE-T was supported by the expected high correlations with related measures (UEMS, CUE-Q, and SCIM self-care) and lower correlations with dissimilar measures (SCIM mobility). The progression of scores in the enrollment groups also supports validity. Subjects with motor incomplete injuries tested better than those with motor complete injuries and higher scores were achieved by subjects with lower (more caudal) motor levels of injury.

The test procedures for the CUE-T have been designed to minimize the influence of compensatory strategies on task completion, and do not permit use of adaptive equipment to perform an action. Therefore, improvement on the CUE-T should indicate an increase in ability to use the arms and hands, and reflect a decrease in an underlying impairment. It is important to also evaluate any impairments expected to change in order to understand the reason for the difference in function. During the months following SCI, recovery of motor power in the upper extremities would have the most influence on the actions measured by the CUE-T. However, other impairments can also impact functional capabilities of the upper extremity. Rotator cuff pathology, for example, could limit performance on the reaching items while finger contractures could affect the grasping items. Prior or concomitant peripheral nerve dysfunction such as brachial plexus injuries could also impede performance on testing.

Good reliability and agreement are necessary but not sufficient properties of an assessment meant to evaluate change in function. To be useful, the test must be sensitive to meaningful changes in that function.[28] The CUE-T must still be evaluated for sensitivity to change. Given the high levels of agreement for the right and left side and hand scores, we are optimistic that the CUE-T will be responsive to changes in function in these subscales. In addition, studies need to be carried out in children to determine the age range where reliable data can be obtained. Some of the CUE-T items would need to be scaled down for smaller children, and normative data used to score the strength items.

Study limitations

One limitation of this study is that we enrolled fewer subjects with motor complete injuries than planned, particularly in the high cervical (C2–5) and low cervical (C8–T1) levels. As a result, there is limited information on test–retest reliability and agreement for these groups. In addition, this was a single-center study. Whether similar results would be found in a multicenter trial is unknown.

Table 7 CUE-T and UEMS scores by ASIA impairment scale and motor level group

Group	Total CUE-T					Total UEMS			
AIS A–B	*N*	Mean	SD	Min	Max	Mean	SD	Min	Max
C2–C5	2	18.0	15.6	7	29	11.5	4.9	8	15
C6	6	51.8	16.4	26	72	22.8	6.0	16	31
C7	4	59.3	15.0	48	81	29.0	1.2	28	30
C8–T1	1	111.0	…	111	111	46.0	…	46	46
T2–T6	7	117.3	11.3	102	130	49.6	1.1	47	50
AIS C–D									
C2–C5	10	85.2	20.5	52	117	38.5	7.0	24	48
C6	5	93.6	24.7	56	121	38.0	8.7	24	45
C7	7	107.7	17.3	83	126	42.1	4.7	35	48
C8–T1	4	113.8	6.4	106	120	46.3	3.4	42	49
T2–T6	4	128.8	6.2	122	135	47.3	1.9	46	50

CUE-T, = Capabilities of Upper Extremities Test; UEMS, upper extremity motor score.

Conclusion

As per our knowledge, the CUE-T is the only test of upper extremity functional limitations that includes assessment of the entire upper limb. It has excellent test–retest reliability and agreement, and there is some evidence of construct and divergent validity. The CUE-T can be used to evaluate upper extremity functional capabilities in persons with chronic SCI, and could be used to evaluate change in function with the understanding that sensitivity to change has not yet been determined.

Disclaimer statements

Contributors RJM is involved in conceiving and designing the study, obtaining funding and/or ethics approval, interpreting the data, writing the article in whole or in part. SBK is involved in collecting the data, interpreting the data. BL is involved in analysing the data, interpreting the data. MS-R is involved in conceiving and designing the study, collecting the data, writing the article in whole or in part. MJM is involved in conceiving and designing the study, interpreting the data, writing the article in whole or in part.

Funding This manuscript is based on original research funded in part by grant #H133N060011 from the National Institute on Disability and Rehabilitation Research, Office of Special Education and Rehabilitative Services, US Department of Education, Washington DC.

Conflicts of interest None.

Ethics approval Ethical approval was obtained from the Institutional Review Board of Thomas Jefferson University.

References

1 Tohda C, Kuboyama T. Current and future therapeutic strategies for functional repair of spinal cord injury. Pharmacol Ther 2011; 132(1):57–71.
2 Steeves JD, Kramer JK, Fawcett JW, Cragg J, Lammertse DP, Blight AR, *et al.* Extent of spontaneous motor recovery after traumatic cervical sensorimotor complete spinal cord injury. Spinal Cord 2011;49(2):257–65.
3 Marino RJ, Burns S, Graves DE, Leiby BE, Kirshblum S, Lammertse DP. Upper- and lower-extremity motor recovery after traumatic cervical spinal cord injury: an update from the national spinal cord injury database. Arch Phys Med Rehabil 2011;92(3):369–75.
4 Steeves JD, Lammertse DP, Kramer JLK, Kleitman N, Kalsi-Ryan S, Jones L, *et al.* Outcome measures for acute/subacute cervical sensorimotor complete (AIS-A) spinal cord injury during a phase 2 clinical trial. Top Spinal Cord Inj Rehabil 2012;18(1):1–14.
5 Kalsi-Ryan S, Curt A, Verrier MC, Fehlings MG. Development of the Graded Redefined Assessment of Strength, Sensibility and Prehension (GRASSP): reviewing measurement specific to the upper limb in tetraplegia. J Neurosurg Spine 2012;17(Suppl 1):65–76.
6 Kapadia N, Zivanovic V, Verrier M, Popovic M. Toronto rehabilitation institute-hand function test: assessment of gross motor function in individuals with spinal cord injury. Top Spinal Cord Inj Rehabil 2012;18(2):167–86.
7 Marino RJ, Patrick M, Albright W, Leiby BE, Mulcahey M, Schmidt-Read M, *et al.* Development of an objective test of upper-limb function in tetraplegia: the capabilities of upper extremity test. Am J Phys Med Rehabil 2012;91(6):478–86.
8 Marino RJ. Domains of outcomes in spinal cord injury for clinical trials to improve neurological function. J Rehabil Res Dev 2007; 44(1):113–22.
9 Verbrugge LM, Jette AM. The disablement process. Soc Sci Med 1994;38(1):1–14.
10 Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehab 1998;12(3):187–99.
11 American Spinal Injury Association. International standards for neurological classification of spinal cord injury, revised 2000, reprinted 2002. Chicago, IL: American Spinal Injury Association; 2002.
12 Marino RJ, Shea JA, Stineman MG. The Capabilities of Upper Extremity instrument: reliability and validity of a measure of functional limitation in tetraplegia. Arch Phys Med Rehabil 1998; 79(12):1512–21.
13 Catz A, Itzkovich M, Tesio L, Biering-Sørensen F, Weeks C, Laramee MT, *et al.* A multicenter international study on the spinal cord independence measure, version III: Rasch psychometric validation. Spinal Cord 2007;45(4):275–91.
14 Kirshblum SC, Burns SP, Biering-Sørensen F, Donovan W, Graves DE, Jha A, *et al.* International standards for neurological classification of spinal cord injury (revised 2011). J Spinal Cord Med 2011;34(6):535–46.
15 Oleson CV, Marino RJ. Responsiveness and concurrent validity of the revised Capabilities of Upper Extremity-Questionnaire (CUE-Q) in patients with acute tetraplegia. Spinal Cord 2014;52(8):625–8.
16 Itzkovich M, Gelernter I, Biering-Sørensen F, Weeks C, Laramee MT, Craven BC, *et al.* The Spinal Cord Independence Measure (SCIM) version III: reliability and validity in a multi-center international study. Disabil Rehabil 2007;29(24):1926–33.
17 Kalsi-Ryan S, Beaton D, Curt A, Duff S, Popovic MR, Rudhe C, *et al.* The graded redefined assessment of strength sensibility and prehension: reliability and validity. J Neurotrauma 2012;29(5): 905–14.
18 Rudhe C, Van Hedel HJA. Upper extremity function in persons with tetraplegia: relationships between strength, capacity, and the spinal cord independence measure. Neurorehabil Neural Repair 2009;23(5):413–21.
19 Anderson K, Aito S, Atkins M, Biering-Sørensen F, Charlifue S, Curt A, *et al.* From the 2006 NIDRR SCI measures meeting functional recovery measures for spinal cord injury: an evidence-based review for clinical practice and research. J Spinal Cord Med 2008;31(2): 133–44.
20 Fekete C, Eriks-Hoogland I, Baumberger M, Catz A, Itzkovich M, Lüthi H, *et al.* Development and validation of a self-report version of the Spinal Cord Independence Measure (SCIM III). Spinal Cord 2013;51(1):40–7.
21 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–74.
22 Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd ed. Oxford: Oxford University Press; 1995.
23 Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8(2):135–60.
24 Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001; 10(7):571–8.
25 Marino RJ, Jones L, Kirshblum S, Tal J, Dasgupta A. Reliability and repeatability of the motor and sensory examination of the international standards for neurological classification of spinal cord injury. J Spinal Cord Med 2008;31(2):166–70.
26 Savic G, Bergström EMK, Frankel HL, Jamous MA, Jones PW. Inter-rater reliability of motor and sensory examinations performed according to American Spinal Injury Association standards. Spinal Cord 2007;45(6):444–51.
27 Anderson KD, Acuff ME, Arp BG, Backus D, Chun S, Fisher K, *et al.* United States (US) multi-center study to assess the validity and reliability of the Spinal Cord Independence Measure (SCIM III). Spinal Cord 2011;49(8):880–5.
28 Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38(1):27–36.