

ASSESSING MOTOR IMPAIRMENT OF THE TRUNK IN PATIENTS WITH TRAUMATIC BRAIN INJURY: RELIABILITY AND VALIDITY OF THE TRUNK IMPAIRMENT SCALE

ABSTRACT: *Introduction:* Literature regarding trunk assessment after Traumatic Brain Injury (TBI) is limited. The Trunk Impairment Scale (TIS) is a newly developed tool which is intended to assess static and dynamic sitting balance and trunk co-ordination.

Aim: It was the aim of this study to examine the reliability and validity of the TIS in TBI patients.

Methods: Thirty TBI subjects were recruited from within a rehabilitation setting. Two researchers observed each subject simultaneously, but scored independently. Each subject was re-examined by one of the raters.

Results: Kappa and weighted kappa values for all items ranged from 0.34 to 1. All percentages of agreement were 70% or higher. Intraclass correlation (ICC) coefficients for the sub-scale scores were between 0.72 and 0.88. Test-retest and inter-rater reliability for the total TIS score (ICC) was 0.88 and 0.95, respectively. The 95% limits of agreement for the test-retest and interexaminer measurement error interval were -4,4 and -3,3, respectively. The construct validity was evaluated by means of the Spearman rank correlation coefficient between the TIS and the Barthel Index ($r=0.59$, $p=.0007$).

Discussion and conclusion: Fair to perfect item agreement was found but the reliability of certain items requires further attention. Acceptable sub-scale and total TIS reliability and validity justify the use of the TIS in TBI treatment and research.

KEY WORDS: TRAUMATIC BRAIN INJURY, OUTCOME ASSESSMENT, REPRODUCIBILITY OF RESULTS

INTRODUCTION

Traumatic brain injury (TBI) is a significant cause of morbidity in the South African context (Reed and Welsh 2002). Although statistics regarding TBI in South Africa are limited, a study in Johannesburg in 1991 recorded an incidence of 316 TBI patients per 100,000 inhabitants annually (Nell and Brown 1991).

Following TBI, there is often an associated loss of trunk control and balance (Davies 1994) which are considered as some of the most disabling aspects following TBI (Black et al 1999). Selective trunk control is required for balance, limb function, gait, respiration and speech (Davies 1990). Furthermore, sitting balance has been cited as an important predictor of functional outcome following TBI (Black et al 1999).

The condition of TBI patients is often characterized by poor concentration,

attention and memory. These patients are also frequently confused, disoriented and agitated (Sohlberg and Mateer 1989, Quinn and Sullivan 2000). It is therefore important that evaluation instruments should be brief and not complex.

Although a reliable and accurate assessment instrument of trunk function is required to define appropriate aims of rehabilitation (Mazaux et al 2001) there are few instruments which have been developed for measuring this in the TBI population. The Clinical Outcomes Variable Scale is a scale which consists of 13 items, one of which assesses sitting balance using a 7-point ordinal scale. In a study on 16 TBI patients, the sitting balance item was found to be reliable (Low-Choy et al 2002). The results should be interpreted with caution however, due to the small sample size and the use of intraclass correlation

coefficients which may not be the appropriate statistical analysis to evaluate rater agreement for an ordinal item. The validity of the instrument was not examined.

The Trunk Impairment Scale (TIS) was developed by Verheyden et al (2004) as a comprehensive tool to assess impairments in trunk control after stroke. The TIS contains three sub-sections, which assess static sitting balance, dynamic sitting balance and trunk co-ordination.

CORRESPONDENCE TO:

Geert Verheyden
Katholieke Universiteit Leuven,
Department of Rehabilitation Sciences,
Tervuursevest 101, B-3001
Leuven, Belgium
Tel: +32 16 32 91 17 (work)
Geert.Verheyden@faber.kuleuven.be

Verheyden G, MSc Physiotherapy¹;
Hughes J, BSc (Hons) Physiotherapy²;
Jelsma J, PhD³; Nieuwboer A, PhD⁴;
De Weerd W, PhD⁴

¹ Research Assistant, Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Belgium.

² Lecturer, Division of Physiotherapy, University of Cape Town, South Africa.

³ Associate Professor, Division of Physiotherapy, University of Cape Town, South Africa.

⁴ Professor, Department of Rehabilitation Sciences, Katholieke Universiteit Leuven, Belgium.

Static sitting balance (scoring range 0-7) evaluates the ability to remain in the sitting position with a) both feet on the floor and b) with the legs crossed. The dynamic sitting balance sub-scale (scoring range 0-10) assesses lateral flexion of the trunk, initiated from the upper and lower part of the trunk. Co-ordination (scoring range 0-6) scores rotation from the shoulder and pelvic girdle in the horizontal plane. The total TIS score ranges from 0 to 23, a higher score indicating better trunk function. The time to complete the TIS was found to range from 2 to 18 minutes. It was concluded that the TIS is sufficiently reliable and valid for use in both clinical practice and research involving stroke patients (Verheyden et al 2004). However, the reliability and validity of this measure in other conditions, such as TBI has not been established.

If the TIS was found to be reliable and valid in this group of patients, the use of the instrument might contribute to a more consistent provision of care and a better treatment planning and monitoring of progress. It would also ensure that an improvement in a patient's scores on the TIS is due to an actual improvement of performance and not from an inaccurate assessment. It was the aim of this study to determine the reliability and construct validity of the TIS when used in the assessment of TBI patients.

METHODS

Subjects

The subjects for this study were recruited from consecutive admissions to a rehabilitation and long-term care centre in the Western Cape. Inclusion criteria included: a traumatic brain injury (confirmed from medical records); over 18 years of age; medically stable; and ability to understand and speak English, Afrikaans or Xhosa. Subjects were excluded if they had other medical complaints or injuries that would affect their trunk function. The medical notes of the patients were used to obtain medical and demographic information, including age, gender, affected side, date and type of injury.

Instrumentation

The Barthel Index is a known reliable

and valid measure of disability (Collin et al 1988) with scores ranging from a minimum of zero to a maximum of 100 points. It was chosen as the "measure" against which the validity of the TIS was measured. The TIS, as developed by Verheyden et al (2004) was administered to all subjects (See Appendix 1). In the case of non-English speaking subjects, the services of a translator were utilised.

Ethical approval (ref: 096/2004) was obtained from the Research Ethics Committee of the University of Cape Town. Informed consent was obtained from all subjects. Patients who were unable to give their consent due to cognitive impairments were excluded.

Procedure

The TIS was used to evaluate each subject on two separate occasions. On one occasion, two research assistants (final year physiotherapy students) (rater 1 and rater 2) scored the TIS simultaneously but independently. On a second occasion, rater 1 re-assessed the patient alone. Rater 1 instructed the patient verbally (or demonstrated if needed) on both occasions. The two consecutive evaluations took place on the same day, separated by one or two hours of rest. During this time no treatment was offered. To minimise recall bias by the raters at least two other patients were evaluated before re-assessing the patient. In addition, raters filled in the score sheet but did not add up the scores to calculate a total score. The order of the evaluations (first and second occasion) was randomised.

To avoid possible scoring bias and to maintain a standardised procedure, every item of the scale was performed three times in each assessment, even if a patient reached a maximum score after one or two attempts. For each item, the best score of the three attempts was recorded. After all evaluations were completed, the researchers calculated the total scores for each assessment. The raters were introduced to the TIS by means of an instruction video before the beginning of the study.

Statistical Analysis

Test-retest agreement was examined by comparing the results of rater 1, who

tested the subjects twice. To determine inter-rater reliability, the results of both raters, who observed the subject simultaneously, were compared. Construct validity was determined by comparing total TIS scores with the Barthel Index scores.

To evaluate test-retest and inter-rater agreement of individual items, kappa or weighted kappa statistics were calculated and percentages of agreement were reported. Kappa and weighted kappa statistics are used to evaluate the agreement of dichotomous and ordinal items respectively. The value of the (weighted) kappa statistic indicates the strength of agreement. According to Landis and Koch (1977), values less than zero indicate poor agreement from 0-0.20 suggest slight, from 0.21 to 0.40 fair, from 0.41 to 0.60 moderate, from 0.61 to 0.80 substantial and from 0.81 to 1.00 almost perfect agreement. Percentage of agreement reflects the amount of agreement between the two observations.

Intraclass correlation (ICC) coefficients were used to determine test-retest and inter-rater agreement for the three sub-sections of the scale and the total TIS score. ICC values of less than 0.40 were considered low, 0.40 to 0.59 moderate, 0.60 to 0.79 moderately high and values above 0.80 were considered very high (Katz et al 1992).

The 95% test-retest and interexaminer measurement error interval was determined according to the method proposed by Haas (1991). This analysis offers the possibility of interpreting clinical changes when evaluating patients. An improvement in a patient from the upper limit of the 95% interval or higher can be seen as an increase without reproducibility bias. This is important information for therapists. If the patient improves after a period of treatment more than the upper limit of the 95% interval or higher, an actual improvement in performance took place and the increase in the score is not due to rater bias.

Construct validity was assessed by calculating the Spearman rank correlation coefficient between total TIS and Barthel Index scores. Construct validity refers to the extent to which obtained results from one measurement correlate with the results from a second measurement, identifying an underlying theore-

tical model or construct (Wade 1992). According to Hinkle et al (1988), a correlation between zero and 0.30 shows little if any correlation, between 0.30 and 0.50 low, between 0.50 and 0.70 moderate, between 0.70 and 0.90 high and between 0.90 and 1.00 very high correlation. The level of significance selected throughout was $p < .05$.

Statistical analyses were conducted with the statistical software programme SAS 8.2 and SAS Enterprise Guide 2.0.

RESULTS

Thirty subjects (three females and 27 males) participated in the study. Of the subjects, 10 were affected on their right

side, 11 on their left side and nine bilaterally. Ages of the subjects ranged from 19 to 69, with a median of 32 years (interquartile range (IQR): 26-41). The median number of days since injury was 84 (IQR: 57-248, range 39-1496). In terms of mechanism of injury, 10 subjects sustained either assault or blunt head trauma, one person fell, one subject was shot and the remaining 18 were involved in motor vehicle accidents.

The results of the test-retest and inter-observer agreement for the different items of the TIS indicate that for the static sitting balance sub-scale, two out of three items have perfect agreement (Table 1). Item three (crossing the legs actively) has a moderate weighted kappa

value of 0.52 with 67% of inter-observer agreement. The kappa values for the dynamic sitting balance sub-scale vary between 0.34 and 1. Item seven (lifting the pelvis at the weakest side) shows fair agreement. Eight kappa values have moderate agreement, six substantial agreement and two an almost perfect and another two a perfect agreement. All percentages of agreement are 70% or higher. The kappa and weighted kappa values for the co-ordination sub-scale vary between 0.44 and 0.78. Three kappa values are moderate, the remaining five show substantial agreement. All percentages of agreement exceed 76%.

The values for the test-retest and inter-observer agreement for the three

Table 1: Kappa or weighted Kappa, lower value of the 90% confidence limit of the Kappa or weighted Kappa and percentage of agreement for test-retest and inter-observer agreement.

Item	K/Kw ^a	Value ^b	Test-retest agreement		Value ^b	Inter-observer agreement	
			90%lcl ^c	% ^d		90%lcl ^c	% ^d
Static sitting balance							
1 Remain seated position	K	1.00	1.00	100%	1.00	1.00	100%
2 Crossing legs passively	K	1.00	1.00	100%	1.00	1.00	100%
3 Crossing legs actively	Kw	* ^e		53%	0.52	0.27	67%
Dynamic sitting balance							
1 Bring elbow down at weak side	K	0.78	0.44	97%	0.47	-0.03	93%
2 Observe trunk movement	K	0.57	0.31	80%	0.71	0.49	87%
3 Observe compensations	K	0.59	0.34	80%	0.65	0.41	83%
4 Bring elbow down at strong side	K	1.00	1.00	100%	1.00	1.00	100%
5 Observe trunk movement	K	0.42	0.03	87%	0.71	0.40	93%
6 Observe compensations	K	0.52	0.17	87%	0.45	0.12	83%
7 Lift pelvis at weakest side	K	0.34	0.05	70%	0.37	0.07	73%
8 Observe compensations	K	0.60	0.36	80%	0.80	0.62	90%
9 Lift pelvis at strongest side	K	0.59	0.32	83%	0.84	0.66	93%
10 Observe compensations	K	0.73	0.53	87%	0.87	0.72	93%
Co-ordination							
1 Rotate shoulder girdle	Kw	0.76	0.56	87%	0.58	0.33	77%
2 Rotate shoulder girdle fast	K	0.61	0.38	80%	0.62	0.40	80%
3 Rotate pelvic girdle	Kw	0.44	0.11	80%	0.64	0.36	83%
4 Rotate pelvic girdle fast	K	0.51	0.12	90%	0.78	0.44	97%

^a indicates the use of a Kappa (K) or weighted Kappa (Kw) for statistical analysis.

^b shows the value of the calculated Kappa or weighted Kappa.

^c displays the lower value of the 90% confidence limit of the Kappa or weighted Kappa.

^d gives the percentage of agreement.

^e means no weighted Kappa could be calculated because not all scoring possibilities were present.

Table 2: ICC values for test-retest and inter-observer agreement

Total	Test-retest agreement	Inter-observer agreement
Static sitting balance	0.81 (0.67)	0.87 (0.78)
Dynamic sitting balance	0.72 (0.54)	0.88 (0.80)
Co-ordination	0.77 (0.61)	0.73 (0.55)
Trunk Impairment Scale	0.88 (0.78)	0.95 (0.91)

Values are presented as ICC (90% lower confidence limit).

sub-scales and total TIS score are shown in Table 2. ICC coefficients for the sub-scales were all moderately high to very high and ranged between 0.72 and 0.88. ICC values for the test-retest and inter-rater reliability for the total TIS score are 0.88 and 0.95, respectively.

The 95% test-retest examiner measurement error interval on the total TIS score was -4,4 and the 95% interexaminer measurement error interval -3,3.

The median Barthel Index was 88 points (IQR: 63-95, range 0-100). The median total TIS score was 16 out of 23 points (IQR: 14-19, range 0-21). The Spearman rank correlation of the total TIS score with the Barthel Index was 0.59 ($p=0.0007$).

DISCUSSION

It was the aim of this study to examine the reliability and validity of the Trunk Impairment Scale in patients with traumatic brain injury.

For item one and two of the static sitting balance sub-scale, a perfect test-retest and inter-observer agreement was found. No weighted kappa value for test-retest agreement could be calculated because the score of one was not given to any of the 30 patients on one occasion of the test procedure. The inter-rater agreement was moderate and the percentages of agreement for both test-retest and inter-observer agreement appeared to be somewhat low and the reliability of this item could be questioned. The low percentage of agreement can be explained because the item is scored on a 4-point ordinal scale. Perfect agreement is more difficult to achieve than on a dichotomous item because of the extended scoring possibilities. However, the moderate weighted kappa value still indicates sufficient reliability. It is believed that discrepancies between raters in interpretation of the options for scoring item three may be responsible for the moderate inter-rater

agreement. In item three, the patient is asked to actively cross the legs while maintaining the seated position without backward trunk displacement. Differences in perception of trunk movement between raters might have occurred. They had to judge if this displacement was more or less than 10 cm. Adding a standardised guideline to the TIS procedure wherein this distance is measured before the patient performs this item and a firm pillow is positioned at 10 cm to monitor if the patient touches this object whilst crossing the legs could enhance the reliability of this item. Kappa and weighted kappa values in the study by Verheyden et al (2004) on stroke patients ranged from 0.51 to 1 and are comparable to the present findings.

Kappa values for the items of the dynamic sitting balance sub-scale ranged from fair to perfect. Although some low kappa values were noted, percentages of agreement ranged from 70% to 100% indicating an acceptable reliability. The discrepancy between a low kappa and a high percentage of agreement is a known weakness of the kappa statistic in case of poor inter-subject variation (Feinstein and Cicchetti 1990). In this population, distribution problems were also observed in the Barthel Index scores and total TIS scores. They are both skewed towards the maximum value of the scales, indicating that the subjects were less severely impaired. The inclusion of more severely impaired TBI patients is recommended for future studies. The same is true for the items of the co-ordination sub-scale where kappa and weighted kappa values ranged from 0.44 to 0.78 but percentages of agreement were between 77% and 97%. Kappa and weighted kappa values for the TIS used in a stroke population (Verheyden et al 2004) ranged for the items of the dynamic sitting balance and co-ordination sub-scales from 0.46 to 1 and are similar to these results.

Moderately high and very high ICC values were obtained for all sub-scale totals and total TIS score. An ICC value of 0.98 was reported for the sitting balance item of the Clinical Outcomes Variable Scale (Low Choy et al 2002). But as mentioned before, the use of an ICC for examining the reliability of an ordinal item with 16 subjects should be questioned. Very high ICC coefficients for the summed scores of the three sub-scales and total TIS score (between 0.85 and 0.99) were also found for stroke patients (Verheyden et al 2004). The ICC values of the present study are comparable and sufficiently high to apply the different sub-scales scores and total TIS score in rehabilitation practice and research.

A 95% test-retest examiner measurement error interval of -4 to +4 was found for the total TIS score. This would mean that an increase or decrease of four points or more on the TIS can be seen as a genuine change and does not simply reflect measurement bias. The test-retest interval is clinically more important than the inter-rater interval (-3 to +3) since the test-retest interval is more realistic in the clinical setting where a patient is evaluated after a period of treatment by the same therapist.

In order to assess construct validity of the TIS, total scores were correlated with those of the Barthel Index. A moderate Spearman rank correlation of 0.59 ($p=0.0007$) was found. This indicates that the level of trunk function is moderately related to the level of activities of daily living (ADL). Sitting balance has been attributed as a useful predictor of functional recovery following TBI (Black et al 2000) and a stronger correlation between these two scores was therefore anticipated. This could be due to health care resources in South Africa. Due to the high patient turnover, rehabilitation aims may focus on patient function for early discharge rather than ideal movement sequences. As a result, level of functioning may be high due to compensatory mechanisms, even though underlying trunk function is still impaired.

CONCLUSION

This study evaluated the reliability and validity in the assessment of TBI patients of the Trunk Impairment Scale (TIS) which measures static and dynamic sitting balance as well as trunk co-ordination.

Item per item reliability showed kappa and weighted kappa values from

fair to perfect. The reliability of some items needs further attention. Suggestions are made as to how to improve the reliability of these items. The ICC values for the sub-scales and total TIS scores were moderately high to very high indicating the applicability of the TIS as a measure of trunk performance in TBI patients. In determining construct validity a significant correlation between total TIS and Barthel Index scores was found.

To our knowledge, this is the first study to assess the reliability and validity of a scale that evaluates motor impairment of the trunk in TBI patients. TBI research could benefit from further studies that investigate the psychometric properties of the Trunk Impairment Scale. It is suggested that the TIS has been found to be a reliable, valid and simple instrument with which to evaluate TBI patients and document treatment. Clinical practice will improve if standardized measurements are used for evaluation purposes and identifying therapy guidelines.

ACKNOWLEDGEMENT

The researchers would like to thank patients and staff at the involved centre for their assistance in the project.

The researchers would also like to express their thanks to the following research assistants: A Goldstein, L Johnston, C Sparrius and C Cumming.

REFERENCES

- Black K, Zafonte R, Millis S, Desantis N, Harrison-Felix C, Wood D, Mann N 1999 Sitting balance following brain injury: does it predict outcome? *Brain Injury* 14:141-152
- Collin C, Wade DT, Davies S, Horne V 1988 The Barthel ADL Index: a reliability study. *International Disability Studies* 10:61-63
- Davies PM 1990 Problems associated with the loss of selective trunk activity in hemiplegia. In: *Right in the middle*. pp 31-65. Springer-Verlag, Berlin
- Davies PM 1994 Starting Again. Early rehabilitation after traumatic brain injury or other severe brain lesion. pp 391-393. Springer-Verlag, Germany
- Feinstein AR, Cicchetti DV 1990 High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology* 43:543-549
- Haas M 1991 Statistical methodology for reliability studies. *Journal of Manipulative and Physiological Therapeutics* 14:119-132
- Hinkle DE, Wiersma W, Jurs SG 1988 Applied statistics for the behavioral sciences. pp 117-121. Houghton Mifflin Company, Boston
- Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH 1992 Comparing measurement sensitivity of short and longer health status instruments. *Medical Care* 30:917-925
- Landis RJ, Koch GG 1977 The measurement of observer agreement for categorical data. *Biometrics* 33:159-174
- Low-Choy N, Kuys S, Richards M, Isles R 2002 Measurement of functional ability following traumatic brain injury using the Clinical Outcomes Variable Scale: a reliability study. *Australian Journal of Physiotherapy* 48:35-39
- Mazaux JM, de SÈze M, Joseph PA, Barat M 2001 Early rehabilitation after severe brain injury: a french perspective. *Journal of Rehabilitation Medicine* 33:99-109
- Nell V, Brown SO 1991 Epidemiology of traumatic brain injury in Johannesburg - II. Morbidity, mortality and aetiology. *Social Science & Medicine* 33:289-296
- Quinn B, Sullivan JS 2000 The identification by physiotherapists of the physical problems resulting from a mild traumatic brain injury. *Brain Injury* 14:1063-1076
- Reed AR, Welsh DG 2002 Secondary injury in traumatic brain injury patients - a prospective study. *South African Medical Journal* 92: 221-224
- Sohlberg MM, Mateer CA 1989 Introduction to cognitive rehabilitation: theory and practice. pp 93-96. The Guilford Press, New York
- Verheyden G, Nieuwboer A, Mertin J, Preger R, Kiekens C, De Weerd W 2004 The Trunk Impairment Scale: a new tool to measure motor impairment of the trunk after stroke. *Clinical Rehabilitation* 18:326-334
- Wade DT 1992 Measurement in neurological rehabilitation. pp 37-38. Oxford University Press, Oxford