Title:

DEVELOPMENT OF THE ITALIAN VERSION OF THE TRUNK IMPAIRMENT SCALE IN SUBJECTS WITH ACUTE AND CHRONIC STROKE.

Cross-cultural adaptation, reliability, validity and responsiveness.

Running head: The Italian TIS

Article category: Research paper

Authors names: Marco Monticone¹, Emilia Ambrosini^{2,3}, Geert Verheyden⁴, Flavia Brivio³, Roberto Brunati³, Luca Longoni³, Gaia Mauri³, Alessandro Molteni³, Claudia Nava³, Barbara Rocca³, Simona Ferrante²

1.Department of Medical Sciences and Public Health, University of Cagliari, Cagliari, Italy

2.Neuroengineering and Medical Robotics Laboratory, Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

3.Physical Medicine and Rehabilitation Unit, Scientific Institute of Lissone, Clinical and Scientific Institutes Maugeri, Institute of Care and Research (IRCCS), Lissone, Italy

4.Geert Verheyden, KU Leuven, Department of Rehabilitation Sciences, Leuven, Belgium

Name and address for correspondence

Marco Monticone

Department of Medical Sciences and Public Health, University of Cagliari.

Cittadella Universitaria, Strada Statale, 554 - Monserrato, Cagliari, Italy - 09042

Phone: 0039 070 51096487; Fax: 0039 070 51096487

Email: marco.monticone@unica.it

Declaration of interest :

The authors report no declarations of interest.

ABSTRACT (200 words)

Purpose. To cross-culturally adapt and psychometrically analyse the Italian version of the Trunk Impairment Scale on acute (*cohort 1*) and chronic stroke patients (*cohort 2*).

Methods. The Trunk Impairment Scale was culturally adapted in accordance with international standards. The psychometric testing included: internal consistency (Cronbach's alpha), inter- and intra-rater reliability (intraclass correlation coefficient; standard error of measurement and minimal detectable change), construct validity by comparing Trunk Impairment Scale score with Barthel Index, motor subscale of Functional Independence Measure, and Trunk Control Test (Pearson's correlation), and responsiveness (Effect Size, Effect Size with Guyatt approach, standardized response mean, and Receiving Operating Characteristics curves).

Results. The Trunk Impairment Scale was administered to 125 and 116 acute and chronic stroke patients, respectively. Internal consistency was acceptable ($\alpha > 0.7$), inter- and intra-rater reliability (ICC > 0.9 , Minimal Detectable Change for total score < 1.6 in *cohort 1* and < 1.1 in *cohort 2*) were excellent. The construct validity showed acceptable correlations ($r > 0.4$) with all scales but the motor Functional Independence Measure in *cohort 2*. Distribution-based methods showed large effects in *cohort 1* and moderate to large effects in *cohort 2*. The Minimal Important Difference was 3.5 both from patient's and therapist's perspective in *cohort 1* and 2.5 and 1.5 from patient's and therapist's perspective, respectively, in *cohort 2*.

Conclusion. The Trunk Impairment Scale was successfully translated into Italian and proved to be reliable, valid, and responsive. Its use is recommended for clinical and research purposes.

Keywords (5-8)

Trunk Impairment Scale; stroke; outcome measures; psychometric properties; responsiveness; rehabilitation.

INTRODUCTION

Trunk control is an essential part of balance and postural control, constituting an important prerequisite for daily activities and function. Impairment of trunk control represents a common problem in stroke subjects leading to restricted balance, impaired postural control, increased risk of falls, motor disability and dependency in their activities of daily living [1,2].

Over the course of time, clinical measures to evaluate trunk performance after stroke have evolved and, based on a recent systematic review on clinical tools evaluating trunk performance after stroke, the Trunk Impairment Scale (TIS) primarily attracted our attention as an interesting evaluation tool for neurorehabilitation programmes [3,4]. It was published by Verheyden in 2004 and it is a measure that specifically evaluates static and dynamic sitting balance and trunk co-ordination in stroke subjects [4].

The TIS proved to be reliable and internally consistent, as well as satisfactorily valid in comparison with other instruments used to evaluate a mixed population of subjects with acute and chronic stroke [4]. Item per item test-retest and inter-rater reliability were established by means of kappa and weighted kappa values (ranging from 0.46 to 1) and percentage of agreement (ranging from 82 to 100%). Intraclass correlation coefficients (ICC) were calculated for test-retest and inter-rater reliability for subscale and total scores (ranging from 0.85 to 0.99) [4]. Cronbach's alpha varied from 0.65 to 0.89 and construct and concurrent validity was shown by close correlations with the Barthel Index (0.86) and the Trunk Control Test (0.83) [4]. A strong relation with measures of balance, gait and functional ability were also found in a later study (model $R^2 = 0.55-0.62$) involving a mixed population of acute and chronic stroke [5]. Another study conducted on a similar sample evaluated the discriminant ability of the scale by comparing the scores of subjects with stroke with those of age- and sex-matched healthy subjects ($P < 0.001$) [6]. A subsequent study showed the TIS score on admission (acute sample) was a stronger predictor of the Barthel Index score at six months (chronic sample) than the Barthel Index score itself on admission, suggesting the importance of trunk performance when predicting functional outcome after stroke (partial $R^2 = 0.52$, $p < 0.001$) [7]. A Norwegian translation of the TIS has been investigated in subjects after stroke [8]. The scale was modified in the sense that the subscale "Static sitting balance" was removed but still showed to be internally consistent, reliable and valid [8].

Recently, an Italian version of the TIS has been developed and psychometrically analysed [9]. However, given the paucity and the characteristics of the sample which included 41 subjects with acute stroke, additional studies are recommended to strengthen the applicability of this measure in populations characterized by wider clinical features and to provide a psychometric assessment.

Hence, the aim of this study was to investigate the psychometric properties (reliability, validity and responsiveness) of the TIS in a large population of Italian subjects with acute and chronic stroke. Both populations were evaluated separately since the psychometric properties of any scale have to be considered population-specific. Above all, we expect the effect of an intervention to be larger in the acute phase after stroke than in the chronic phase. Therefore, we hypothesized a lower measurement error and a lower minimal important change (MIC) for chronic stroke patients.

METHODS

The study was part of a population-based cohort study, approved by the hospital's Institutional Review Board prior to patient enrolment and conducted in conformity with ethical and humane principles of research. The patients gave their written consent.

Patients

Inpatients and outpatients attending the Physical Medicine and Rehabilitation Unit of our Hospital were recruited between July 2012 and December 2015. The inclusion criteria were: adult age (45–80 years old), a cerebrovascular stroke occurring either <30 days before study entry (*cohort 1*) or >6 months before study entry (*cohort 2*), a good understanding of Italian, and an ability to sit upright without any help. For both cohorts, exclusion criteria were major cognitive deterioration (Mini Mental State Examination score ≤ 24), cardiovascular instability (New York Heart Association class III, recent ischemic heart attack, uncontrolled arterial hypertension, and respiratory instability as revealed by an oxygen saturation of <90%), systemic diseases, and neurodegenerative or neuromuscular diseases, co-morbidities that might affect the progress from stroke, and communication problems that could interfere with the test protocol.

Cross-cultural adaptation

After receiving formal approval by the original developer (GV), cross-cultural translation was done in accordance with the protocol issued by the American Association of Orthopedic Surgeon Outcomes Committee [10], taking into account the principles described in the International Society for Pharmacoeconomics and Outcomes Research task force report [11].

Step 1 Translation into Italian: the items were translated with the aim of retaining the concepts of the original version. Two translations were made independently by two professional Italian translators. Keeping the language compatible with a reading age of 14 years, the discrepancies were resolved by means of reconciliation; step 1 ended when a common adaptation was agreed.

Step 2 Back-translation: two bilingual translators independently back-translated the initial translation. The principal investigator (MM) reviewed these translations and ensured the Italian version reflected the same item content as the original version.

Step 3 Expert Committee: the translations were submitted to a bilingual committee of clinicians, methodologists and the translators chaired by the principal investigator, who explored the semantic, idiomatic and conceptual equivalence of the items and answers to identify any discrepancies or mistakes. This phase ended when a pre-final version was agreed.

Step 4 Test of the pre-final version: the pre-final version was tested in order to assess the comprehensibility and cognitive equivalence of the translation. This was done by means of cognitive interviews conducted by trained psychologist (BR) and physiotherapists (GM, AM) who administered the TIS to 10 subjects with acute and chronic stroke, respectively. The Expert Committee reviewed the results with the aim of identifying any modification to improve the Italian form.

Procedures

All of the participants were provided with written information concerning the procedures by two research assistants.

Those satisfying the entry criteria underwent: an inpatient neurorehabilitation programme (*cohort 1*) and an outpatient neurorehabilitation programme (*cohort 2*). Both programmes involved neuromotor techniques, articular mobilisation, stretching exercises, consisting of ‘sit-to-stand’ exercises with and without impediments, load transfer exercises under active static conditions, step division exercises, and ambulatory training, with or without aids, on different types of surfaces and treadmills. Both intervention groups underwent 30 neuromotor rehabilitation sessions on 30 consecutive days, receiving the same total amount of time in the therapy. The programmes had already been tested for efficacy (control group in [12]).

In *cohort 2*, patients could have received other treatments before being enrolled into the study while no other treatments were advised at home during the execution of the study. In both cohorts mild analgesics and non-steroidal anti-inflammatory drugs (NSAIDs) were permitted.

Outcome measures

During the assessment phases (i.e. before beginning the treatment and at the end of treatment, e.g. after one month), the scales were administered to the subjects enrolled (*cohort 1* and *cohort 2*) by the physiotherapists involved in the study. Data of TIS, Barthel Index, Functional Impairment Measure,

Trunk Control Test and Global Perceive Effect were gathered by trained research assistants who completed an electronic database for subsequent statistical analyses.

Trunk Impairment Scale. This scale consists of three sub-scales and comprises 17 items scored on a 2-, 3- or 4-point ordinal scale. The first subscale named Static sitting balance (score range 0–7) evaluates the ability to remain in the sitting position with both feet on the floor and with the legs crossed; patients should cross the non-affected leg over the affected leg. The second sub-scale named Dynamic sitting balance (score range 0–10) assesses lateral flexion of the trunk, initiated from the upper and lower part of the trunk. The third subscale named Co-ordination (score range 0–6) scores rotation from the shoulder and pelvic girdle in the horizontal plane. The total score ranges from 0 to 23, with a higher score indicating a better performance [4]. All participants were examined by two observers who were familiarized with the test protocol of the TIS by means of an instruction video. Every subject was evaluated when sitting on the edge of a bed or treatment table and from the starting position as previously described [4].

Barthel Index (BI). It describes 10 ADLs, with the total score ranging from 0 (the greatest limitation) to 20 (no limitation) [13].

Functional Independence Measure (FIM). It describes 18 ADLs associated with motor, cognitive, and sphincteral problems, with the total score ranging from 18 (greatest limitation) to 126 (no limitation), motor score from 13 to 91 and cognitive score from 5 to 35 [14].

Trunk Control Test (TCT). It consists of four items which are assessed on a 3-point ordinal scale. The items are rolling from supine to the affected and unaffected side, sitting up from lying down and maintaining balance in the sitting position on the side of the bed. The total score for the Trunk Control Test ranges from 0 to 100, with a higher score indicating a better performance [15].

Global Perceived Effect (GPE). At the end of treatment, we evaluated the GPE using the question: “Overall, how much did the treatment you received help your trunk impairment due to stroke?”; the GPE was determined using a seven-level Likert scale with three improvement levels (helped a lot=1, helped=2, helped only a little=3), one no change level (did not help=4) and three worsening levels (made things worse in part=4,

made things worse=6, made things worse a lot=7) [16]. Both the patients' and therapists' perception were assessed.

Scale properties investigated

Acceptability

The time needed to complete the TIS scale was recorded. All of the data were checked for missing or multiple responses.

Floor/ceiling effects

Descriptive statistics were calculated in order to identify floor/ceiling effects, which were considered to be present when <15% of the subjects obtained the lowest or highest possible scores, indicating the proportion for whom, respectively, no meaningful deterioration or improvement in their condition could be detected since they are already at the extreme of the range [17].

Reliability

Reliability was assessed as follows:

-internal consistency, which reflects the inter-relatedness among the items (Cronbach's alpha computed on subscales and total score, with a value >0.70 being considered acceptable) [17];

-Inter-rater reliability: two observers score the scale concurrently but independently and the results of both therapists who observe the patient simultaneously are compared (ICCs computed on subscales and total score, with good and excellent reliability being, respectively, indicated by values of 0.70–0.85 and >0.85) [17].

The Standard Error of Measurement (SEM) was also computed to evaluate absolute reliability as follows:

$$SEM = SD\sqrt{1 - ICC}$$

where SD is the pooled standard deviation.

From the SEM, the minimum detectable change (MDC), i.e. the smallest change in score that is likely to reflect a true change rather than a measurement error, was estimated as follows:

$$MDC = SEM * 1.96 * \sqrt{2}$$

where $\sqrt{2}$ accounts for the variance associated with two independent sessions and 1.96 represents the 95% confidence interval.

-Intra-rater reliability (test-retest). After 2 hours, one of the two observers assesses the patient alone (ICCs computed on subscales and total score, with good and excellent reliability being, respectively, indicated by values of 0.70–0.85 and >0.85) [17]. SEM and MDC were computed also to evaluate intra-rater absolute reliability.

The inter- and intra-rater reliability of scores on individual test items of the scale were assessed using the kappa statistic (for dichotomous items), the weighted kappa statistic (for ordinal items), and the percentage of agreement. It was considered acceptable when the kappa or weighted kappa statistic was >0.60.

Validity

Construct validity

The extent to which an instrument's score relates to the score of the theoretical construct of another instrument, to the expected degree, was investigated by means of hypothesis testing (8). It was hypothesised a priori that the TIS and corresponding scales would achieve a level of correlation ranging from 0.40 to 0.80 (Pearson's correlation coefficient):

- the TIS and the BI;
- the TIS and the motor subscale of the FIM;
- the TIS and the TCT.

We defined the construct validity of the TIS as good if >75% of the hypotheses were confirmed.

Responsiveness

The ability of the instrument to detect change over time in the construct being measured was investigated by means of both distribution and anchor-based methods [18,19].

The distribution methods included the effect size (ES), also using the Guyatt approach, and the standardized response mean (SRM). The ES is a standardized measure of change over time calculated on the whole sample by dividing the difference between the pre-test and post-test scores by the pre-test SD; in the case of the Guyatt approach, the change computed on the whole sample is divided by the pre-test SD calculated only for stable subjects whose clinical status remained unchanged (GPE = 3;4). The ES therefore represents individual change in terms of the number of pre-test SDs, with values of 0.20, 0.50, and 0.80, respectively, representing

small, moderate, and large changes. The SRM is the ratio between individual change and the SD of that change. It has been suggested that SRM values of 0.20, 0.50, and 0.80, respectively, represent small, moderate, and large changes [20,21]; the SRM allows a group-level interpretation of the study population undergoing treatment [22].

- Receiving Operating Characteristics (ROC) curves are calculated on the total TIS score-change that is examined in relation to a dichotomous variable (the external criterion) able to discriminate between good and poor outcomes. This criterion variable represents the therapist's and the patient's perspective on the obtained change in TIS score. The question on the "global outcome of treatment" will be dichotomized into improved (treatment helped a lot or helped) and stable outcomes (treatment helped only little, did not help). Using the ROC curves the responsiveness is described in terms of sensitivity (probability of the measure of correctly classifying patients who demonstrate change on an external criterion of clinical change) and specificity (probability of the measure of correctly classifying patients who do not demonstrate change on an external criterion). Values for sensitivity and for false-positive rates (1-specificity) are plotted on the y and the x axis of the curve and the area under the ROC curve (AUC) represents the probability that a measure correctly classifies patients as improved or unchanged. This area theoretically ranges from 0.5 (non accuracy in discriminating) to 1.0 (perfect accuracy), and an AUC of at least 0.70 is considered to be acceptable (14). The optimal cut off point was computed using the Youden index and taken as the MIC, which indicates the change score associated with the least misclassification [23].

The statistical analysis was entirely performed with IBM SPSS Statistics v23 software.

RESULTS

Subjects

A total of 125 (*cohort 1*) and of 116 (*cohort 2*) patients agreed to participate and, of them, 103 and 100 satisfied the inclusion criteria, respectively; these were 47 females (46%) and 56 males (54%) with a mean age of 74 ± 11 years in *cohort 1* and 42 females (42%) and 58 males (58%) with a mean age of 67 ± 12 years in *cohort 2*. The median duration of the time after brain lesion at the time of recruitment was 22 days (range 16-38) in *cohort 1* and 12 months (range 6-108) in *cohort 2*. Table 1 shows the patients' characteristics for the two cohorts.

[Please insert here TABLE 1]

All of the subjects completed the rehabilitation programmes and were assessed with the TIS scale at both pre- and post-treatment assessments. No specific issues were raised during the study neither by the patients nor by the physiotherapists. An excessive use of medicines to relieve pain was not reported. The treatment proposed was well-accepted by all participants as confirmed by a full compliance rate (100%).

Translation and cross-cultural adaptation

The translation procedure took three months to reach a culturally adapted version and all of the items were easily forward and back-translated, and no difficulties were showed during the review of the back-translations. The correctness of the process, the content of the items, and the concepts expressed were confirmed by the experts. The cognitive interviews confirmed the comprehensibility and the cognitive equivalence of the translation; no other issues causing confusion were pointed out. Finally, the principal investigator and the Expert Committee confirmed the work done.

The adapted TIS is reproduced in the Appendix.

Scale properties

Acceptability. All the individual items of the TIS required were well accepted. The TIS scale was completed in 9.5 ± 2.8 minutes in *cohort 1* and in 8.7 ± 2.7 minutes in *cohort 2*. There were no missing items.

Floor/ceiling effects. The mean and standard deviation values of the TIS scale and its subscales for the two cohorts are reported in Figure 1. A ceiling effect was observed for the static sitting balance subscale: 34

subjects (34%) in *cohort 1* and 32 subjects (31%) in *cohort 2* achieved the highest score at baseline. For the remaining subscales and the TIS total score, less than 15% of subjects achieved the highest possible score at baseline.

[Please insert here FIGURE 1]

Reliability. Cronbach's α was acceptable in both *cohort 1* (0.794 ± 0.072) and *cohort 2* (0.726 ± 0.056). Inter- and intra-rater relative reliability were excellent for both *cohort 1* and *cohort 2*, as reported by ICC values in Table 2. The measurement error was overall quite low for both populations in terms of subscales and total score (<0.7 in *cohort 1* and <0.4 in *cohort 2*). MDC values for total score were 1.57 (inter-rater) and 1.64 (intra-rater) for acute stroke, and became even lower (1.10-0.67) for chronic stroke patients.

The inter- and intra-rater reliability of scores on individual test items were always acceptable in both *cohort 1* (0.695-1.000; 0.741-1.000) and *cohort 2* (0.823-1.000; 0.662-1.000) with percentages of agreement in the range of 93-100 (*cohort 1*) and 95-100 (*cohort 2*). Tables 3 and 4 show the results for *cohort 1* and *cohort 2*, respectively.

[Please insert here TABLE 2]

[Please insert here TABLE 3]

[Please insert here TABLE 4]

Construct validity. The Pearson's correlation coefficients between the TIS scale and the other scales (TCT, FI-motor subscale, and BI) are reported in Table 5. For *cohort 2*, a low correlation was obtained between the TIS and the motor subscale of the FIM both at baseline and after treatment. Thus, the a priori criterion chosen for the construct validity was met only for *cohort 1*.

[Please insert here TABLE 5]

Responsiveness. Table 6 shows the results of the distribution-based and anchor-based methods for determining responsiveness and MIC values for the total TIS score. Based on the GPE score 80 subjects (78%) in *cohort 1* and 65 subjects (65%) in *cohort 2* were classified as improved; 23 subjects (22%) in *cohort 1* and 35 (35%) in *cohort 2* as stable; no subject in either *cohort 1* or *2* considered their clinical condition as worse.

All the distribution based responsiveness parameters showed that TIS scale was highly responsive for *cohort 1* (>0.9) and moderately to highly responsive for *cohort 2* (>0.6). As for the ROC analyses, acceptable values of AUC (>0.7) were found for both study populations.

[Please insert here TABLE 6]

DISCUSSION

This paper describes the adaptation, validation and responsiveness of the TIS in Italian subjects with acute and chronic stroke. Analysing the psychometric properties of an outcome measure is a continuous process recommended in order to strengthen its properties and expand its applicability in specific clinical contexts and countries [24,25].

The results of the adaptation process were successfully in agreement with international guidelines. The experts played an important role during the re-evaluation of the process and confirmed the quality of the work done. The test of the pre-final version confirmed the comprehensibility of the items, leading to a valid measure of another culture's conception of health that allows data comparability and cross-national studies. The TIS scale required less than 10 minutes to be completed; it responded satisfactorily to the requirements of relevance and completeness, and seemed to be fully applicable in everyday clinical practice. Most subscales did not show floor/ceiling effects, which suggests that the TIS had enough discriminating power in subjects with acute and chronic stroke. The only exception was represented by the static sitting balance subscale which showed a ceiling effect in both cohorts, suggesting a reduced capability of discriminating subjects with highest scores. In a Norwegian acute stroke populations an even higher ceiling effect was found and the researchers suggested to maintain a prerequisite of sitting for 10 s in the starting position, and to remove the Static sitting balance subscale from the test [8]. Since we preferred to maintain the scale as it was originally developed, the high ceiling effect of this subscale might have altered the results in terms of responsiveness. To partially limit this drawback, we evaluated the responsiveness for the total score but also for all the individual subscales. Using the individual subscales scores, researchers might decide if they prefer to use the original TIS scale or eliminate the static sitting balance subscale.

Our analysis demonstrated that the Italian TIS was internally consistent with similar estimates to the original findings in both cohorts [4]. These were in line with another study conducted in a Norwegian acute stroke population ($\alpha=0.85$) [8] as well as in an Italian acute stroke population ($\alpha=0.79-0.83$) [9].

Inter-rater agreement and intra-rater reliability were satisfactory suggesting good repeatability between operator and over time in these populations; our findings were similar to those originally found [4] and higher

than Norwegian (total score, ICC of 0.77 and 0.85 for inter and intra-rater reliability, respectively) [8], and Italian study (total score, ICC of 0.93 and for 0.91 for inter and intra-rater reliability, respectively) [9]. The inter- and intra-rater reliability of scores on individual test items were also in line with the original findings confirming satisfactory repeatability between operators and over time [4].

The MDC was previously evaluated only in a sample of acute subjects reporting higher estimates both in the total and partial scores (from 1.29 to 3.08 and from 1.44 to 4.55 for inter- and intra-rater respectively) due to worst reliability values. The MDC values obtained in our study were quite low (<1.7 and <1.1 for total score in cohort 1 and cohort 2, respectively), confirming the possibility to use the TIS for evaluative purposes in longitudinal studies involving both acute and chronic stroke patients.

The Italian TIS was associated with related constructs such as trunk control and activities of daily living, supporting its construct validity and suggesting that subjects who have difficulties in trunk impairment are more physically impaired and generally show and report more physical-related limitations. Our findings are in line with those found by the original developers, giving further support to the construct validity of this scale [4]. Lower correlations were achieved with motor subscale of the FIM, suggesting items evaluated by the FIM motor subscale were only partially related to trunk control.

Distribution-based methods showed large responsiveness to the rehabilitation programme for acute stroke group, and a large to moderate responsiveness in the chronic stroke group. In previous studies estimates of ES and SRM were not provided therefore comparisons cannot be made.

Distribution-based methods should be used cautiously as they tend to measure the magnitude of change scores rather than their validity [25]. When a general measure of change, such as the GPE, is available and can be dichotomized into representative groups of improved and stable subjects, an anchor-based method, such as ROC analysis, is preferred as the AUC measures the ability of an instrument to discriminate between improved and stable subjects [24]. For both populations, the values of AUC were acceptable. The optimal cut-off point estimated on the basis of ROC analysis was 3.5 for acute and 2.5 for chronic patients, on the basis of patient's perspective. When the therapist's perspective was used for determining the MIC, a lower value (1.5) was found for chronic patients, indicating that the patients have a higher expectation of

improvement with respect to the therapists. MIC values were always higher than MDC values. Future studies might be devoted to compare the responsiveness of the TIS with other scales in order to evaluate whether it is a more responsive tool than other assessments currently used in clinics.

A 2.0 version of the TIS was recently proposed with the static sitting balance subscale omitted after Rasch analysis [26]. However, we decided to use the original scale version and to conduct a basic validation study based upon Classical Test Theory in order to compare our psychometric findings to the existing ones. Given also the ceiling effects we found when investigating the static sitting balance subscale, further analyses in Italian stroke subjects using the 2.0 TIS would be of interest to allow additional comparisons and to choose the best measure to use.

When comparing acute and chronic populations as per construct validity, higher estimates were found in the former sample, suggesting a poorer relationship between performances in trunk control and ADL in the latter population. In other terms, caution is recommended when treating chronic subjects as improvements in trunk mobility could not necessarily lead to better ADL and vice versa. Rehabilitative training that target both aims are therefore advocated. When comparing *cohort 1* and 2 in terms of minimal important change, there were again differences between acute and chronic subjects based on patients' and therapists' perspective: higher values of MIC were found for acute stroke patients than for chronic stroke. This result is in line with the higher improvements that patients experienced in an early phase after stroke and has to be taken into account when planning treatment, paying commensurate attention to functional improvement as well as trunk mobility.

Our study has some limitations. Firstly, the relationships between trunk control and physical tests, e.g. trunk strength, or flexion/relaxation response of back muscles, were not considered since only clinical scales were used. Secondly, some of the internationally most established measures used to conduct validation studies like the Short Form Health Survey-36 were not used, but researchers are recommended to analyse them in future studies on the TIS in order to further investigate its properties. Thirdly, the TIS might not have been responsive to worsening outcomes as the patients who worsened were unavailable. Fourthly, the final investigation was performed just after the end of intervention, and a significant placebo effect cannot be

excluded, limiting the clinical interpretation of the estimates found. Fifthly, the number of patients in the stable outcome group was rather low (<50 [27]), which might have limited the validity of the responsiveness analysis. Finally, the applicability of this study is limited to an Italian population and similar studies are recommended for other Countries.

CONCLUSIONS

The TIS administered in Italian subjects with acute and chronic stroke was reliable, valid and responsive. This new measure is expected to help clinicians and researchers in terms of diagnosis and therapy by identifying key functional processes related to disability in people with acute and chronic stroke.

ACKNOWLEDGMENTS

The authors would like to thank all of the health professionals and patients who took part in the study.

DECLARATION OF INTEREST STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

1. Di Monaco M, Trucco M, Di Monaco R, Tappero R, Cavanna A. The relationship between initial trunk control or postural balance and inpatient rehabilitation outcome after stroke: a prospective comparative study. *Clin Rehabil.* 2010;24:543–54.
2. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? *Age Ageing.* 2006;35 Suppl 2:ii7–ii11.
3. Verheyden G, Nieuwboer A, Van de Winckel A, De Weerd W. Clinical tools to measure trunk performance after stroke: a systematic review of the literature. *Clin Rehabil.* 2007;21:387–94.
4. Verheyden G, Nieuwboer A, Mertin J, Preger R, Kiekens C, De Weerd W. The Trunk Impairment Scale: a new tool to measure motor impairment of the trunk after stroke. *Clin Rehabil.* 2004;18:326–34.
5. Verheyden G, Vereeck L, Truijen S, Troch M, Herregodts I, Lafosse C, Nieuwboer A, De Weerd W. Trunk performance after stroke and the relationship with balance, gait and functional ability. *Clin Rehabil.* 2006;20:451–8.
6. Verheyden G, Nieuwboer A, Feys H, Thijs V, Vaes K, De Weerd W. Discriminant ability of the Trunk Impairment Scale: A comparison between stroke patients and healthy individuals. *Disabil Rehabil.* 2005;27:1023–8.
7. Verheyden G, Nieuwboer A, De Wit L, Feys H, Schuback B, Baert I, Jenni W, Schupp W, Thijs V, De Weerd W. Trunk performance after stroke: an eye catching predictor of functional outcome. *J Neurol Neurosurg Psychiatry.* 2007;78:694–8.
8. Gjelsvik B, Breivik K, Verheyden G, Smedal T, Hofstad H, Strand LI. The Trunk Impairment Scale - modified to ordinal scales in the Norwegian version. *Disabil Rehabil.* 2012;34:1385–95.
9. Lombardi B, Orioli A, Casavola D, Paci M. The Italian version of the Trunk Impairment Scale: development and psychometric properties. *Eur J Phys Rehabil Med.* 2017 April 4 [EPUB].
10. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25:3186–91.
11. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of Good

Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Heal.* 2005;8:94–104.

12. Monticone M, Ambrosini E, Ferrante S, Colombo R. “Regent Suit” training improves recovery of motor and daily living activities in subjects with subacute stroke: a randomized controlled trial. *Clin Rehabil.* 2015;27:792–802.
13. Galeoto G, Lauta A, Palumbo A, Sf C, Mollica R, Santilli V, MI S, Galeoto G. The Barthel Index: Italian Translation, Adaptation and Validation. *Int J Neurol Neurother.* 2015;2.
14. Pallicino P, Snyder W, Granger C. The NIH stroke scale and the FIM in stroke rehabilitation. *Stroke.* 1992;23:919.
15. Franchignoni FP, Tesio L, Ricupero C, Martino MT. Trunk control test as an early predictor of stroke rehabilitation outcome. *Stroke.* 1997;28:1382–5.
16. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidemiol.* 2010;63:760–766.e1.
17. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42.
18. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102–109.
19. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* 2000;53:459–68.
20. Cohen J. *Statistical power analysis for the behavioral sciences.* Academic Press; 1977.
21. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* 1989;27:S178-89.
22. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol.*

2001;54:1204–17.

23. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39:561–77.
24. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol*. 2010;10:22.
25. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011.
26. Verheyden G, Kersten P. Investigating the internal validity of the Trunk Impairment Scale (TIS) using Rasch analysis: the TIS 2.0. *Disabil Rehabil*. 2010;32:2127–2137.
27. Schuller W, Ostelo RW, Janssen R, de Vet HC. The influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the neck disability index. *Health Qual Life Outcomes*. 2014;12:53.

Table 1. Socio-demographic characteristics of subjects with acute (N=103) and chronic stroke (N=100).

	Acute stroke	Chronic Stroke
Age, years ^a	74.50 (10.79)	67.18 (12.23)
Males / females	56/47	58/42
Body mass index (kg/m ²) ^a	22.73(4.72)	25.77(3.70)
Time post brain lesion ^b	22 days (range 16-38)	12 months (range 6-108)
Type of lesion		
Ischemic	83	81
Haemorrhagic	20	19
Localization of lesion		
Right	49	42
Left	47	50
Bilateral	7	8
Most affected body half		
Right	48	49
Left	50	44
Bilateral	5	7
Married (Yes/no)	83/20	84/16
Education		
Primary school	58	36
Middle school	21	30
High school	19	27
University	5	7
Working status		
Employed	7	12
Domestic works	5	3
Retired	91	85
Smoking (Yes/no)	11/92	17/83

^a Mean values (standard deviation)^b Median values (range)

Table 2. Results of the ICC reliability analysis

	Inter-rater Reliability ICC [95% CI]; SEM; MDC	Intra-rater Reliability ICC [95% CI]; SEM; MDC
<i>Cohort 1</i>		
Static Sitting Balance	0.998 [0.998 0.999]; 0.26; 0.73	0.991 [0.986 0.994]; 0.72; 2
Dynamic Sitting Balance	0.977 [0.966 0.984]; 0.39; 1.07	0.980 [0.970 0.986]; 0.36; 1
Co-ordination	0.921 [0.886 0.946]; 0.41; 1.14	0.968 [0.953 0.978]; 0.26; 0.72
Total	0.989 [0.983 0.992]; 0.57; 1.57	0.988 [0.983 0.992]; 0.59; 1.64
<i>Cohort 2</i>		
Static Sitting Balance	0.996 [0.994 0.997]; 0.10; 0.28	0.975 [0.963 0.983]; 0.25; 0.70
Dynamic Sitting Balance	0.974 [0.962 0.983]; 0.34; 0.94	0.992 [0.989 0.995]; 0.19; 0.52
Co-ordination	0.979 [0.968 0.985]; 0.23; 0.64	0.998 [0.997 0.999]; 0.07; 0.20
Total	0.989 [0.984 0.993]; 0.40; 1.10	0.996 [0.994 0.997]; 0.24; 0.67

ICC, intraclass correlation coefficient

SEM, standard error of measurement

MDC, minimal detectable change

Table 3. Kappa or weighted kappa, lower value of the 90% confidence limit of the kappa or weighted kappa and percentage of agreement for inter- and intra-rater agreement in *cohort 1*.

Inter-rater agreement						Intra-rater agreement			
	k/ wk ^a	Value ^b	SE	90% IC_lb ^c	% ^e	Value ^b	SE	90% IC_lb ^c	% ^e
Static Sitting Balance									
Item1	k	d			100	d			100
Item2	k	1.000	0.000	1.000	100	0.977	0.023	0.939	99
Item3	wk	0.992	0.008	0.980	99	0.947	0.027	0.903	96
Dynamic Sitting Balance									
Item1	k	0.957	0.030	0.907	98	0.979	0.021	0.944	99
Item2	k	0.847	0.056	0.755	93	0.895	0.045	0.821	95
Item3	k	NaN ^d			98	NaN ^d			100
Item4	k	0.904	0.047	0.827	96	1.000	0.000	1.000	100
Item5	k	0.961	0.027	0.916	98	0.902	0.043	0.832	95
Item6	k	NaN ^d			94	NaN ^d			98
Item7	k	0.975	0.025	0.933	99	0.897	0.050	0.815	96
Item8	k	NaN ^d			98	NaN ^d			98
Item9	k	0.909	0.044	0.837	96	0.911	0.043	0.840	96
Item10	k	NaN ^d			98	NaN ^d			98
Co-ordination									
Item1	wk	0.858	0.043	0.788	90	0.958	0.024	0.919	97
Item2	k	NaN ^d			93	NaN ^d			98
Item3	wk	0.940	0.030	0.891	96	0.912	0.036	0.852	94
Item4	k	NaN ^d			99	NaN ^d			99

^a k/wk indicates the use of a kappa (k) or weighted kappa (wk) for statistic analysis.

^b Value of the calculated kappa or weighted kappa.

^c Lower value of the 90% confidence limit of the kappa or weighted kappa.

^d NaN, not available. No Kappa or weighted Kappa could be calculated because of the skewed distribution of the data.

^e Percentage of agreement

Table 4. Kappa or weighted kappa, lower value of the 90% confidence limit of the kappa or weighted kappa and percentage of agreement for inter- and intra-rater agreement in *cohort 2*.

Inter-rater agreement						Intra-rater agreement			
	k/ wk ^a	Value ^b	SE	90% IC_lb ^c	% ^e	Value ^b	SE	90% IC_lb ^c	% ^e
Static Sitting Balance									
Item1	k	NaN ^d			100	NaN ^d			99
Item2	k	NaN ^d			100	NaN ^d			100
Item3	wk	0.944	0.032	0.891	98	0.954	0.032	0.902	98
Dynamic Sitting Balance									
Item1	k	NaN ^d			100	NaN ^d			100
Item2	k	0.960	0.028	0.914	98	0.980	0.020	0.947	99
Item3	k	NaN ^d			96	NaN ^d			99
Item4	k	NaN ^d			100	NaN ^d			100
Item5	k	0.955	0.032	0.902	98	0.977	0.023	0.939	99
Item6	k	0.903	0.047	0.826	96	0.901	0.049	0.820	96
Item7	k	0.975	0.024	0.936	99	0.899	0.049	0.818	96
Item8	k	NaN ^d			97	NaN ^d			98
Item9	k	0.918	0.040	0.853	96	0.899	0.044	0.827	95
Item10	k	NaN ^d			99	NaN ^d			99
Coordination									
Item1	wk	0.896	0.057	0.803	97	0.982	0.018	0.951	99
Item2	k	0.945	0.034	0.889	98	1.000	0.000	1.000	100
Item3	wk	0.967	0.023	0.930	98	1.000	0.000	1.000	100
Item4	k	NaN ^d			100	NaN ^d			100

^a k/wk indicates the use of a kappa (k) or weighted kappa (wk) for statistical analysis.

^b Value of the calculated kappa or weighted kappa.

^c Lower value of the 90% confidence limit of the kappa or weighted kappa.

^d NaN, non available. No Kappa or weighted Kappa could be calculated because of the skewed distribution of the data.

^e Percentage of agreement

Table 5. Results of the construct validity analysis: correlation between the total score of the TIS and TCT, FIM motor subscale, and BI.

	Means (SD)		Pearson's coefficient	
	Pre-training	Post-training	Pre-training	Post-training
<i>Cohort 1</i>				
TCT	57.8 (31.3)	83.1 (23.9)	0.852**	0.669**
FIM, motor subscale	32.4 (12.9)	64.8 (18.7)	0.567**	0.731**
BI	5.4 (3.2)	13.8 (5.6)	0.669**	0.698**
<i>Cohort 2</i>				
TCT	75.1 (23.3)	87.8 (18.7)	0.673**	0.677**
FIM, motor subscale	56.5 (12.9)	71.9 (12.2)	0.010	0.310**
BI	13.7 (3.5)	17.4 (3.6)	0.477**	0.460**

** p<0.01

TCT, Trunk Control Test

FIM, Functional Independent Measure

BI, Barthel Index

TIS, Trunk Impairment Scale

SD, standard deviation

Table 6. Results of the responsiveness analysis for the total TIS score.

	ES	SRM	ES Guyatt		MIC [AUC; Sensitivity; Specificity]	
			Patient's perspective	Therapist's perspective	Patient's perspective	Therapist's perspective
Cohort 1	0.997	1.634	1.390	1.386	3.5 [0.709; 70; 52]	3.5 [0.709; 70; 52]
Cohort 2	0.633	0.985	0.624	0.727	2.5 [0.728; 49; 82]	1.5 [0.713; 73; 57]

ES, Effect Size
SRM, Standardized Response Mean
MIC, Minimal important Change
AUC, Area Under the Curve

FIGURE CAPTIONS

Figure 1. Mean and standard deviation values of the TIS subscale and total score for the two study populations (*cohort 1*, acute stroke, in light grey; *cohort 2*, chronic stroke, in dark grey). “Rater 1” and “Rater 2” indicate the values obtained by the two raters at baseline, “Retest” shows the values obtained at the second evaluation of one of the two raters at baseline. Finally, “Post” indicate the values achieved after training.