

The Postural Assessment Scale for Stroke Patients: Translation into Norwegian, cultural adaptation and examination of reliability

Abstract

Background: To translate the Modified Swedish version of Postural Assessment Scale for Stroke (SwePASS) into Norwegian (SwePASS-NV) and assess its intra- and interrater reliability. **Methods:** Translation and cross-cultural adaptation was accomplished according to international guidelines. Forty-seven adults, suffering a stroke within the last six months, were strategically recruited and allocated into groups according to five different levels of function. Their performance on SwePASS-NV was video-recorded and then scored by three raters twice, with a minimum of 4 weeks between sessions. Relative reliability was investigated by calculating intraclass correlation coefficients ($ICC_{1,1}$ and $ICC_{3,1}$). Absolute reliability was analysed using within-subject standard deviation (S_w) and smallest detectable difference (SDD). For individual items, Cohen's kappa (k) and percentage of agreement were calculated. **Results:** The study showed excellent intra- and interrater reliability ($ICC_{1,1}$ and $ICC_{3,1} \geq .99$) for SwePASS-NV. K -values for the individual items ranged between 0.68–1.00. Percentages of agreement ranged from 77-100%. SDD at 95% confidence interval was ≤ 2 points for intrarater assessments and ≤ 3 points for interrater assessments. **Conclusion:** This study demonstrated very high intra- and interrater reliability of the SwePASS-NV in adults within the first six months after stroke. All items showed very high or high agreement.

Keywords: SwePASS-Norwegian version; intrarater reliability; interrater reliability; balance; balance measure; stroke.

Introduction

Stroke is the third leading cause of death and a major cause of lasting disability among the elderly, and in Norway approximately **15,000** persons suffer a stroke each year (1). This number is expected to rise due to a presumed increase in the number of older people in the decades to come (2). Impaired balance is a common problem amongst these patients, affecting up to 80% in the initial stages (3). Reduced balance can severely influence the patient's ability to perform activities of daily living and is associated with reduced quality of life (4). There is a need for reliable measurements designed to evaluate and monitor balance in the initial phase after stroke and throughout the rehabilitation process (5, 6). Good measurements of balance may predict later disability, assist in selecting the most appropriate therapy, and assess changes (7). Several different tests exist but there is no consensus on a 'gold standard' (8). Many of the tests used today are not optimal for patients with poor balance, due to a floor effect (9, 10).

The Postural Assessment Scale for Stroke Patients (PASS) was developed specifically to evaluate balance after stroke and is applicable for all including those with poor postural control (9). It has shown high intrarater and interrater reliability (9), good individual item agreement (11), and acceptable or high test-retest reliability (7, 12). PASS has been examined for different aspects of validity in several studies, showing good to excellent validity (9, 11, 13-15). Clinical instruments should be used in the patient's native language to ensure a common understanding of the original intention with the test. PASS has previously been translated into English, Swedish, Portuguese and Spanish (9, 16-18). The Swedish version of PASS, SwePASS, is a new and modified version of the test (16). The aim of the present study was to translate the SwePASS into Norwegian and to examine its intrarater and interrater reliability in the acute and sub-acute phases after stroke.

Methods

Design

The present study was conducted in two stages. First, the English back-translation of SwePASS, the SwePASS-EV, was translated into Norwegian, SwePASS-NV, and cross-culturally adapted according to international guidelines. Second, a methodological study of intrarater and interrater reliability of SwePASS-NV was performed. The participants' performance of SwePASS-NV was video recorded and the recordings used to assess intrarater

and interrater reliability. **The Norwegian Social Science Data Services and the data protection officer at HUH approved the study. The participants gave their written informed consent. If the participant was not able to understand the information, the next of kin gave informed consent.**

Setting

The study was conducted from October 2013 to December 2014 at the Department of Physical Medicine and Rehabilitation (DPMR) and the Department of Neurology at Haukeland University Hospital (HUH), Bergen, Norway.

PASS and SwePASS

PASS is used to assess the ability to maintain or change positions in lying, sitting and standing, and contains items with varying degrees of difficulty (9). PASS consists of twelve items (scoring 0-3, 3=highest/best) with a maximum total score of 36 points. The SwePASS modifications include a change in the order of the items, which in the SwePASS are listed in the same order as they are logically performed in a clinical situation. Moreover, in item 10 of PASS the participant is asked to pick up an object/pencil, which in SwePASS is changed to a shoe/slipper, to minimise a possible impact of fine motor skills. In addition, the scoring criteria have been clarified, where 'much help' is defined as 'support of two persons' and 'little help' as 'support of one person'. These modifications and clarifications are in accordance with our clinical experience and in line with a previous study (19).

Translation and cross-cultural adaptation

SwePASS-EV was translated into Norwegian following international guidelines (20) after consent from the developers of PASS (9) and SwePASS (16). Two physiotherapists and a speech therapist, all bi-lingual (**Norwegian/English**), translated SwePASS-EV into Norwegian separately. The three versions were compared and consensus reached for a first draft. In a pilot study including five stroke patients, the test was tried out clinically. The first draft together with the experiences and questions from the pilot study were discussed by an expert panel consisting of the forward translators and the first (KB) and second (BEBG) authors. Consensus was reached for a second draft of SwePASS-NV. A bi-lingual (**English/Norwegian**) colleague with no previous experience with PASS or SwePASS

translated the second version back into English.

Reliability study

Study sample

Acute and sub-acute patients with stroke were recruited by physiotherapists working at the two departments. Inclusion criteria: diagnosed with stroke and ≥ 18 years of age. Stroke was defined according to the World Health Organisation criteria (21), and the diagnosis confirmed with MRI or CT. Individuals that could not follow test instructions or had impaired balance due to other diagnoses were not included. In order to enable evaluation of the whole range of scores of the SwePASS-NV, the participants were strategically selected based on their level of function and allocated to the following groups by their treating physiotherapist:

- Group 1 Cannot sit without support.
- Group 2 Can sit without support for 10 seconds; need much support/help of two persons when transferring.
- Group 3 Can sit without support for more than 10 seconds; need only slight support/supervision with transfers.
- Group 4 Can transfer independently, in a secure manner.
- Group 5 Can stand on the affected leg without support or walk independently without a walking aid.

Age, sex and information about the stroke (date of incident, type of stroke, brain localisation and most affected body half) were registered for all participants.

Testing procedures

A research assistant video-recorded the participants during their performance of SwePASS-NV in accordance with a pre-standardised procedure. The testers were all physiotherapists who worked at the two recruiting departments. Their experience with neurological patients and with PASS/SwePASS varied from none to extensive. To ensure that they all had a basic degree of common understanding, they were given an hour training session by the first author before testing commenced.

At the respective departments, the same rooms were used during testing. The angle and distance of the video-recordings and the equipment used (a wide treatment plinth and a

stopwatch) were standardised. Testing was conducted in accordance with a manual developed for the SwePASS-NV and the items were performed in the same order each time. If two persons were needed to assist the participant, the research assistant would help. The participants had the opportunity to take breaks if necessary. Time taken to complete the test varied from eight to fifteen minutes.

Raters and rating procedure

Three physiotherapists who were not involved in the testing, rated the participants' performance on SwePASS-NV based on the video recordings, which were not edited in any way. These raters, labelled A, B and C, were selected among the staff working with stroke patients at DPMR. Raters A and B had 12 and 10 years clinical experience, respectively, and mostly with neurological patients. They had previous experience with the PASS from clinical work and other studies. Rater C was a junior physiotherapist with no experience with neurorehabilitation or PASS or SwePASS.

Prior to the main study, the raters participated in two one-hour training sessions to become familiar with the test and scoring instructions, and create a common understanding. This was achieved by rating five stroke patients (those included in the pilot study) on their performance of SwePASS-NV. The recordings were scored individually and then discussed between them.

The three raters scored the video recordings of the participants twice, with a minimum of four weeks between the first (A1, B1, C1) and second (A2, B2, C2) rating. The raters used a stopwatch each when scoring items 7, 8 and 9. Each rater assigned scores independently, and was instructed not to discuss the scores with each other during or between the two ratings. The order of the video recordings differed between the two ratings. With the intention to make the rating as similar as possible to direct scoring in the clinic, the raters watched the video recordings once and scored during the viewing. A fresh SwePASS-NV assessment form was distributed for each participant and the forms were unavailable for the raters after the assessment was completed. The raters were not informed about the participants' group allocation.

Data analysis

Characteristics of the participants are presented as means, standard deviation (SD), and range

for continuous variables or percentages for dichotomous variables. The correlation between the total score on SwePASS-NV (mean value of all six ratings) and the group allocation was examined using Spearman's rho statistics.

Intraclass correlation coefficients (ICC) were calculated to assess relative intrarater and interrater reliability of the total score of SwePASS-NV. ICCs were calculated to assess pairwise correlation between raters (A1-B1, A1-C1, B1-C1), within the group of raters (A1-B1-C1), and within the two ratings by the same rater (A1-A2, B1-B2, C1-C2). ICC_{1,1} was used because the raters were strategically chosen (22). In addition, ICC_{3,1} was used to examine whether systematic errors occurred. When ICC_{1,1} equals ICC_{3,1}, no systematic errors are present (22). The ICC ranges from 0-1, where 0 indicates no agreement and 1 indicates perfect agreement. Reference values for ICC: .00-.25=little if any, <0.50=low, 0.50-0.69=moderate; 0.70-0.89 high, and 0.90-1.00 very high (23). Portney and Watkins (24) recommend that for many clinical measurements, reliability should exceed .90 to ensure reasonable validity.

Within subject standard deviation (S_w) was calculated to investigate absolute reliability. S_w is used to investigate how a given sum of the SwePASS-NV is related to a 'true' score for that person, and the variability in total scores with repeated observations, expressed in scores on SwePASS-NV (25). S_w was calculated using analysis of variance (ANOVA), where S_w equals the square root of the within-people residual mean square (25). The difference between a score and the true value for an individual is expected to be less than 1.96 S_w for 95% of the observations (25). The difference between two scores for the same participant is expected to be $<\sqrt{2} \times 1.96 S_w$ of the pairwise observations. This value is an estimate of the minimum change in scores needed to ensure that the change in scores is greater than the measurement error, and is referred to as the smallest detectable difference (SDD) (26).

The reliability of each item of the SwePASS-NV was assessed by kappa (k) statistics. The reference values are: $k < 0.20$ =poor, $k 0.21-0.40$ =weak, $k 0.41-0.60$ =moderate, $k 0.61-0.80$ =high, and $k 0.81-1.0$ =very high (27). A prerequisite for the use of k is that all scoring alternatives for an item are used. Percentage agreement was calculated for all items in addition to k . In accordance with Gjelsvik (28) and Verheyden (29) we considered 80% agreement as acceptable.

The data were analysed with the 22. version of IBM SPSS for Windows (IBM SPSS, Armonk, New York).

Results

Translation and cross-cultural adaptation

We chose to translate the English back-translation of the SwePASS, SwePASS-EV, into Norwegian, as this was the version of SwePASS approved by the original developer, Charles Benaim.

During the translation process we drew attention to the fact that neither SwePASS nor PASS stated whether the participants should be barefoot or wear shoes during testing. The translation process also detected some inconsistencies between the SwePASS and their English back translation. The developers were kept up to date and given the opportunity to respond to our remarks and questions during the process to insure agreement between the different versions. Several new versions of SwePASS were posted online during this time (30). Since the changes corresponded to our desired corrections, we chose to include these in the final version of SwePASS-NV (appendix 1 and 2). An overview of particular challenges encountered during the process of translation and information on how they were approached can be obtained from the corresponding author. SwePASS-NV formed the basis for the next part of the study, examination of intra- and interrater reliability.

Intrarater and interrater reliability study

Characteristics of the study sample are presented in Table I. Twenty participants were included at the Department of Neurology and 27 participants at DPMR, HUH. Of the 47 participants, 55% were male and 81% suffered ischemic infarction. Time between stroke onset and date of testing ranged between 1 and 136 days. The sample size was just short of the 50 recommended for reliability studies (31).

(Insert Table I about here)

The total scores on SwePASS-NV ranged from 3-36 (Table II) and correlated significantly with the participants' group allocation based on their level of function (Spearman's $\rho=0.88$, $P<0.001$). The relationship between group allocation and mean total score is illustrated in Figure 1. The distribution between the groups was somewhat skewed with a large proportion in group 3 ('needs only slight support with transfers', $n=18$) and a small proportion in group 1 ('no independent sitting balance', $n=3$).

(Insert Table II and Fig. 1 about here)

A large proportion of the sample (77%) received a total score on the upper half of the scale with 29% having a score between 31-33, and seven percent with a score between 34 and 36.

Relative reliability of both intrarater and interrater assessments was high ($ICC \geq 0.99$) (Table III).

(Insert Table III about here)

S_w and SDD for intrarater and interrater reliability of the total score are reported in Table III. ANOVA calculation of within-subject residual mean square was 0.688 for interrater analyses for all raters (A1-B1-C1), and was calculated to 0.83 by use of the equation $\sqrt{0.688 S_w}$. The difference between a participant's total score and the 'true' measurement value was expected to be <1,62 points for 95% of the scores ($\pm 1,96 \times 0.83$). The SDD of the total score between two measurements for the same participant ranged between 1.7 and 2.7 points (95% CI) when scored by the different raters and 1.4-1.9 (95% CI) when scored by the same rater. k ranged from 0.68-1.00 (Table IV). For both intrarater and interrater assessments the items where k could be calculated (item 5-12) showed very high or high agreement. Percentages of agreement ranged from 77-100%, with item 7 showing the lowest values. Because all score options were not used, k could not be estimated for item 1-4. Instead, the percentage agreement was calculated, showing a range from 94-100%.

(Insert Table 4 about here)

Time between the first and second scoring (rating) was a median of 34 days (range 27-168 days).

Discussion

The SwePASS-NV total score for both intrarater and interrater assessments demonstrated very high relative and absolute reliability. All items showed high to very high agreement.

Translation and cross-cultural adaptation

SwePASS was translated according to internationally recommended guidelines (20, 32). The back-translation was thoroughly discussed by the expert panel and corresponded well with the original, and was considered satisfactory. The differences primarily dealt with wording, not affecting content. For example, SwePASS-EV uses the wording 'without any help' while we use the more exact wording 'independently'. The test procedure (appendix 2) also differs

slightly: While SwePASS-EV writes: 'In items 1-3, 5-6 and 10-12 the patient's postural balance/control should be scored according to different degrees of support (verbal, tactile or supervision)' the expert panel chose the following wording: 'Items 1-3, 5-6 and 10-12 score the patient's balance with respect to the degree of support/help described in the test. Help/support can be in form of verbal prompting, physical facilitation or supervision.' This change was done with the intention to make the test as unambiguous as possible to ensure that it is understood and used in the same way by therapists throughout Norway (cultural adaptation). PASS was developed in French and later translated into English. SwePASS was developed in Sweden, Norway's neighbouring country, with a very similar culture. We believe that we achieved a satisfactory translation reflecting the intentions of the developers of both PASS and SwePASS.

Reliability study

The high relative reliability ($ICC \geq .99$) implies that most participants maintained their positions within their initial allocated groups with repeated measurements (33). Moreover, since $ICC_{1,1}$ and $ICC_{3,1}$ were almost identical, one can assume that no systematic errors influenced the calculations of relative reliability (22). These results are similar to another study using video to assess the reliability of the Spanish version of PASS ($ICC = .99$) (18) and similar to or higher than studies using direct scoring (interrater $ICC = .97$) (11). The SwePASS has also been found to have high reliability (16). However, because different statistical methods were used direct comparison of the results are impossible.

The low S_w for all raters (range 0.7-1.0 point) indicates low measurement error of SwePASS-NV. A low measurement error is essential both for clinical and research purposes and especially for the assessment of real changes in balance over time (31). Calculations of SDD imply that a change in score of at least 1.9 (2) points (intrarater) and 2.7 (3) points (interrater) can be interpreted as a real change when two measurements of the same participant are compared (25). In a study by Liaw et al (12) examining patients with chronic stroke using PASS and direct scoring a higher SDD was found ($SDD = 4$ points). This may be due to more precise wording in the scoring options in SwePASS-NV making it easier to score than PASS, or because the use of video is not affected by a potential difference in the participants' performance. The measurement error may thus be considered to be artificially

low. The lower S_w and SDD interrater values from the second scoring indicate that experience with the test may further lower the measurement error.

All individual items had high or very high k values for both intrarater and interrater agreement. All but one item (item 7, interrater B-C= 77%) had values above 80%. The lowest interrater agreement was found for items 6 (standing with support) and 7 (standing without support). It might be difficult to differentiate between level 2 and 3 regarding the amount of support needed on item 6; and for item 7 both time and degree of asymmetry must be considered. These differentiations may be easier to observe for the more experienced raters (A and B). Despite a tendency for lower k values for rater C on these items, the values are still high, suggesting that the reliability of SwePASS-NV may be independent of experience.

Because all scoring alternatives were not used, k could not be estimated for items 1-4. We were not able to recruit more than three participants with the lowest level of function. Therefore, the reliability of the lower part of the scale was examined to a lesser extent than the other parts. It could be argued that the high agreement in our study may partly be due to the fact that a score of 0 (not able to perform) on items 1-3 is rare, because most patients can be turned on their sides by two persons. Therefore, the scoring option 0 on these items may be redundant. However, PASS was specifically developed to evaluate patients with poor postural control. Changing the scoring options could potentially lead to a higher floor effect, which would negate its use for these patients. Also, we found a strong relationship between levels of function of the participants and total score on SwePASS-NV (fig. 1).

A floor or ceiling effect is considered to be present if 15% or more of the participants achieve the lowest or highest score. A floor or ceiling effect indicates low reliability and limits the responsiveness of the test as a change of performance in these participants cannot be measured (31). Seven percent of our sample received a score between 34-36, indicating no ceiling effect. However, with a large proportion of the sample (29%) receiving a score between 31-33, a higher measurement error than shown in this study could indicate a ceiling effect. With a median of 19 days between onset and testing and the fact that postural control is likely to improve during the rehabilitation process, it is also possible that SwePASS-NV may have a ceiling effect later on. This is in line with studies on both PASS and SwePASS showing that the tests are especially suited to measure change during the first three months (9, 11, 34).

Strengths and limitations of the study

The present study has several strengths, as highly standardised procedures following current recommendations for translation and cultural adaptation as well as recommendations with regard to methodological evaluation of balance measures combined with a comprehensive analysis of rater reliability were used (20, 26, 32, 35).

Concerning the examination of intrarater reliability, there are no standard criteria regarding the time interval between assessments. We consider that the time in our study (median 34 days) is sufficient to minimise the influence of the raters' memory. With direct scoring at an early stage there is a real probability that the patient's performance changes between assessments. By using video we ensured that variability in the scoring is not a result of this. A reliability study allowing the use of three raters would not be possible using direct scoring. Due to the fact that the validity of PASS has been extensively investigated in other versions/languages (9, 13-15) an examination of validity was not included in this study. However, due to inconclusive results on the validity of SwePASS further studies may be called for, including a larger sample with poorer postural control (36).

The assessment made by a therapist based on video may not be a complete reflection of the judgment made in live situations. We aimed to make the rating situation as similar as possible to direct scoring by allowing the raters to view the recording only once. Thereby, we ensured that agreement was not a result of watching the participants' performance several times. The scorings were based on the testers' evaluation regarding the amount of personal assistance/supervision needed to perform the items. Therefore, the raters were not completely free to make their own judgement when assigning scores. However, when using direct scoring in the clinic, experience with neurological patients could potentially influence the assignment of scores to a greater degree than in this study.

The use of strategic sampling of participants and raters may limit the study's external validity (37). On the other hand, the variety of the participants' level of function making it possible to assess almost the whole range of SwePASS-NV scores and using testers and raters with different degrees of clinical experience with neurological patients, add to the generalisability of the results. Individuals with major cognitive impairments were not included in the study and results can therefore not be generalised to this group.

Clinical relevance

PASS and SwePASS are considered as tests of high clinical relevance. In a study from 2013 including 77 outcome measures for stroke, PASS was selected as one of 14 measurements highly recommended for use in the clinic and in research (38). PASS and SwePASS include examination of activities that are essential to daily living, and thereby an evaluation of a wide spectre of activities that has been identified as important predictors for functional outcome (14, 39, 40). Moreover, PASS has been found to predict ambulation ability in patients with ischemic stroke after rehabilitation (41). Fewer changes in test positions in SwePASS make this test less exhausting for patients and less time consuming than PASS. SwePASS requires equipment available at most physiotherapy departments, it is free online and needs no extensive training.

In conclusion, SwePASS-NV is a reliable balance test for adults with stroke within the first six months, and has no floor effect. Thereby Norwegian physiotherapists **and other health practitioners** have a new tool to assess a wide range of stroke patients and evaluate the effect of interventions.

References

1. Ellekjær H, Holmen J, Indredavik B, Terent A. Epidemiology of stroke in Innherred, Norway, 1994 to 1996. Incidence and 30-day case-fatality rate. *Stroke; a journal of cerebral circulation*. 1997;28(11):2180-4.
2. Mackay Judith, Mensah George A., Mendis Shanthy, Greenlund Kurt. The Atlas of Heart Disease and Stroke WHO global digital library: Geneva : World Health Organization; 2004 [cited 2016 06.06.16]. Published in collaboration with the U. S. Centers for Disease Control and Prevention]. Available from: http://who.int/cardiovascular_diseases/en/cvd_atlas_15_burden_stroke.pdf?ua=1.
3. Tyson SF, Hanley M, Chillala J, Selley A, Tallis RC. Balance disability after stroke. *Physical therapy*. 2006;86(1):30-8.
4. Schmid AA, Van Puymbroeck M, Altenburger PA, Miller KK, Combs SA, Page SJ. Balance is associated with quality of life in chronic stroke. *Topics in stroke rehabilitation*. 2013;20(4):340-6.
5. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Reliability. Measurement in Medicine*. Cambridge: Cambridge University Press; 2011. p. 96-149.
6. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disability and rehabilitation*. 2005;27(6):315-40.
7. Chien CW, Hu MH, Tang PF, Sheu CF, Hsieh CL. A comparison of psychometric properties of the smart balance master system and the postural assessment scale for stroke in people who have had mild stroke. *Archives of physical medicine and rehabilitation*. 2007;88(3):374-80.
8. Tyson SF, Connell LA. How to measure balance in clinical practice. A systematic review of the psychometrics and clinical utility of measures of balance activity for neurological conditions. *Clinical rehabilitation*. 2009;23(9):824-40.
9. Benaim C, Pérennou DA, Villy J, Rousseaux M, Pelissier JY. Validation of a standardized assessment of postural control in stroke patients: The Postural Assessment Scale for Stroke Patients (PASS). *Stroke; a journal of cerebral circulation*. 1999;30(9):1862-8.
10. de Oliveira CB, de Medeiros IR, Frota NA, GreTERS ME, Conforto AB. Balance control in hemiparetic stroke patients: main tools for evaluation. *Journal of rehabilitation research and development*. 2008;45(8):1215-26.
11. Mao HF, Hsueh IP, Tang PF, Sheu CF, Hsieh CL. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke; a journal of cerebral circulation*. 2002;33(4):1022-7.
12. Liaw LJ, Hsieh CL, Lo SK, Chen HM, Lee S, Lin JH. The relative and absolute reliability of two balance performance measures in chronic stroke patients. *Disability and rehabilitation*. 2008;30(9):656-61.
13. Chien CW, Lin JH, Wang CH, Hsueh IP, Sheu CF, Hsieh CL. Developing a Short Form of the Postural Assessment Scale for people with Stroke. *Neurorehabilitation and neural repair*. 2007;21(1):81-90.
14. 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. *Clinical rehabilitation*. 2010;24(6):543-54.
15. Wang CH, Hsueh IP, Sheu CF, Yao G, Hsieh CL. Psychometric properties of 2 simplified 3-level balance scales used for patients with stroke. *Physical therapy*. 2004;84(5):430-8.
16. Persson CU, Hansson PO, Danielsson A, Sunnerhagen KS. A validation study using a modified version of Postural Assessment Scale for Stroke Patients: Postural Stroke Study in Gothenburg

- (POSTGOT). Journal of neuroengineering and rehabilitation [Internet]. 2011 3203036]; 8:[57 p.]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21978462>
<http://www.jneuroengrehab.com/content/pdf/1743-0003-8-57.pdf>.
17. Yoneyama S, Roiz R, Oliveira T, Oberg T, Lima N. Validation of the Brazilian version of the Postural Assessment Scale for Stroke Patients. *Acta Fisiatr.* 2008;15(2):96-100.
 18. Cabanas-Valdes R, Girabent-Farres M, Canovas-Verge D, Caballero-Gomez FM, German-Romero A, Bagur-Calafat C. [Spanish translation and validation of the Postural Assessment Scale for Stroke Patients (PASS) to assess balance and postural control in adult post-stroke patients]. *Revista de neurologia.* 2015;60(4):151-8.
 19. Hofstad H, Gjelsvik BE, Næss H, Eide GE, Skouen JS. Early supported discharge after stroke in Bergen (ESD Stroke Bergen): three and six months results of a randomised controlled trial comparing two early supported discharge schemes with treatment as usual. *BMC neurology.* 2014;14(1):239.
 20. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186-91.
 21. Hatano S. Experience from a multicentre stroke register: a preliminary report. *Bulletin of the World Health Organization.* 1976;54(5):541-53.
 22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin.* 1979;86(2):420-8.
 23. Munro BH. Correlation. *Statistical methods for health care research.* 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2005. p. 223-45.
 24. Portney LG, Watkins MP. *Statistical Measures of Reliability. Foundations of clinical research : applications to practice.* 3rd ed. Upper Saddle River: Pearson Prentice Hall; 2009. p. 585-618.
 25. Bland JM, Altman DG. Measurement error. *BMJ.* 1996;313(7059):744.
 26. Moe-Nilssen R, Nordin E, Lundin-Olsson L. Criteria for evaluation of measurement properties of clinical balance measures for use in fall prevention studies. *Journal of evaluation in clinical practice.* 2008;14(2):236-40.
 27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.
 28. Gjelsvik B, Breivik K, Verheyden G, Smedal T, Hofstad H, Strand LI. The Trunk Impairment Scale - modified to ordinal scales in the Norwegian version. *Disability and rehabilitation.* 2012;34(16):1385-95.
 29. 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. *Clinical rehabilitation.* 2004;18(3):326-34.
 30. Sahlgrenska akademien, Institutionen för neurovetenskap och fysiologi, Göteborgs Universitet. SwePASS [updated 01.04.2016; cited 2017 18.04.]. Available from: http://neurophys.gu.se/sektioner/klinisk-neurovetenskap/forskning/rehab_med/swepass.
 31. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology.* 2007;60(1):34-42.
 32. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. 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 in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2005;8(2):94-104.

33. Portney LG, Watkins MP. Reliability of Measurements. Foundations of clinical research : applications to practice. 3rd ed. Upper Saddle River: Pearson Prentice Hall; 2009. p. 77-96.
34. Persson CU, Sunnerhagen KS, Danielsson A, Grimby-Ekman A, Hansson PO. Responsiveness of a modified version of the postural assessment scale for stroke patients and longitudinal change in postural control after stroke- Postural Stroke Study in Gothenburg (POSTGOT). Journal of neuroengineering and rehabilitation. 2013;10:8.
35. Kucukdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine. 2011;43(8):661-72.
36. Persson CU, Sunnerhagen KS, Lundgren-Nilsson A. Rasch analysis of the modified version of the postural assessment scale for stroke patients: postural stroke study in Gothenburg (POSTGOT). BMC neurology. 2014;14:134.
37. Portney LG, Watkins MP. Sampling. Foundations of clinical research : applications to practice. 3rd ed. Upper Saddle River: Pearson Prentice Hall; 2009. p. 143-61.
38. Sullivan JE, Crowner BE, Kluding PM, Nichols D, Rose DK, Yoshida R, et al. Outcome measures for individuals with stroke: recommendations from the American Physical Therapy Association neurology section task force. Physical therapy. 2013;93(10):1383-96.
39. Tyson SF, Hanley M, Chillala J, Selley AB, Tallis RC. The relationship between balance, disability, and recovery after stroke: predictive validity of the Brunel Balance Assessment. Neurorehabilitation and neural repair. 2007;21(4):341-6.
40. Tung FL, Yang YR, Lee CC, Wang RY. Balance outcomes after additional sit-to-stand training in subjects with stroke: a randomized controlled trial. Clinical rehabilitation. 2010;24(6):533-42.
41. Huang YC, Wang WT, Liou TH, Liao CD, Lin LF, Huang SW. Postural Assessment Scale for Stroke Patients Scores as a predictor of stroke patient ambulation at discharge from the rehabilitation ward. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine. 2016;48(3):259-64.

Table I: Characteristics of the study sample (N=47).

Variables			
Sex; n (%)			
Male		26	(55)
Female		21	(45)
Age in years; mean (SD) range	68	(10.2)	40-86
Diagnosis; n (%)			
Ischemic		38	(81)
Haemorrhagic		5	(10,5)
Ischemic and haemorrhagic		4	(8,5)
Most affected body half; n (%)			
Right		24	(51)
Left		21	(45)
Bilateral		2	(4)
Days between date of ictus and date of testing; mean (SD) range (median)	25	(28.2)	1-136 (19)
Distribution based on level of function; n (%)			
Group 1. No independent sitting balance		3	(7)
Group 2. Can sit independently for 10 seconds		8	(17)
Group 3. Needs only slight support with transfers		18	(38)
Group 4. Safe and independent in transfers		9	(19)
Group 5. Can stand on the affected leg for 10 seconds or walk without aids		9	(19)

Table II: Total score of the Norwegian version of SwePASS (N=47) according to each rater.

Rater	Mean	Standard deviation	Range
A1	24.7	8.7	4-36
A2	24.7	8.7	3-36
B1	24.4	8.7	3-36
B2	24.4	8.7	3-36
C1	24.9	8.8	5-36
C2	24.9	8.7	5-36

A1, B1, C1 refer to the first scoring of the three raters. A2, B2, C2 refer to the second scoring (with a median of 34 days after the first scoring)

Table III: Intrarater and interrater reliability of the total score on the Norwegian version of SwePASS for the participants (N=47).

Raters	ICC (1,1)	95% CI	ICC (3,1)	95% CI	S _w	SDD
<u>Intrarater</u>						
A1-A2	1.00	.994-.998	1.00	.994-.998	0.5	1.4
B1-B2	.99	.989-.997	.99	.989-.997	0.7	1.9
C1-C2	.99	.990-.997	.99	.990-.997	0.7	1.8
<u>Interrater 1. scoring</u>						
A1-B1	.99	.988-.996	.99	.990-.997	0.7	1.9
A1-C1	.99	.985-.995	.99	.985-.995	0.8	2.2
B1-C1	.99	.979-.993	.99	.981-.994	1.0	2.7
A1-B1-C1	.99	.985-.995	.99	.986-.995	0.8	2.3
<u>Interrater, 2. scoring</u>						
A2-B2	1.00	.992-.997	1.00	.993-.998	0.6	1.6
A2-C2	1.00	.994-.998	1.00	.994-.998	0.5	1.4
B2-C2	.99	.990-.997	1.00	.992-.998	0.7	1.8
A2-B2-C2	1.00	.993-.997	1.00	.994-.998	0.6	1.6

A1, B1, C1 refer to the first scoring of the three raters. A2, B2, C2 refer to the second scoring.

Abbreviations: ICC=intraclass correlation coefficient, CI=confidence interval, S_w=within subject standard deviation, SDD=smallest detectable difference for 95 % of pairs of observations

Table IV: Intrarater and Interrater reliability of each item on SwePASS-NV for the participants (N=47)

Items	1st-2nd scoring						1st scoring					
	Intrarater A1-A2		Intrarater B1-B2		Intrarater C1-C2		Interrater A1-B1		Interrater A1-C1		Interrater B1-C1	
	Kappa	% agreement	Kappa	% agreement	Kappa	% agreement	Kappa	% agreement	Kappa	% agreement	Kappa	% agreement
1. Supine to side-lying affected side	*	100	*	100	*	100	*	98	*	100	*	98
2. Supine to side-lying unaffected side	*	98	*	100	*	98	*	94	*	96	*	94
3. Supine to sitting on the edge of the bed	*	100	*	100	*	100	*	100	*	98	*	98
4. Sitting without support	*	94	*	96	*	98	*	96	*	100	*	96
5. Sitting to standing up	.83	89	.90	94	.89	94	.97	98	.86	91	.83	89
6. Standing with support	.91	96	.95	98	.70	85	.87	94	.70	85	.73	87
7. Standing without support	.97	98	.97	98	.88	91	.91	94	.76	83	.68	77
8. Standing on unaffected leg	.92	96	.84	91	.92	96	.88	94	.88	94	.76	87
9. Standing on affected leg	1.00	100	.91	96	.95	98	1.00	100	.95	98	.95	98
10. Standing picking up a shoe from the floor	.87	91	.79	85	.91	94	.82	87	.81	87	.82	87
11. Sitting down from standing up	.89	94	.86	91	.89	94	.96	98	.82	89	.85	91
12. Sitting on the edge of the bed to supine	.95	98	.91	96	1.00	100	.86	94	.85	94	.90	96
Mean	.92	96	.89	95	.89	96	.91	95	.83	93	.82	91

* Kappa could not be calculated

Caption

Figure 1: Mean total score on the Norwegian version of SwePASS for the 47 participants with regard to their level of function. Group 1: no independent sitting balance, group 2: can sit independently for 10 seconds, group 3: needs only slight support with transfers, group 4: safe and independent in transfers, group 5: can stand on the affected leg for 10 seconds or walk without aids.