# Test-retest reliability of the ABILHAND Questionnaire in persons with chronic stroke.

Ekstrand, Elisabeth; Lindgren, Ingrid; Lexell, Jan; Brogårdh, Christina

*Published in:*
PM&R

*DOI:*
10.1016/j.pmrj.2013.09.015

2014

# Test-retest reliability of the ABILHAND Questionnaire in persons with chronic stroke

**ABSTRACT**

**Background:** To be able to evaluate recovery, effects of rehabilitation interventions and changes over time, reliable and valid outcome measures are needed. The ABILHAND Questionnaire is a measure of self-reported ability to perform complex daily hand activities. It is commonly used in stroke rehabilitation, but data about the measurement variability are missing.

**Objective**: To assess the test-retest reliability of the ABILHAND Questionnaire in chronic stroke and to define limits for the smallest change that indicates a real change, both for a group of individuals and a single individual.

**Design**: A test-retest reliability study.

**Settings:** University Hospital.

**Participants**: A convenience sample of 43 individuals (11 women and 32 men; mean age 64 years) with mild to moderate impairments of hand function 6 to 48 months after stroke.

**Intervention:** Not applicable.

**Main Outcome Measurements**: The ABILHAND Questionnaire is Rasch analysed, enabling ordinal data to be converted into an interval scale (logits) and the use of parametric statistical analyses. The participants responded to 23 items in the ABILHAND Questionnaire on two occasions, two weeks apart. Reliability was assessed with the intraclass correlation coefficient ($ICC2_{2.1}$), the mean difference between the test sessions (đ) together with the 95% confidence intervals for đ, the standard error of measurement (SEM and SEM%), the smallest real difference (SRD and SRD%) and a Bland & Altman graph.

**Results**: Four outliers, with high mean logit scores (>4.0), were identified in the sample. The results are therefore presented for the whole sample (n=43) and without the four outliers (n=39). The test-retest agreement was high, $ICC_{2,1}$ = 0.85 (n=43) and 0.91 (n=39). The SEM%, representing the smallest change that indicates a real improvement for a group of individuals, was 21% (n=43) and 15% (n=39). The SRD%, representing the smallest change that indicates a real clinical improvement for a single individual, was 59% (n=43) and 42% (n=39), respectively.

**Conclusion*:*** The ABILHAND Questionnaire is reliable in persons with chronic stroke and can be recommended to evaluate recovery, rehabilitation interventions and changes over time in a group of individuals but is less suitable for a single individual.

**Key Words**: stroke; outcome assessment; rehabilitation; reproducibility of results; upper extremity.

## INTRODUCTION

Stroke is one of the most common causes of disability in the world [1]. It often leads to a variety of sensorimotor deficits that affect motor control. Various impairments of arm and hand are present in 70 to 80% of the stroke patients in the acute phase and still in about 40% in the chronic phase [2, 3]. These impairments impact on the ability to perform daily activities and to participate in society, according to the International Classification of Functioning Disability and Health (ICF) [4, 5]. Regaining arm and hand function and the ability to perform daily hand activities is therefore an important goal during stroke rehabilitation [6, 7]. However, much research remains to establish effective rehabilitation interventions that enable the recovery of the arm and hand and the ability to perform daily activities after stroke.

To evaluate recovery, effects of rehabilitation interventions or changes over time, reliable and valid outcome measures are needed. Several outcome measures are available to assess different aspects of impairments and activity limitations in the arm and hand after stroke [8-10]. The majority of these are based on observations and ratings of the ability to perform different tasks [8-13], whereas some are self-reported outcome measures. The advantage of such measures is that they provide a better understanding of the everyday difficulties that disabled persons may perceive [14, 15], which enable the clinicians to design more individually targeted rehabilitation interventions.

One self-reported outcome measure is the ABILHAND [16], which is an interview-based questionnaire that measures self-reported ability to perform complex hand activities in 23 daily situations. It was initially developed and validated for persons with rheumatoid arthritis [16] and then for persons with chronic stroke [17] as well as for several other diagnosis [18-22]. The ABILHAND Questionnaire is reliable in a test-retest situation in persons with rheumatoid arthritis and in persons with systemic sclerosis [19, 23, 24]. It has also been shown to have high internal consistency reliability after stroke [17]. In several recent studies it has been used to assess recovery of the arm and hand and changes following interventions [25-30]. However, to the best of our knowledge no study has assessed the test-retest reliability of the ABILHAND Questionnaire in persons with stroke.

To fully assess the reliability of a measure or instrument, several statistical methods are required. These should cover agreement between measurements, systematic changes in the mean and measurement variability [31]. Such a comprehensive reliability analysis can also be used to define limits for the smallest change that indicates improvements following an intervention or changes over time, both for a group of individuals or a single individual.

The aim of this study was to assess the test-retest reliability of the ABILHAND Questionnaire as a measure of daily hand activities in chronic stroke and to define limits for the smallest change that indicates a real change, both for a group of individuals and a single individual.

## MATERIAL AND METHODS

*Participants*

A convenience sample of 43 community-dwelling individuals (32 men and 11 women) with chronic stroke and mild to moderate impairments of hand function was recruited from a university hospital in the south of Sweden during January to August 2012. Inclusion criteria were: (i) persistent upper extremity paresis but remaining motor function in the affected arm and hand (i.e., ability to grasp and release an object); and (ii) 6 to 48 months post stroke. Exclusion criteria were: (i) other diseases or disorders that could affect arm and hand function; (ii) difficulty to communicate or to understand test instructions; and (iii) obvious signs and symptoms of cognitive impairment. In Table I, the participants' characteristics are described. Their mean age was 64 years (SD 8) and the mean time from stroke onset to first test occasion was 16 months (SD 8). Their median motor function of the more affected arm and hand was 11 points (range 3 to 15), as assessed by the upper extremity part of the Swedish version of the Modified Motor Assessment Scale [10, 32].

*Ethics*

Prior to inclusion, information about the purpose of the study was provided and each individual gave their written consent to participate. The principles of the Declaration of Helsinki were followed, and the study was approved by the Regional Ethical Review Board, Lund, Sweden (Dnr 2011/742).

*The ABILHAND Questionnaire*

The ABILHAND Questionnaire for chronic stroke [17] consists of 23 common complex uni- and bilateral daily hand activities (see Table II) that are rated on a 3-point scale (0=impossible; 1=difficult; 2=easy). Activities not performed during the last 3 months are scored as missing responses. The ABILHAND is Rasch analysed, enabling raw (ordinal) data to be converted into an interval scale and the use of parametric statistical analyses [17]. The Rasch model estimates the item difficulty on a linear scale based on the responses for each item within a probabilistic model. The model automatically includes missing data in the analysis by using available data to estimate a person's probable rating and converts the raw score into a Rasch score. The Rasch score is presented in logits (i.e. log odds units), a linear unit defined as the natural logarithm of the odds of successful achievement of any item by the participant. The model assumes that participants with a higher ability to perform daily hand activities have a higher probability to succeed in accomplishing any item compared to participants with a lower ability. The logit scale ranges from plus to minus with the centre of the scale set to zero, where higher logits values represent better self-perceived ability [33]. The ABILHAND Questionnaire logits scale for chronic stroke ranges from approximate -6 to +6 [17].

*Procedures*

In this study, the Swedish version of the ABILHAND Questionnaire was used [34]. The participants responded to the ABILHAND Questionnaire on two occasions at the same time of the day, two weeks apart. The interviews were performed in a quiet and separate room of the hospital by the same physiotherapist (IL), who has more than 10 years' experience of stroke rehabilitation. The questionnaire took about 15 minutes to complete. According to the instructions for the ABILHAND Questionnaire, the participants were guided through the first five items in order to understand each level of the scale. They were asked to rate their ability to perform the items regardless of how much the more affected hand was used during the activity. The participant rated the remaining 18 items without guidance from the physiotherapist, but the instructions could be repeated if needed. Due to cultural differences, four items were slightly modified and rephrased after recommendations from the researchers that have translated the ABILHAND Questionnaire into Swedish [25, 34]. Item number 3 *"Peeling potatoes with a knife"* was changed to *"Peeling potatoes"* (as a potato-peeler is usually used). For item number 12 *"Tearing open a pack of chips"* the phrase *or a candy-bag* was added (since older persons rarely eat chips). For item number 4 *"Cutting one's nails"* and item number 7 *"Filing one's nails*" the phrase *on both hands* was added.

At the second test occasion, the participants responded to the items in a different order. The ABILHAND Questionnaire is accessible online in sets of 10 sheets, with the 23 items in a random order, to avoid any systematic bias. According to the instructions the examiner selected the next one of the 10 sheets for each new assessment.

After each assessment, the raw scores were entered into an online Rasch based data analysis module (www.rehab-scales.org) which automatically converted the data into logits by using the RUMM Rasch measurement software program (www.rummlab.com) according to the calibration of the scale established for chronic stroke patients [17].

*Statistical methods*

Descriptive statistics were calculated for each participant's characteristics, for the response frequencies and differences in raw scores between the two measurements of the ABILHAND Questionnaire.

To determine the test-retest reliability several statistical methods were applied; a detailed account of the analyses, a rational for their use, and all equations have been presented previously [31]. Agreement between the measurements of the logit values (i.e., parametric data) was evaluated by the intra-class correlation coefficient (ICC). Different subforms of ICCs are available for different study designs and the most commonly used are $ICC_{1,1}$, $ICC_{2,1}$ and $ICC_{3,1}$. In practice, their values are often very similar. For a test-retest design both $ICC_{1,1}$ or $ICC_{2,1}$ can be used [35]. In this study $ICC_{2,1}$ was chosen because it is calculated from a two-way repeated-measures analysis of variance that provides the basics for other calculations of reliability, i.e. the measurement variability [31]. However, the ICC can give misleading results as the values are highly sensitive to the spread of the measurements between subjects. Therefore, the analysis should be complemented by an analysis of the change in mean and measurement variability [35].

The change in the mean was calculated from the logit values obtained from the two test occasions. A Bland & Altman graph was formed to display the data graphically (the difference from the two test occasions plotted against the mean of the two test occasions for each participant), and to reveal any systematic bias and outliers. To assess if participants with higher logits had a higher dispersion than those with lower logits (heteroscedasticity) the mean of the two test occasions was correlated to the difference between the two test occasions. The mean difference (đ) was calculated between the two measurements (test 2 minus test 1), and a 95% interval for the đ was formed to determine if there was a true systematic difference between the logit values from the two test occasions.

Measurement variability was assessed by the standard error of measurement (SEM) and the smallest real difference (SRD). The SEM gives the measurement variability in absolute values and represents the limit for the smallest change that indicates a real change for a group of individuals. From the SEM, the SEM% was calculated, representing the change in relative terms. The smallest real difference (SRD), representing the limits for the smallest change that indicates a real change for a single individual, was calculated together with an 95% confidence interval around the mean difference of the two measurements (đ). From the SRD, the SRD% was also calculated. Probability values less than 0.05 were considered statistically significant. Data were analysed using the IBM SPSS Statistics version 20 (IBM Corporation, Armonk, New York, United States).

**RESULTS**

In Table II the number of participants who responded to the 23 items (using the response options 0, 1 or 2) in the ABILHAND Questionnaire is presented; a value less than 100% indicates that some participants responded that they had not performed an item during the last three months. Most participants rated the same response options at the second test occasion as during the first (i.e., the difference in scores was zero). Eleven items (items 6, 10, 11, 13, 15, 16, 18, 19, 20, 21 and 23) had a high response frequency (≥95%) at both test occasions, whereas five items had a low response frequency at both test occasions (items 1, 2, 5, 9 and 14). In total, for both test occasions, there were 18% missing responses. Two items, item 9 (*Shelling hazelnuts*) and item 14 (*Sharpening a pencil*), had a particularly high rate of missing responses.

The mean number (SD; range) of missing items for each individual on both test occasions were 4 (2; 1-9). The men had significantly more missing responses compared to the women (*P*=.036).

The Bland & Altman graph (Figure I) displays the difference between the logit scores between the two test occasions (test 2 minus test 1) plotted against the mean of the two test occasions. As can be seen in Figure I, four participants had high means and large differences in logits between the two test occasions and were considered as outliers. The results of the reliability analyses are therefore presented for the whole sample (n=43) and for the 4 outliers excluded (n=39).

As can be seen in Table III, the mean values in logits (i.e., converted raw scores) for the whole sample were 2.28 at the first test occasion and 2.52 at the second test occasion (n=43). When the four outliers were excluded (n=39) the mean values were 2.14 and 2.21, respectively.

In Table IV, the reliability of the ABILHAND Questionnaire is summarized. Test-retest agreement was high: $ICC_{2,1}$=0.85 for the whole sample (n=43) and $ICC_{2,1}$=0.91 when the four outliers were excluded (n=39). The calculation of the change in the mean, obtained from the two measurements, showed that the đ value was positive for the whole sample (n=43). Zero was not included in the 95% confidence interval of đ, indicating that the score from the second test session was significantly larger than during the first test session. If the four outliers were removed from the analysis the đ value was still positive but zero was included in the 95% confidence interval of đ (i.e., there was no longer any systematic change in the mean between the test occasions). The Bland & Altman graph (Figure I) revealed that participants with higher logits had a higher dispersion than those with lower logits, i.e., a heteroscedasticity (r=0.49, $P<.001$). If the four outliers were excluded from the analysis there were no longer a heteroscedasticity (r=0.04; $P=.79$).

The measurement variability (Table IV) for a group of individuals, SEM and SEM%, was 0.51 logits and 21.2% for the whole sample (n=43). When the four outliers were excluded (n=39) the SEM and SEM% decreased to 0.33 logits and 15.2%, respectively. The measurement variability for a single individual, SRD and SRD%, was 1.41 logits and 58.8% for the whole sample (n=43). When the four outliers were excluded (n=39) the SRD and SRD% decreased to 0.92 logits and 42.1%.

**DISCUSSION**

In this study we have assessed the test-retest reliability of the ABILHAND Questionnaire as a measure of daily hand activities in a group of persons with mild to moderate impairments of hand function after stroke. The main findings were that the test-retest agreement was high and the measurement variability was acceptable to assess changes in a group of individuals. Four participants had particularly large differences in their mean logits between the first and second test occasion (see Figure I) and were considered as outliers. Therefore, the reliability analysis was performed both with the whole sample (n=43) and without the four outliers (n=39). The results were affected accordingly, which have some important implications for the use of the ABILHAND Questionnaire as an outcome measure.

Our findings of high intra-class correlation coefficients ($ICC_{2,1}$=0.85 to 0.91) are in agreement with previous studies in persons with rheumatoid arthritis (ICC=0.74 to 0.86) [23, 24] and systemic sclerosis (ICC=0.96) [19]. ICC is one of the most commonly used methods to evaluate reliability, but it is insufficient to fully assess the reliability of measurements [31].

ICC assesses the agreement between repeated test occasions and thereby only the variance between individuals. It is also important to detect a variance within individuals by calculating the change in mean from the two test occasions. Therefore, we have expanded our analysis and included analyses of changes of the mean and limits for the smallest change that indicates a real improvement, both for a group of individuals and for a single individual [31, 36].

When analyzing the data for the whole sample (n=43), a systematic change in the mean logits was revealed. The mean difference (đ) was positive and zero was not included in the 95% confidence interval of đ (Table IV) demonstrating that the participants systematically rated their ability to perform daily hand activities better on the second test occasion than on the first. Possible reasons could be a change in behaviour or a learning effect. The first test occasion may have prompted some participants to perform specific activities or to practice them at home before the second test occasion. However, when the four outliers were removed from the analysis there was no longer any significant systematic change in the mean.

The measurement variability, SEM and SEM%, for a group of individuals was 0.51 logits (21.2%) for the whole sample, and decreased to 0.33 logits (15.2%) when the four outliers were excluded from the analysis. These values could be considered acceptable and indicates that the ABILHAND Questionnaire can detect changes in a group of individuals over time or after an intervention. Other studies that have used the ABILHAND Questionnaire as an outcome measure have reported similar or higher changes in scores. Liao et al. [26] and Wang et al. [37] found mean changes of 0.25 to 0.88 logits after robotic training in persons with stroke, and Batcho et al. [24] reported a mean change of 0.48 logits in a stable group of persons with rheumatoid arthritis over a period of one year.

To quantify the size of the measurement variability between two test occasions for a single individual, the SRD and SRD% were calculated. The SRD and SRD% was 1.41 logits (58.8%) for the whole sample. When the four outliers were excluded from the analysis the SRD and SRD% decreased to 0.92 logits (42.1%), respectively. These values are fairly high and indicate that the ABILHAND is less suitable to detect changes for single individuals with chronic stroke.

Several factors may have influenced the measurement variability, for example the number of missing responses, ambiguity in item description and instructions, and the extreme values in the Rasch analysis. Missing responses can influence the result in the Rasch analysis. Even if the Rasch model estimates a person’s missing responses, missing data can reduce the precision and thus the reliability [38]. In the present study the missing responses were relatively high (between 2% to 86% per item, or 18% in total). This is higher as compared to other studies in persons with other diagnosis. In the study by Barrett et al. [22] missing responses ranged from 2% to 21% per item. In the study by Vanthuyne et al. [19], missing responses were 6% in total and in the study by Vandervelde et al. [21] missing responses were 3% in total. However, two items in our study (item 9, 14) had a high rate of missing responses (see Table II) and if these two items were deleted from the analysis the missing data decreased to 12% in total. Item number 9 "*Shelling hazel nuts"* is an uncommon activity, which has been pointed out also in other studies [22, 34], and item number 14 "*Sharpening a pencil"* could be considered outdated as mechanical pencils are now most often used.

We also found a difference in missing responses between men and women, but not related to other variables such as age, paretic side or arm- and hand function. This indicates that some items in the ABILHAND Questionnaire are traditionally more performed by women than men. Missing data are always difficult to handle when using a rating scales. As missing data automatically are estimated in the Rasch analysis, the participants in our study were asked to rate only those activities that were commonly performed in real life during the last three months, i.e., the participants should actually have done the activities and not anticipated how these activities could possibly have been performed. One way to reduce missing responses could be to ask more follow-up questions for all items and not just for the first five, to verify the participant's answers.

Another factor that could have influenced the measurement variability is the brief description of many items in the ABILHAND Questionnaire, which could make them difficult to interpret. The response options – easy, difficult and impossible – might also be too imprecise and the distinction between them is unclear, which also has been described by others [22, 34]. Furthermore, the timespan of three months could have made it difficult for the participants to remember which activities they actually performed during this time period.

Moreover, a change in raw score at both ends of the Rasch scale leads to a higher change in logits as compared to a change in the centre of the scale (around zero). The standard errors vary across the scale and are smallest at the centre of the scale and largest at the end of the scale. Thus, measures that are located at the ends of the Rasch scale (i.e., high or low logits) are therefore less precise [22]. This was actually the case for the four outliers in our study with mean logits greater than 4.0. A recently published study has pointed out the same limitation and recommends that the range of measurement in the ABILHAND Questionnaire could be extended by adding more difficult items to increase the precision in the ends of the scale [22].

Despite these limitations it seems like the ABILHAND Questionnaire in practice works fairly well and most of our participants rated the same response options at the second test occasion as the first (Table II). The test-retest agreement was high and measurement variability acceptable on a group level. Although the measurement variability for a single individual was fairly high the ABILHAND Questionnaire could still be useful in the clinical setting as a guide to individual goal setting and treatment planning of complex hand activities in daily life. However, for individuals with 'extreme logits', careful consideration must be taken when the effects of different interventions are evaluated to be confident that a real improvement has occurred.

In this study, we included only individuals with mild to moderate impairments of hand function in a chronic phase, an ability to use both hands in daily activities (i.e., to grasp and release an object), no difficulty to communicate and no major cognitive impairments. Therefore, the results cannot not be generalized to the entire stroke population. A strength of the study is that we included 43 participants, which can be considered sufficiently large when the reliability is examined [39]. Moreover, great care was taken to standardize the test situation; the tests were performed at the same time of the day, at the same location and with the same time interval between the tests which strengthens the result.

## CONCLUSIONS

The ABILHAND Questionnaire, as a measurement of self-reported ability to perform daily hand activities, is reliable in persons with chronic stroke and can be recommended to evaluate recovery, rehabilitation interventions and changes over time in a group of individuals but is less suitable for a single individual.

## ACKNOWLEDGMENT

We thank the persons who volunteered to participate and Ulla-Britt Flansbjer, PT, PhD for statistical guidance. This study was supported by grants from Skane county council's research and development foundation and the Norrbacka Eugenia Foundation.

## REFERENCES


1. WHO.int. Geneva: World Health Organization. Available at: http://www.who.int/topics/cerebrovascular_accident/en/. Accessed September 3, 2012.

2. Broeks JG, Lankhorst GJ, Rumping K, Prevo AJ. The long-term outcome of arm function after stroke: results of a follow-up study. Disabil Rehabil. 1999;21:357-64.

3. Nakayama H, Jorgensen HS, Raaschou HO, Olsen TS. Recovery of upper extremity function in stroke patients: the Copenhagen Stroke Study. Arch Phys Med Rehabil. 1994;75:394-8.

4. International Classification of Functioning, Disability and Health (ICF). Geneva: WHO 2001. Available at: http://www.who.int/classifications/icf/en/. Accessed September 5, 2012.

5. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. Lancet. 2011;377:1693-702.

6. Dobkin BH. Motor rehabilitation after stroke, traumatic brain, and spinal cord injury: common denominators within recent clinical trials. Curr Opin Neurol. 2009;22:563-9.

7. Carter AR, Connor LT, Dromerick AW. Rehabilitation after stroke: current state of the science. Curr Neurol Neurosci Rep. 2010;10:158-66.

8. Mathiowetz V, Volland G, Kashman N, Weber K. Adult norms for the Box and Block Test of manual dexterity. Am J Occup Ther. 1985;39:386-91.

9. Mathiowetz V, Weber K, Kashman N, Volland G. Adult Norms for the 9 Hole Peg Test of Finger Dexterity. Occup Ther J Res. 1985;5:24-38.

10. Carr JH, Shepherd RB, Nordholm L, Lynne D. Investigation of a new motor assessment scale for stroke patients. Phys Ther. 1985;65:175-80.

11. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. Int J Rehabil Res. 1981;4:483-92.

12. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. Stroke. 2001;32:1635-9.

13. Sollerman C, Ejeskar A. Sollerman hand function test. A standardised method and its use in tetraplegic patients. Scand J Plast Reconstr Surg Hand Surg. 1995;29:167-76.

14. Stewart JC, Cramer SC. Patient-Reported Measures Provide Unique Insights Into Motor Function After Stroke. Stroke 2013, published online ahead of print Feb 19 doi:10.1161/STROKEAHA.111.674671.

15. Ashford S, Slade M, Malaprade F, Turner-Stokes L. Evaluation of functional outcome measures for the hemiparetic upper limb: a systematic review. J Rehabil Med. 2008;40:787-95.

16. Penta M, Thonnard JL, Tesio L. ABILHAND: a Rasch-built measure of manual ability. Arch Phys Med Rehabil. 1998;79:1038-42.

17. Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. Stroke. 2001;32:1627-34.

18. Arnould C, Penta M, Renders A, Thonnard JL. ABILHAND-Kids: a measure of manual ability in children with cerebral palsy. Neurology. 2004;63:1045-52.

19. Vanthuyne M, Smith V, Arat S, Westhovens R, de Keyser F, Houssiau FA, et al. Validation of a manual ability questionnaire in patients with systemic sclerosis. Arthritis Rheum. 2009;61:695-703.

20. Burger H, Franchignoni F, Kotnik S, Giordano A. A Rasch-based validation of a short version of ABILHAND as a measure of manual ability in adults with unilateral upper limb amputation. Disabil Rehabil. 2009;31:2023-30.

21. Vandervelde L, Van den Bergh PY, Penta M, Thonnard JL. Validation of the ABILHAND questionnaire to measure manual ability in children and adults with neuromuscular disorders. J Neurol Neurosurg Psychiatry. 2010;81:506-12.

22. Barrett L, Cano S, Zajicek J, Hobart J. Can the ABILHAND handle manual ability in MS? Mult Scler. 2012:1-10.

23. Durez P, Fraselle V, Houssiau F, Thonnard JL, Nielens H, Penta M. Validation of the ABILHAND questionnaire as a measure of manual ability in patients with rheumatoid arthritis. Ann Rheum Dis. 2007;66:1098-105.

24. Batcho CS, Durez P, Thonnard JL. Responsiveness of the ABILHAND Questionnaire in Measuring Changes in Rheumatoid Arthritis Patients. Arthritis Care Res. 2011;63:135-41.

25. Alt Murphy M, Persson HC, Danielsson A, Broeren J, Lundgren-Nilsson A, Sunnerhagen KS. SALGOT - Stroke Arm Longitudinal study at the University of Gothenburg, prospective cohort study protocol. BMC Neurol. 2011;11:56.

26. Liao WW, Wu CY, Hsieh YW, Lin KC, Chang WY. Effects of robot-assisted upper limb rehabilitation on daily function and real-world arm activity in patients with chronic stroke: a randomized controlled trial. Clin Rehabil. 2012;26:111-20.

27. Piron L, Turolla A, Agostini M, Zucconi C, Cortese F, Zampolini M, et al. Exercises for paretic upper limb after stroke: a combined virtual-reality and telemedicine approach. J Rehabil Med. 2009;41:1016-102.

28. Bovolenta F, Goldoni M, Clerici P, Agosti M, Franceschini M. Robot therapy for functional recovery of the upper limbs: a pilot study on patients after stroke. J Rehabil Med. 2009;41:971-5.

29. Michielsen ME, Selles RW, van der Geest JN, Eckhardt M, Yavuzer G, Stam HJ, et al. Motor recovery and cortical reorganization after mirror therapy in chronic stroke patients: a phase II randomized controlled trial. Neurorehabil Neural Repair. 2011;25:223-33.

30. Wu CY, Huang PC, Chen YT, Lin KC, Yang HW. Effects of Mirror Therapy on Motor and Sensory Recovery in Chronic Stroke: A Randomized Controlled Trial. Arch Phys Med Rehabil 2013, published online ahead of print Feb 15 doi:10.1016/j.apmr.2013.02.007.

31. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil. 2005;84:719-23.

32. Barkelius K, Johansson A, Körm K, Lindmark B. Reliabilitets-och validitetsprövning av modifierad Motor Assessment Scale enligt Uppsala Akademiska sjukhus-95. (In Swedish) Nordisk Fysioterapi. 1997;1:121-6.

33. Rasch G. Probabilistic models for some intelligence and attainment tests. Expanded ed. Chicago: University of Chicago Press; 1980.

34. Gustafsson S, Stibrant Sunnerhagen K, Dahlin-Ivanoff S. Occupational therapists' and patients' perceptions of ABILHAND, a new assessment tool for measuring manual ability. Scand J Occup Ther. 2004;11:107-17.

35. Downham DY, Holmbäck AM, Lexell J. Reliability of measurements in medical research and clinical practice. In: Ray P, Laura AM, editors. Studies in Multidisciplinarity: Elsevier; 2006. p. 147-63.

36. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res. 2001;10:571-8.

37. Wang TN, Lin KC, Wu CY, Chung CY, Pei YC, Teng YK. Validity, responsiveness, and clinically important difference of the ABILHAND questionnaire in patients with stroke. Arch Phys Med Rehabil. 2011;92:1086-91.

38. Institute for Objective Measurement, Inc. Available at: http://www.rasch.org/ . Accessed January 22, 2013.

39. Hobart JC, Cano SJ, Warner TT, Thompson AJ. What sample sizes for reliability and validity studies in neurology? J Neurol. 2012;259:2681-94.

**LEGENDS**

**Figure I**. The difference (logits) between the measurements from the two test occasions (test 2 minus test 1) plotted against the mean (logits) of the two test occasions for the ABILHAND (n=43). Four participants (outliers) had high means (>4.0) and large differences in logits between the two test occasions (located to the right of the vertical dashed line).

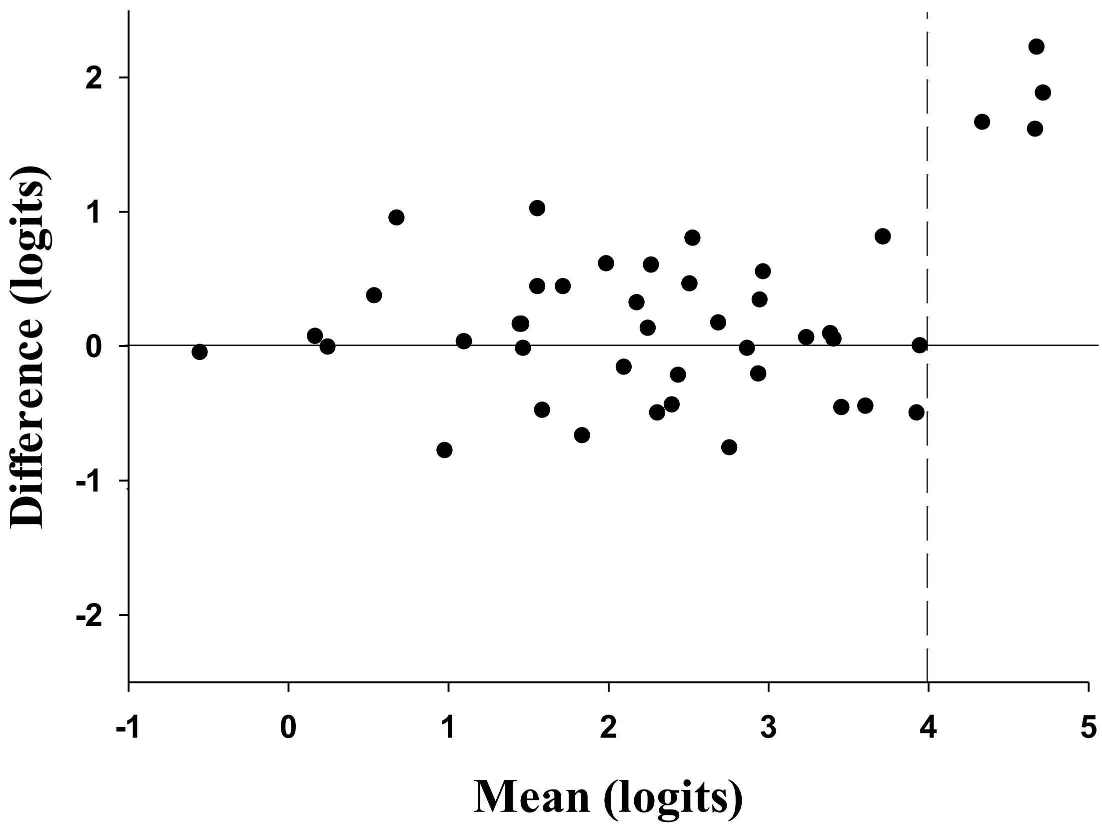


*Figure 1*

**Table I. Characteristics of the participants with chronic stroke (n=43).**

| | |
|---|---|
| Gender, n (%) | |
| Male | 32 (74) |
| Female | 11 (26) |
| Age, mean years (SD; range) | 64 (8; 45 to 81) |
| Type of stroke, n (%) | |
| Cerebral infarction | 34 (79) |
| Cerebral haemorrhage | 9 (21) |
| Months from stroke onset to first test occasion, mean (SD) | 16 (8.0) |
| Affected side, n (%) | |
| Right | 24 (56) |
| Left | 19 (44) |
| Handedness, n (%) | |
| Right | 41 (95) |
| Left | 2 (5) |
| Living situation n (%) | |
| Living alone | 17 (40) |
| Living with another person | 26 (60) |
| Vocational situation, n (%) | |
| Not working (retired, disability pension) | 35 (81) |
| Working | 8 (19) |
| Modified Motor Assessment Scale, median (range) | |
| More affected arm | 11 (3 to 15) |
| Less affected arm | 15 (7 to 15) |

The Swedish version of the Modified Motor Assessment Scale assesses arm and hand function as well as advanced hand activities. The three subscales range from 0 to 5, where 5 = normal or almost normal motor function and 0 = no motor function.

**Table II. Number of participants with chronic stroke that responded to each item in the ABILHAND Questionnaire at both test occasions and the differences in scores between the two test occasions (n=43).**

| Item | Manual activities | Responses at both test occasions[1] n (%) | Differences in scores between test occasions (2 minus 1)[2] | | | | |
|---|---|---|---|---|---|---|---|
| | | | -2 | -1 | 0 | +1 | +2 |
| 1 | Hammering a nail | 20 (47) | 0 | 2 | 17 | 1 | 0 |
| 2 | Threading a needle | 16 (37) | 0 | 4 | 11 | 1 | 0 |
| 3 | Peeling potatoes with a knife | 30 (70) | 0 | 3 | 26 | 1 | 0 |
| 4 | Cutting one's nails | 38 (88) | 0 | 4 | 26 | 8 | 0 |
| 5 | Wrapping up gifts | 18 (42) | 0 | 2 | 15 | 1 | 0 |
| 6 | Cutting meat | 41 (95) | 0 | 3 | 34 | 4 | 0 |
| 7 | Filing one's nails | 26 (60) | 0 | 2 | 17 | 6 | 1 |
| 8 | Peeling onions | 28 (65) | 0 | 2 | 23 | 3 | 0 |
| 9 | Shelling hazel nuts | 6 (14) | 0 | 0 | 6 | 0 | 0 |
| 10 | Opening a screw-topped jar | 41 (95) | 0 | 2 | 36 | 2 | 1 |
| 11 | Fastening the zipper of a jacket | 42 (98) | 0 | 1 | 36 | 5 | 0 |
| 12 | Tearing open a pack of chips | 37 (86) | 2 | 3 | 23 | 7 | 2 |
| 13 | Buttoning up a shirt | 42 (98) | 0 | 6 | 33 | 3 | 0 |
| 14 | Sharpening a pencil | 15 (35) | 0 | 0 | 15 | 0 | 0 |
| 15 | Taking the cap off a bottle | 42 (98) | 1 | 4 | 33 | 4 | 0 |
| 16 | Spreading butter on a slice of bread | 41 (95) | 0 | 2 | 39 | 0 | 0 |
| 17 | Fastening a snap (jacket, bag) | 38 (88) | 0 | 2 | 34 | 2 | 0 |
| 18 | Buttoning up trousers | 42 (98) | 0 | 5 | 35 | 2 | 0 |
| 19 | Opening mail | 43 (100) | 0 | 4 | 36 | 3 | 0 |
| 20 | Pulling up the zipper of trousers | 42 (98) | 0 | 1 | 40 | 1 | 0 |
| 21 | Squeezing toothpaste on a toothbrush | 42 (98) | 0 | 1 | 39 | 2 | 0 |
| 22 | Unwrapping a chocolate bar | 37 (86) | 0 | 3 | 27 | 7 | 0 |
| 23 | Washing one's hands | 43 (100) | 0 | 0 | 42 | 1 | 0 |

[1]A value less than 100% indicates that some participants responded that the item was not performed during the last three months.
[2]Zero indicates that participants rated the same response option at both test occasions; a positive or negative value indicates that the rating at the second test occasion was higher or lower, respectively, than at the first test occasion.

**Table III. Mean logits for the two test occasions of the ABILHAND Questionnaire.**

| | n=43 | n=39 |
|---|---|---|
| Test occasion 1, mean (SD; range) | 2.28 (1.18; -0.52 to 4.18) | 2.14 (1.14; -0.52 to 4.18) |
| Test occasion 2, mean (SD); range) | 2.52 (1.42; -0.58 to 5.80) | 2.21 (1.09; -0.58 to 4.13) |

**Table IV. Reliability of the ABILHAND Questionnaire logits.**

| | n=43 | n=39 |
|---|---|---|
| Measurement agreement | | |
| Intraclass correlation coefficient ($ICC_{2,1)}$ | 0.85 | 0.91 |
| 95% CI for $ICC_{2.1}$ | 0.73 to 0.92 | 0.84 to 0.95 |
| Change in the mean | | |
| Mean difference (đ) | 0.24 | 0.07 |
| 95% confidence interval for đ | 0.03 to 0.45 | -0.08 to 0.23 |
| Measurement variability | | |
| Standard error of measurement (SEM) | 0.51 | 0.33 |
| SEM% | 21.2 | 15.2 |
| SRD | 1.41 | 0.92 |
| 95% CI for smallest real difference (SRD) | -1.17 to 1.65 | -0.84 to 0.99 |
| SRD% | 58.8 | 42.1 |

CI=confidence interval