Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2018;99:2513-22

ORIGINAL RESEARCH

# The Adult Assisting Hand Assessment Stroke: Psychometric Properties of an Observation-Based Bimanual Upper Limb Performance Measurement

Annick Van Gils, MSc,[a] Sarah Meyer, PhD,[a] Margaretha Van Dijk, MSc,[b] Liselot Thijs, MSc,[c] Marc Michielsen, MSc,[c] Christophe Lafosse, PhD,[d] Veronik Truyens, MSc,[e] Kristine Oostra, PhD,[f] Andre Peeters, MD,[g] Vincent Thijs, PhD,[h,i] Hilde Feys, PhD,[a] Lena Krumlinde-Sundholm, PhD,[j] Daphne Kos, PhD,[a] Geert Verheyden, PhD[a]

*From the [a]Department of Rehabilitation Sciences, KU Leuven-University of Leuven, Leuven, Belgium; [b]Department of Physical Medicine and Rehabilitation, University Hospitals Leuven, Leuven, Belgium; [c]Jessa Hospital, Rehabilitation Campus Sint Ursula, Herk-de-Stad, Belgium; [d]Rehabilitation Hospital RevArte, Edegem, Belgium; [e]Rehabilitation and MS Centre Overpelt, Overpelt, Belgium; [f]Department of Physical and Rehabilitation Medicine, Ghent University Hospital, Ghent, Belgium; [g]Department of Neurology, Cliniques Universitaires Saint-Luc, Brussels, Belgium; [h]Florey Institute of Neuroscience and Mental Health, University of Melbourne, Victoria, Australia; [i]Department of Neurology, Austin Health, Heidelberg, Victoria, Australia; and [j]Neuropediatric Research Unit, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden.*

**Abstract**

**Objective:** To investigate interrater and intrarater reliability, measurement error, and convergent and discriminative validity of the Adult Assisting Hand Assessment Stroke (Ad-AHA Stroke).

**Design:** Cross-sectional observational study.

**Setting:** A total of 7 stroke rehabilitation centers.

**Participants:** Stroke survivors (reliability sample: n=30; validity sample: N=118) were included (median age 67y; interquartile range [IQR], 59-76); median time poststroke 81 days (IQR 57-117).

**Interventions:** N/A.

**Main Outcome Measures:** Ad-AHA Stroke, Action Research Arm Test (ARAT), upper extremity Fugl-Meyer Assessment (UE-FMA). The Ad-AHA Stroke is an observation-based instrument assessing the effectiveness of the spontaneous use of the affected hand when performing bimanual activities in adults poststroke. Reliability of Ad-AHA Stroke was examined using intraclass correlation coefficients (ICCs), Bland—Altman plots, and weighted kappa statistics for reliability on item level. SEM was calculated based on Ad-AHA units. Convergent validity was assessed by calculating Spearman rank correlation coefficients between Ad-AHA Stroke and ARA test and UE-FMA. Comparison of Ad-AHA Stroke scores between subgroups of patients according to hand dominance, neglect, and age evaluated discriminative validity.

**Results:** Intrarater and interrater agreement showed an ICC of 0.99 (95% confidence interval, 0.99-0.99), an SEM of 2.15 and 1.64 out of 100, respectively, and weighted kappa for item scores were all above 0.79. The relation between Ad-AHA and other clinical

Presented to the World Conference of Neurorehabilitation, May 10-13, 2016, Philadelphia, PA, and the COTEC-ENOTHE Congress, June 15-19, 2016, Galway, Ireland.

Supported in part by grants from the Promobilia Foundation, Sweden (grant no. 15060), and from the Foundation Van Goethem-Brichant, Belgium.

Disclosures: Vincent Thijs reports personal fees from Medtronic, Pfizer, Boehringer/Ingelheim, Bayer, and BMS, outside the submitted work. Lena Krumlinde-Sundholm is involved in the company Handfast AB (Inc), which supplies training courses and test material for the Assisting Hand Assessment. She collaborated on the design of this specific study and reviewed the draft manuscript. Scoring of the Ad-AHA video observations, data analysis, and interpretation were independently performed. The other authors have nothing to disclose.



https://doi.org/10.1016/j.apmr.2018.04.025

assessments was strong (ρ=0.9). Patients with neglect had significantly lower Ad-AHA scores compared to patients without neglect ($P=.004$).

**Conclusions:** The Ad-AHA Stroke captures actual bimanual performance. Therefore, it provides an additional aspect of upper limb assessment with good to excellent reliability and low SEM for patients with subacute stroke. High convergent validity with the ARA test and UE-FMA and discriminative validity were supported.

Archives of Physical Medicine and Rehabilitation 2018;99:2513-22



Stroke is a major cause of adult-acquired disability leading to loss of independence during self-care and activities of daily living.[1,2] This is not surprising as 70% of stroke survivors present with upper limb impairments.[3] Upper limb recovery is found to be essential to regain independence in daily activities.[4-6] Most daily activities require the use of both hands simultaneously.[7-9] In stroke rehabilitation, several upper limb assessments are described in literature.[10] According to the International Classification of Functioning, Disability and Health, upper limb assessment occurs at 3 levels, namely body function, activities, and participation.[11] In the latter 2, a differentiation can be made between the qualifiers capacity and performance.[11] On the one hand, the qualifier capacity refers to the highest probable level of functioning that a person may reach in a standardized environment and describes an individual's ability to execute a task (eg, Action Research Arm Test [ARAT]).[11] On the other hand, the performance qualifier describes what an individual does in his or her current environment.[11] Self-report measures, such as the Motor Activity Log, are commonly used to measure performance in stroke rehabilitation.[10,12] The Motor Activity Log is a structured interview during which patients are questioned about the amount of use and the quality of movement of the affected arm.[12] However, self-report measures reflect perceived performance rather than actual performance.[10] Although the patients' perspective is very valuable in determining goals and interventions in stroke rehabilitation, a comprehensive assessment should involve a measure that captures actual real-world performance as well.[7,10,13,14] The Actual Amount of Use Test is a covert video-observed measure with 17 daily tasks used to observe spontaneous arm use, but sound psychometric properties have not been reported yet.[15] On the contrary, accelerometry has been proven a valuable and valid measure to quantify real-world performance of the affected upper limb poststroke, although it is limited to the amount of use of the affected upper limb and fails to inform about the quality of upper limb use.[7,13,14]

The Adult Assisting Hand Assessment Stroke (Ad-AHA Stroke) is a novel assessment tool with the purpose to measure how the affected hand collaborates spontaneously with the other hand when performing bimanual activities.[16] It is an observation-based assessment tool, with standardized activities and materials, however, semistructured concerning how to perform the activity. It is the patient's spontaneous way of using the affected hand that is assessed, both the amount and the quality of upper limb use are taken into account.[16] The participant is asked to use both hands in a way that feels natural for him or her and not to specifically use the paretic hand. This allows observation of the spontaneous use of the hand in a bimanual activity.[17] Therefore, the Ad-AHA Stroke reflects more closely how an individual will use the affected hand in his or her current environment and captures actual performance rather than capacity, and consequently it represents an essential aspect of assessment in stroke rehabilitation.[11,18]

The Assisting Hand Assessment (AHA) was originally developed for children with cerebral palsy or brachial plexus palsy (Kids-AHA). The Kids-AHA has shown excellent psychometric properties in earlier studies[9,16-19] and has been adapted for adults with hemiparesis poststroke. An important feature of the AHA is the use of Rasch measurement analysis to produce interval level measures for reporting outcomes.[8,16] Besides the stronger statistical properties, this also is a strong asset in clinical research and in clinical practice regarding evaluation of change. Furthermore, the Rasch measurement analysis generates an item difficulty order. This hierarchy of effective hand use can be used to generate an ability profile, showing which items are performed effectively and which items are not yet performed effectively, but come next in difficulty order.[8,16] Thereby tailor-made intervention goals for improved hand use can be set based on the individual ability profiles.[8,16]

The development of the Ad-AHA Stroke was conducted in several steps, first by creating a test situation which allowed observation of bimanual performance, second by generating test items with a starting point in the Kids-AHA 5.0 version, and third by evaluating internal scale validity through Rasch measurement analysis.[16] Data from 144 Ad-AHA Stroke assessments were evaluated, and the scale was adjusted using Rasch measurement analysis until unidimensionality was demonstrated.[16] Test items showing misfit were removed from the scale. The remaining 19 items showed satisfactory fit with the assertions of the Rasch model.[16] Aspects of convergent validity were evaluated in a pilot study with 24 participants, and high correlations were demonstrated for outcomes of the Jebsen-Taylor Hand Function Test (ρ= −0.93) and ABILHAND for stroke (ρ=0.80 [present task] and ρ=0.77 [sandwich task]).[16]

Although internal scale validity and some evidence of convergent validity (between Ad-AHA Stroke and Jebsen-and Taylor Hand Function Test and ABILHAND for stroke) have been demonstrated, other psychometric properties of the Ad-AHA Stroke need to be evaluated such as reliability, measurement error, and other types of validity.[16] The aim of this study is, therefore, to determine interrater and intrarater reliability, measurement error, and convergent and discriminative validity of the Ad-AHA Stroke in subacute stroke.

In the present study, we hypothesize a high correlation between unilateral and bimanual performance, in line with previous

***List of abbreviations:***

| | |
|---|---|
| **Ad-AHA Stroke** | **Adult Assisting Hand Assessment Stroke** |
| **AHA** | **Assisting Hand Assessment** |
| **ARAT** | **Action Research Arm Test** |
| **CI** | **confidence interval** |
| **ICC** | **intraclass correlation coefficient** |
| **IQR** | **interquartile range** |
| **Kids-AHA** | **Kids Assisting Hand Assessment** |
| **UE-FMA** | **upper extremity Fugl-Meyer Assessment** |

findings.[16] Because neglect is linked to more severe impairment in unilateral motor function,[20-22] a similar relation between neglect and bimanual performance is expected. Furthermore, we hypothesize a better bimanual performance when the dominant hand is affected. Differences in bimanual performance are also hypothesized between older and younger participants.

## Methods

### Participants

In a cross-sectional observational study, 118 participants were recruited from 7 rehabilitation centers in Belgium and assessed in an inpatient setting (n=101) or in the home environment of the patient (n=17).[20] Participants were included if diagnosed with a first-ever stroke according to the World Health Organization criteria,[23] with a unilateral motor and/or somatosensory deficit in the upper limb (upper extremity Fugl-Meyer assessment [UE-FMA] score<60 [scores range from 0 to 66])[24] and/or Erasmus MC modification of the Nottingham Sensory Assessment<7 (scores range from 0 to 8)[25] and/or Thumb Finding Test score≥1 (scores range from 0 to 3)[26], as described elsewhere[20]; age≥18 years and <6 months poststroke onset. Exclusion criteria were a prestroke Barthel Index score ≤95 out of 100[27]; other neurologic diseases with permanent damage, such as multiple sclerosis or Parkinson disease; and serious cognitive or communication deficits, hampering the evaluation. A trained physical therapist (S.M.) evaluated clinically whether the cognitive and communication deficits limited participants to understand and conduct the study protocol. Participants signed a written informed consent form prior to participation. All procedures followed were in accordance with the Declaration of Helsinki. Ethical approval for this study was granted by the Ethics Research Committee of the University Hospital of Leuven and all participating centers.

### Assessment

All participants were assessed in a 1-hour single test session following a standardized procedure by 1 trained researcher (S.M.). First, the patients' characteristics were obtained, including age at stroke onset, sex, time since stroke, type of stroke (ischemic or hemorrhagic), lateralization of symptoms, hand dominance, and comorbidities (Cumulative Illness Rating Scale[28]). Then assessments were performed in the following order: (1) UE-FMA[24]; (2) ARAT[29]; (3) Star Cancellation Test[30]; and (4) Ad-AHA Stroke. A short rest was allowed between the tests. Unilateral upper limb motor function was assessed by the UE-FMA,[24] and activities level was assessed by the ARAT.[29] Higher scores reflect better motor function and activity. Validity and reliability were established for both measures in a subacute stroke population.[24,31,32] The Star Cancellation Test was used to assess the presence of visuospatial neglect.[30] This paper-and-pencil measure was found to be sensitive in detecting visuospatial neglect in stroke patients.[30] A cutoff score of <44 (out of 54) indicates the presence of visuospatial neglect.[30]

### Ad-AHA Stroke

Administering the Ad-AHA Stroke involved 2 steps. First, the performance of a semistructured task was videotaped according to a standardized procedure.[16] In the Ad-AHA Stroke, 2 tasks were developed and both provided a different context with subtasks that are intended to encourage and generate bimanual performance, that is, wrapping a present and making a sandwich.[16] Although both tasks can be used interchangeably because both tasks produce equivalent results,[16] the task of wrapping a present was chosen in this study for practical reasons. This task, which takes about 13 minutes to complete,[16] involved unwrapping a box, opening a letter, opening the box, and then wrapping up a new present. This involved opening a jar and taking a piece of candy out of the jar to put it into the box, wrapping the box including cutting wrapping paper, applying sticky tape, and tying a string. Another subtask is writing a short greeting, including handling a pencil case and pen with lid, folding paper, and placing it in an envelope.[16] All materials required for the task were placed on the table on the side of the affected hand and objects, and subtasks were selected to elicit the use of the affected hand by holding, grasping, and reaching for objects.[16] Prior to the assessment, following instruction was given: This test relates to how you use your two hands together in such a way that it feels most natural to you. You do not have to make extra effort to use the hand as much as possible, but you should use it in a way that feels most natural to you.[16,17]

In the second step, the videotaped task performance was analyzed and scored.[16] The effectiveness of the use of the affected hand in this bimanual task was scored based on 19 items covering 5 domains: general use, arm use, grasp-release, fine motor adjustments, and coordination. An overview of the 19 items is shown in tables 1 and 2. Every item was rated on a 4-point rating

**Table 1** Intrarater reliability for item and sum scores of the Ad-AHA Stroke

| Ad-AHA Stroke Items | Weighted Kappa or ICC (95% CI) |
|---|---|
| Ad-AHA units 0-100 logit scale (ICC) | 0.99 (0.99-0.99) |
| General use items | |
| Amount of use | 0.81 (0.68-0.95) |
| Initiates use | 0.82 (0.70-0.94) |
| Chooses affected hand | 0.94 (0.85-1.00) |
| Arm use items | |
| Stabilizes by weight | 0.87 (0.75-0.98) |
| Reaches | 0.94 (0.97-1.00) |
| Moves upper arm | 0.84 (0.72-0.96) |
| Moves forearm | 0.87 (0.77-0.98) |
| Grasp-release items | |
| Grasps | 0.82 (0.69-0.94) |
| Holds | 0.94 (0.86-1.00) |
| Stabilizes by grasp | 0.90 (0.80-1.00) |
| Varies types of grasp | 0.88 (0.77-0.98) |
| Releases | 0.86 (0.75-0.97) |
| Fine motor adjustments items | |
| Moves fingers | 0.90 (0.83-0.99) |
| Readjusts grasps | 0.93 (0.85-1.00) |
| Manipulates | 0.87 (0.76-0.97) |
| Grip force regulation | 0.97 (0.92-1.00) |
| Coordination items | |
| Coordinates | 0.90 (0.81-0.99) |
| Orients objects | 0.85 (0.75-0.96) |
| Flow in bimanual performance | 0.87 (0.77-0.98) |

NOTE. Values are weighted kappa or as otherwise indicated.

**Table 2** Interrater reliability for item and sum scores of the Ad-AHA Stroke

| Ad-AHA Stroke Items | Weighted Kappa or ICC (95% CI) |
|---|---|
| Ad-AHA units 0-100 logit scale (ICC) | 0.99 (0.99-0.99) |
| General use items | |
| Amount of use | 0.79 (0.65-0.92) |
| Initiates use | 0.79 (0.65-0.93) |
| Chooses affected hand | 1.00 (1.00-1.00) |
| Arm use items | |
| Stabilizes by weight | 0.88 (0.74-1.00) |
| Reaches | 0.87 (0.77-0.97) |
| Moves upper arm | 0.84 (0.73-0.96) |
| Moves forearm | 0.83 (0.71-0.94) |
| Grasp-release items | |
| Grasps | 0.89 (0.80-0.99) |
| Holds | 0.82 (0.71-0.94) |
| Stabilizes by grasp | 0.95 (0.88-1.00) |
| Varies types of grasp | 0.93 (0.85-1.00) |
| Releases | 0.79 (0.65-0.92) |
| Fine motor adjustments items | |
| Moves fingers | 0.90 (0.80-0.99) |
| Readjusts grasps | 0.85 (0.75-0.96) |
| Manipulates | 0.81 (0.69-0.93) |
| Grip force regulation | 0.87 (0.76-0.98) |
| Coordination items | |
| Coordinates | 0.82 (0.70-0.94) |
| Orients objects | 0.85 (0.74-0.96) |
| Flow in bimanual performance | 0.90 (0.80-0.99) |

NOTE. Values are weighted kappa or as otherwise indicated.

scale with specific criteria for all items and qualifiers. A general meaning of the steps in the rating scale is (4) effective, (3) somewhat effective, (2) ineffective, (1) does not do.[16] An example of scoring criteria is provided in table 3. The raw sum score, with a range of 19-76, was transformed through the use of Rasch measurement analysis to interval measures, called the logit-based 0-100 Ad-AHA units.[16] A higher score represents a better performance.

## Procedure

Reliability properties of the Ad-AHA Stroke were studied in a subsample of 30 videotaped Ad-AHA Stroke sessions. We conducted a stratified random selection procedure with strata based on participants' ability level. This aim of this approach was to ensure that the full score range of the Ad-AHA Stroke was represented.[33,34] Intrarater reliability and SEM were determined with 1 certified Ad-AHA Stroke rater (A.V.G.), scoring on 30 videotaped Ad-AHA Stroke sessions twice, each with at least 1-month interval in between. Interrater reliability and measurement error were studied with 2 certified Ad-AHA Stroke raters (S.M. and A.V.G.) independently scoring on 30 videotaped Ad-AHA Stroke sessions. A single test occasion was preferred because we included participants in the subacute phase poststroke with ongoing recovery. Results might be affected in a test–retest design due to spontaneous recovery. The validity aspects were investigated in all 118 videotaped Ad-AHA Stroke sessions scored by 1 certified rater (S.M.).

## Statistical Analysis

Participants' clinical and demographic characteristics were displayed as frequencies with percentage and median with interquartile range (IQR).[33] Normality analysis using the Shapiro–Wilk normality test indicated that all variables were not normally distributed.

To determine reliability, intraclass correlations coefficients (ICCs) were calculated for sum scores (AHA units)[33] and weighted kappa was calculated for item scores.[35] Intrarater reliability was conducted by a 2-way mixed model, ICC (3,1), which is commonly used to establish intrarater reliability,[33] or in other words to demonstrate that a specific rater (A.V.G.) is reliable in scoring the videotaped Ad-AHA Stroke sessions. In this model, participants are considered to be randomly chosen and are expected to represent the population from which they were drawn.[33] Interrater reliability was conducted by a 2-way random effects model, ICC (2,1).[33] This model is suitable when each participant is assessed by the same set of raters.[33] Participants and raters are randomly chosen, and both are expected to represent the population from which they were drawn.[33] Values for ICC above 0.90 indicate excellent reliability, from 0.90 to 0.75 good reliability, and below 0.75 poor to moderate reliability.[33] Values of weighted kappa above 0.80 represent excellent agreement, above 0.60 substantial agreement, from 0.40 to 0.60 moderate, and below 0.40 poor agreement.[36] For ICC and weighted kappa, 95% confidence intervals (CIs) were calculated.

According to Portney and Watkins, the SEM expresses response stability in the same unit of measurement as the original measures of the scale, which allows for clinically useful interpretation of results.[33] The SEM was calculated for intrarater and interrater agreement ($SEM = SD \times \sqrt{1 - ICC}$).[33] To show systematic differences within the same rater (intrarater) and between different raters (interrater), Bland–Altman plots were created. The spread of scores around the zero point and across the range of values shows possible variability of difference within the same rater or between raters. A range of 95% limits of agreement was applied by mean difference $\pm$ 1.96 $\times$ SD of the difference.[33] Visual analysis was applied to detect potential systematic differences.

**Table 3** Example of scoring criteria of the Ad-AHA Stroke

| Item | Score | Specification |
|---|---|---|
| Grasps | | |
| | (4) Effective | Grasps objects from the table, automatically and with ease |
| | (3) Somewhat effective | Grasps a number of objects from the table but most often grasps objects from the nonaffected hand |
| | | Grasps objects from the table with delay, difficulty or awkward grasp |
| | (2) Ineffective | Almost always grasps objects form the nonaffected hand or from the assessor's hand |
| | (1) Does not do | Does not keep objects in the hand or objects are placed in the affected hand |

Table 4 Participants' clinical and demographic characteristics

| Characteristics | Reliability Sample (n=30) | Validity Sample (N=118) |
|---|---|---|
| Age at stroke onset (y), median (IQR) | 61 (49-68) | 67 (59-76) |
| Sex, n (%) | | |
| Men | 20 (67) | 74 (63) |
| Women | 10 (33) | 44 (37) |
| Time since stroke (d), median (IQR) | 138 (100-181) | 81 (57-117) |
| Type of stroke, n (%) | | |
| Ischemia | 26 (87) | 104 (88) |
| Hemorrhage | 4 (13) | 14 (12) |
| Lateralization, n (%) | | |
| Right hemiparesis | 12 (40) | 47 (40) |
| Left hemiparesis | 18 (60) | 71 (60) |
| Hand dominance, n (%) | | |
| Left | 2 (7) | 7 (6) |
| Right | 28 (93) | 110 (93) |
| Both | 0 (0) | 1 (1) |
| Affected side, n (%) | | |
| Dominant side | 14 (47) | 49 (41) |
| Nondominant side | 16 (53) | 68 (58) |
| Ambidexterity | 0 (0) | 1 (1) |
| Ad-AHA Stroke units, median (IQR) | 57 (26-77) | 50 (14-80) |
| CIRS (/42), median (IQR) | 4 (3-7) | 6 (4-8) |
| Neglect, n (%) with score<44 out of 54 | 8 (27) | 26 (22) |
| UE-FMA (/66), median (IQR) | 37 (10-57) | 40 (8-59) |
| UE-FMA hand function (/14), median (IQR) | 11.5 (1-14) | 12 (0-14) |
| ARAT (/57), median (IQR) | 11 (3-54) | 26 (3-54) |

Abbreviation: CIRS, Cumulative Illness Rating Scale.

As for convergent validity, the relation of the Ad-AHA Stroke with both the ARAT and UE-FMA was examined by using scatter plots and Spearman rank correlation coefficient (ρ).[33,37] The strength of the relation was interpreted as follows: very low, ρ=0.01-0.24; low, ρ=0.25-0.49; moderate, ρ= 0.50-0.69; high, ρ=0.70-0.89; very high, ρ=0.90-1.00.[38] Discriminative validity was studied by determining differences in the Ad-AHA Stroke scores between the following subgroups: (1) dominant hand affected versus nondominant hand affected; (2) age≥65 versus age≤64; and (3) visuospatial neglect versus no visuospatial neglect. To study discriminative validity, the scores on the Ad-AHA Stroke were compared between subgroups of patients using a Mann-Whitney *U* test.[33,37]

All statistical analyses were computed with SPSS, version 23.0.[a] The alpha level was set at $P=.05$.

## Results

In the reliability sample (n=30), participants had a median age of 61 years (IQR: 49-68), 67% of the participants were men, and the median time since stroke was 138 days (IQR: 100-181). The median score on the UE-FMA was 37 (IQR: 10-57), representing mild to severe affected upper limb motor function. In the validity sample (N=118), participants had a median age of 67 years (IQR: 59-76), 63% of the participants were men, and the median time since stroke was 81 days (IQR: 57-117). Median score on the UE-FMA was 40 (IQR: 8-54), representing mild to severe affected upper limb motor function. Participants' characteristics for both the reliability and validity sample are displayed in table 4.

### Reliability and Measurement Error

Intrarater reliability of the total score in Ad-AHA units resulted in an ICC of 0.99 (95% CI, 0.99-0.99). The SEM for intrarater agreement was 2.15 out of 100 Ad-AHA units. Weighted kappa values for item scores ranged all above 0.81 (0.68-1.00), representing substantial to excellent agreement (see table 1).

Interrater reliability of the total score in Ad-AHA units resulted in an ICC of 0.99 (95% CI, 0.99-0.99). The SEM for interrater agreement was 1.64 out of 100 Ad-AHA units. Weighted kappa values for item scores were between 0.79 (substantial) and 1 (excellent agreement) (0.65-1.00) (see table 2).

Bland−Altman plots showed no systematic differences within 1 rater or between raters, and symmetrical limits of agreement were shown: −6 to 6 for intrarater agreement and −4.4 to 4.7 for interrater agreement (fig 1).

### Validity

A very high and significant relation was found between scores on the Ad-AHA Stroke and ARAT: ρ=0.93 and between scores on the Ad-AHA Stroke and UE-FMA: ρ=0.92, supporting convergent validity. The scatterplots of the correlation between Ad-AHA Stroke and ARAT (fig 2A) and UE-FMA (fig 2B) reveal a noteworthy distribution of scores for better and poorer performance. Although we found a very high correlation between bimanual and unimanual performance, there is relatively more variability in the higher and lower ends of the Ad-AHA Stroke in relation to unimanual performance. Patients with visuospatial neglect had significantly lower Ad-AHA Stroke scores (median 11, IQR 0-75)

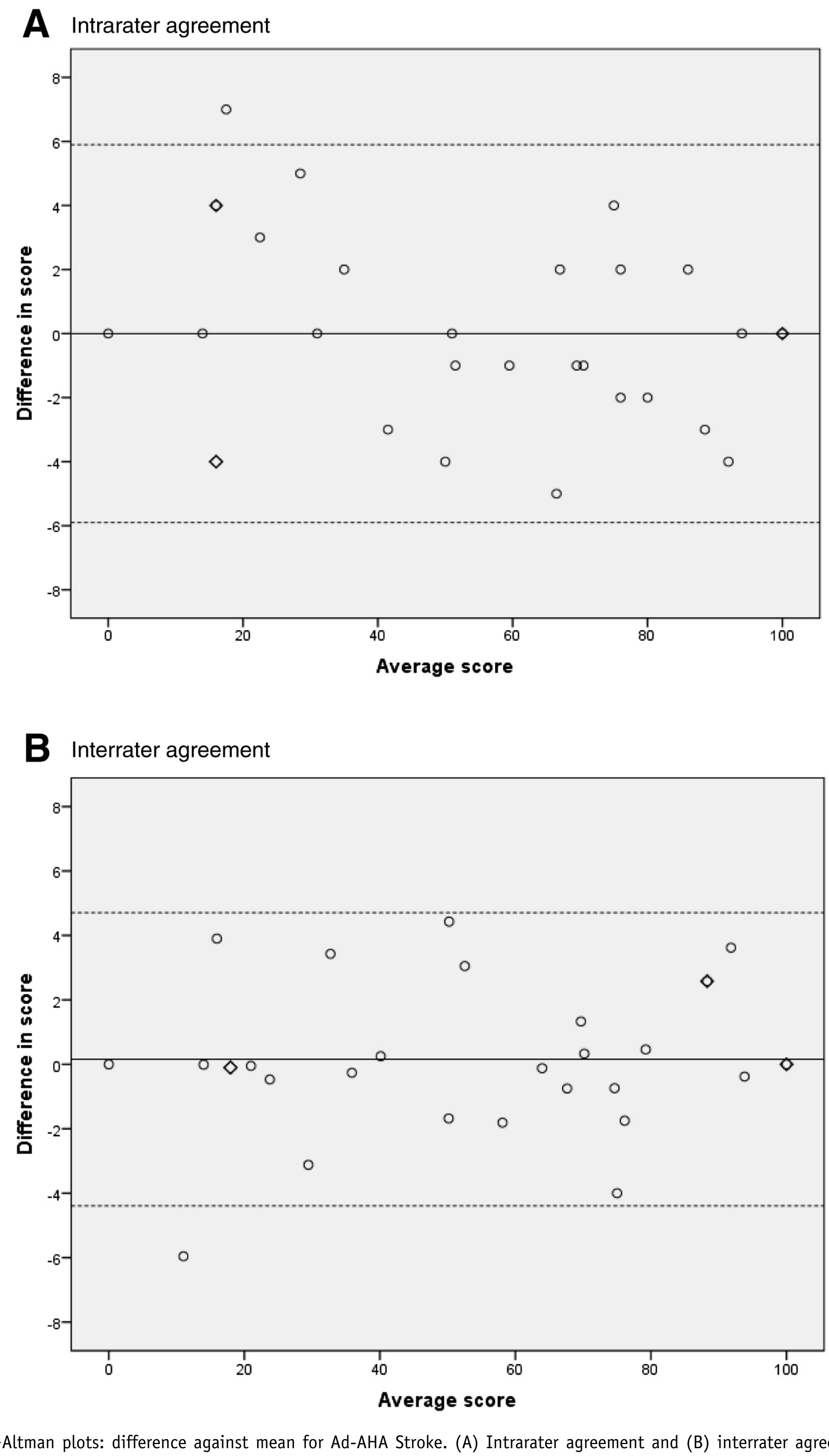


**Fig 1** Bland–Altman plots: difference against mean for Ad-AHA Stroke. (A) Intrarater agreement and (B) interrater agreement. Solid line: group mean difference. Dotted lines: limits of agreement. ◇: 2 values.

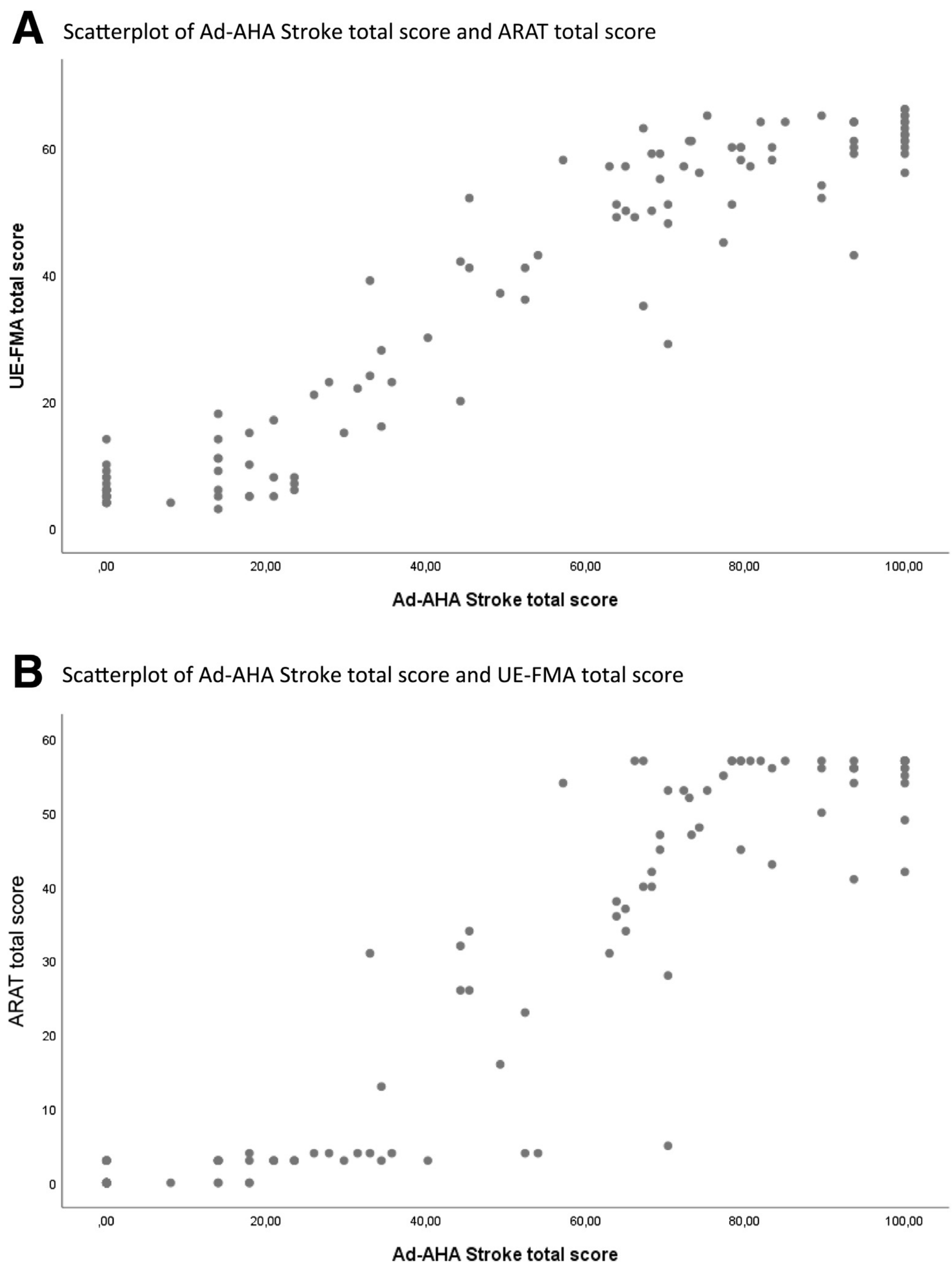


**Fig 2** Scatterplot of Ad-AHA Stroke total score and upper limb measures. (A) Scatterplot of Ad-AHA Stroke total score and ARAT total score and (B) scatterplot of Ad-AHA Stroke total score and UE-FMA total score.

than patients without visuospatial neglect (median 63, IQR 18-88) ($P$=.004). No significant differences in Ad-AHA Stroke scores were found between the 2 age groups ($P$=.530). Finally, individuals with the dominant hand affected had relatively higher Ad-AHA Stroke scores (median 65, IQR 18-79) than individuals with the nondominant hand affected (median 38, IQR 3-81), although this difference was not significant ($P$=.303) (table 5).

## Discussion

The results of our study provide new evidence that the Ad-AHA Stroke is a reliable and valid outcome measurement with an established measurement error to assess the effectiveness of the spontaneous use of the affected hand in bimanual activities in subacute stroke rehabilitation and supports the use of the Ad-AHA Stroke in research and clinical practice.

The first aim of this study was to evaluate reliability of the Ad-AHA Stroke. For clinical measurements, reliability should exceed 0.90 to ensure reasonable agreement.[33] This was demonstrated in the present study with substantial to excellent intrarater and interrater agreement found for the Ad-AHA Stroke. Although this was the first study investigating reliability properties of the Ad-AHA Stroke, similar results were found in the Kids-AHA: excellent interrater ICC (ICC=0.98 for 2-rater design and ICC=0.97 for 20-rater design) and excellent intrarater ICC (ICC=0.99).[18] Further, weighted kappa

**Table 5** Ad-AHA Stroke scores for different subgroups

| Characteristics | Ad-AHA Units 0-100 Logit Scale, Median (IQR) | P |
|---|---|---|
| Neglect | | |
| Participants with neglect (n=26) | 11 (0-75) | .004 |
| Participants without neglect (n=92) | 63 (18-88) | |
| Affected side | | |
| Dominant side affected (n=49) | 65 (18-79) | .303 |
| Nondominant side affected (n=68) | 38 (3-81) | |
| Age | | |
| Older participants (age≥65y) (n=67) | 68 (0-82) | .530 |
| Younger participants (age≤64y) (n=51) | 34 (18-75) | |

statistics was used in our study to determine reliability for item scores. Substantial to excellent agreement was found. No weighted kappa statistics but ICCs as an equivalent of weighted kappa[33] were calculated in studies using the Kids-AHA. However, the results are comparable, supporting acceptable reliability for both measurement tools; ICCs for item scores were calculated in 2 trials; a 2-rater study with 18 randomly selected video observations (ICCs between 0.35 and 1.00) and a 20-rater study with 8 video observations selected by a stratified random selection (ICCs between 0.70 and 0.92).[18] In the present study, a larger sample of video observations was used, as well as a stratified selection procedure to ensure sufficient variation in Ad-AHA Stroke performance. Although it is suggested that this procedure inflates reliability values, we believe it is important that reliability is sufficiently demonstrated for the full score range of the Ad-AHA Stroke.[33,34]

According to Portney and Watkins, the SEM reflects the response stability of outcomes on the Ad-AHA Stroke.[33] We calculated an SEM of 2.15 and 1.64 out of 100 Ad-AHA Stroke units for intra- and interrater agreement, respectively. An SEM of 1.2 (intrarater) and 1.5 (interrater) was found for the Kids-AHA,[18] and an SEM of 2.3 (interrater) was found in the adolescent AHA.[39] Our SEM results are only slightly higher but in a comparable range. These small values for SEM imply that the Ad-AHA Stroke is able to produce outcomes of bimanual performance that are reliable, precise, and stable measures of bimanual performance, and thus the Ad-AHA Stroke can be used with confidence in clinical practice.

To determine convergent validity, the relation between the Ad-AHA Stroke and other upper limb outcome measures was investigated. High correlations were expected and indeed a strong relation (ρ=0.92-0.93) between the Ad-AHA Stroke and unimanual upper limb outcome measures was demonstrated. Previously, high correlations were also found between the Ad-AHA Stroke and the Jebsen-Taylor Hand Function Test, and between the Ad-AHA Stroke and the ABILHAND for stroke.[16] The results of the latter and our study indicate a strong relation between unimanual motor function and Ad-AHA Stroke performance. This is in line with previous studies where motor function was found to affect the amount of use of the affected hand in real-world activity.[40,41] Even though we found high correlations between the Ad-AHA Stroke and unimanual upper limb measurement tools, the increased variability in the higher and lower ends of the Ad-AHA Stroke suggests that different constructs are being measured, especially in these subsections of the scoring range. Thrane et al[13] reported a similar variation in the relation between UE-FMA and arm use measured with accelerometry, especially in the upper range of the UE-FMA. And very recently Essers et al[42] showed that although an overall high correlation was demonstrated between measures of observed and perceived upper limb function poststroke, distinct subgroups of patients exist showing either a match or mismatch between observed and perceived function. Thus, a high correlation coefficient does not capture the complexity of the relation between those measures and could conceal different constructs being measured. In contrast to the UE-FMA, which captures motor impairment, and in contrast to the ARAT, in which patients are asked to demonstrate their best ability to execute a task with the affected hand, the Ad-AHA Stroke measures and describes how the affected hand is used spontaneously to perform tasks requiring the use of both hands. Therefore, the Ad-AHA Stroke reflects actual performance rather than capacity, although the Ad-AHA Stroke does not reflect performance in the strict sense as defined by the ICF. A comprehensive assessment in stroke rehabilitation should involve motor function and unimanual upper limb capacity, as well as bimanual performance measures. Moreover, the Ad-AHA has some important advantages compared to other upper limb outcome measures. First, the use of Rasch measurement analysis to produce interval level measures for reporting outcomes is a strong asset in clinical research and practice regarding evaluation of change.[16] Furthermore, the individual ability profile, based on the hierarchy of items, provides direct guidance to tailor-made interventions goals for improved bimanual performance.[16] However, some limitations of the Ad-AHA Stroke must also be acknowledged, namely that the Ad-AHA Stroke can be administered only by a certified rater, and to date, only 2 activities are developed and can be used to administer the Ad-AHA Stroke.

We expected poorer Ad-AHA Stroke performance from individuals with visuospatial neglect. We found that participants with visuospatial neglect indeed scored lower on the Ad-AHA Stroke. These finding are in line with previous studies, demonstrating more severe impairment in unimanual motor function and decreased functional outcome in patients with visuospatial neglect.[20-22] Also, we hypothesized a better performance on the Ad-AHA Stroke in individuals with the dominant hand affected, based on previous studies demonstrating less upper limb impairment in individuals with the dominant hand affected.[43] Yet, there was no significant difference between Ad-AHA Stroke scores for participants with the dominant hand affected and those with the nondominant hand affected, albeit higher Ad-AHA Stroke scores were found in individuals with the dominant hand affected compared to individuals with the nondominant hand affected. Although age at stroke onset is found to be related to upper limb motor recovery,[20] age did not affect performance on the Ad-AHA Stroke. A possible explanation could be the goal-focused approach of the Ad-AHA Stroke, intended to encourage spontaneous bimanual performance, which is different from measuring motor function capacity. The first relies on acquired basic skills, and the latter is more dependent of age-related decline.

This study was the first to establish excellent reliability for item scores and sum scores. The SEM was only 1.64-2.15 out of 100; therefore, the Ad-AHA Stroke can be considered as a useful outcome measure in intervention studies. Another strength of this study was the use of a large sample size for the validity properties. Participants' upper limb motor impairment in both the reliability and validity sample varied from mild to severe, so the results of this study can be generalized to the stroke rehabilitation population as included in our sample. Furthermore, because the Ad-AHA Stroke was developed for all degrees of upper limb impairment, the variability in upper limb impairment demonstrated in this study can be considered as a strength. Moreover, reliability was demonstrated over the full score range of the Ad-AHA Stroke. The way in which the reliability subsample was selected made it representative for the stroke rehabilitation population.[8]

### Study Limitations

Some limitations of our study need to be acknowledged. First, the Ad-AHA Stroke includes 2 possible activities, wrapping a present and making a sandwich. In the present study, all Ad-AHA Stroke sessions involved *wrapping a present*. Although it can be assumed that both activities will result in equivalent Ad-AHA Stroke scores, caution is needed when generalizing psychometric properties demonstrated in this study to the alternate activity *making a sandwich*. Further research should study reliability of both activities of the Ad-AHA Stroke. Second, reliability was demonstrated based on 2 ratings of 1 single assessment occasion for each participant. Therefore, it was possible to capture the raters' variability but not the participants' variability. Although this is a valid approach to demonstrate inter- and intrarater reliability, this procedure might enhance reliability values and should be taken into account when interpreting the excellent values for ICC and SEM. Next steps in investigating the psychometric properties of the Ad-AHA Stroke should involve a test−retest design that covers the assessment of 2 subsequent sessions in 1 participant, preferably in the chronic phase poststroke. In the present study, a strong relation between the Ad-AHA Stroke and upper limb motor function and capacity measures was demonstrated. Future research should explore the relation between Ad-AHA Stroke and lifelike measures of upper limb performance like self-report measures and accelerometry.

## Conclusion

Assessment of the upper limb in stroke rehabilitation should involve both motor impairment and activity as well as bimanual performance. The Ad-AHA Stroke captures actual bimanual performance, and thereby it provides an additional aspect of upper limb assessment. The results of this study demonstrate that the Ad-AHA Stroke has good to excellent reliability and low SEM for patients with subacute stroke. High convergent validity with ARAT and UE-FMA and discriminative validity were supported. The established psychometric properties of the Ad-AHA Stroke allow application in clinical practice and research.

## Supplier

a. SPSS, version 23.0; IBM Corp.

## Keywords

Outcome measure; Rehabilitation; Stroke; Task performance; Upper extremity

## Corresponding author

Annick Van Gils, MSc, Department of Rehabilitation Sciences, Catholic University of Leuven, Tervuursevest 101, bus 1501, 3001 Leuven, Belgium. *E-mail address:* annick.vangils@kuleuven.be.

## References

1. Mozaffarian D, Benjamin EJ, Go AS, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation 2015;131:e29-322.
2. Miller EL, Murray L, Richards L, et al; American Heart Association Council on Cardiovascular Nursing and the Stroke Council. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. Stroke 2010;41:2402-48.
3. Nakayama H, Jørgensen 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. Chen CM, Tsai CC, Chung CY, Chen CL, Wu KP, Chen HC. Potential predictors for health-related quality of life in stroke patients undergoing inpatient rehabilitation. Health Qual Life Outcomes 2015;13:118.
5. Veerbeek JM, Kwakkel G, Van Wegen EH, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. Stroke 2011;42:1482-8.
6. Bouffioulx D, Arnould C, Thonnard JL. Satisfaction with activity and participation and its relationships with body functions, activities, or environmental factors in stroke patients. Arch Phys Med Rehabil 2011;92:1404-10.
7. Bailey RR, Klaesner JW, Lang CE. An accelerometry-based methodology for assessment of real-world bilateral upper extremity activity. PLoS One 2014;9:e103135.
8. Holmefur M, Krumlinde-Sundholm L. Psychometric properties of a revised version of the Assisting Hand Assessment (Kids-AHA 5.0). Dev Med Child Neurol 2016;58:618-24.
9. Krumlinde-Sundholm L, Holmefur M, Kottorp A, Eliasson AC. The Assisting Hand Assessment: current evidence of validity, reliability, and responsiveness to change. Dev Med Child Neurol 2007; 49:259-64.
10. Lemmens RJ, Timmermans AA, Janssen-Potten YJ, Smeets RJ, Seelen HA. Valid and reliable instruments for arm-hand assessment at ICF activity level in persons with hemiplegia: a systematic review. BMC Neurol 2012;12:21.
11. World Health Organization. International Classification of Functioning, Disability and Heatlh. Geneva, Switzerland: World Health Organization; 2001.
12. Uswatte G, Taub E, Morris D, Light K, Thompson PA. The Motor Activity Log-28 assessing daily use of the hemiparetic arm after stroke. Neurology 2006;67:1189-94.
13. Thrane G, Emaus N, Askim T, Anke A. Arm use in patients with subacute stroke monitored by accelerometry: association with motor impairment and influence on self-dependence. J Rehabil Med 2011;43: 299-304.
14. Bailey RR, Klaesner JW, Lang CE. Quantifying real-world upper-limb activity in nondisabled adults and adults with chronic stroke. Neurorehabil Neural Repair 2015;29:969-78.

15. Chen S, Wolf SL, Zhang Q, Thompson PA, Winstein CJ. Minimal detectable change of the actual amount of use test and the Motor Activity Log: the EXCITE trial. Neurorehabil Neural Repair 2012;26:507-14.
16. Krumlinde-Sundholm L, Lindkvist B, Plantin J, Hoare B. Development of the assisting hand assessment for adults following stroke: a Rasch-built bimanual performance measure. Disabil Rehabil 2017 Oct 31 [Epub ahead of print].
17. Krumlinde-Sundholm L, Eliasson A-C. Development of the Assisting Hand Assessment: a Rasch-built measure intended for children with unilateral upper limb impairments. Scand J Occup Ther 2003;10:16-26.
18. Holmefur M, Krumlinde-Sundholm L, Eliasson A-C. Interrater and intrarater reliability of the Assisting Hand Assessment. Am J Occup Ther 2007;61:79-84.
19. Holmefur M, Aarts P, Hoare B, Krumlinde-Sundholm L. Test-retest and alternate forms reliability of the Assisting Hand Assessment. J Rehabil Med 2009;41:886-91.
20. Meyer S, De Bruyn N, Lafosse C, et al. Somatosensory impairments in the upper limb poststroke: distribution and association with motor function and visuospatial neglect. Neurorehabil Neural Repair 2016;30:731-42.
21. Nijboer T, van de Port I, Schepers V, Post M, Visser-Meily A. Predicting functional outcome after stroke: the influence of neglect on basic activities in daily living. Front Hum Neurosci 2013;7:182.
22. Nijboer TC, Kollen BJ, Kwakkel G. The impact of recovery of visuospatial neglect on motor recovery of the upper paretic limb after stroke. PLoS One 2014;9:e100584.
23. WHO MONICA Project Principal Investigators. The World Health Organization MONICA project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. J Clin Epidemiol 1988;41:105-14.
24. See J, Dodakian L, Chou C, et al. A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. Neurorehabil Neural Repair 2013;27:732-41.
25. Stolk-Hornsveld F, Crow JL, Hendriks EP, van der Baan R, Harmeling-van der Wel BC. The Erasmus MC modifications to the (revised) Nottingham Sensory Assessment: a reliable somatosensory assessment measure for patients with intracranial disorders. Clin Rehabil 2006;20:160-72.
26. Prescott R, Garraway W, Akhtar A. Predicting functional outcome following acute stroke using a standard clinical examination. Stroke 1982;13:641-7.
27. Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability? Int Disabil Stud 1988;10:64-7.
28. Fortin M, Hudon C, Dubois MF, Almirall J, Lapointe L, Soubhi H. Comparative assessment of three different indices of multimorbidity for studies on health-related quality of life. Health Qual Life Outcomes 2005;3:74.
29. 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.
30. Halligan P, Wilson B, Cockburn J. A short screening test for visual neglect in stroke patients. Disabil Rehabil 1990;12:95-9.
31. Connell LA, Tyson SF. Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. Arch Phys Med Rehabil 2012;93:221-8.
32. Hsueh IP, Hsu MJ, Sheu CF, Lee S, Hsieh CL, Lin JH. Psychometric comparisons of 2 versions of the Fugl-Meyer Motor Scale and 2 versions of the Stroke Rehabilitation Assessment of Movement. Neurorehabil Neural Repair 2008;22:737-44.
33. Portney LG, Watkins MP. Foundations of clinical research: application to practice. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
34. van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the Action Research Arm test: a practical test of upper extremity function in patients with stroke. Arch Phys Med Rehabil 2001;82:14-9.
35. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257-68.
36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
37. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: practical guides to biostatistics and epidemiology. New York: Cambridge; 2011.
38. Munro BH. Statistical methods for health care research. Philadelphia: Lippincott Williams & Wilkins; 2005.
39. Louwers A, Beelen A, Holmefur M, Krumlinde-Sundholm L. Development of the Assisting Hand Assessment for adolescents (Ad-AHA) and validation of the AHA from 18 months to 18 years. Dev Med Child Neurol 2016;58:1303-9.
40. Rand D, Eng JJ. Predicting daily use of the affected upper extremity 1 year after stroke. J Stroke Cerebrovasc Dis 2015;24:274-83.
41. Bailey RR, Birkenmeier RL, Lang CE. Real-world affected upper limb activity in chronic stroke: an examination of potential modifying factors. Top Stroke Rehabil 2015;22:26-33.
42. Essers B, Meyer S, De Bruyn N, et al. Mismatch between observed and perceived upper limb function: an eye-catching phenomenon after stroke. Disabil Rehabil 2018 Mar 22 [Epub ahead of print].
43. Harris JE. Individuals with the dominant hand affected following stroke demonstrate less impairment than those with the nondominant hand affected. Neurorehabil Neural Repair 2006;20:380-9.