

ORIGINAL REPORT

RELIABILITY AND VALIDITY OF A NEW DIAGNOSTIC DEVICE FOR QUANTIFYING HEMIPARETIC ARM IMPAIRMENTS: AN EXPLORATORY STUDY

Levinia Lara van der VELDEN, MSc^{1,3}, Joyce Lisanne BENNER, PhD¹, Bram ONNEWEER, PhD^{1,3}, Claudia Josephina Wilhelmina HAARMAN, MSc², Ruud Willem SELLES, PhD^{1,3,4}, Gerard Maria RIBBERS, MD, PhD^{1,3} and Marij Eugenie ROEBROECK, PhD¹

From ¹Erasmus MC University Medical Center Rotterdam, Department of Rehabilitation Medicine, Rotterdam, ²Hankamp Rehab, Enschede, ³Rijndam Rehabilitation, Rotterdam and ⁴Erasmus MC University Medical Center Rotterdam, Department of Plastic and Reconstructive Surgery, The Netherlands

Objective: To assess test-retest reliability and validity of a new diagnostic device, the Shoulder Elbow Perturbator, to quantify muscle weakness, abnormal synergy, (muscle activity-related) spasticity, and changes in viscoelastic joint properties of the elbow.

Subjects: Stroke patients, adults with cerebral palsy and healthy controls.

Methods: Test-retest reliability was evaluated using intra-class correlations (ICC) and assessment of measurement error. The device's validity was evaluated by demonstrating differences between patients and healthy controls, and correlations of spasticity and abnormal synergy outcomes using the clinical Modified Tardieu Scale, the Fugl-Meyer Assessment, and the Test of Arm Selective Control. **Results:** Reliability was excellent, with an ICC > 0.75 for synergy and ICCs > 0.90 for all other impairments, with relatively small measurement errors. Validity was confirmed by group differences between patients and healthy controls for muscle weakness, spasticity, and viscoelastic joint properties, but not for abnormal synergy. Correlation analysis with clinical scales confirmed validity for spasticity, while, for synergy, correlations were found in the patients with stroke, but not those with cerebral palsy.

Conclusion: This new diagnostic device is a reliable and valid instrument to assess multiple upper limb impairments in patients with neurological conditions, supporting its use in clinical practice. Further studies are needed to confirm these findings.

Key words: reliability and validity; muscle spasticity; muscle strength; synergy; viscoelastic joint properties; robotics.

Accepted March 8, 2022; Epub ahead of print April 1, 2022

J Rehabil Med 2022: jrm00283

DOI: 10.2340/jrm.v54.12

Correspondence address: Levinia van der Velden, Erasmus MC University Medical Center Rotterdam, Department of Rehabilitation Medicine, 3000 CA Rotterdam, The Netherlands. E-mail: l.vandervelden@erasmusmc.nl

LAY ABSTRACT

Reliable and valid quantitative measurement of arm impairments due to neurological conditions is important for clinical decision-making. Current clinical measurement tools are subjective and not reliable. The aim of this study was therefore to assess the reliability and validity of a new diagnostic device to quantify multiple arm impairments (muscle weakness, abnormal synergy, spasticity, changes in viscoelastic joint properties) in patients with stroke, adults with cerebral palsy (CP), and healthy controls. Reliability was excellent for all impairments and was analysed by testing the consistency of the results on 2 occasions in the same subject. Validity was confirmed by group differences between patients and healthy controls for all impairments, except for synergy. Also, correlations of the results with clinical measurement tools were found for spasticity and synergy, except for synergy in patients with CP. In conclusion, this new diagnostic device is a reliable and valid instrument to assess multiple impairments of the elbow in patients with neurological conditions. Further studies are needed to confirm these findings.

Assessing sensorimotor impairments resulting from brain damage, such as muscle weakness, abnormal synergy, spasticity, or changes in viscoelastic joint properties, is essential for prognosis, treatment selection, and evaluation of interventions (1, 2). Ideally, such assessment is rater-independent, reliable, valid, and responsive to change. However, common clinical tests, such as the Modified Ashworth Scale (MAS) and the Modified Tardieu Scale (MTS) to measure spasticity, and the Fugl-Meyer Assessment (FMA) or the Test of Arm Selective Control (TASC) to assess abnormal synergy (3–8) use ordinal scales, suffer from low reliability, and lack sensitivity. For example, the ordinal MAS and MTS show intra-class correlation (ICC) values of 0.58 (5). Despite the FMA's high ICC values, it is criticized for its sensitivity to change, due to a large minimal important change, and its responsiveness, due

to a ceiling effect (9, 10). The TASC is a relatively new, not widely used, clinical test. To date, only 1 study has investigated this new measure, and reported high reliability and validity scores (8). However, the TASC uses ordinal scales and its sensitivity to change is unknown.

Robotic devices that impose precise and standardized perturbations quantitatively measure sensorimotor responses, allowing them to overcome the limitations of current clinical tests and enabling better and objective differentiation of the impairments in a paretic extremity (11, 12). Test-retest reliability of such robotic diagnostic devices is good, with high ICC values (13–15). However, these devices are often limited to measuring only 1 or 2 impairments and require extensive measurement time with a high burden for patients (15, 16).

In a previous study, we demonstrated content validity of the Shoulder-Elbow Perturbator (SEP), an innovative single diagnostic device to measure muscle weakness, abnormal flexion synergy, (muscle activity-related) spasticity, and changes in viscoelastic joint properties of the elbow in patients with upper limb impairments (17). The results showed that patient impairment values were outside the reference intervals of healthy controls, and that specific patient profiles can be identified. This study investigates the SEP test-retest reliability in chronic stroke patients and adults with cerebral palsy (CP). Furthermore, we assess the validity of SEP measurements, addressing: (i) construct validity by exploring differences in SEP outcomes between chronic stroke patients, adults with CP, and healthy controls for all impairments, and (ii) criterion validity by investigating correlations between SEP outcomes and assessments using clinical tests.

METHODS

Participants

Chronic stroke (>6 months after stroke) and CP patients were recruited from an outpatient rehabilitation clinic based on impaired upper limb function documented in their medical records. Inclusion criteria were: (a) self-reported upper limb impairment; (b) ability to actively abduct their shoulder up to 80° and extend the elbow with a minimum of 5°; (c) minimal passive range of motion in the shoulder joint of 0–90° abduction and 0–45° anterior flexion; (d) ability to follow instructions; (e) aged 18 years and older. Exclusion criteria were: patients with severe hemiplegic shoulder pain, those receiving pharmacological drugs for spasticity treatment (i.e. botulinum toxin, per-oral baclofen, or shockwave therapy), or history of pre-existing neuromusculoskeletal disorders that would influence upper limb function. As a reference, adult healthy controls

with no known history of neurological or orthopaedic disorders were recruited. In accordance with the Declaration of Helsinki, the Medical Ethical Committee of the Erasmus University Medical Center Rotterdam approved the study protocol, and all participants provided written informed consent.

Procedures

Data were extracted from the medical records regarding the type of stroke and time post-stroke for stroke patients, and the Gross Motor Function Classification System (GMFCS) level and Manual Ability Classification System (MACS) level for adults with CP. All participants were assessed twice. During the first visit, age, sex, dexterity, body mass, and height were recorded, followed by SEP measurements, and, if applicable, the paretic arm side was recorded and clinical tests were performed prior to SEP measurements: FMA of the upper limb (FMA-UL) in patients with stroke or TASC in patients with CP, and MTS in both groups. The order of the different measurements with the SEP was randomized over participants to prevent order and fatigue effects. A resting period of at least 5 min was scheduled between measurements, with 5 s between repetitions. SEP measurements were repeated 7–10 days after the first visit, at the same part of the day (morning or afternoon).

Measurement instruments

Shoulder-Elbow Perturbator: The SEP (Hankamp Rehab, Enschede, The Netherlands) was used to measure isometric muscle strength, synergy, spasticity, and viscoelastic joint properties of the elbow (Fig. 1) (17).

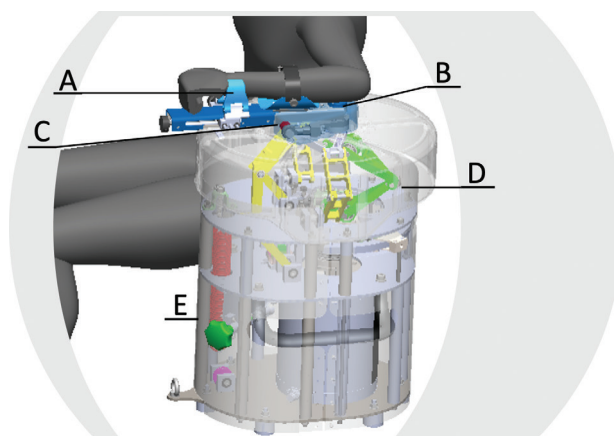


Fig. 1. The measurement device: the Shoulder Elbow Perturbator (SEP), to quantify muscle weakness, abnormal synergy, spasticity, and changes in viscoelastic joint properties of the elbow. (A) Fixation of the wrist. (B) Servo motor rotation point aligned with the elbow (medial epicondyle of the humerus). (C) Position force sensor. (D) Sarrus mechanism (yellow) and torque link (green). (E) Passive spring mechanism with spring in red and cable in pink.

The device can manipulate the elbow joint angle while partially or fully compensating the weight of the arm. A lever arm supports the forearm, and a high torque (Nm) direct-drive servo motor (“HIWIN TMS3C”, Offenburg, Germany) aligned with the medial epicondyle of the humerus controls the elbow joint angle. A fixed-point position force sensor measures elbow resistance (N) during rotation, which we converted to elbow torque (Nm). An encoder recorded the lever arm’s angular position. Vertical displacement was controlled by the Sarrus linkage mechanism (Fig. 1D, *yellow*), while the torque link allowed the elbow to rotate (Fig. 1D, *green*). The weight of the arm can be (partially) compensated with a passive spring mechanism to allow weight support. A cable (Fig. 1E, *pink*) attached to the spring (Fig. 1E, *red*) was routed over several pulleys to allow an independent upward force during the elbow rotation. The amount of arm compensation was adjusted in increments of 25% using the green knob (Fig. 1).

Several procedures ensured participant safety throughout SEP measurements: (i) restriction in elbow rotation with both mechanical and software end-stops; (ii) limitation at 66 Nm of maximum torque; (iii) down-regulated maximum torque limit to 11 Nm during fast elbow rotation; (iv) an emergency button strapped around the leg to stop the motor; and (v) a quick-release system on the wrist fixation.

Measurement protocols and data processing

For all measurements, participants were seated with the upper body fixed with Velcro straps to the chair and the hemiparetic forearm or non-dominant arm (in healthy controls) fixed to the SEP with a clamp (Fig. 1).

Isometric muscle strength (Nm) was measured in 80° shoulder abduction and 90° elbow flexion. Participants relaxed their arm for 5 s to establish resting torque and were subsequently requested to extend or flex their elbow to maximum force in 5 s, with 3 repetitions for each measure. If a repetition deviated by >10% from the largest peak torque, it was omitted, and an additional repetition was performed. The mean of 3 repetitions maximum torques was calculated for elbow flexion and extension as the outcome parameters for muscle strength.

Synergy was quantified using maximum active elbow extension angles (°) under different arm weight support levels. Participants started in the maximal elbow flexion position and were requested to extend their elbow slowly as far as possible (to minimize the velocity-dependent resistance and prevent reflex-activity), under randomly assigned different arm weight support levels (100%, 75%, 50%, 25% and 0%). To relate the maximum extension angle to support levels as a measure of synergy, the maximum elbow exten-

sion angle at 100% support was taken at zero and a linear regression line was fitted through all 5 extension angles. The linear regression line was set to zero when a negative slope value appears due to hyperextension or compensation strategies.

Spasticity was quantified as maximum elbow torque (Nm) during a fast passive elbow extension at 100°/s (14, 16). Within the maximum elbow torque, the inertial components of the SEP and the human arm, as well as reflex and non-reflex properties are included. First, passive range of motion was determined by manually moving the forearm to maximum flexion and extension positions. Then, 3 repetitions of fast elbow extension were imposed by the SEP with a constant velocity of 100°/s. The force sensor measured the resistance applied by the arm of participants. If a predefined maximum resistance (11 Nm) was attained before reaching the maximum extension position, the extension movement was stopped and held in position for 5 s. The highest maximum torque was averaged over 3 repetitions and used as outcome parameter for spasticity.

Viscoelastic joint properties were quantified as the slope (Nm/°) of resistance at different joint angles during a slow passive elbow extension and flexion. While the participants’ arm was rotated 3 times at a constant velocity of 6°/s from the maximum flexion to extension positions and back, the force sensor measured its resistance. For each repetition, the mean torque was extracted at 10 evenly spaced elbow positions; a linear regression line was fitted. Mean slopes over repetitions was used as the outcome parameter for viscoelastic joint properties.

Clinical measurement instruments

The FMA-UL assesses upper limb abnormal synergy by evaluating 32 items (6). Each item is scored using a 3-point ordinal scale: 0 (cannot perform), 1 (can partially perform), and 2 (can fully perform). For this study, to compare clinical measurement instruments with SEP outcomes, only stadium 2, 3, and 4 were used to evaluate abnormal synergy (flexion pattern) of the shoulder, elbow, and forearm. The FMA-UL score is the sum of 12 items totalling 24 points, with 24 indicating no abnormal synergies.

In patients with CP, we used the TASC to assess abnormal synergy by evaluating movement patterns, coordination, fluency, mirroring, and speed of upper limb motions (8). The TASC comprises 8 motions of the shoulder, elbow, wrist, hand, and finger joints of the hemiparetic arm, of which we used 4: shoulder flexion/extension, shoulder abduction/adduction, elbow flexion/extension, and forearm supination/pronation, further indicated as partial-TASC scores. Participants were instructed to actively move their

arm to an extreme position from the starting position with the elbow, wrist, and fingers fully extended, or the latter with the elbow flexed to 90°, forearm in a neutral position, wrist and fingers extended. Movements were performed using verbal cues to maintain correct pace and graded using a 3-point ordinal scale: 0 (no selective voluntary motor control), 1 (impaired selective voluntary motor control), and 2 (intact selective voluntary motor control). The partial-TASC score was the sum of 4 items, with 8 indicating no abnormal synergies.

The MTS is a clinical measurement instrument frequently used to assess spasticity. It assesses spasticity by moving the hemiparetic arm at slow velocity to measure the maximum elbow extension angle, referred to as R2. Angles are measured using a goniometer. Next, the hemiparetic arm is moved manually as fast as possible to measure final elbow extension angle, referred to as R1. The quality of muscle reaction was scored on a 4-point ordinal scale: 0 (no resistance), 1 (slight resistance), 2 (clear catch), 3 (fatigable clonus), 4 (infatigable clonus), and the angle of muscle reaction was calculated by subtracting the 2 angles from each other, defined as R2–R1 (18).

Statistical analysis

All SEP measurement calculations were performed using Matlab (The MathWorks, Inc., Natick, Massachusetts, United States). Statistical analyses were performed in R version 4.0.2. According to the Shapiro-Wilk test, data were normally distributed for muscle strength data, hence parametric tests were used. For synergy, spasticity, and viscoelastic joint properties, non-parametric tests were used as data were not normally distributed. For all impairments, variance between repeated measurements was normally distributed, therefore Bland & Altman and analysis of variance (ANOVA) was allowed.

Reliability

To calculate test-retest reliability, we first visualized differences between test and re-test scores using Bland-Altman plots, including limits of agreement (LOA). The LOA was calculated as $\text{LOA} = \text{mean}_{\text{diff}} \pm 1.96 * \text{SD}_{\text{diff}}$, and indicates the measurement error (19).

Next, we performed an ANOVA for each impairment measured with the SEP. Using a random-effects model, we estimated the variance components between participants (σ_p^2), test occasions (σ_o^2), and residual variance, including the interaction between the 2 factors, and residual random error. Intraclass correlation coefficients (ICC_{2,1}) were calculated as the ratio of the variance between participants to the variance between participants plus error variance. ICCs above 0.75 indicated excellent reliability, those from 0.40 to 0.75,

fair to good reliability, and those below 0.40, poor reliability (20).

In addition to ICC, the standard error of measurement ($\text{SEM}_{\text{agreement}}$) was calculated as the square root of the error variance, and the smallest detectable change (SDC) was calculated as $\text{SDC} = 1.96 * \text{SEM} * \sqrt{2}$ (19). SDC was also presented relative to the total measurement range, i.e. $\text{SDC}\% = (\text{SDC}/\text{range}) * 100\%$ to compare the SDC between the different parameters. We assume a measurement instrument with an SDC% smaller than 20% is adequate to assess changes in clinical practice (21, 22).

Validity

Construct validity was evaluated using a predefined hypothesis that patients differ from healthy controls for all SEP outcomes and, more specifically, for subgroups of patients (stroke or CP) vs healthy controls (19). Between-group differences were tested using the independent *t*-test for muscle strength and the unpaired 2-sample Wilcoxon test for synergy, spasticity, and viscoelastic joint properties. To determine between-group differences across 3 subgroups, we used 1-way ANOVA and the post-hoc Tukey's honestly significant difference (Tukey HSD) for muscle strength and Kruskal-Wallis with Mann-Whitney *U* as a post-hoc test for synergy, spasticity, and viscoelastic joint properties. A *p*-value <0.05 was considered statistically significant, since the analyses were explorative, and the number of primary comparisons was limited to 5 SEP outcomes and 2 groups (patients vs controls).

Criterion validity were evaluated by correlating SEP outcomes for synergy and spasticity with clinical tests of the same impairments using Spearman's correlation coefficients (19). According to COSMIN recommendations, we hypothesized that SEP spasticity and synergy outcomes would correlate moderately (between 0.50 and 0.70) with clinical tests of these impairments (23).

Sample size calculation

Sample size estimation was based on the ICC for the impairment outcomes. Following Walter et al. (24), we tested whether the expected ICC (ρ_1) was equal (null hypothesis) or higher (alternative hypothesis) than the acceptable ICC (ρ_0). A ρ_0 value of 0.60 was used, based on literature, and a ρ_1 of 0.85 (5, 25). The number of observations (*n*) was fixed at 2. Using a significance level (α) of 0.05 and a power ($1-\beta$) of 0.80, a sample size of 21 participants per group (patient and healthy control group) is required. To account for an expected dropout of 15–30%, the current study aimed at sample sizes of 25–30 participants.

Table I. Demographics and clinical characteristics of participants per subgroup of stroke, cerebral palsy (CP), or healthy controls

	Stroke	CP	Healthy
Participants, <i>n</i> (%)	9 (17)	20 (37)	25 (46)
Sex, males, <i>n</i> (%)	8 (89)	9 (45)	8 (32)
Age, years: mean (SD); range	71 (8.6); 53–82	40 (14.3); 23–68	48 (18.9); 21–75
Type of stroke/CP, <i>N</i> (%)			NA
Ischaemic	5 (56)		
Haemorrhagic	3 (33)		
Both	1 (11)		
Spastic			
Unilateral		11 (55)	
Bilateral		9 (45)	
Time since stroke in months: mean (SD); range	72 (46); 15–142	NA	NA
MACS level, <i>n</i> (%)	NA		NA
I		7 (35)	
II		4 (20)	
III		6 (30)	
IV		3 (15)	
V		0 (0)	
GMFCS level, <i>n</i> (%)	NA		NA
I		5 (25)	
II		6 (30)	
III		5 (25)	
IV		3 (15)	
V		1 (5)	
Measured arm, left, <i>n</i> (%)	5 (56)	9 (45)	24 (96)
MTS, angle of muscle reaction (°): median (IQR); range	0 (0–80); 0–90	0 (0–0); 0–49	NA
MTS, quality of muscle reaction <i>n</i> (%)			NA
0	6 (67)	12 (60)	
1	0 (0)	5 (25)	
2	3 (33)	3 (15)	
3	0 (0)	0 (0)	
4	0 (0)	0 (0)	
5	0 (0)	0 (0)	
FMA-UL flexion: median (IQR); range	20 (12–24); 8 to 24	NA	NA
Partial-TASC: median (IQR); range	NA	8 (5–8); 3–8	NA

CP: cerebral palsy; FMA-UL: Fugl-Meyer Assessment of the upper limb; GMFCS: Gross Motor Function Classification System; IQR: interquartile range; MACS: Manual Ability Classification System; MTS: Modified Tardieu Scale; angle of muscle reaction (R2–R1); *n*: Number; NA: not applicable; SD: standard deviation; TASC: Test of Arm Selective Control; Q: Quality of muscle reaction (scale 0–5).

RESULTS

Participant characteristics

A total of 54 adults participated in the study: 9 with a stroke, 20 with CP, and 25 healthy controls. Mean

age for stroke patients was 71 years (SD±9 years), for CP 40 years (SD±14 years), and for healthy controls 48 years (SD±19 years). Demographics and clinical characteristics of all participants are shown in Table I.

Reliability

All measurements indicated excellent reliability between test and re-test (Table II); all ICCs were above 0.90, except for synergy, which was 0.78. In line with this, the Bland-Altman plots (Fig. 2) and the SDC indicated relatively small limits of agreement, also relative to the mean values of these outcomes (SDC%<20). Only synergy showed an SDC% above 20, indicating that this instrument is less sensitive for measuring synergy in clinical practice. Calculating test-retest reliability in specific groups (patients and healthy control separately) the results showed a good reliability, except for viscoelasticity in patients due to high SEM values (see Additional file 1).

The Bland-Altman plots also indicated that the difference between test and re-test measurements was not related to the value of the impairment, again except for synergy. For synergy, the Bland-Altman plot has an unusual “<”-shaped pattern, especially for the participants with low slopes (Fig. 2). The reason for this is that a large number of participants had a zero or negative slope (corrected to zero) on either the test or re-test. If either the test or re-test value was zero, the value of these participants is displayed on the $y=(-)2\times$ line in the Bland-Altman plot, explaining the “<”-shaped pattern.

Validity

The construct validity results (Fig. 3) were in line with the predefined hypothesis that muscle strength, spasticity, and viscoelastic properties differ between the hemiparetic upper limb group (stroke and CP combined) and healthy controls. Only for synergy, no between-group difference was found ($p=0.38$). When comparing patients with stroke or patients with CP separately with healthy controls, between-group differences were also found (for all, $p<0.01$) except for extension muscle strength and synergy. Extension muscle strength showed no difference between stroke

Table II. Test-retest reliability outcomes (intraclass correlation coefficient; ICC), standard error of measurement (SEM), smallest detectable change (SDC) and SDC relative to measurement range in percentage (SDC%) of Shoulder-Elbow Perturbator (SEP) measurement for the 4 upper limb impairments (muscle strength in 2 directions)

	ICC _{2,1}	SEM	SDC	SDC % range
MVTF [Nm]	0.90	4.0	11.2	18.8
MVTE [Nm]	0.90	3.8	10.7	17.4
Synergy [slope: °/25% arm weight support]	0.78	0.6	1.7	25.2
Spasticity [Nm]	0.95	0.6	1.6	12.7
Viscoelasticity [slope: Nm/°]	0.97	0.008	0.02	6.6

CP: cerebral palsy; ICC_{2,1}: intraclass correlation coefficient; MVTE: maximum voluntary torque for extension; MVTF: maximum voluntary torque for flexion.

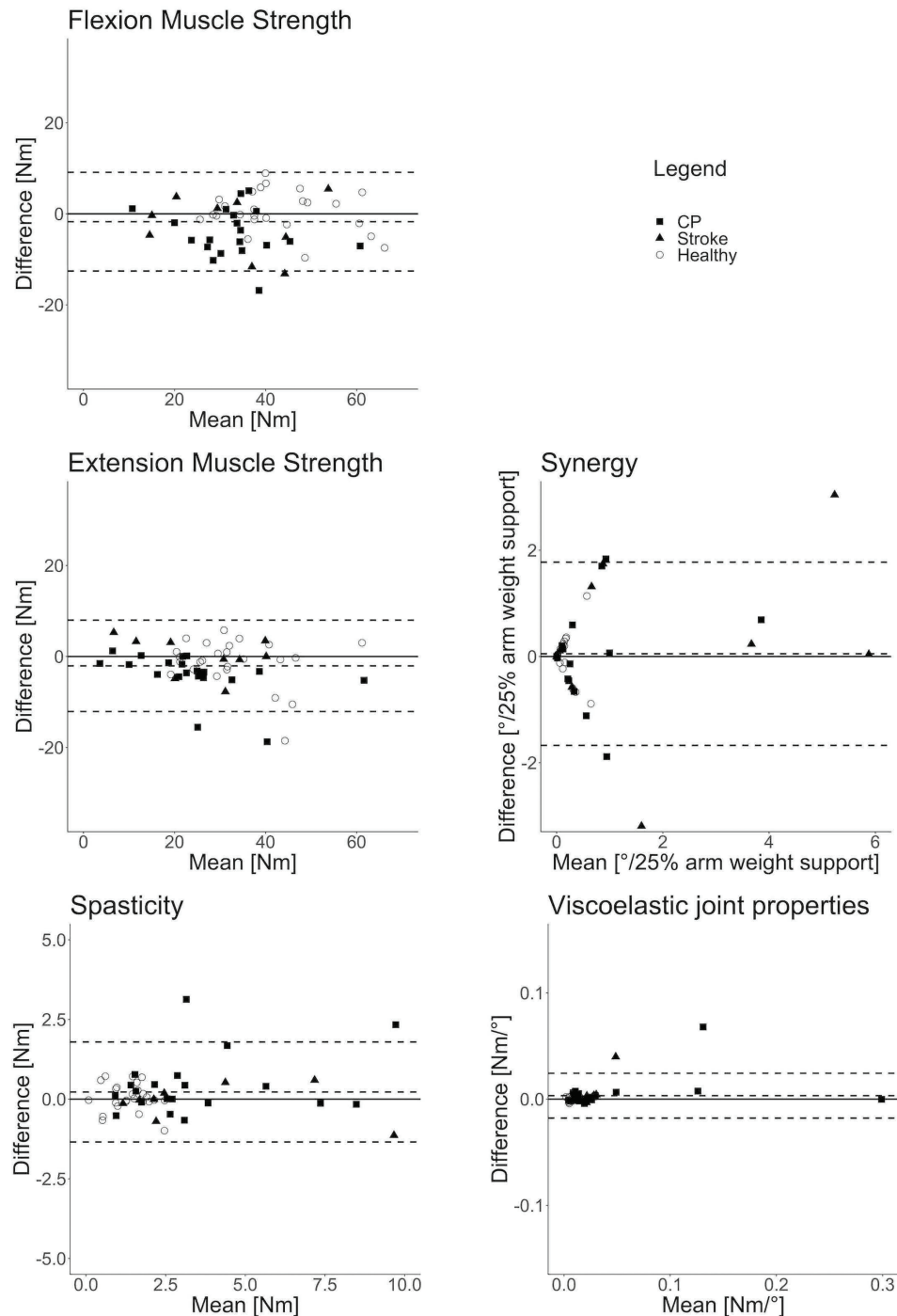


Fig. 2. Bland-Altman plots indicating agreement between test and re-test measurements for the 4 upper limb impairments (muscle strength in 2 directions). *Dashed lines* indicate the mean systematic difference and upper and lower limits of agreement. The *solid line* is the zero-line. There is no pattern of the difference values between test and re-test of all impairments relative to the mean, except for synergy. A $y=2x$ or $y=-2x$ line occurs for synergy due to negative slope values in either the test or re-test measurement, which were corrected to zero, while a positive slope value was measured in the other measurement. CP: cerebral palsy.

patients and healthy controls ($p=0.23$), and synergy showed no difference between patients with CP and healthy controls ($p=1$).

The results for criterion validity (Fig. 4) showed that, for synergy, FMA-UL and synergy SEP outcomes correlated moderately in stroke patients ($r=0.69$, $p=0.04$).

In adults with CP, a weak correlation was found between partial-TASC and synergy SEP outcomes ($r=-0.3$, $p=0.19$). Furthermore, SEP spasticity outcomes correlated well with the MTS in stroke patients (angle of muscle reaction: $r=-0.78$, $p=0.01$ and quality of muscle reaction: $r=0.78$, $p=0.01$), and moderate in adults with

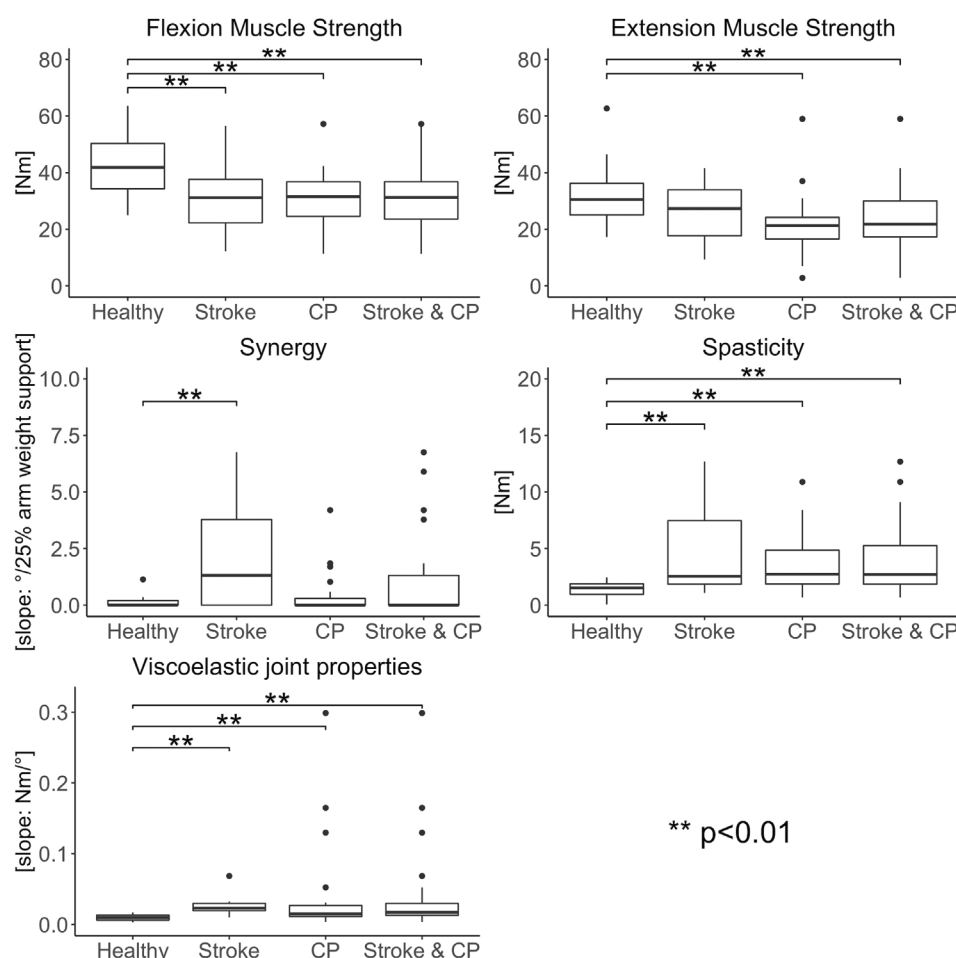


Fig. 3. Differences between the entire hemiparetic upper limb group (stroke and cerebral palsy (CP) combined), subgroups of stroke and CP separately, and healthy controls for all recorded elbow impairments. Group differences were found between the total hemiparetic upper limb group and healthy controls for all impairments except for synergy. Between-group differences for subgroups were significant for 3 Shoulder-Elbow Perturbator (SEP) impairments: muscle strength, spasticity, and viscoelastic joint properties. Only for synergy and extension muscle strength, no subgroup difference was found between healthy controls and patients with CP.

CP (angle of muscle reaction: $r = -0.66$, $p = 0.001$ and quality of muscle reaction: $r = 0.71$, $p = 0.001$).

DISCUSSION

The SEP shows excellent test-retest reliability and valid levels of measurement error for measuring upper limb muscle strength, spasticity, and viscoelastic joint properties in patients with stroke and those with CP. Reliability was moderate for synergy measurements. good construct and content validity were found; however, the validity of the synergy measurements was less favourable in adults with CP. Findings support that the SEP can reliably and validly quantify elbow impairments in stroke and CP patients, but requires further optimization to quantify synergy.

Test-retest reliability was excellent with ICCs above 0.9 and small measurement errors for all impairments, except synergy. This indicates that day-to-day varia-

bility and measurement error do not hamper use of the SEP (26). In addition, SDC as a percentage of the measurement range was below 20%, indicating that relevant changes in clinical practice can be distinguished for all impairments except synergy (21, 22).

Reliability results are comparable to previously reported robotic measurement instruments assessing the same impairments. For example, muscle strength measurements in stroke patients with the Biodex showed ICCs between 0.79–0.99 for elbow extension and 0.64–0.97 for elbow flexion (27). For spasticity, other robotic measurement instruments reported ICCs between 0.66–0.95 in stroke patients (14, 28, 29). For viscoelastic joint properties, the current ICC was in line with reliability (ICC 0.64–0.91) of comparable measurements in joints other than the elbow in stroke and CP patients (13, 15). The advantage of the SEP is that a single device can measure all outcomes with excellent test-re-test reliability.

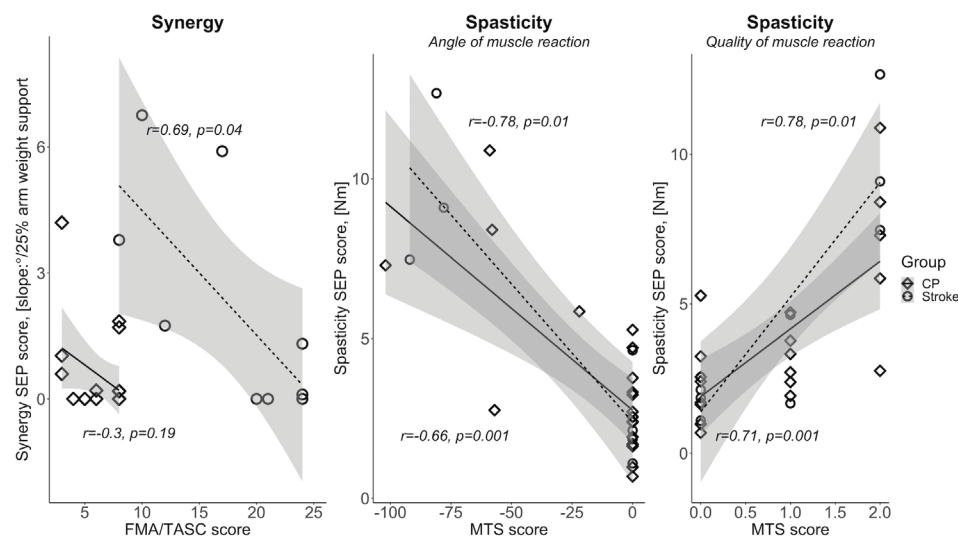


Fig. 4. Correlations (Spearman's rho) between Shoulder-Elbow Perturbator (SEP) outcomes and clinical test scores for synergy and spasticity in subgroups of stroke patients and patients with cerebral palsy (CP). FMA-UL: Fugl-Meyer Assessment of the upper limb; TASC: Test of Arm Selective Control; MTS: Modified Tardieu Scale.

Construct and content validity of SEP outcomes are good, with at least moderate correlations with clinical tests to evaluate spasticity and synergy. Only validity to quantify synergy in adults with CP is low, due to a lack of anticipated difference between adults with CP and healthy controls and a weak correlation with the clinical test. It is possible that test movements of the TASC do not measure the same construct as synergy measured with the SEP protocol. The SEP protocol focuses on the coupled movement between shoulder and elbow, typical for flexion synergy, while the TASC instructs participants to perform isolated movements, while scoring the occurrence of associated movements in other joints (8). The FMA-UL in stroke patients measures coupled movements and thus is a more similar construct, presumably explaining the correlation found in that patient group. Another explanation might be that the small variation in synergy between adults with CP is too small.

For synergy measurements, reliability and validity were less optimal than for other impairments, but sufficient for patients with stroke. For adults with CP, synergy measurements need further evaluation regarding within-group variety. Direct comparison with other studies on robotic quantification of synergy (12, 30, 31) is complicated, since they use other measures of reliability or measurement error (19). Ellis et al. (32) reported no systematic effect between repeated measurements in patients with stroke when evaluating the work area during reaching under 9 different arm support levels as a measure of synergy, without reporting ICCs or measurement error.

The occurrence of compensation strategies during synergy measurements may explain its lower reliability.

For example, in the current SEP design, participants may lower their arm at decreased arm weight support levels, reducing the amount of shoulder activity and allowing participants to further extend their arm. Adjusting the design to avoid or detect unwanted arm lowering is expected to decrease this source of measurement error. Secondly, elbow hyper-extension can occur, as can elbow extension increases due to external rotation of the shoulder when arm support decreases. Further standardization of these aspects may improve synergy measurements.

The strength of this study lies in the fact that not only were we able to assess 4 impairments evaluated through use of a single device, but we were also able to report on their test-retest reliability and validity. With this information, a reliable and valid patient profile of upper limb impairments can be created, which could have added clinical value compared with currently available tests. The study was further strengthened by the inclusion of a heterogeneous sample with multiple impairments, representing patients with varying levels of severity.

This study has some limitations. First, regarding standardization of measurements, intersession errors can occur because participants' elbow positioning in the SEP is determined using a goniometer, which has a measurement error of $\pm 8^\circ$ (33, 34). Also considered in the measurement protocol is normal daily variation because the study imposes no behavioural rules on patients (e.g. rest before measurement). A second limitation could be a lack of heterogeneity of synergy impairment in the subsample of adults with CP, despite adequate distributions of MACS levels and other impairments. Perhaps this pattern is typical for adults with CP and indicates

the long-term consequences of CP, which requires further investigation. Furthermore, due to the measurement set-up, patients with severe reduction in range of motion or severe muscle weakness in the shoulder or elbow can technically not be measured with the SEP. This is because these patients are not able to perform the measurements. Finally, these measurements are more time-consuming and more costly than current clinical measurement tools. However, they will provide more accurate measurement data, thus contributing to more individualized treatment plans.

CONCLUSION

Measurements with the SEP are reliable and valid for assessing muscle strength, spasticity, and viscoelastic joint properties around the elbow in patients with stroke and adults with CP. Reliability and validity of synergy measurements with the SEP are sufficient, but need further optimization. Overall, the study results support use of the SEP to quantify upper limb impairments in these patient populations. Next steps would be to evaluate the sensitivity and specificity of the SEP and determine its added value in clinical practice to improve personalized medicine and predict treatment effects.

ACKNOWLEDGEMENTS

The authors thank all the study participants for their contribution, and Wilma van der Slot (MD PhD), Marjolein van de Spek (OT), Marc Evers (PT), and Raphaela Imkamp (research assistant) of Rijndam Rehabilitation for recruiting adults with CP and post-stroke patients for this study.

Ethics approval and consent to participate

In this study, all methods were performed in accordance with the relevant guidelines and regulations. The study was approved by the Medical Ethics Committee of Erasmus University Medical Center Rotterdam, and all the included subjects gave written informed consent.

Competing interests

The authors have no conflicts of interest to declare.

Funding

This project was supported by Hankamp Rehab, Rijndam Rehabilitation and a PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health (grant number LSHM16030-H002).

Authors' contributions

LL included and measured all participants for this study. In addition, LL analysed and interpreted all data, and made a major contribution to writing the manuscript. JL conducted measurements and interpreted the data as well and contributed to writing the manuscript. BO helped to interpret the data. CJW developed and manufactured the SEP and contributed to designing the experiments. GM assisted in recruiting the patients and interpreted the data from a clinical viewpoint. RW assisted with the analysis process and interpreting the data and contributed to writing the manuscript. ME helped with data analysis process, interpretation of the data and contributed to writing the manuscript. The final version of the manuscript was read and approved by all authors.

REFERENCES

1. Patel AT, Duncan PW, Lai SM, Studenski S. The relation between impairments and functional outcomes poststroke. *Arch Phys Med Rehabil* 2000; 81: 1357–1363.
2. Simon-Martinez C, Jaspers E, Mailleux L, Desloovere K, Vanrenterghem J, Ortibus E, et al. Negative influence of motor impairments on upper limb movement patterns in children with unilateral cerebral palsy. A statistical parametric mapping study. *Front Hum Neurosci* 2017; 11.
3. Fleuren JF, Voerman GE, Erren-Wolters CV, Snoek GJ, Rietman JS, Hermens HJ, et al. Stop using the Ashworth Scale for the assessment of spasticity. *J Neurol Neurosurg Psychiatry* 2010; 81: 46–52.
4. Meseguer-Henarejos AB, Sánchez-Meca J, López-Pina JA, Carles-Hernández R. Inter- and intra-rater reliability of the Modified Ashworth Scale: a systematic review and meta-analysis. *Eur J Phys Rehabil Med* 2018; 54: 576–590.
5. Li F, Wu Y, Li X. Test-retest reliability and inter-rater reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in hemiplegic patients with stroke. *Eur J Phys Rehabil Med* 2014; 50: 9–15.
6. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scand J Rehabil Med* 1975; 7: 13–31.
7. Brunnstrom S. Motor Testing Procedures in Hemiplegia: Based on Sequential Recovery Stages. *Physical Therapy* 1966; 46: 357–375.
8. Sukal-Moulton T, Gaebler-Spira D, Krossschell KJ. The validity and reliability of the Test of Arm Selective Control for children with cerebral palsy: a prospective cross-sectional study. *Dev Med Child Neurol* 2018; 60: 374–381.
9. van der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients. *J Rehabil Med* 2001; 33: 110–113.
10. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. *Neurorehabil Neural Repair* 2002; 16: 232–240.
11. Trumbower RD, Ravichandran VJ, Krutky MA, Perreault EJ. Contributions of altered stretch reflex coordination to arm impairments following stroke. *J Neurophysiol* 2010; 104: 3612–3624.
12. Ellis MD, Lan Y, Yao J, Dewald JP. Robotic quantification of upper extremity loss of independent joint control or flexion synergy in individuals with hemiparetic stroke: a review of paradigms addressing the effects of shoulder abduction loading. *J Neuroeng Rehabil* 2016; 13: 95.

13. van der Krogt H, Klomp A, de Groot JH, de Vlugt E, van der Helm FC, Meskers CG, et al. Comprehensive neuro-mechanical assessment in stroke patients: reliability and responsiveness of a protocol to measure neural and non-neural wrist properties. *J Neuroeng Rehabil* 2015; 12: 28.
14. Starsky AJ, Sangani SG, McGuire JR, Logan B, Schmit BD. Reliability of biomechanical spasticity measurements at the elbow of people poststroke. *Arch Phys Med Rehabil* 2005; 86: 1648–1654.
15. Sloot LH, van der Krogt MM, de Gooijer-van de Groep KL, van Eesbeek S, de Groot J, Buizer AI, et al. The validity and reliability of modelled neural and tissue properties of the ankle muscles in children with cerebral palsy. *Gait Posture* 2015; 42: 7–15.
16. McPherson JG, Stienen AHA, Schmit BD, Dewald JPA. Bio-mechanical parameters of the elbow stretch reflex in chronic hemiparetic stroke. *Exp Brain Res* 2019; 237: 121–135.
17. Abstracts from the 2019 Annual Meeting. *Neurorehabil Neural Repair* 2019; 33: 1061–1108.
18. Morris S. Ashworth And Tardieu Scales: Their clinical relevance for measuring spasticity in adult and paediatric neurological populations. *Phys Ther Rev* 2002; 7: 53–62.
19. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine - A practical guide*. New York: U.S.A., Cambridge University Press, New York, 2011: 1–338.
20. Fleiss JL. *Reliability of Measurement. The Design and analysis of clinical experiments*. John Wiley & Sons; 1999, p. 1–32.
21. Buffart LM, Roebroek ME, Van Heijningen VG, Pesch-Batenburg JM, Stam HJ. Evaluation of arm and prosthetic functioning in children with a congenital transverse reduction deficiency of the upper limb. *J Rehabil Med* 2007; 39: 379–386.
22. Miller GA. The magical number seven, plus or minus two: some limitations on our capacity for processing information. *Psychol Rev* v63 1994: 81–97.
23. Domholdt E. *Rehabilitation research: principles and applications*. St Louis, MO.: Elsevier Saunders; 2005.
24. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998; 17: 101–110.
25. Mutlu A, Livanelioglu A, Gunel MK. Reliability of Ashworth and Modified Ashworth scales in children with spastic cerebral palsy. *BMC Musculoskelet Disord* 2008; 9: 44.
26. Sköld C. Spasticity in spinal cord injury: self- and clinically rated intrinsic fluctuations and intervention-induced changes. *Arch Phys Med Rehabil* 2000; 81: 144–149.
27. Kristensen OH, Stenager E, Dalgas U. Muscle strength and poststroke hemiplegia: a systematic review of muscle strength assessment and muscle strength impairment. *Arch Phys Med Rehabil* 2017; 98: 368–380.
28. Centen A, Lowrey CR, Scott SH, Yeh TT, Mochizuki G. KAPS (kinematic assessment of passive stretch): a tool to assess elbow flexor and extensor spasticity after stroke using a robotic exoskeleton. *J Neuroeng Rehabil* 2017; 14: 59.
29. Condliffe EG, Clark DJ, Patten C. Reliability of elbow stretch reflex assessment in chronic post-stroke hemiparesis. *Clin Neurophysiol* 2005; 116: 1870–1878.
30. Beer RF, Ellis MD, Holubar BG, Dewald JP. Impact of gravity loading on post-stroke reaching and its relationship to weakness. *Muscle Nerve* 2007; 36.
31. Sukal TM, Ellis MD, Dewald JP. Shoulder abduction-induced reductions in reaching work area following hemiparetic stroke: neuroscientific implications. *Exp Brain Res* 2007; 183.
32. Ellis MD, Acosta AM, Yao J, Dewald JP. Position-dependent torque coupling and associated muscle activation in the hemiparetic upper extremity. *Exp Brain Res* 2007; 176: 594–602.
33. Youdas JW, Carey JR, Garrett TR, Suman VJ. Reliability of goniometric measurements of active arm elevation in the scapular plane obtained in a clinical setting. *Arch Phys Med Rehabil* 1994; 75: 1137–1144.
34. Wolf SL, Segal RL, Catlin PA, Tschorn J, Raleigh T, Kontos H, et al. Determining consistency of elbow joint threshold angle in elbow flexor muscles with spastic hypertonia. *Phys Ther* 1996; 76: 586–600.