



## Upper limb motor and sensory impairments in children with hemiplegic cerebral palsy. Can they be measured reliably?

K. Klingels, P. De Cock, G. Molenaers, K. Desloovere, C. Huenaearts, E. Jaspers & H. Feys

To cite this article: K. Klingels, P. De Cock, G. Molenaers, K. Desloovere, C. Huenaearts, E. Jaspers & H. Feys (2010) Upper limb motor and sensory impairments in children with hemiplegic cerebral palsy. Can they be measured reliably?, *Disability and Rehabilitation*, 32:5, 409-416, DOI: [10.3109/09638280903171469](https://doi.org/10.3109/09638280903171469)

To link to this article: <http://dx.doi.org/10.3109/09638280903171469>



Published online: 25 Jan 2010.



Submit your article to this journal [↗](#)



Article views: 522



View related articles [↗](#)



Citing articles: 5 View citing articles [↗](#)

RESEARCH PAPER

## Upper limb motor and sensory impairments in children with hemiplegic cerebral palsy. Can they be measured reliably?

K. KLINGELS<sup>1</sup>, P. DE COCK<sup>2</sup>, G. MOLENAERS<sup>3</sup>, K. DESLOOVERE<sup>1,4</sup>, C. HUENAERTS<sup>4</sup>, E. JASPERS<sup>1</sup> & H. FEYS<sup>1</sup>

<sup>1</sup>Department of Rehabilitation Sciences, Faculty of Kinesiology and Rehabilitation Sciences, <sup>2</sup>Department of Paediatrics, Faculty of Medicine, <sup>3</sup>Department of Musculoskeletal Sciences, Faculty of Medicine, Katholieke Universiteit Leuven, Leuven, Belgium, and <sup>4</sup>Clinical Motion Analysis Laboratory, University Hospital Leuven, Pellenberg, Belgium

Accepted July 2009

### Abstract

**Purpose.** To establish interrater and test–retest reliability of a clinical assessment of motor and sensory upper limb impairments in children with hemiplegic cerebral palsy aged 5–15 years.

**Method.** The assessments included passive range of motion (PROM), Modified Ashworth Scale (MAS), manual muscle testing (MMT), grip strength, the House thumb and Zancolli classification and sensory function. Interrater reliability was investigated in 30 children, test–retest reliability in 23 children.

**Results.** For PROM, interrater reliability varied from moderate to moderately high (correlation coefficients 0.48–0.73) and test–retest reliability was very high (>0.81). For the MAS and MMT, total score and subscores for shoulder, elbow, and wrist showed a moderately high to very high interrater reliability (0.60–0.91) and coefficients of >0.78 for test–retest reliability. The reliability for the individual muscles varied from moderate to high. The Jamar dynamometer was found to be highly reliable. The House thumb classification showed a substantial reliability and the Zancolli classification an almost perfect reliability. All sensory modalities had a good agreement.

**Conclusions.** For all motor and sensory assessments, interrater and test–retest reliability was moderate to very high. Test–retest reliability was clearly higher than interrater reliability. To improve interrater reliability, it was recommended to strictly standardize the test procedure, refine the scoring criteria and provide intensive rater trainings.

**Keywords:** *Upper limb impairments, reliability, hemiplegic cerebral palsy*

### Introduction

In children with hemiplegia, upper limb function depends on several factors, including the severity of paresis, the degree of spasticity and the extent of sensory loss [1]. Various therapeutic, surgical and pharmacological treatment strategies are used to improve upper limb function. To guide and delineate treatment strategies and to evaluate the efficacy, reliable and valid measurements are indispensable. According to the International Classification of Functioning, Disability and Health model, upper limb function can be measured on the body function level and on the activity level [2]. On the activity level, several measurement scales have been

developed, such as the Melbourne Assessment of Unilateral Upper Limb Function [3], the Quality of Upper Extremity Skills Test [4] and the Assisting Hand Assessment [5], which have been tested for their psychometric properties. In contrast, the reliability and validity of the measurements on the body function level, including passive range of motion (PROM), muscle tone, strength and sensibility measurements, have scarcely been evaluated. Still, these assessments are systematically used in clinical practice as well as in upper limb intervention studies.

PROM measurements are routinely used to evaluate muscle shortening and joint contracture and to monitor the effects of tone reduction

treatment. Several studies have been published on the reliability of PROM measurements in the lower limbs in children with cerebral palsy (CP) [6–9]. They all reported moderate to high interrater and intrarater reliability. The reliability of goniometric measurements for the upper limb in hemiplegic children has only been reported in one abstract [10]. In that preliminary study, good test–retest reliability of PROM was found for elbow, wrist, and thumb extension and forearm supination in young children with upper limb spasticity. Shoulder abduction and flexion and wrist radial deviation indicated poor test–retest reliability.

To assess spasticity in children with CP, many different methods are used which vary from subjective, ordinal clinical scales to complex electrical or orthogenetic devices. The clinical scale routinely used in children with CP is the Modified Ashworth Scale (MAS). Several studies have been published on the reliability of the MAS in the lower limbs in children with CP [6,9,11,12]. The results indicated a variable interrater and intrarater reliability. Clopton et al. evaluated the interrater and intrarater reliability in the upper limb, but only for the elbow flexors [13]. They reported good interrater and moderate intrarater reliability.

Muscle weakness has been recognized as a primary clinical feature in CP [14]. There is a wide range of procedures available to evaluate muscle strength but there is limited research proving reliability and validity, especially in children [15]. To assess muscle strength in children with CP, manual muscle testing (MMT) and grip strength measurements are most routinely used. The reliability of MMT has been reported in children with Duchenne's muscular dystrophy [16], but not yet in children with CP. Also, the reliability of grip strength measurement in children with CP has not been reported before.

The Zancolli classification provides insight on the grip and release patterns [17]. Analysis of thumb deformities is facilitated by the House thumb classification that portrays the static and the dynamic components of the deformities [18]. Both classifications are generally accepted and commonly used in children with CP. However, to the best of our knowledge, the reliability of the Zancolli classification has not been evaluated. The reliability of the House thumb classification has only been investigated in 10 children. Interrater reliability was only fair and intrarater reliability was substantial [19].

Sensory dysfunction has been reported to occur in more than 50% of the children with hemiplegia [20], and the impact of a disturbed sensibility on hand function has been emphasized [21]. In children with hemiplegia, stereognosis and two-point discrimination (TPD) are the sensory modalities most

commonly impaired [22]. In literature, the reliability of a sensory assessment has only been investigated in normal children [23, 24]. TPD, touch-pressure and kinaesthesia measurements have shown to be sufficiently reliable. Although several studies have assessed sensory function to document deficits in children with CP, the reliability of a sensory assessment for the upper limb has not been evaluated in this patient group.

In conclusion, there is a clear lack of reliability studies on the clinical assessment of motor and sensory impairments in children with CP. Nevertheless, reliability is considered as the basic psychometric criterion for assessment tools. Therefore, the aim of this study was to evaluate the interrater and test–retest reliability of a clinical test battery to assess motor and sensory upper limb impairments in hemiplegic children.

## Method

### *Participants*

Participants were included in the study if they were diagnosed with hemiplegic CP and aged between 5 and 15 years. Children who were not mentally capable of cooperating with and comprehending the assessments were excluded. The study population consisted of 30 children. Eighteen children were recruited from the Clinical Motion Analysis Laboratory of the University Hospital of Leuven, 12 children from special education schools. Ethical approval was obtained from the Ethics Committee of the University Hospital of Leuven and informed consent was obtained from the parents.

### *Procedure*

To evaluate interrater reliability, two physiotherapists ( $O_1$  and  $O_2$ ) routinely involved in the clinical evaluation of children with CP, performed the complete test procedure independently. The physiotherapist who performed the examination first, was randomly chosen. There was a time interval of ~1 h between both examinations during which no treatment was offered. The first assessment was performed at the place of recruitment. To determine test–retest reliability,  $O_1$  performed the clinical examination a second time after a 2-week interval. Twenty-three children were re-examined, of which 12 children were tested at the same place of their first assessment and 11 children at home. A detailed instruction manual was developed and the assessors practised jointly to standardize the test procedure.

### Assessment

The clinical evaluation included first an assessment of motor impairments involving, in order of testing, measurements of PROM, muscle tone, MMT, grip strength and two classifications. These assessments were evaluated in supine position, except for grip strength and the classifications which were tested in sitting position. Secondly, a sensory assessment was performed in sitting position, involving, in order of testing exteroception, proprioception, TPD, and stereognosis.

### Motor assessment

PROM of shoulder flexion and abduction, elbow extension, supination and wrist extension with finger extension was measured using a universal goniometer.

Muscle tone was evaluated in eight muscle groups by using the 6-point ordinal MAS, ranging from 0 to 4 [25]. The score 1+ was converted to 1.5. A total score for the hemiplegic side was calculated (range 0–32) as well as subscores for the muscle groups of the shoulder (adductors, internal/external rotators; range 0–12), the elbow (flexors/extensors; range 0–8) and the wrist and hand (pronators, wrist and finger flexors; range 0–12).

To assess muscle strength, MMT was used according to the 8-point ordinal scale of Daniels and Worthingham, ranging from 0 to 5 [26]. MMT was performed for shoulder flexion and abduction/adduction, elbow flexion/extension, wrist supination/pronation and flexion/extension. The score 2– was converted to 1.5 and 3+ to 3.5. A total score was calculated (range 0–45) as well as subscores for the muscle groups of the shoulder (range 0–15), elbow (range 0–10) and wrist (range 0–20).

A calibrated Jamar dynamometer was used to assess grip strength. The position of the dynamometer was adapted to the child's hand size. The child was sitting with the arm adducted, the elbow flexed at 90°, and the forearm and wrist in neutral position (if possible). The mean of three maximum voluntary contractions was recorded for each hand. Also, the ratio of the grip strength of the hemiplegic hand to the grip strength of the non-hemiplegic hand was calculated to eliminate the correlation between age and grip strength [27].

To assess selective extension in wrist and fingers, the Zancolli classification was used [17]. Three patterns were distinguished, determined by the severity of spasticity and paresis. Pattern I corresponds to a complete extension of the fingers with the wrist in neutral position or less than 20° of flexion. Pattern II is a complete extension of the fingers with

more than 20° of wrist flexion. Pattern II is further subdivided into IIa where active extension of the wrist is possible with fingers flexed and IIb where no active extension of the wrist is possible even with fingers flexed. Pattern III refers to the inability to actively extend the fingers even with maximal wrist flexion.

Analysis of thumb deformities was facilitated by the use of the House thumb classification which distinguishes four types of contractures [18]. Type I refers to a simple metacarpal adduction contracture, type II to a metacarpal adduction contracture with metacarpophalangeal flexion deformity, type III to a metacarpal adduction contracture with a metacarpophalangeal hyperextension deformity or instability and type IV to a metacarpal adduction contracture with metacarpophalangeal and interphalangeal flexion deformities.

### Sensory assessment

Exteroception was evaluated by assessing tactile sense of the index finger. The finger tip pulp was lightly touched three times and the child was asked to say 'yes' when the touch was perceived. Exteroception was scored as absent (not felt in any attempt), impaired (not felt in one or more attempt(s)) or intact (all three attempts correct).

Proprioception was evaluated by assessing the passive movement sense at the index finger. The child was asked to say 'yes' when the movement was perceived. If not felt within a small amplitude, the examiner moved within larger amplitudes. Proprioception was scored as absent (not felt in any attempt), impaired (only felt within a larger amplitude), intact (correctly felt within a small amplitude in all three attempts).

To assess TPD, either one or two points of a Discriminator were placed longitudinally and perpendicularly on the finger tip pulp of the index. The minimal distance at which a subject could distinguish one or two discrete points in five consecutive correct trials was recorded. The examiner started with five trials at a distance of 4 mm and progressively decreased (until 2 mm) or increased (until 10 mm) the distance.

Stereognosis was evaluated by tactile identification of familiar objects. Of the 12 objects, six were matched in pairs similar in size and shape (pencil/pen, coin/button, paperclip/safety pin) and six objects clearly differed from each other (key, clothespin, marble, comb, spoon, ball). O<sub>1</sub> and O<sub>2</sub> randomly selected six objects of which three from the similar pairs and three from the non-similar objects. The number of objects correctly identified with the hemiplegic hand was recorded. The total score ranged from 0 to 6.

For exteroception, proprioception and TPD, the unaffected hand was tested first to verify the ability to comprehend the examination. For stereognosis, the affected hand was tested first and the unaffected hand was only tested if a child could not identify the object with the affected hand.

### Statistical analysis

Interrater and test-retest reliability was calculated by the percentage of agreement, weighted kappa ( $\kappa_w$ ) and intraclass correlation coefficients (ICC), dependent on the type of data. The 95% confidence limits for  $\kappa$ ,  $\kappa_w$ , and ICC were calculated. According to the guidelines of Landis and Koch [28],  $\kappa$  values of more than 0.80 were considered as an excellent reliability, 0.61–0.80 as a substantial, 0.40–0.60 as a moderate and less than 0.40 as a poor to fair reliability. For the ICC, values of more than 0.80 were considered as a very high reliability, 0.60–0.79 as a moderately high, 0.40–0.59 as a moderate, and below 0.40 as a low reliability [29].

## Results

### Subject characteristics

Thirty children (13 girls and 17 boys) with hemiplegia were included. The mean age was 10 years 6 months (SD 2 years 7 months, range 5–15 years). Eighteen children had a right hemiplegia and 12 children a left hemiplegia. Seventeen children attended special education schools, 13 attended mainstream schools. The children were sampled across the first two levels of the Gross Motor Function Classification System (GMFCS) (level I in 22 children, level II in eight children) [30] and across all four patterns of the Zancolli Classification (pattern I in 23 children, pattern IIa in three children, pattern IIb in one child, pattern III in three children).

### Motor assessment

*Passive range of motion.* PROM of shoulder flexion and abduction was limited in only three and four children, respectively and therefore these movements were not included for statistical analysis. Interrater and test-retest reliability was evaluated for elbow extension, supination and wrist extension with finger extension (Table I). Coefficients for interrater reliability were moderately high for elbow extension and wrist supination with ICCs of 0.69 and 0.73, respectively, but only moderate for wrist extension (ICC = 0.48). For test-retest reliability, all

correlations were very high with ICCs ranging from 0.81–0.94.

*Muscle tone.* Interrater and test-retest reliability for the MAS is presented in Table II. Interrater reliability for the individual muscles was very high for elbow extensors (ICC = 0.83) and pronators (ICC = 0.82) and moderately high for shoulder adductors (ICC = 0.64), external rotators (ICC = 0.77), elbow flexors (ICC = 0.72) and wrist flexors (ICC = 0.65). ICC values were only moderate for internal rotators (ICC = 0.52) and finger flexors (ICC = 0.58). For the composite scores of the muscle groups, ICCs ranged from 0.64 to 0.83 and for the total score, the ICC was 0.88 (CI 0.76–0.94.)

Test-retest reliability for the individual muscles was very high for elbow flexors (ICC = 0.85) and wrist flexors (ICC = 0.80). For all other individual muscles ICCs were moderately high (ICC = 0.70–0.79) except for the internal rotators (ICC = 0.57).

Table I. Intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) between raters and test-retest for PROM.

	Interrater (N = 30)		Test-retest (N = 23)	
	ICC	95% CI	ICC	95% CI
Elbow				
Extension	0.69	0.44–0.84	0.94	0.86–0.97
Wrist				
Supination	0.73	0.51–0.86	0.81	0.61–0.91
Extension	0.48	0.15–0.71	0.88	0.74–0.95

Table II. Intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) between raters and test-retest for the MAS.

	Interrater (N = 30)		Test-retest (N = 23)	
	ICC	CI	ICC	CI
Shoulder				
Adductors	0.64	0.37–0.81	0.70	0.42–0.86
Internal rotators	0.52	0.21–0.74	0.57	0.22–0.79
External rotators	0.77	0.58–0.88	0.76	0.52–0.89
Elbow				
Flexors	0.72	0.50–0.86	0.85	0.69–0.93
Extensors	0.83	0.67–0.91	0.79	0.56–0.90
Wrist and fingers				
Pronators	0.82	0.66–0.91	0.75	0.51–0.89
Wrist flexors	0.65	0.39–0.82	0.80	0.60–0.91
Finger flexors	0.58	0.29–0.78	0.75	0.48–0.88
Composite scores				
Shoulder	0.64	0.38–0.81	0.78	0.55–0.90
Elbow	0.83	0.67–0.91	0.85	0.67–0.93
Wrist	0.81	0.64–0.90	0.85	0.68–0.93
Total	0.88	0.76–0.94	0.90	0.78–0.96

For the composite scores of the muscle groups, test-retest reliability was high (ICC = 0.78–0.85) and for the total score the ICC was 0.90 (CI 0.78–0.96).

**Muscle strength.** Table III presents interrater and test-retest reliability for MMT. Interrater reliability for muscle strength in the individual muscles ranged from 0.53 to 0.93 with higher ICC values in the more distal muscle groups. Composite scores of the shoulder and the elbow had a moderately high reliability with ICCs of 0.60 and 0.62, respectively. The muscle group of the wrist had a very high reliability with an ICC of 0.91. The ICC for the total score was 0.90 (CI 0.80–0.95).

Test-retest reliability for muscle strength was very high for all individual muscles (ICC = 0.80–0.98) except for a moderately high agreement for shoulder abductors (ICC = 0.69). Composite scores of the muscle groups of the shoulder, elbow and wrist showed ICCs above 0.88. The ICC for the total score was 0.96 (CI 0.91–0.98).

Table III. Intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) between raters and test-retest for manual muscle strength.

	Interrater (N = 30)		Test-retest (N = 23)	
	ICC	CI	ICC	CI
Shoulder				
Flexors	0.54	0.23–0.75	0.81	0.60–0.91
Abductors	0.53	0.22–0.74	0.69	0.40–0.85
Adductors	0.66	0.40–0.82	0.94	0.86–0.97
Elbow				
Extensors	0.68	0.44–0.84	0.88	0.75–0.95
Flexors	0.49	0.16–0.72	0.80	0.60–0.91
Wrist				
Supinators	0.93	0.85–0.96	0.98	0.95–0.99
Pronators	0.70	0.47–0.85	0.92	0.82–0.96
Flexors	0.79	0.61–0.90	0.85	0.69–0.93
Extensors	0.77	0.58–0.88	0.81	0.62–0.92
Composite scores				
Shoulder	0.60	0.32–0.79	0.93	0.84–0.97
Elbow	0.62	0.34–0.80	0.88	0.74–0.95
Wrist	0.91	0.82–0.96	0.96	0.91–0.98
Total	0.90	0.80–0.95	0.96	0.91–0.98

For grip strength, ICCs for the mean scores on the hemiplegic side were calculated. For interrater reliability, the ICC was 0.95 (CI 0.89–0.97). Comparing test and retest showed an ICC of 0.96 (CI 0.90–0.98). For the ratio of mean grip force on the hemiplegic side to the non-hemiplegic side, ICC for interrater reliability was 0.88 (CI 0.76–0.94) and for test-retest reliability 0.91 (CI 0.80–0.96).

**Classifications.** For the interrater reliability of the House thumb classification, the percentage of agreement was 83% and weighted  $\kappa$  value was 0.73 (CI 0.51–0.95). Test-retest reliability was also substantial (83% agreement,  $\kappa_w = 0.74$ , CI 0.50–0.98).

For the Zancolli classification the weighted  $\kappa$  value for the interrater reliability was 0.95 with an almost perfect agreement (93%). For the test-retest reliability, there was perfect agreement (100% agreement,  $\kappa_w = 1.00$ ).

#### Sensory assessment

As shown in Table IV, all modalities of the sensory assessment had a good agreement. Exteroception showed an almost perfect agreement for both interrater (96.43% agreement,  $\kappa_w = 0.88$ ) and test-retest reliability (95.24% agreement,  $\kappa_w = 0.89$ ). Proprioception showed a substantial agreement for interrater reliability (83% agreement,  $\kappa_w = 0.69$ ) and an excellent agreement for test-retest reliability (91.30% agreement,  $\kappa_w = 0.83$ ). TPD had a very high interrater and test-retest reliability with ICCs of 0.92 and 0.81. Stereognosis had an ICC of 0.78 for interrater reliability and 0.86 for test-retest reliability.

#### Discussion

Test-retest reliability of PROM was very high for elbow extension, supination and wrist extension which is in agreement with the results of Glazier et al. [10]. They reported good test-retest reliability for these goniometric measurements in young children with upper limb spasticity. In contrast,

Table IV. Percentage of agreement (%), weighted kappas ( $\kappa_w$ ), intraclass correlation coefficients (ICC) and 95% confidence intervals (CI) between raters and test-retest for exteroception, proprioception, two-point discrimination (TPD) and stereognosis.

	Interrater (N = 28)				Test-retest (N = 21)			
	%	$\kappa_w$	ICC	CI	%	$\kappa_w$	ICC	CI
Exteroception	96	0.88		0.63–1.00	95	0.89		0.68–1.00
Proprioception	83	0.69		0.42–0.96	91	0.83		0.60–1.00
TPD			0.92	0.83–0.96			0.81	0.60–0.92
Stereognosis			0.78	0.59–0.89			0.86	0.71–0.94

interrater reliability for PROM at the elbow and wrist was only moderate to moderately high. This is in agreement with the variable interrater reliability results reported for the lower limb in children with CP [6–9]. In literature, test–retest has consistently been reported to be more reliable than interrater reliability [8]. This is confirmed in our results. The lower ICC value for interrater reliability of wrist extension, could also be partly explained by a low variability in the total range of measurements as only seven children presented with tight wrist flexors and a limited PROM. This small variability between subjects may consequently have contributed to a poorer ICC. Based on our results, PROM measurements in elbow and wrist can be suggested for status and follow-up assessments in children with hemiplegia, though ideally these should be assessed by the same rater. Further study is warranted to re-evaluate the reliability in a study group with more movement restrictions in the upper limb by including children with a more severe type of CP (e.g. quadriplegia).

Although the MAS has traditionally been used to quantify spastic muscle tone, its reliability has been debated [13]. In our study, interrater reliability of the MAS scores for the individual muscles of the upper limb was moderate to very high and test–retest reliability was moderately high to very high. Only one study assessed the reliability of the MAS in the elbow flexors in children with CP [13]. They found good interrater reliability and moderate intrarater reliability. In the lower limbs in children with CP, reliability results of the MAS scores varied from poor to good [6,9,11,12]. Our findings and those of Clopton et al. [13] suggest a higher reliability of the MAS in the upper limb compared to the lower limb in children with CP. This is in agreement with the results of most studies in stroke patients [31]. The composite scores of the shoulder, elbow and wrist joint, as well as the total score showed a moderately high to very high interrater and test–retest reliability. This is comparable to the study of Platz et al. in stroke patients [32]. In this study, the Resistance to Passive Movement (REPAS) has been presented as a rating scale for resistance to passive movement, based on the Ashworth Scale. The REPAS is a summary rating scale with a total score and regional body subtest scores. Although composite scores are less informative for treatment delineation, their reliability is better than the reliability for the individual muscle groups. The sum scores for the different joints can be useful for both cross-sectional assessment and follow-up purposes, e.g. treatment evaluation.

For almost all individual muscles and all composite scores, test–retest reliability was higher than interrater reliability. Thus, measurement variability might be more attributable to the variability between

raters than to the variability within the child. We included general instructions on how to perform the test movements as well as specific instructions for individual joint motions, such as hand placement, alignment and movement range. However, to improve interrater reliability, we hypothesize that the original scoring criteria have to be amended by explanatory details for each ordinal response category for the different muscle groups in children with CP, similar to the study of Platz et al. [32].

Test–retest reliability for MMT was high for all individual muscles as well as for the composite scores and the total score. Interrater reliability, on the contrary, was significantly lower but reliability in the distal muscle groups was better than in the proximal muscle groups. Visual inspection of the data showed that for the shoulder and elbow muscle groups, all children obtained relatively good strength scores (between 3+ and 5) whereas for the wrist muscle strength, the children scored in at least six of the eight possible classes. Thus, the greater scoring variability of the distal muscle groups might explain the higher ICC calculation. Nevertheless, test–retest reliability clearly exceeded interrater reliability. We would therefore recommend the examiners to follow a standardized protocol that specifies patient position, the precise alignment of the muscle being tested, the movement range, the verbal instruction and demonstration to the patient. Although we provided general scoring instructions, more detailed scoring descriptions should also be appended for each muscle group of the upper limb. We further believe that interrater sessions should regularly be organized to practise jointly and to discuss the test performance and scoring.

For grip strength measured with the Jamar dynamometer, results showed a very high interrater and test–retest reliability for the hemiplegic hand and for the ratio between the hemiplegic and the non-hemiplegic hand. Reliability of grip strength measurements with the Jamar dynamometer have been established in normal children [33]. Our study confirms the high reliability of this measurement method in children with hemiplegia.

The House thumb classification revealed a substantial interrater and test–retest reliability. In a previous study of Waters et al., interrater reliability was only fair, whereas videotaped intrarater reliability was substantial [19]. Agreement between observers for the Zancolli classification was high and test–retest agreement was perfect. Our results support the reliability of these commonly used classifications.

In this study, all sensory modalities had a good agreement. In literature, the reliability of different modalities of sensory assessment has been investigated in normal children [23, 24]. In the study of Thibault et al. [24], touch–pressure was evaluated

with the Semmes Weinstein monofilaments and for the kinaesthesia measurement the child had to define the direction in which the limb was moved passively. Although the test procedure in this study was different, results were comparable. No other studies have investigated the reliability of a sensory assessment in children with CP.

Although the clinical assessments applied in this study are widely used in clinical practice in children with CP, the reliability of most of these assessments has not been investigated in this study population. Our results show that the reliability of motor and sensory assessments is acceptable for interrater measurements and even better for test–retest measurements. Still, some critical reflections are warranted. The study population was limited to children with hemiplegia recruited from mainstream schools and special education schools to obtain a representative sample. However, generalizations to other types of CP should be made with caution because of the limited scoring range for some assessments of PROM and proximal muscle strength. Further study is necessary to assess the reliability in a test group including children with a more severe type of CP. Secondly, the sample size was only modest, particularly for test–retest reliability. Still, this sample size is comparable to other studies that investigated reliability of impairment assessments in children with CP [7–13,19]. Thirdly, the selection of the children who were included for the test–retest reliability was arbitrary and unrelated to the severity of hemiplegia or age. Comparison of the test and retest group showed no significant differences with respect to age, sex, hemiplegic side or GMFCS.

## Conclusion

This was the first study that assessed interrater and test–retest reliability of a clinical assessment of motor and sensory upper limb impairments in children with hemiplegia. For all motor and sensory assessments, interrater and test–retest reliability was moderate to very high. An important finding was that test–retest reliability was clearly higher than interrater reliability. Thus, assessments should preferably be performed by single raters rather than different raters for follow-up assessments and treatment evaluation. To improve interrater reliability it was recommended to strictly standardize the test procedure, to refine the scoring criteria, and to provide an intensive rater training.

From this study, we conclude that upper limb motor and sensory impairments can be measured reliably in children with hemiplegia, but great efforts have to be made for standardization and rater trainings.

**Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## References

1. Fedrizzi E, Pagliano E, Andreucci E, Oleari G. Hand function in children with hemiplegic cerebral palsy: prospective follow-up and functional outcome in adolescence. *Dev Med Child Neurol* 2003;45:85–91.
2. International Classification of Functioning Disability and Health. Internet. 2001. World Health Organisation. Geneva, Switzerland. Available from: <http://www3.who.int/icf/icftemplate.cfm>. Last accessed July 2009.
3. Randall M, Johnson L, Reddihough D. The Melbourne Assessment of Unilateral Upper Limb Function: test administration manual. Melbourne: Royal Children's Hospital; 1999.
4. DeMatteo C, Law M, Russell D, Pollock N, Rosenbaum P, Walters S. Quality of upper extremity skill test. . Ontario: Neurodevelopmental Clinical Research Unit, 1992.
5. Krumlinde-Sundholm L, Eliasson AC. 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.
6. Platz T, Eickhof C, Nuyens G, Vuadens P. Clinical scales for the assessment of spasticity, associated phenomena, and function: a systematic review of the literature. *Disabil Rehabil* 2005;27:7–18.
7. Mutlu A, Livanelioglu A, Gunel MK. Reliability of goniometric measurements in children with spastic cerebral palsy. *Med Sci Monit* 2007;13:323–329.
8. McWhirk LB, Glanzman AM. Within-session interrater reliability of goniometric measures in patients with spastic cerebral palsy. *Pediatr Phys Ther* 2006;18:262–265.
9. Fosang AL, Galea MP, McCoy AT, Reddihough DS, Story I. Measures of muscle and joint performance in the lower limb of children with cerebral palsy. *Dev Med Child Neurol* 2003;45:664–670.
10. Glazier JN, Fehlings DL, Steele C. Test–retest reliability of upper extremity goniometric measurements of passive range of motion and sphygmomanometer measurements of grip strength in children with cerebral palsy and upper extremity spasticity. *Dev Med Child Neurol* 1995;75:33–34 (Scientific Poster).
11. 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.
12. Yam WK, Leung MS. Interrater reliability of Modified Ashworth Scale and Modified Tardieu Scale in children with spastic cerebral palsy. *J Child Neurol* 2006;21:1031–1035.
13. Clopton N, Dutton J, Featherston T, Grigsby A, Mobley J, Melvin J. Interrater and intrarater reliability of the Modified Ashworth Scale in children with hypertonia. *Pediatr Phys Ther* 2005;17:268–274.
14. Damiano DL, Quinlivan J, Owen BF, Shaffrey M, Abel MF. Spasticity versus strength in cerebral palsy: relationships among involuntary resistance, voluntary torque, and motor function. *Eur J Neurol* 2001;8:40–49.
15. Jones, Stratton G. Muscle function assessment in children. *Acta Paediatr* 2000;89:753–761.
16. Escolar DM, Henricson EK, Mayhew J, Florence J, Leshner R, Patel KM, Clemens PR. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. *Muscle Nerve* 2001;24:787–793.

17. Zancolli EA, Goldner LJ, Swanson AB. Surgery of the spastic hand in cerebral palsy: report of the committee on spastic hand evaluation (International Federation of Societies for Surgery of the Hand). *J Hand Surg [Am]* 1983;8:766–772.
18. House JH, Gwathmey FW, Fidler MO. A dynamic approach to the thumb-in palm deformity in cerebral palsy. *J Bone Joint Surg [Am]* 1981;63:216–225.
19. Waters PM, Zurakowski D, Patterson P, Bae DS, Nimec D. Interobserver and intraobserver reliability of therapist-assisted videotaped evaluations of upper-limb hemiplegia. *J Hand Surg* 2001;29:328–334.
20. Van Heest AE, House J, Putnam M. Sensibility deficiencies in the hands of children with spastic hemiplegia. *J Hand Surg [Am]* 1993;18:278–281.
21. Cooper J, Majnemer A, Rosenblatt B, Birnbaum R. The determination of sensory deficits in children with hemiplegic cerebral palsy. *J Child Neurol* 1995;10:300–309.
22. Krumlinde-Sundholm L, Eliasson A. Comparing tests of tactile sensibility: aspects relevant to testing children with spastic hemiplegia. *Dev Med Child Neurol* 2002;44:604–612.
23. Menier C, Forget R, Lambert J. Evaluation of two-point discrimination in children: reliability, effects of passive displacement and voluntary movements. *Dev Med Child Neurol* 1996;38:523–537.
24. Thibault A, Forget R, Lambert J. Evaluation of cutaneous and proprioceptive sensation in children: a reliability study. *Dev Med Child Neurol* 1994;9:796–812.
25. Bohannon RW, Smith MB. Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987;67:206–207.
26. Hislop HJ, Montgomery J. Daniels and Worthingham's Muscle Testing: techniques of manual examination. Philadelphia: Harcourt Brace & Company 1995.
27. De Smet L, Vercammen A. Grip strength in children. *J Pediatr Orthop B* 2001;10:352–354.
28. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;86: 420–428.
29. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparing measurement sensitivity of short and longer health status instruments. *Med Care* 1992;30:917–925.
30. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997;39:214–223.
31. Pandyan AD, Johnson GR. A review of the properties and limitations of the Ashworth and modified Ashworth Scales as measures of spasticity. *Clin Rehabil* 1999;13:373–383.
32. Platz T, Vuadens P, Eickhof C, Arnold P, Van Kaick S, Heise K. REPAS, a summary rating scale for resistance to passive movement: item selection, reliability and validity. *Disabil Rehabil* 2008;30:44–53.
33. Molenaar HM, Zuidam JM, Selles RW, Stam HJ, Hovius SE. Age-specific reliability of two grip-strength dynamometers when used by children. *J Bone Joint Surg Am* 2008;90:1053–1059.