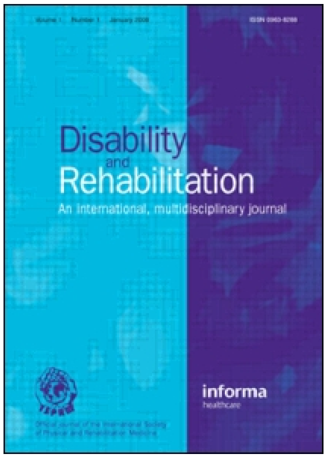

Disability & Rehabilitation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713723807

Reliability of hand-held dynamometry and functional strength tests for the lower extremity in children with Cerebral Palsy

Olaf Verschuren [abc]; Marjolijn Ketelaar [acd]; Tim Takken [b]; Marco Van Brussel [b]; Paul J. M. Helders [b]; Jan Willem Gorter [acd]

[a] Centre of Excellence, Rehabilitation Centre 'De Hoogstraat', Utrecht
[b] Department of Pediatric Physical Therapy & Exercise Physiology, University Hospital for Children and Youth 'Het Wilhelmina Kinderziekenhuis', University Medical Center Utrecht, Utrecht
[c] Partner of NetChild, Network for Childhood Disability Research, The Netherlands
[d] Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht

First Published on: 06 November 2007
To cite this Article: Verschuren, Olaf, Ketelaar, Marjolijn, Takken, Tim, Van Brussel, Marco, Helders, Paul J. M. and Gorter, Jan Willem (2007) 'Reliability of hand-held dynamometry and functional strength tests for the lower extremity in children with Cerebral Palsy', Disability & Rehabilitation, 1 - 9
To link to this article: DOI: 10.1080/09638280701639873
URL: http://dx.doi.org/10.1080/09638280701639873

Disability and Rehabilitation, 2007; 1–9, i*First* article

informa
healthcare

RESEARCH PAPER

Reliability of hand-held dynamometry and functional strength tests for the lower extremity in children with Cerebral Palsy

OLAF VERSCHUREN[1,2,4], MARJOLIJN KETELAAR[1,3,4], TIM TAKKEN[2], MARCO VAN BRUSSEL[2], PAUL J. M. HELDERS[2] & JAN WILLEM GORTER[1,3,4]

[1]*Centre of Excellence, Rehabilitation Centre 'De Hoogstraat', Utrecht,* [2]*Department of Pediatric Physical Therapy & Exercise Physiology, University Hospital for Children and Youth 'Het Wilhelmina Kinderziekenhuis', University Medical Center Utrecht, Utrecht,* [3]*Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, and* [4]*Partner of NetChild, Network for Childhood Disability Research, The Netherlands*

Accepted August 2007

Abstract

Purpose. To evaluate the intertester reliability of two methods for measuring lower-limb strength in children with cerebral palsy (CP).

Method. Twenty-five subjects with CP (7–17 years of age) participated in this study. Lower-limb muscle strength was measured on 2 occasions using a Hand-held Dynamometer (HHD; break-method and make-method) and a 30-sec Repetition Maximum (RM) during three functional strength tests for the lower extremities. Reliability was measured using the intraclass correlation coefficients (ICCs), the standard error of measurement (SEM) and the coefficient of variation (CV).

Results. The intertester reliability of strength measurement using a HHD was questionable with ICC values ranging from 0.42–0.73 for the break-method, and from 0.49–0.82 for the make-method. The SEM and CV (%) values ranged from 27.9–58.9 and 22.2–35.3% for the break-method, and from 30.6–52.7 and 16.2–56.2% for make-method. The intertester reliability of strength measurement using the 30-sec RM was acceptable with ICC values ranging from 0.91–0.96, and SEM and CV (%) values ranging from 1.1–2.6 and 10.9–39.9% for the functional exercises.

Conclusion. The intertester reliability of measuring muscle strength of the lower extremities using a hand-held dynamometer is questionable. The intertester reliability of the 30-sec RM for the lower extremity is acceptable.

Keywords: *Muscles, exercise test, reliability, cerebral palsy*

Introduction

The assessment of muscle strength in children with cerebral palsy (CP) has become standard in clinical practice and research. Many children and adolescents with CP have a different distribution of muscle strength, and this can be associated with difficulties performing everyday functional activities [1]. Recent studies show that children with CP can benefit from strength training programs [2–6]. Therefore, it is important that clinicians and researchers who are interested in the efficacy of strength training programs have knowledge of the psychometric properties of their measurements of strength.

In general, there are two different muscle strength measurements that are used in children and adolescents with CP: Isometric and isokinetic. An isometric-based test measures the ability of a muscle group to produce force without a change in overall muscle-tendon length. Isokinetic literally means 'same velocity' and refers to tests that are performed at a predetermined constant velocity.

In clinical practice, muscle strength is usually assessed by isometric resistance-based methods. Two methods to assess muscle strength are manual muscle testing (MMT) and the hand-held dynamometry (HHD). MMT uses an ordinal 6-point (0–5) Medical Research Council (MRC) scale. The HHD

Correspondence: Olaf Verschuren, BSc, PT, Rehabilitation Centre 'De Hoogstraat', Rembrandtkade 10, 3583 TM Utrecht, The Netherlands. Tel: +31 30 2561366. Fax: +31 30 2511344. E-mail: o.verschuren@dehoogstraat.nl

ISSN 0963-8288 print/ISSN 1464-5165 online
DOI: 10.1080/09638280701639873

consists of a simple hand-held device equipped with a small internal load cell capable of measuring muscular force (in Newton).

Isokinetic dynamometry, most often used in a laboratory setting, uses computer-controlled equipment to measure the muscular force generated throughout a controlled movement. Isokinetic assessment has several advantages over isometric testing. Most activities of daily living involve phases of dynamic muscle action, and in this sense isokinetic testing may provide more specificity in terms of muscle action type than isometric testing [7]. A maximal isometric contraction is only indicative of the capacity to produce force in that condition and at that particular muscle length, and can not necessarily be extrapolated to conditions where the muscle length is different or changing throughout the task [8]. Perhaps most important, however, is the inherent safety of isokinetic actions, afforded by the computer-controlled mechanism of accommodating resistance.

An alternative approach to measure muscle strength in a dynamical way is the use of the Repetition Maximum (RM) during functional exercises [9]. Since many lower extremity training and rehabilitation programs have been centred on functional exercises [10,11], the RM during functional strength tests, based upon the specificity of training principle, may be better suited to detect actual improvements in lower extremity performance. Two ways in which this kind of functional strength tests may be used to assess lower extremity performance are by counting the number of repetitions performed at a specific activity over a specific period of time and by assessing the time necessary to complete a specified number of repetitions at a specific activity [12,13]. Because of the great diversity of functional abilities in young people with CP, counting the numbers of repetitions performed over a specific period of time seems the most appropriate. Using this way of testing all children can complete the test.

With respect to reliability of muscle strength measures it has been found that the reliability of MMT is fair when assessed for individual muscles [14]. Considerable training is required to achieve this modest reliability [14]. No data are available on the reliability of this method in children with CP. Moreover, this method is often not sufficiently sensitive to assess muscle strength of MRC grade 4 (active movement against moderate resistance) and grade 5 (normal strength; active movement against maximum resistance) or to detect small or moderate increases of strength over the course of rehabilitation [15].

For the HHD, the topic of reliability is more complex. Many researchers use the HHD to measure muscle strength. However, in children and adolescents with CP there is no consensus about the way of testing that should be used. In published studies the break-method [16,17], and the make-method [10,18] were used to measure muscle strength. In three studies that assessed intratester reliability in a group of children with CP, the used method (break or make) was different, and intraclass correlations were >0.79 for most muscle groups [16,18,19]. Damiano et al. [8] suggested that the peak value, achieved in a series of trials during the make-method, should be used to measure muscle strength in children with CP. Before using the HHD in general practice, training is generally recommended. This method might enhance reliability of measurement in this population. To date, there is no study in which the 'break and make' method in a group of children and adolescents with CP has been compared.

Ayalon et al. [20] reported high intraclass correlation coefficients for test-retest reliability for isokinetic strength measurement in children with CP. Although these results are encouraging, many clinicians do not have access to expensive isokinetic dynamometry equipment.

Measuring muscle strength using a HHD, the most common way to assess muscle strength in studies involving children with CP, is difficult. The examiners need experience in hand-held dynamometry. HHD measures muscle strength in a single joint movement, whereas the functional strength tests are more related to the activity level and consist of multiple-joint movements. Moreover, the functional strength tests not only measure muscle strength, but also balance, coordination and some endurance and therefore correspond with the contemporary task-oriented approach [21]. The functional strength measures could be more appropriate for the use in everyday clinical practice. Testing is easily performed, functional and does not take a long time. Moreover, it seems that formal training is not required. To date, the reliability of the RM during functional exercises has not been studied in a group of children and adolescents with CP.

The consistency of measures of muscle strength is important for clinicians. In research and clinical practice it is preferable that a test can be reliably performed by different assessors. Therefore, we aimed to determine the intertester reliability of measuring lower-limb strength with the HHD (break vs. make) and the functional strength tests.

Methods

Subjects

A convenience sample of 28 children and young adults from a school for special education were invited to participate in the study. To be included, subjects were required to be within the age range of 7 – 20 years, had to be diagnosed with CP, and, due to the physical demands of the tests, classified as level I or II on the Gross Motor Function Classification System (GMFCS) [22]. Cognitively, they had to be capable of understanding and following simple commands. Subjects had to have the degree of passive joint range of motion in the hips and knees that allowed them to assume the test positions (Table I). The study was approved by the Institutional Review Board of the University Medical Center Utrecht.

Procedures

Prior to testing, each child was weighed on electronic scales (Seca, Hamburg, Germany). Height measurements were taken on the same visit while standing against a wall. Body composition was assessed using the sum of 7 skin folds method according to Pollack et al. [23]. The measurements were taken at 7 sites (on the right side of the body: triceps, biceps, subscapular, suprailiac, mid-abdominal, medial calf, and front thigh) by the investigators (OV and TT) in accord with the American College of Sports Medicine guidelines [24].

For the reliability study of the hand-held dynamometry, two examiners (OV and MvB), previously trained in hand-held dynamometry, performed hand-held dynamometry measurements on 5 muscle groups of the lower extremity (hip extensors, hip abductors, knee extensors, knee flexors, ankle plantar flexors), using break- and make- techniques. Normally, a 3-min recovery period is sufficient to repeat short running sprints without substantial fatigue [25]. Because of the number of measurements, the child was allowed a 5 minute rest between the break- and make-test, assessed by the first examiner in that order. Before the child performed the tests, which were assessed in the same order by the second examiner, he/she had a 10-min rest-period.

For the reliability study of the functional strength tests, eight paediatric physical therapists performed the functional closed chain strength tests. For each child two therapists were randomly chosen out of a sample of eight to perform the tests. All therapists were experienced paediatric physical therapists (4 – 20 years of experience). Prior to data collection, the therapists were given written instructions in the application and scoring of the functional strength measures, and had no formal training in how to perform functional strength measures. Since a decrease in muscle strength can be expected after five consecutive all-out functional tests the children were given a minimum and maximum time between the functional strength tests of respectively 24 and 72 h.

All HHD and functional strength assessments were performed within one week. The minimum time between the HHD and the functional strength tests was, respectively, 24 and 96 h. All tests were performed during school hours at the school of Special Education ‘Ariane de Ranitz, Utrecht, the Netherlands’.

Measures

The GMFCS. The Dutch translation of the GMFCS [22] was used by a paediatric physical therapist (OV), who was experienced in using the GMFCS, to classify the children and adolescents with CP into groups based on their functional ability. Level I represents the highest level of functional abilities and level V the lowest. In this study, only children with GMFCS levels I or II were recruited (Level I: Children walk indoors and outdoors, and climb stairs without limitation, Level II: Children walk indoors and outdoors, and climb stairs holding onto a railing, but experience limitations in walking on uneven surfaces and inclines, and walking in crowds or confined spaces).

Table I. Standardized muscle test positions.

Muscle group	Posture	Subject position	Dynamometer position
Hip extensors	Supine	Hip in 90° flexion; pelvis stabilized	Anterior mid-thigh
Hip abductors*	Supine	Hips in neutral position and 0°abduction/adduction; pelvis stabilized to prevent sliding	Lateral mid-thigh
Knee extensors*	Sitting	Knee and hip flexed 90°; pelvis stabilized to prevent hip extension	Anteriorly 5 cm proximal to lateral malleoli
Knee flexors*	Sitting	Knee and hip flexed 90°; thigh support to prevent hip flexion	Posteriorly 5 cm proximal to lateral malleoli
Ankle plantar flexors¶	Supine	Knee extended and resting on a bench; lower leg stabilized	Dorsum of foot at level of metatarsal heads

*Muscle test positions according to Berry et al. [18]; ¶Muscle test position according to Wiley and Damiano [17].

Hand-held dynamometry. Hand-held dynamometry was used to quantify subjects' isometric muscle force production. Children were given two practice trials for each test until the investigators were confident that they understood the task. Each child performed three trials for each muscle group, and the peak force values from the dynamometer were recorded for break and make-method [26]. Clear standardized instructions regarding the testing procedure and strong verbal encouragement during the trials were used to produce maximal effort.

In the *break* test, the examiner gradually applied force with the dynamometer for 1 sec to allow the subject to adjust and to recruit the maximum number of muscle fibres. After 1 sec, the force of the examiner exceeds the force of the patient very slightly [27,28].

The *make* test, operationally defined as a maximal effort exerted against a stationary HHD was used. The children were instructed to gradually 'push as hard as possible' against the dynamometer (held rigidly by the examiner perpendicular to the child's limb segment) over a period of approximately 5 sec until the examiner told them to relax [27,28].

The HHD (type CT3001, Citec, C.I.T. Technics BV, Groningen, The Netherlands) that was used, could measure forces up to 500 Newton (N). The HHD was calibrated with weights before beginning data collection and was found to be accurate within $\pm$0.89 N. The HHD was equipped with a padded end piece provided by the manufacturer and was not modified for this study.

Test positions, except for the hip extensors, were standardized according to Berry et al. [18] and Wiley and Damiano [17]. The position of the testers was not standardized. The following muscle groups were evaluated bilaterally: hip extensors and abductors, knee flexors and extensors and ankle plantar flexors. All muscle groups were assessed in the same order as shown in Table I. The point on the extremity were resistance is applied was standardized. Absolute strength values in Newton were recorded and used in the analyses.

Functional strength tests. To assess functional performance in children with CP we chose functional exercises in which the large muscle groups that are important for standing and walking are being tested [29]. Three closed kinetic chain exercises were chosen:

(1) *The Lateral Step-up Test* (*on a 20 cm bench*). This step test is a closed kinetic chain test that has been utilized to assess lower extremity muscular performance [30]. For each test, the subjects were asked to stand with the extremity being tested on the step with their feet parallel and shoulder width apart. Appropriate lateral step-up technique was defined as achieving a position within 15° of knee extension for the tested extremity during the extension phase of the test. Repetitions were counted each time the heel or toes of the extremity not being tested touched the floor. Test-retest reliability in young healthy adult subjects was very good [12]. This protocol has never been examined before in subjects with CP.

(2) *Sit-to-Stand* (*from 90° flexion of the knee and hip to standing position*). This test is a functional item in which the child has to attain stand without using the arms. The child was positioned on a small bench, and seated with feet flat on the floor and knees flexed at 90°. The child had to achieve standing, arms free, without any assistance from their arms on the bench or their body in the transition. Repetitions were counted each time the child's legs and hips were within 15° of the extended position.

(3) *Attain stand through half kneel, without using arms.* This test is a functional item in which the child has to attain stand without using the arms. The child was positioned on a mat in high kneeling, arms free. This means that weight bearing is on one knee and the opposite foot, and that the alignment may vary as long as the buttocks are clear of the lower legs and/or the weight bearing surface. The child was instructed to assume standing without using any external support such as furniture or the floor. Repetitions were counted each time the child achieved a standing position, and both legs and hips are within 15° of the extended position.

Exercise 1 and 3 were assessed bilaterally. During exercise 2 both extremities were used to perform the task. Total scores for the left and right side were calculated from the repetition maximums for each side, so in total 5 scores were calculated; Lateral Step-up Test (scores for left and right side), Sit-to-Stand (score) and left and right Attain stand through half kneel. (scores for the left and right side). Following the command, 'Ready, set, go,' subjects started each test and timing began on a stopwatch accurate to 0.01 sec. Subjects were given 2 practice repetitions per extremity prior to formal testing. Subjects were given 30-sec rest following the practice session and 180-sec between tests. The exercises were assessed in the following order: (i) Lateral Step-up Test left, (ii) Lateral Step-up Test right, (iii) Sit-to Stand, (iv) Attain stand through half kneel left, and (v) Attain stand through half kneel right. The subjects were instructed to perform as many repetitions as possible in 30 sec and they were verbally encouraged. The number of repetitions performed over 30 sec were recorded and used in the analyses.

Data analysis

Data gathered from both tests and extremities were analysed using the statistical software package (SPSS, version 13.0; SPSS; Chicago, Ill) and MS Excel 2005 for Windows. Intra-class correlations (ICC; two way mixed) were computed for the intertester reliability of the HHD-measurements and the functional strength measures. The ICC gives a relative index of the ratio of variance between subjects to the variance between subjects plus error variance. Acceptable reliability criteria for ICC values were values > 0.80 [31].

In order to assess the amount of error associated with repeated measurements, the standard error of measurement (SEM) was calculated by means of the following equation: $SEM = SD\sqrt{1\text{-}ICC}$. Values were computed for each muscle group of each limb (HHD) and for each test of the functional strength tests.

Moreover, limits of agreement (LOA) were calculated to conform to the procedure described by Bland and Altman [32]. Bland-Altman analysis describes the level of agreement between two measurements. In this analysis, the 'bias' is an estimate of how closely on average the two measurements agree and the 'precision; indicates how well the methods agree for an individual. By multiplying the precision by 1.96, the 'limits of agreement' are calculated. The coefficient of variation (CV) was calculated to compare the reliability of the different measurement tools [33]. The level of statistical significance was set at $p = 0.05$.

Results

In total, 25 subjects (15 males, 10 females) and their parents agreed to participate and signed the informed consent form. Three subjects refused to participate because they were involved in another study. All included subjects were able to assume the test positions as described in Table I. Group characteristics are described in Table II. All 25 subjects successfully performed all four test-sessions. The following results refer to both sexes and GMFCS levels combined.

Hand-held dynamometry

Break test. The ICC and SEM for the break-test can be found in Table III. ICC values ranged from 0.42 for left knee flexors to 0.73 for right hip abductors. The precision of our measures, represented as CV, ranged from 22.2% for the right hip abductors to 35.3% for the left hip abductors and left knee flexors. In most subjects the tested muscle groups could be measured bilaterally. For the ankle plantar flexors the strength could only be measured in 10 subjects for the left side and 8 (tester 1) and 5 (tester 2) subjects for the right side. In this study the break test has questionable reliability for measuring muscle strength using hand-held dynamometry for most muscle groups. The SEM values, the standard error of the difference values, are listed in Table III. SEM values ranged from 30.7 N to 58.9 N for hip extensors and hip abductors, knee extensors and knee flexors.

Make test. The ICC and SEM for the make-test can be found in Table III. ICC values ranged from -0.04 for left ankle plantar flexors to 0.82 for right hip extensors. The precision of our measures, represented as CV, ranged from 16.2% for the right hip extensors to 56.2% for the left ankle plantar flexors.

In all subjects the tested muscle groups could be measured bilaterally, except for the ankle plantar flexors in which for 25 subjects one right ankle plantar flexor by tester 1, two left and three right ankle plantar flexors could not be tested by tester 2. While specific muscle groups measured by hand-held dynamometry using the make test have questionable reliability, the tool also has questionable reliability for measuring the strength of knee flexors and ankle plantar flexors in this population.

Table II. Subject characteristics ($n = 25$).

Number of subjects	GMFCS I[a] 14				GMFCS II[b] 11			
Variables	Mean	SD	Median	Range	Mean	SD	Median	Range
Age (y.m)	11.11	2.8	12.6	7.8 – 16.8	10.11	2.7	12.2	7.5 – 17.4
Height (cm)	149.6	15.3	149.0	125.0 – 175.0	148.9	18.9	145.1	123.0 – 175.0
Body mass (kg)	40.3	12.4	35.1	23.8 – 60.8	38.6	12.1	32.7	24.0 – 59.7
Sum of seven skin folds	77.5	44.5	62.0	36.5 – 197	74.5	28.5	70.5	33.5 – 132

SD, standard deviation; GMFCS, Gross Motor Function Classification System; [a]12 hemiplegia: 2 diplegia; [b]1 hemiplegia; 10 diplegia.

Table III. Mean scores, intertester reliability and standard error of measurements for Break- and Make-method of lower-limb strength, using a Hand held Dynamometer ($n = 25$).

	T1 NT/T	T2 NT/T	T1 Mean (N)	T1 SD (N)	T2 Mean (N)	T2 SD (N)	ICC	SEM	CV (%)	LOA
Break method										
Hip extensors										
Left	3/22	2/23	209.6	81.7	187.4	57.4	0.53	55.5	28.5	134.6
Right	3/22	3/22	204.8	64.6	188.0	62.4	0.63	38.9	23.3	107.5
Hip abductors										
Left	1/24	1/24	176.4	66.1	189.2	67.3	0.49	47.2	35.3	131.7
Right	1/24	1/24	181.8	59.9	199.0	67.6	0.73	30.7	22.2	92.6
Knee extensors										
Left	1/24	1/24	206.9	91.8	168.8	77.0	0.60	57.5	32.7	148.4
Right	2/23	2/23	198.9	87.2	169.4	79.9	0.62	54.7	33.1	142.7
Knee flexors										
Left	3/22	3/22	178.5	78.0	171.0	66.4	0.42	58.9	35.3	152.6
Right	3/22	1/24	187.3	73.7	176.7	56.4	0.46	54.1	34.2	133.8
Ankle plantar flexors										
Left	15/10	15/10	170.7	56.9	158.0	79.5	0.70	27.9	29.4	105.1
Right	17/8	20/5	169.0	29.7	142.5	31.8	N/A	N/A	N/A	N/A
Make method										
Hip extensors										
Left	0/25	0/25	236.2	70.3	158.8	70.4	0.67	40.4	22.4	111.2
Right	0/25	0/25	209.4	72.7	164.3	61.4	0.82	30.9	16.2	79.4
Hip abductors										
Left	0/25	0/25	136.0	55.5	127.0	53.0	0.69	30.6	27.4	83.2
Right	0/25	0/25	137.7	55.2	127.5	56.0	0.68	31.7	27.4	86.9
Knee extensors										
Left	0/25	0/25	188.7	73.8	143.9	62.0	0.70	40.4	21.6	103.3
Right	0/25	0/25	198.4	77.9	150.5	63.4	0.70	42.7	28.1	108.1
Knee flexors										
Left	0/25	0/25	154.5	67.2	139.5	47.9	0.49	40.7	27.5	115.2
Right	0/25	0/25	169.0	66.9	155.9	68.0	0.64	40.1	31.2	112.6
Ankle plantar flexors										
Left	0/25	2/23	128.3	53.9	137.3	45.5	−0.04	52.7	56.2	141.0
Right	1/24	3/22	127.3	47.2	136.7	51.2	0.48	35.6	28.8	98.3

NT/T, not testable/testable; N, Newton; SD, standard deviation; ICC, intra class correlation; LOA, limits of agreement; SEM, Standard Error of Measurement; CV, coefficient of variation; N/A, not calculated because of the small sample size.

The SEM values, the standard error of the difference values, are also listed in Table III. SEM values ranged from 30.6 N to 42.7 N for hip extensors and hip abductors, knee extensors and knee flexors.

Functional strength tests

The reliability statistics of all functional measures are shown in Table IV. ICC values ranged from 0.91 – 0.96, and SEM values ranged from 1.1 – 2.6 for the three functional exercises, and from 3.7 – 4.5 for the total scores on the right and left side. The test-retest data are based on the two assessors that tested the subject on both tests. The precision of our measures, represented as CV, for the total number of repetitions for the left and right lower extremity were respectively 10.9% and 16.2%. As can be appreciated from the Bland-Altman plot (see Figure 1), there were some obvious outliers. Although these outliers are included in the calculations, the reliability statistics are still acceptable.

Discussion

In this study the intertester reliability of strength measurements in a group of children with CP was assessed. Functional tests showed acceptable reliability values ($ICC \geq 0.91$; CV 10.9 – 39.9%) for all three functional strength measures: (i) The Lateral Step-Up Test; (ii) Sit-to-Stand; and (iii) Attain stand through half kneel; and variable reliability values ($ICC = -0.04$ to 0.82; CV 16.2 – 56.2%) for HHD measurements between two assessors.

In the present study different statistics to examine reliability (i.e., ICC, SEM and CV) were used. The ICC indicates the proportion of the total variance in the measurements which is due to the true difference between subjects. The SEM and CV are methods of analysing absolute reliability [34]. The SEM, that is unaffected by the range of measurements, is used if the goal is to determine the stability of a child's performance. The coefficients of variation (CV) are particularly useful for comparing the reliability between performance tests. Thus, when measuring the reliability, the ICC, SEM and CV yield useful, although quite different information. Previous reliability studies in children with CP have used the Pearson correlation as a measure of reliability [35–37]. However, the Pearson correlation is not very suited for the purpose since it cannot assess systematic bias and it depends greatly on the range of values in the sample [38].

Measurements taken by an individual are typically more reliable than those taken by multiple examiners [39]. However, in research and clinical practice it is preferable that a test can be reliably performed by different assessors. Taking this into account, it is not surprising that in the studies, investigating the intratester reliability of HHD in subjects with CP [16,18,19], the ICCs are higher than those found in the present study. The results of our study show that agreement among two observers with experience in the assessment of muscle strength using a HHD is low. It would be interesting to examine agreement in a group of observers with no or limited experience in HHD strength testing.

For rehabilitation practice, it is desirable that a measurement tool is reliable enough to be used on individuals. For example, a clinician may need to

Table IV. Mean scores, intertester reliability and standard error of measurements of functional lower-limb strength measurement ($n=25$).

Test positions	NT/T	Mean	SD	ICC	SEM	CV (%)	LOA
Lateral step up test							
Left	0/25	13.2 reps	10.5	0.94	2.4 reps	17.8	7.24
Right	0/25	12.6 reps	10.4	0.94	2.6 reps	22.7	6.83
Sit-to stand	0/25	14.4 reps	5.0	0.91	1.5 reps	12.1	4.04
Half kneel to stand							
Left	0/25	7.5 reps	5.5	0.96	1.1 reps	28.6	3.16
Right	0/25	6.0 reps	5.3	0.93	1.4 reps	39.9	3.87
Total left	0/25	35.4 reps	18.9	0.96	3.7 reps	10.9	10.38
Total right	0/25	33.1 reps	18.2	0.94	4.5 reps	16.2	12.16

NT/T, not testable/testable; ICC, intra class correlation; LOA, limits of agreement, CV, coefficient of variation; SEM, Standard Error of Measurement; reps, repetitions.

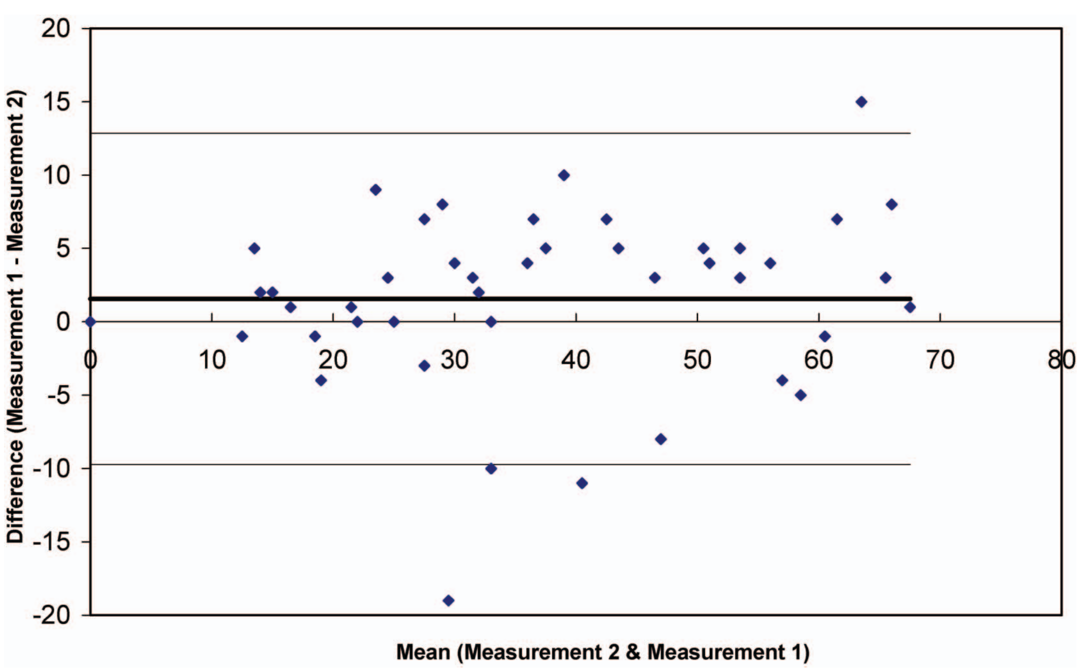

Figure 1. Bland-Altman plot for measurement 1 and measurement 2 for total repetitions for the left & right lower extremity. The bold-type line shows the mean difference between the 2 measurements, and the 2 thin lines indicate the limits of agreement (±1.96 standard deviations). On the X-axes, the average repetitions from both tests are displayed. On the Y-axes, the difference between the total repetitions for the left & right lower extremity during measurement 1 and measurement 2 is displayed.

know whether an improvement in muscle performance following a rehabilitation program is real or merely due to measurement error. The results of the functional strength tests showed that they could be assessed in a highly reliable way between two assessors in children with CP. The calculated SEM can be used to determine the range in which a person's 'true score' could be expected to lie, considering the amount of error associated with repeated measures. For example, we can be 95% confident that the 'true score' for people performing the functional strength tests lies within ± 1.96 SEM. Based on the data in Table IV, total RM increases of 7.3 (1.96×3.7) and 8.8 (1.96×4.5) for the left and right side respectively can be attributed to real change with 95% confidence. Since the SEM values for each of the functional strength tests were small, it appears that these tests represent stable measures and may be used in the clinical evaluation of children and adolescents with CP.

The coefficients of variation (CV) for the hip extensors and abductors, and the knee extensors and flexors HHD measures vary from 22.2 – 35.3% and from 16.2 – 31.2% for, respectively, the break and make method. The CV for the individual functional exercises vary from 12.1 – 39.9% for, respectively, the Sit-to-Stand and the Attain stand through half kneel (right) item. The CVs for the total score of the functional strength test for the left and right lower extremities are 10.9% and 16.2%, respectively. Since the CVs for the sit-to-stand test and the total scores for the lower extremities are smaller than for the other functional strength tests and HHD, these tests might be the most sensitive to change.

Poor selective control in some muscle groups may prevent an individual from being able to perform the task [8]. In the current study this was only found in the ankle plantar flexors during the make-method. This may be due to the fact that in children and adolescents classified as a level I or II on the GMFCS (the ones with the best motor abilities), motor control limitations are not likely to be a substantial factor in the ability to generate force. In the break-method the examiner attempts to exceed the force of the child. In this way of testing the examiner's strength can be inadequate to overpower the tested muscle. In this group of children the make-method has been shown to be more reliable than the break-method for the following muscle groups: hip extensors, hip abductors, knee extensors and knee flexors. We therefore recommend the use of the make method for measuring muscle strength in young people with CP.

In previous studies that assessed the intratester reliability for strength testing using HHD in children with CP, experienced examiners were used. In our study, that assessed the intertester reliability for HHD we also used experienced examiners. We found questionable reliability for the make and break method. As mentioned in the introduction it is preferable that a test can be reliably performed by different testers. However, it is clear from this and previous studies [40] that HHD suffers from questionable intertester reliability. Intratester reliability is high or moderate for most muscle groups and therefore it is recommended to use the same examiner if longitudinal measurements have to be done.

For the functional strength tests physical therapists who had no formal training were used to assess the intertester reliability. So, to perform the functional strength tests in a reliable way, a special training is not required. Therefore, this strength testing method is available and applicable for a variety of professionals working with children with CP. If a researcher or clinician is interested in the functional gain in task-related activities, the functional strength test is an appropriate and reliable measure. Moreover, the functional strength test is reliable when used by different testers.

There are some limitations in our study. In the present study the order in which all muscle groups and functional exercises were examined were fixed. Fatigue, concentration and distraction can play an important role in these kind of test protocols. The present study provides no information on the effect of order of testing and if these results apply to other testing orders or the testing of a single muscle or functional movement. As can be appreciated from the Bland-Altman plot (Figure 1) there were some outliers in muscle strength scores. Both muscle strength tests are influenced by the motivation or co-operation of a patient. Moreover, children with CP show a large variability in the generation of muscle strength [41]. Cognitive impairment has been cited as a possible reason for large within-participant variability when testing isokinetic strength [35]. Unfortunately, cognition of the children that participated in this study were not formally assessed.

In conclusion, the results of this study show that intertester reliability of HHD is low, and therefore it is advised to use the same examiner when HHD is the preferred muscle strength test method. When a task-related outcome is preferred the 30-sec RM total scores for the left and right lower extremity can be used by different examiners to obtain highly reliable measures of lower extremity performance when used conform a standardized protocol.

Acknowledgements

The authors are grateful for the help of the Dr W.M. Phelps Foundation which funded this study.

In addition, the children, their parents, and the (paediatric) physical therapists from the School for special education 'Ariane de Ranitz' who gave their time and assistance are appreciated.

References

1. Styer-Acevedo J. Physical therapy for the child with cerebral palsy. In: Tecklin J, editor. Pediatric Physical Therapy. Philadelphia: Lippincott Williams & Wilkins; 1999. pp 107–162.
2. Damiano DL, Vaughan CL, Abel MF. Muscle response to heavy resistance exercise in children with spastic cerebral palsy. Dev Med Child Neurol 1995;37:731–739.
3. Damiano DL, Abel MF. Functional outcomes of strength training in spastic cerebral palsy. Arch Phys Med Rehabil 1998;79:119–125.
4. Darrah J, Wessel J, Nearinburg P, O'Connor M. Evaluation of a community fitness program for adolescents with cerebral palsy. Ped Phys Ther 1999;11:18–23.
5. Dodd KJ, Taylor NF, Damiano DL. A systematic review of the effectiveness of strength-training programs for people with cerebral palsy. Arch Phys Med Rehabil 2002;83:1157–1164.
6. MacPhail HEA, Kramer JF. Effect of isokinetic strength-training on functional ability and walking efficiency in adolescents with cerebral palsy. Dev Med Child Neurol 1995;37:763–775.
7. Armstrong N, Mechelen van W. Paediatric exercise science and medicine. New York: Oxford University Press, Inc.; 2000.
8. Damiano DL, Dodd KJ, Taylor NF. Should we be testing and training muscle strength in cerebral palsy? Dev Med Child Neurol 2002;44:68–72.
9. Gray GW. Lower extremity functional profile. Adrian, MI: Wynn Marketing, Inc.; 1995. pp 109–110.
10. Blundell SW, Sheperd RB, Dean CM, Adams RD. Functional strength training in cerebral palsy: a pilot study of a group circuit training class for children aged 4–8 years. Clin Rehabil 2003;17:48–57.
11. Liao HF, Liu YC, Liu WY, Lin YT. Effectiveness of Loaded Sit-to-Stand Resistance Exercise for children with mild spastic diplegia: A randomized clinical trial. Arch Phys Med Rehabil 2007;88:25–31.
12. Ross M. Test-retest reliability of the lateral step-up test in young adult healthy subjects. JOSPT 1997;25(2):128–132.
13. Bohannon RW. Sit-to-stand test for measuring performance of lower extremity muscles. Percept Mot Skills 1995;80:163–166.
14. Mendell JR, Florence J. Manual muscle testing. Muscle Nerve 1990;13:s16–20.
15. Noreau L, Vachon J. Comparison of three methods to assess muscular strength in individuals with spinal cord injury. Spinal Cord 1998;36:716–723.
16. Taylor NF, Dodd KJ, Graham HK. Test-retest reliability of hand-held dynamometric strength testing in young people with cerebral palsy. Arch Phys Med Rehabil 2004;85:77–80.
17. Wiley ME, Damiano DL. Lower-extremity strength profiles in spastic cerebral palsy. Dev Med Child Neurol 1998;40:100–107.
18. Berry ET, Giuliani CA, Damiano DL. Intrasession and intersession reliability of handheld dynamometry in children with cerebral palsy. Ped Phys Ther 2004;16:191–198.
19. Crompton J, Galea MP, Phillips B. Hand-held dynamometry for muscle strength measurement in children with cerebral palsy. Dev Med Child Neurol 2007;49:106–111.
20. Ayalon M, Ben-Sira D, Gilad T. Reliability of isokinetic strength measurements of the knee in children with cerebral palsy. Dev Med Child Neurol 2000;42:398–402.
21. Larin HM. Motor learning theories and strategies for the practicioner. In: Campbell SK, Linden DWV, Palisano RJ, editor. Physical therapy for children. Philadelphia: W.B. Saunders company; 1994. pp 170–197.
22. Palisano RJ, Rosenbaum P, Walter S. The development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 1997;39:214–223.
23. Pollack ML, Schmidt DH, Jackson AS. Measurement of cardio-respiratory fitness and body composition in the clinical setting. Compr Ther 1980;6:12–27.
24. Latin RW. Surface anatomy. Roitman JL, editor. Baltimore: Williams and Wilkins; 1998. pp 89–100.
25. Ratel S, Williams CA, Oliver J, Armstrong N. Effects of age and recovery duration on performance during multiple treadmill sprints. Int J Sports Med 2006;27(1):1–8.
26. Bohannon RW. Make tests and break tests of elbow flexor muscle strength. Phys Ther 1988;68:193–194.
27. Ploeg RJvd, Oosterhuis HJ, Reuvekamp J. Measuring muscle strength. J Neurol 1984;231:200–203.
28. Ploeg RJvd, Oosterhuis HJ. The 'make/break test' as a diagnostic tool in functional weakness. J Neurol Neurosurg Psychiatry 1991;54:248–251.
29. Gage JR. Gait analysis: an essential tool in the treatment of cerebral palsy. Clin Orthopedics 1993;288:126–134.
30. Worrell TW, Borchert B, Erner K, et al. Effect of a Lateral Step-up exercise protocol on quadriceps and lower extremity performance. JOSPT 1993;18(6):646–653.
31. Portney LG, Watkins MP. Foundations of clinical research: Applications to practice. Norwalk: Appleton and Lange; 2000.
32. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–310.
33. Felz CJ, Miller GE. An asymptomatic test for the equality of coefficients of variation from k populations. Stat Med 1996;16(6):646–658.
34. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998;26(3):217–238.
35. Berg-Emons van den RJ, van Baak MA, de Barbanson DC, et al. Reliability of tests to determine peak aerobic power, anaerobic power and isokinetic muscle strength in children with cerebral palsy. Dev Med Child Neurol 1996;38:1117–1125.
36. Stuberg WA, Metcalf WK. Reliability of quantitative muscle testing in healthy children and children with Duchenne muscular dystrophy using a hand-held dynamometer. Phys Ther 1988;68:977–982.
37. Bohannon RW. Test-retest reliability of hand-held dynamometry during a single-session of strength assessment. Phys Ther 1986;66:206–209.
38. Bates BT, Zhang S, Dufek JS. The effects of sample size and variability on the correlation coefficient. Med Sci Sports Exerc 1996;28(3):386–391.
39. Andrews AW. Hand-held dynamometry for measuring muscle strength. J Hum Muscle Perform 1991;1:35–50.
40. Bohannon RW. Intertester reliability of hand-held dynamometry: A concise summary of published research. Percept Mot Skills 1999;88(3 Pt 1):899–902.
41. Valvano J, Newell KM. Practice of a precision isometric grip-force task by children with spastic cerebral palsy. Dev Med Child Neurol 1998;40(7):464–473.