Int J Clin Exp Med 2016;9(1):199-208
www.ijcem.com /ISSN:1940-5901/IJCEM0015678

Original Article

A simple way to improve the relative and absolute reliability of handheld dynamometer measurements using learners

Genlin Liu[1], Jianjun Li[2], HongjunZhou[1], Ying Zheng[1], Chunxia Hao[1], Ying Zhang[1], Bo Wei[1], Yiji Wang[1], Haiqiong Kang[1], Xiaolei Lu[1]

[1]*Department of Spinal Cord Injury Rehabilitation, China Rehabilitation Research Center, Capital Medical University School of Rehabilitation Medicine, Beijing, People's Republic of China;* [2]*Department of Spinal and Neural Function Reconstruction, China Rehabilitation Research Center, Capital Medical University School of Rehabilitation Medicine, Beijing, People's Republic of China*

Received September 6, 2015; Accepted September 23, 2015; Epub January 15, 2016; Published January 30, 2016

Abstract: Background: To investigate a simple way to improve the reliability of handheld dynamometer (HHD) measurements using learners.Methods: Maximal voluntary contractions (MVCs) of 21 healthy young adults were tested by 2 HHD learners with the subjects in supine position. The MVCs of the elbow flexor, elbow extensor, hip flexor, and knee extensor were tested bilaterally at two different joint angles. The intraclass correlation coefficients (ICCs) and the smallest real differences (SRD) were used to examine the relative and absolute reliabilities. Bland-Altman analyses were used to assess differences between rater measurements, and paired t-tests were used to check for systematic bias.Results: Reliability was better when the muscle was tested in the weaker position. (E1 < E2 (P < 0.0001), ICCs of E1 (0.78) > E2 (0.76), SRD of E1 (29.7) < E2 (34.6); E3 > E4 (P < 0.0001), ICCs of E3 (0.74) < E4 (0.86), SRD of E3 (38.3) > E4 (25.5); H1 > H2 (P < 0.0001), ICCs of H1 (0.19) < H2 (0.62), SRD of H1 (80.4) > H2 (42.1); K1 < K2 (P = 0.0001), ICCs of K1 (0.22) > K2 (0.06), SRD of K1 (59.0) <K2 (94.5)). Systematic bias existed in one of the testing positions (E4). Conclusions: Changing the joint angle of the limb tested could be a simple way to improve the reliability of handheld dynamometer measurements.

Keywords: Muscle strength, joint angle, interrater reliability, measurement, handheld dynamometry

Introduction

Although manual muscle testing is widely used in the clinic for practice, it is not sensitive enough to evaluate subtle changes in motor function [1-3]. As a result, quantitative tests of motor function have been developed for increased accuracy [4, 6]. Isokinetic dynamometry, first introduced in the early 1980s, is one method that was initially used to evaluate muscle strength quantitatively. However, its usage was limited due to its high cost, large space requirement, and long assessment time [6].

Handheld dynamometry is an alternative method of quantitative muscle testing [4-9]. When compared to isokinetic dynamometers, handheld dynamometers are lightweight, portable, and inexpensive, yet still obtain reliable readings [33]. HHD application typically uses either "make" or "break" tests to measure strength, both of which give the most reliable results when the upper limb of the tester is stronger than the muscle group being tested in the subject [7-13, 33]. However, human lower limb strength is typically greater than upper limb strength. As a result, measurements of the subject's lower limb by HHD are typically unreliable and clinical usage is limited.

This research aims to improve the reliability of HHD by exploring differences in the measurement of muscle strength at different joint angles in the same limb. We hypothesize: 1) that the muscle strengths of different joint angles within the same limb are different, with some positions being stronger than others; 2) there is a close correlation between the muscle

Table 1. Subject and rater demographics

	Age (years)	Height (cm)	Body weight (kg)	Grip (kg, right)	Grip (kg, left)	Occupation
Men (n = 10)	39.0 ± 9.5	167.8 ± 4.0	74.0 ± 10.0	73.3 ± 19.5	70.9 ± 18.0	Care worker
Women (n = 11)	30.7 ± 6.3	160.5 ± 3.5	53.5 ± 6.2	39.7 ± 8.9	37.2 ± 7.4	Registered nurse
Rater A (male)	21	178	72	100	100	Intern nurse
Rater B (female)	26	166	61	52	51	Registered nurse

Subject measurements are indicated as Mean ± SD.

Table 2. Test position and dynamometer placement for each muscle action

Muscle Action	Extremity/Joint Positions	Location of Dynamometer Application
Elbow Flexion (E1)	Shoulder neutral, elbow 45°, forearm neutral	Just proximal to styloid processes
Elbow Flexion (E2)	Shoulder neutral, elbow 90°, forearm neutral	Just proximal to styloid processes
Elbow Extension (E3)	Shoulder neutral, elbow 150°, forearm neutral	Just proximal to styloid processes
Elbow Extension (E4)	Shoulder neutral, elbow 90°, forearm neutral	Just proximal to styloid processes
Hip Flexion (H1)	Hip flexed 45°, knee flexed, contralateral hip neutral	Just proximal to femoral condyles
Hip Flexion (H2)	Hip flexed 90°, knee flexed, contralateral hip neutral	Just proximal to femoral condyles
Knee Extension (K1)	Hip flexed 90°, knee flexed 90°, contralateral hip neutral	Just proximal to malleoli
Knee Extension (K2)	Hip flexed 125°, knee flexed 150°, contralateral hip neutral	Just proximal to malleoli

strength at the strong and weak positions within the same limb; and 3) reliability is greater when the weak position is tested. Therefore, it is possible to improve the reliability of handheld dynamometer measurements by changing the joint angle of the tested limb.

Methods

We investigated reliability using a single-group, repeated-measures design.

Subjects

Twenty-one healthy adults were recruited from the department of spinal cord injury (SCI) rehabilitation in the China Rehabilitation Research Center. Subjects were asked to avoid intense exercise in the 24 hours before the test, but light activities were acceptable. The research protocol was approved by the institutional review board of our hospital, and written informed consent was obtained from all subjects. Subject demographics are included in **Table 1**.

Procedures

Muscle strength was tested by 2 handheld dynamometer learners. Rater A, a male intern nurse, and Rater B, a female registered nurse in the department of SCI rehabilitation, participated in a 2-hour training session led by the principal investigator to standardize evaluation procedures. Both raters were randomly assigned and blinded to the strength values obtained by the other. Rater demographics are included in **Table 1**.

A calibrated HBO HHD (Yueqing Haibao Instrument Co., Ltd., Zhejiang Province, China) was used to measure muscle strength of bilateral upper and lower extremities. The tested muscles included the elbow flexor at the position of flex 45° (E1) and 90° (E2), the elbow extensor at the position of flex 150° (E3) and 90° (E4), the hip flexor at the position of flex 45° (H1) and 90° (H2), and the knee extensor at the position of flex 90° (K1) and 135° (K2). For all measurements, the shoulder and hip joints were kept in the neutral position, and the measured joints were stabilized. The moving arm of an angle meter was bound to the patient's body with string to ensure a constant joint angle during muscle strength measurements. Detailed information on individual testing procedures is included in **Table 2**.

Subjects were placed in the supine position during the test (**Figure 1**). The testing sequence began with the side of arm or leg randomized. Subjects were tested twice in a 1-week time span by both raters. The second test sequence was identical to the first.

The make test [5, 6, 15] was used in this study. Subjects were asked to exert maximum force

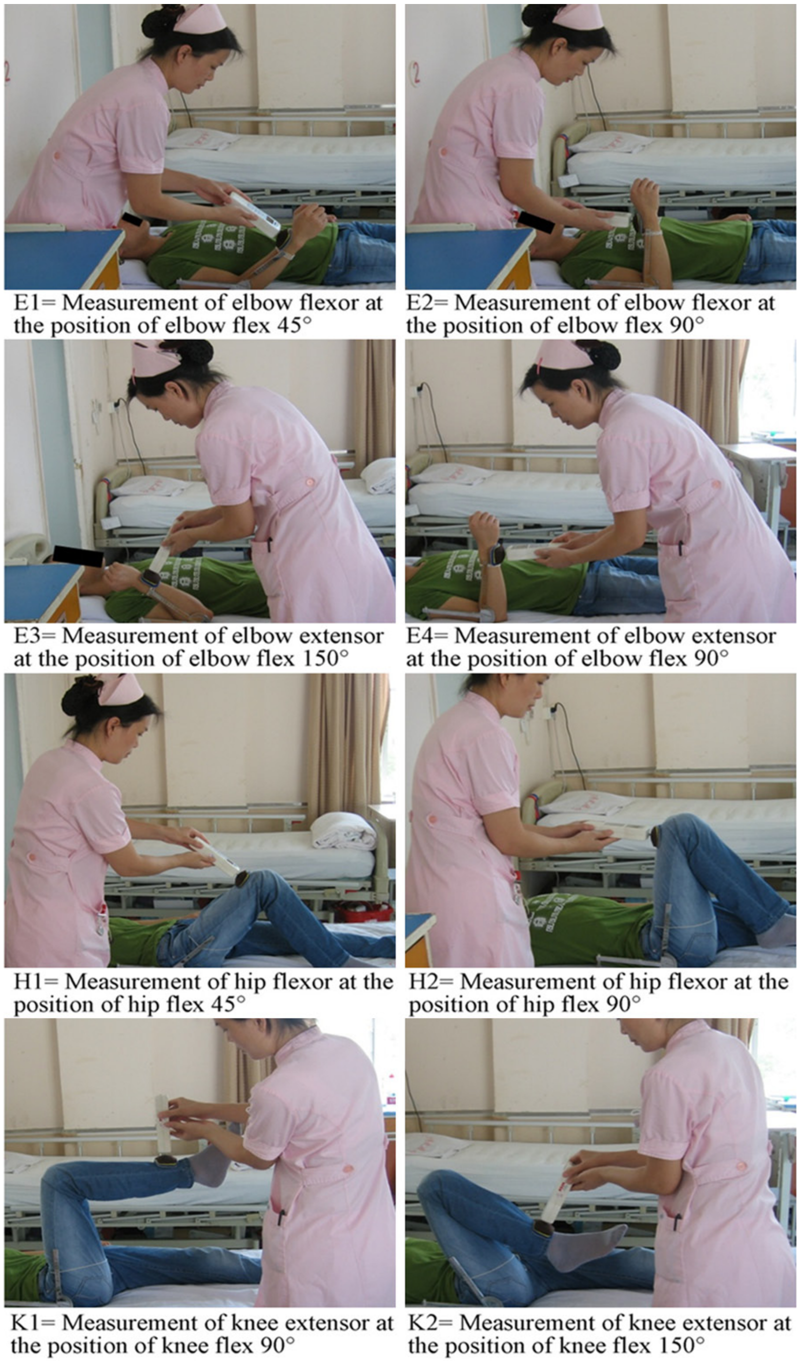

Figure 1. Details of Test Positions and Dynamometer Placements Used During the Testing of 8 Muscle Actions With a Hand-Held Dynamometer.

against the applicator for 3 seconds while the dynamometer was held stationary by the rater. The applicator was immediately removed from the limb and the force was recorded. Before the test, the principle investigator explained and demonstrated the make test and asked subjects to practice once to familiarize themselves with the test method for each muscle. Strength was measured twice for each muscle group during a session, with a 30-second rest period between trials. Averages for the two trials were computed to improve reliability [14, 19].

Data analyses

A MedCalc Version 13.1.2 was used for all statistical analyses. A non-parametric Kolmogorov-Smimov test was used to confirm determines normal distribution of rater measurements. Bland-Altman analyses were conducted, and paired t-tests were to assess systematic bias. ICC and SRD were used to estimate relative and absolute reliability, respectively. SRD was established by a series of statistical procedures [19] (**Table 4**).

Results

All subjects confirmed that they had not exercised intensively in the 24 hours prior to the test. Kolmogorov-Smimov analysis confirmed that limb measurements obtained by both raters were normally distributed. Since there was no significant difference between the mean strength measurements of the left and right limbs (*P > 0.05*, **Table 7**), right and left strength measurements were averaged and used to calculate the interrater reliability of each muscle. For relative reliability, the ICCs for 3 muscle strengths (E1, E2 and E4) were above 0.75 (0.76-0.86), indicating excellent interrater reliability (**Table 3**). The values of the ICCs for E3 and H2 were 0.74 and 0.63, respectively, indicating that interrater reliability was fair to good. The value of the ICCs for H1, K1 and K2 ranged from 0.06 to 0.22, indicating that interrater reliability was poor.

Table 3. Relative interrater reliability of upper and lower extremity strength

Measurement	Rater A (N)	Rater B (N)	Paired t-test	ICC	95% CI for ICC
E1	103.1 ± 23.4	106.7 ± 22.6	*P* = 0.2848	0.78	0.5304 - 0.9078
E2	120.6 ± 22.5	133.7 ± 27.1	*P* = 0.3434	0.76	0.4836 - 0.8960
E3	123.5 ± 24.9	127.7 ± 29.6	*P* = 0.3522	0.74	0.4481 - 0.8867
E4	90.3 ± 21.4	97.7 ± 27.4	*P* = 0.0274	0.86	0.6779 - 0.9413
H1	165.0 ± 35.7	157.3 ± 28.5	*P* = 0.4016	0.19	-0.2596 - 0.5772
H2	127.0 ± 25.7	122.3 ± 23.9	*P* = 0.3434	0.62	0.2512 - 0.8277
K1	112.9 ± 29.8	102.6 ± 15.8	*P* = 0.1689	0.22	-0.2369 - 0.5931
K2	165.6 ± 37.7	152.6 ± 32.0	*P* = 0.1878	0.06	-0.3855 - 0.4758

E1 = strength measurement of elbow flexor at the position of elbow flex 45°; E2 = strength measurement of elbow flexor at the position of elbow flex 90°; E3 = strength measurement of elbow extensor at the position of elbow flex 150°; E4 = strength measurement of elbow extensor at the position of elbow flex 90°; H1 = strength measurement of hip flexor at the position of hip flex 45°; H2 = strength measurement of hip flexor at the position of hip flex 90°; K1 = strength measurement of knee extensor at the position of knee flex 90°; K2 = strength measurement of knee extensor at the position of knee flex 150° (see **Figure 1**). ICC = intraclass correlation coefficient; 95% CI = 95% of confidence interval. Force measurements are indicated as mean ± SD.

Table 4. Absolute interrater reliability of upper and lower extremity strength

Measurement	$\bar{d}$ (N)	$\lvert \bar{d} \rvert$/mean	LOA (d ± 1.96 SD)	SEM (N)	SEM%	SRD (N)	SRD%
E1	3.6	2.2%	33.4 to -26.1	10.7	10.2%	29.7	28.3%
E2	13.1	9.1%	48.0 to -21.8	12.5	9.8%	34.6	27.2%
E3	4.2	0.8%	43.8 to -35.4	13.8	11.0%	38.3	30.5%
E4	7.4	7.3%	35.2 to -20.5	9.2	9.8%	25.5	27.1%
H1	-7.7	9.5%	73.1 to -88.5	29.0	18.0%	80.4	49.9%
H2	-4.6	3.4%	38.1 to -47.3	15.2	12.2%	42.1	33.8%
K1	-10.3	4.8%	54.3 to -74.8	21.3	19.8%	59.0	54.7%
K2	-13.0	4.8%	72.9 to -98.9	34.1	21.4%	94.5	59.4%

$\bar{d}$ = mean difference of the average strength measured by rater A and rater B. A positive value of $\bar{d}$ indicates that rater B obtained a higher mean strength measurement than rater A; $\lvert \bar{d} \rvert$/mean = the absolute mean difference divided by mean strength from rater A and rater B; LOA = limits of agreement = $\bar{d}$ ± 1.96 SD; SEM = standard error of measurement = $\sqrt{(Total\ variance)(1 - ICC)}$; SRD = smallest real differences = 1.96 × $\sqrt{2}$ × SEM; SEM% = SEM/mean strength; SRD% = SRD/mean strength. E1 = strength measurement of elbow flexor at the position of elbow flex 45°; E2 = strength measurement of elbow flexor at the position of elbow flex 90°; E3 = strength measurement of elbow extensor at the position of elbow flex 150°; E4 = strength measurement of elbow extensor at the position of elbow flex 90°; H1 = strength measurement of hip flexor at the position of hip flex 45°; H2 = strength measurement of hip flexor at the position of hip flex 90°; K1 = strength measurement of knee extensor at the position of knee flex 90°; K2 = strength measurement of knee extensor at the position of knee flex 150° (see **Figure 1**).

For absolute reliability, the SRD of each muscle ranged from 25.5 N to 94.5 N (**Table 4**). The SRD% of upper limb muscle groups was between 27.1% and 30.5%, with E4 having the smallest measurement and E3 having the largest. The SRD% of lower limb muscle groups was unsatisfactory, ranging from 33.8 to 59.4%, with H2 having the smallest measurement and K2 having the largest.

Table 5 showed the correlation and regression equation of extremity strength measured at different joint angles. Muscle strength of the same limb was different at different joint angles ($P \leq 0.0001$). Some positions were weak, such as E1, E4, H2 and K1, while others were strong, such as E2, E3, H1 and K2. However, there was a close correlation between strong and weak positions for measurements in the same limb. The ICCs for E1 and E2, E3 and E4, H1 and H2, K1 and K2 were 0.96, 0.83, 0.85 and 0.71 respectively. The regression equation ($P \leq 0.0001$) indicated linear correlation.

The results of Bland and Altman plots are shown in **Figures 2** and **3** for upper and lower

Table 5. The correlation and regression equation of extremity strength measured at different joint angle

Measurement	Paired t-test	ICC	Regression Equation	F-ratio	Significance level
E2 (y) to E1 (x)	$P < 0.0001$	0.96	y = 19.4778 + 1.0269x	216.9077	$P < 0.0001$
E3 (y) to E4 (x)	$P < 0.0001$	0.83	y = 41.4085 + 0.8953x	41.4868	$P < 0.0001$
H1 (y) to H2 (x)	$P < 0.0001$	0.85	y = 41.9443 + 0.9564x	51.580	$P < 0.0001$
K2 (y) to K1 (x)	$P = 0.0001$	0.71	y = 28.5616 + 1.2114x	27.0204	$P = 0.0001$

E1 = strength measurement of elbow flexor at the position of elbow flex 45°; E2 = strength measurement of elbow flexor at the position of elbow flex 90°; E3 = strength measurement of elbow extensor at the position of elbow flex 150°; E4 = strength measurement of elbow extensor at the position of elbow flex 90°; H1 = strength measurement of hip flexor at the position of hip flex 45°; H2 = strength measurement of hip flexor at the position of hip flex 90°; K1 = strength measurement of knee extensor at the position of knee flex 90°; K2 = strength measurement of knee extensor at the position of knee flex 150° (see **Figure 1**). ICC = intraclass correlation coefficient.

Table 6. Comparison of results with previous studies

Measurement	Average of Mean Measurements of Raters A and B (N)	Published Measurements	
		van der Ploeg et al. [27] (N)	Bohannon et al. [18] (N)
E2	127.2 (supine)	> 250 (sitting)	> 274.9 (sitting)
E4	94.0 (supine)	> 154 (sitting)	> 222.7 (sitting)
H2	124.7 (supine)	> 250 (sitting)	> 224.8 (sitting)
K1	107.8 (supine)	> 160 (prone)	> 572.7 (sitting)

E2 = strength measurement of elbow flexor at the position of elbow flex 90°; E4 = strength measurement of elbow extensor at the position of elbow flex 90°; H2 = strength measurement of hip flexor at the position of hip flex 90°; K1 = strength measurement of knee extensor at the position of knee flex 90°.

Table 7. Paired t-test results of left and right limb measurements in both raters

Measurement	Rater A					Rater B				
	Difference of means	SD	SEM	t-value	*P*-value	Difference of means	SD	SEM	t-value	*P*-value
E1	-1.667	8.475	1.849	-0.901	0.378	0.952	11.931	2.604	0.366	0.718
E2	-1.19	12.376	2.701	-0.441	0.664	12.238	16.428	3.585	3.414	0.633
E3	1.19	12.356	2.696	0.442	0.664	-0.667	15.177	3.312	-0.201	0.842
E4	3.952	7.826	1.708	2.314	0.631	0.333	7.895	1.723	0.193	0.849
K1	10.238	14.053	3.067	3.339	0.603	10.19	29.209	6.374	1.599	0.126
K2	-0.571	18.869	4.118	-0.139	0.891	2.143	19.003	4.147	0.517	0.611
H1	-0.952	23.579	5.145	-0.185	0.855	-8.19	21.009	4.584	-1.787	0.089
H2	6.19	22.754	4.965	1.247	0.227	-3.524	13.14	2.867	-1.229	0.233

Right and left limb measurement data obtained by both raters was subjected to the Kolmogorov-Smimov test to confirm normal distribution. T-values were interpreted at 20 degrees of freedom. Difference of means = Mean right limb measurement - mean left limb measurement. SD = standard deviation. SEM = Standard error of mean. E1 = strength measurement of elbow flexor at the position of elbow flex 45°; E2 = strength measurement of elbow flexor at the position of elbow flex 90°; E3 = strength measurement of elbow extensor at the position of elbow flex 150°; E4 = strength measurement of elbow extensor at the position of elbow flex 90°; H1 = strength measurement of hip flexor at the position of hip flex 45°; H2 = strength measurement of hip flexor at the position of hip flex 90°; K1 = strength measurement of knee extensor at the position of knee flex 90°; K2 = strength measurement of knee extensor at the position of knee flex 150° (see **Figure 1**).

limb muscle groups,respectively. The mean difference (d) between raters B and A was positive for upper limb muscle groups and negative for lower limb muscle groups, indicating that rater B obtained higher measurements for upper limb groups only. For 3 strength tests (E1, E3 and H2), there was no significant difference in performance between the 2 raters ($|d| \leq 4.6N$). For heteroscedasticity, larger variability between raters ($|d| \geq 7.4N$) was found for the other

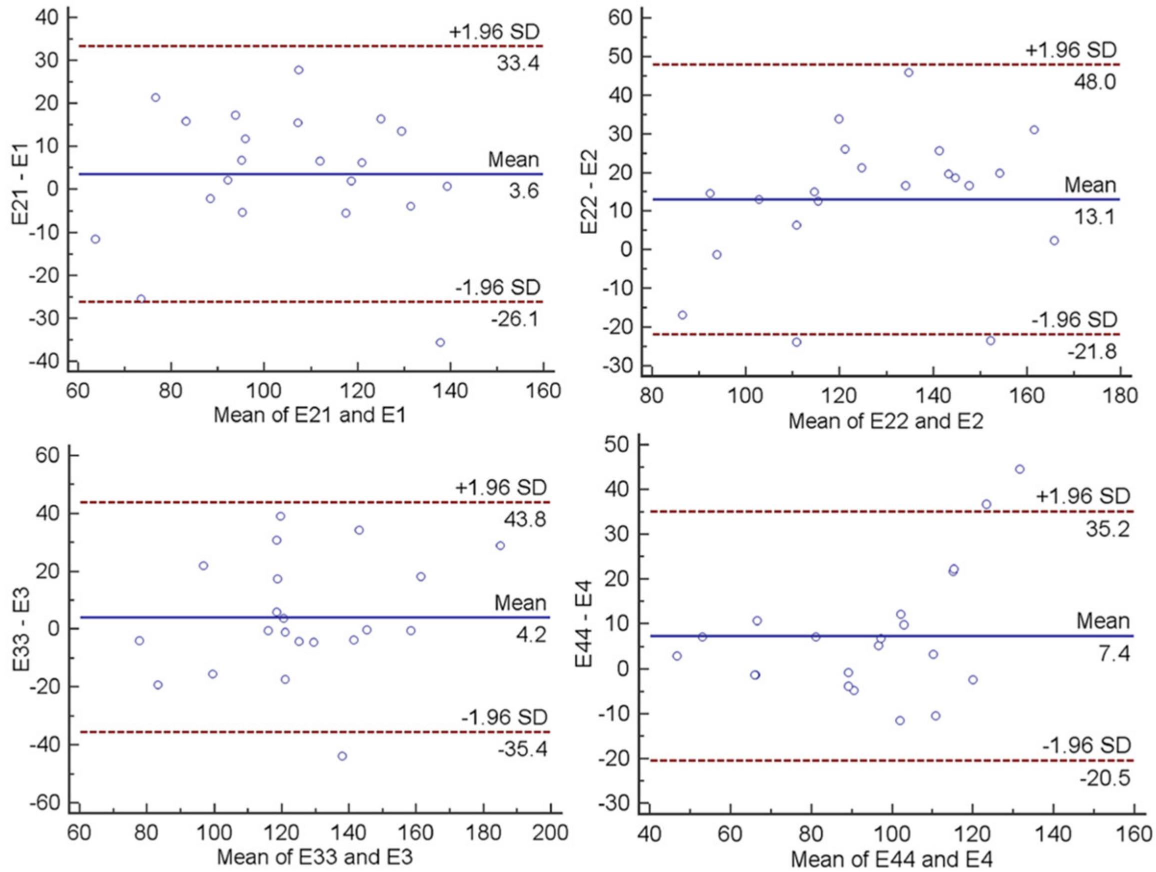

Figure 2. The results of Bland and Altman plots of theupper limb muscle groups. E21 and E1 = strength measurement of elbow flexor at the position of elbow flex 45° by rater B and rater A; E22 and E2 = strength measurement of elbow flexor at the position of elbow flex 90° by rater B and rater A; E33 and E3 = strength measurement of elbow extensor at the position of elbow flex 150° by rater B and rater A; E44 and E4 = strength measurement of elbow extensor at the position of elbow flex 90° by rater B and rater A. The positive value of mean difference ($\bar{d}$) indicated that the performance of rater B was greater than rater A.

5 positions (E2, E4, H1, K1 and K2). Only the E4 testing position showed evidence of systematic bias ($P < 0.05$, **Table 3**).

Discussion

Previous HHD studies [14-19, 22] have emphasized the importance of strong arm strength and clinical experience as prerequisites for testers. However, clinical usage of HHD is not common, and few people had the requisite experience to qualify as testers. Widespread and universal HHD application is contingent on whether beginners demonstrate a high reliability. For these reasons, we chose two learners as testers in this study.

Joint angle differences within the same limb were chosen according to the convenience for clinic. When comparing our results to previously published studies, we found that different testing position have a significant effect on measurement results. Sitting or standing positions are frequently used as postures in previous HHD studies and tend to result in higher measurement values [6-8, 20, 23-25]. However, we chose a supine position for our testing protocol. As a result, our mean measurement value for all joint angles tested was lower than those reported by van der Plogg et al [27] and Bohannon et al [18] (**Table 6**).

The supine position offers two distinct advantages over a sitting position. Firstly, since the sitting position gives larger strength values, it also has the highest requirements for rater arm strength and can make obtaining accurate measurements more difficult. A number of studies have proposed a variety of belt-stabiliz-

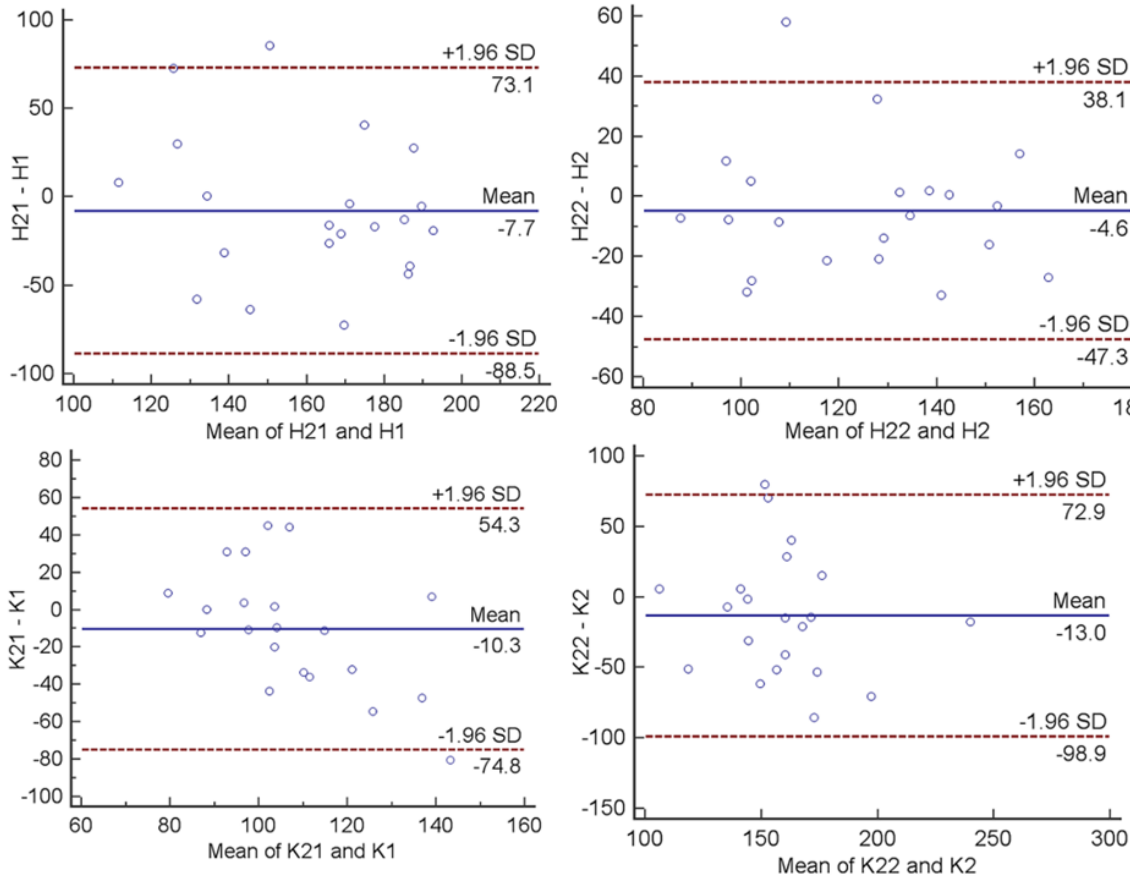

Figure 3. The results of Bland and Altman plots of the lower limb muscle groups. H21 and H1 = strength measurement of hip flexor at the position of hip flex 45° by rater B and rater A; H22 and H2 = strength measurement of hip flexor at the position of hip flex 90° by rater B and rater A; K21 and K1 = strength measurement of knee extensor at the position of knee flex 90° by rater B and rater A; K22 and K2 = strength measurement of knee extensor at the position of knee flex 150° by rater B and rater A. The negative value of mean difference ($\overline{d}$) indicated that the performance of rater B was smaller than rater A.

ing methods to reduce the difficulty of operation, including fixing the measurement tools to a seat [20], a wall [28], the edge of a bed, or a subject's limb [29]. While it is true that these fixing techniques improve the validity and reliability of measurements, they also require the subject to change positions over and over again. This is not convenient for clinical operations, and is especially unsuited for patients with spinal instability or back pain. Thus, belt-fixation methods are limited in the scope of their application.

Secondly, while it is easy for the able-bodied person to maintain a sitting or standing position, it can be exceedingly difficult for patients with spinal cord injuries. The supine position is advantageous because it is suitable for almost all types of patients. Since measuring limb strength is easier for both subjects and raters, the supine position has a wider range of clinical applications.

In accordance with Smidt et al, we considered a SRD% of less than 30% as acceptable [19, 21]. Burns and Spanier identified that a change of > 3.5 kg (34.3 N) likely represents a true change in strength rather than an error [9, 19].The only measurement positions that met both of these criteria were E1, E2, and E4. The other five (E3, H1, H2, K1, K2) were unsatisfactory. We found it was rather difficult to keep the hip or knee stable in the HHD measurements. Overall, the absolute reliabilities of the majority of positions measured were unreliable. Schrama et al [32] speculated that a possible reason for the higher reliability in the elbow is that the elbow can be better stabilized and substitute movements can be prevented. Our results support this finding, since the absolute reliabilities of upper limb measurements were greater overall. Future research could potentially focus on finding a joint angle of a tested limb that provides an even weaker strength measurement, and thus can be more easily stabilized.

We used the interclass coefficient as a measure of relative reliability. The upper limb muscle groups had high ICCs (≥ 0.74) that were deemed acceptable when compared to similar studies [5, 7, 9-12]. However, the ICCs of the lower limb muscle groups were unsatisfactory (≤ 0.62). Five of the testing positions (E2, E4, H1, K1, K2) had very large absolute mean differences in measurements between raters A and B (≥ 7.4N). Analysis by paired t-test showed that systematic bias only existed in the E4 testing position (P < 0.05, **Table 3**).

There was a significant association between muscle strength of the same limb measured at different joint angles (**Tables 3**-**5**). E1, E4, H2, and K1 were weaker positions, while E2, E3, H1, and K2 were stronger positions. ICC showed a close correlation between the strong and weak positions in all muscle groups tested: E1 and E2 were 0.96, E3 and E4 was 0.83, H1 and H2 was 0.85, K1 and K2 was 0.71 (**Table 5**). Furthermore, reliability was significantly increased when the weak muscle position (E1, E4, H2, or K1) was tested. Since there is such a close relationship between strong and weak positions in the same limb, the change in strength at the weak position can be used to speculate the change in strength at the strong position. These results can be extrapolated to determine the change in strength of the entire limb.

In clinical application of HHD, there are noteworthy limitations with movements where subjects can overpower the testers [30]. Wikholm et al [31] suggested that tester strength and reliability of HHD measurements are closely related. Interrater ICCs decreased in magnitude for muscle groups that produced larger amounts of force. Testing for muscle groups that generated a maximum force of 120 N were reliable when clinicians had strengths that were at least equal to that of the subject. For measurement values above 120 N, tester strength appeared to be a major determinant of the magnitude and reliability of forces measured with an HHD. In this study, the joint angle measurements with the highest ICCs (E1 and E4) also had mean force measurements of less than 120 N, verifying this claim.

Rater B consistently obtained larger muscle strength measurements than rater A for upper limbs, and smaller muscle strength measurements than rater A for lower limbs. The greater variability in rater B's measurements can likely be attributed to the interplay between experience and insufficient upper body strength. As an experienced nurse, rater B showed more mastery of the testing protocol than rater A, an intern, and obtained higher readings in upper limb tests. However, human lower limbs are typically stronger than human upper limbs, meaning that upper limb strength of the tester is especially important when obtaining lower limb measurements. Bohannon demonstrated that young females typically have an upper limb strength that is roughly half that of young males [18]. Although she was the strongest registered nurse in our department, rater B had an upper limb strength that was roughly half of rater A's (**Table 1**). The lower limb measurements obtained by rater B were consistently smaller than the measurements obtained by rater A, likely due to the stark difference in tester strength.

Physical therapy is becoming an increasingly common profession for women. For example, over half of the 197 therapists in the China Rehabilitation Research Center at the end of 2013 were women. Female therapists may become tired or exhausted when measuring lower limb muscle strength of male patients, resulting in lower reliability. Measuring muscle strength at the weak position rather than the strong position could conserve the energy of the tester and increase reliability.

Learning effects in the second test, insufficient recovery between tests, inherent biological or mechanical variation, and inconsistencies in the testing protocol are common causes of measurement errors. Our study has two obvious limitations. Firstly, systematic bias was found in one of the testing positions. The learners used require more clinical experience with HHD in order to control posture in a consistent way and standardize muscle strength measurements. Secondly, the sample size used was small. A larger sample size (n = 40) is recommended [26].

Practical applications

Accurate muscle strength assessment in a clinical setting is contingent on strong upper limb strength of the examiner when compared to the subject. Given the large range in tester strength, changing the testing joint angle of the limb to the weaker muscle position can significantly improve the reliability of handheld dynamometer measurements. More accurate assessments are obtained when the elbow flexor is tested at the position of elbow flex 45° instead of 90°, the elbow extensor at the position of flex 90° instead of 150°, the hip flexor at the position of flex 90° instead of 45°, and the knee extensor at the position of flex 90° instead of 135°.

Acknowledgements

This study was supported by China Rehabilitation Research Center (project 2012-CZ-31).

Disclosure of conflict of interest

None.

Abbreviations

HHD, handheld dynamometer; MVCs, Maximal voluntary contractions; ICCs, intraclass correlation coefficients; SRD, smallest real differences; SCI, spinal cord injury.

Address correspondence to: Dr. Jianjun Li, Department of Spinal and Neural Function Reconstruction, China Rehabilitation Research Center, Capital Medical University School of Rehabilitation Medicine,Jiaomenbeilu, Fengtai District, Beijing 100068, China. Tel: +86-10-67563322-3113; Fax: +86-10-67573428; E-mail: crrclijj@126.com

References

[1] Kendall HO, Kendall FP. Care during recovery period in paralytic poliomyelitis. US Public Health Service. Bulletin 242. Rev 1939; 1-96.

[2] Daniels L, Williams M, Worthingham C. Muscle Testing. Techniques of Manual Examination. Philadelphia, PA: W B Saunders Co; 1947.

[3] Lovett RW, Martin EG. The spring balance muscle test. Am J Orthop Surg 1916; 214: 415-24.

[4] Beasley WC. Influence of method on estimates of normal knee extensor force among normal and postpolio children. Phys Ther Rev 1956; 36: 21-41.

[5] Bohannon RW, Andrews AW. Interrater Reliability of Hand-Held Dynamometry. Phys Ther 1987; 67: 931-3.

[6] Sisto SA, Dyson-Hudson T. Dynamometry testing in spinal cord injury. J Rehabil Res Dev 2007; 44: 123-36.

[7] Beasley WC. Quantitative muscle testing: principles and applications to research and clinical services. Arch Phys Med Rehabil 1961; 42: 398-425.

[8] Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. Phys Ther 1996; 76: 248-59.

[9] Burns SP, Spanier DE. Break-technique handheld dynamometry: Relation between angular velocity and strength measurements. Arch Phys Med Rehabil 2005; 86: 1420-26.

[10] Schwartz S, Cohen ME, Herbison GJ, Shah A. Relationship between two measures of upper extremity strength: manual muscle test compared to hand-held myometry. Arch Phys MedRehabil 1992; 73: 1063-8.

[11] Agre JC, Magness JL, Hull SZ, Wright KC, Baxter TL, Patterson R, Stradel L. Strength testing with a portable dynamometer: reliability for upper and lower extremities. Arch Phys Med Rehabil 1987; 68: 454-8.

[12] Kilmer DD, McCrory MA, Wright NC, Rosko RA, Kim HR, Aitkens SG. Strength measurement using hand-held dynamometry. Washington (DC): U.S. Department of Education, National Institute on Disability and Rehabilitation Research, 1997.

[13] Bohannon RW. Test-retest reliability of handheld dynamometry during a single session ofstrength assessment. Phys Ther 1986; 66: 206-9.

[14] Whitley JD, Smith LE. Larger correlations obtained by using "average" rather than "best"strength scores. Research Quarterly 1963; 34: 248-9.

[15] Bohannon RW. Make tests and break tests of elbow flexor muscle strength. Phys Ther 1988; 68: 193-4.

[16] Bohannon RW. Shoulder position influences elbow extension force in healthy individuals. JOSPT 1990; 12: 111-4.

[17] Bohannon RW. Measuring knee extensor muscle strength. Am J Phys Med Rehabil 2001; 80: 13-8.

[18] Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil 1997; 78: 26-32.

[19] Lu YM, Lin JH, Hsiao SF, Liu MF, Chen SM, Lue YJ. The relative and absolute reliability of leg muscle strength testing by a handheld dynamometer. J Strength Cond Res 2011; 25: 1065-71.

[20] Bohannon RW, Bubela DJ, Wang YC, Magasi SR, Gershon RC. Adequacy of belt-stabilized testing of knee extension strength. J Strength Cond Res 2011; 25: 1963-7.

[21] Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Devillé WL, de Winter AF, Bouter LM. Interobserver reproducibility of the assessment of severity of complains, gripstrength, and pressure pain threshold in patients with lateral epicondylitis. Arch Phys Med Rehabil 2002; 83: 1145-50.

[22] Krause DA, Neuger MD, Lambert KA, Johnson AE, DeVinny HA, Hollman JH. Effects of examiner strength on reliability of hip-strength testing using a handheld dynamometer. J Sport Rehabil 2014; 23: 56-64.

[23] Newman BL, Pollock CL, Hunt MA. Reliability of Maximal Isometric Lateral Trunk Flexion Strength Measurement in Athletes Using Hand-held Dynamometry. J Sport Rehabil 2012; Technical Notes (6).

[24] Geary K, Green BS, Delahunt E. Intrarater reliability of neck strength measurement of rugby union players using a handheld dynamometer. J Manipulative Physiol Ther 2013; 36: 444-9.

[25] Shahidi B, Johnson CL, Curran-Everett D, Maluf KS. Reliability and group differences in quantitative cervicothoracic measures among individuals with and without chronic neck pain. BMC Musculoskelet Disord 2012; 13: 215.

[26] Altman DG. Practical Statistics for Medical Research. London, United Kingdom: Chapmanand Hall; 1991.

[27] Van der Ploeg RJO, Fidler V, Oosteruis HJ. Hand-held myometry: reference vaues. J Neurol Neurosurg Psychiatry 1991; 54: 244-247.

[28] Thorborg K, Bandholm T, Holmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. Knee Surg Sports Traumatol Arthrosc 2013; 21: 550-555.

[29] Katoh M. Comparison of reliability of isometric leg muscle strength measurements made using a hand-held dynamometer with and without a restraining belt. J Phys Ther Sci 2009; 21: 37-42.

[30] Kelln BM, McKeon PO, Gontkof LM, Hertel J. Hand-Held Dynamometry: Reliability of Lower Extremity Muscle Testing in Healthy, Physically Active, Young Adults. J Sport Rehabil 2008; 17: 160-170.

[31] Wikholm JB, Bohannon RW. Hand-held Dynamometer Measurements: Tester Strength Makes a Difference. J Orthop Sports Phys Ther 1991; 13: 191-198.

[32] Schrama PPM, Stenneberg MS, Lucas C, van Trijffel E. Intraexaminer Reliability of Hand-Held Dynamometry in the Upper Extremity: A Systematic Review. Arch Phys Med Rehabil 2014; 95: 2444-69.

[33] Kolber MJ, Cleland JA. Strength testing using hand-held dynamometry. Physical Therapy Review 2005; 10: 99-112.