Relative Validity and Reliability of Isometric Lower Extremity Strength Assessment in Older Adults by Using a Handheld Dynamometer

Sports Health

Grootswagers, Pol; Vaes, Anouk M.M.; Hangelbroek, Roland; Tieland, Michael; Loon, Luc J.C. et al

https://doi.org/10.1177/19417381211063847

Relative Validity and Reliability of Isometric Lower Extremity Strength Assessment in Older Adults by Using a Handheld Dynamometer

Pol Grootswagers, PhD,*† Anouk M. M. Vaes, PhD,† Roland Hangelbroek, PhD,† Michael Tieland, PhD,† Luc J. C. van Loon, PhD,‡ and Lisette C. P. G. M. de Groot, PhD†

Background: Handheld dynamometry (HHD) is a practical alternative to traditional testing of lower extremity strength. However, its reliability and validity across different populations and settings are not clear.

Hypothesis: We hypothesize that HHD is a valid and reliable device to assess lower extremity strength in a population of older adults.

Study Design: Cross-sectional/cohort.

Level of Evidence: Level 3.

Methods: This study included 258 older adults (≥65 years). Isometric knee extension and flexion force were measured by 1 examiner, using an HHD (n = 222), including 3 repetitions to calculate within-day intrarater reliability. These measurements were repeated by the examiner in a subgroup (n = 23) to analyze intrarater reliability over a test-retest period of on average 8 weeks. In addition, HHD force measures were performed by a second examiner (n = 29) to analyze interrater reliability. In another subgroup (n = 77), isometric knee extension and flexion torque were measured by 1 examiner using both the HHD and Biodex System 4 to assess relative validity.

Results: HHD and Biodex measurements were highly correlated and showed excellent concurrent validity. HHD systematically overestimated torque as compared with Biodex by 8 N·m on average. Same-day intrarater intraclass correlation coefficients (ICCs) ranged from 0.97 to 0.98. Interrater reliability ICCs ranged from 0.83 to 0.95.

Conclusion: HHD represents a reliable and valid alternative to Biodex to rank individuals on leg strength, or to assess within-person changes in leg strength over time, because of the high validity and reliability. The HHD is less suited for absolute strength assessment because of significant systematic overestimations.

Clinical Relevance: Clinicians are encouraged to use HHD to rank older adults on leg strength, or to assess within-person changes in leg strength over time, but not to compare readings with cut-offs or normative values.

Keywords: leg strength; accuracy; Biodex; MicroFET; intrarater reliability; interrater reliability

During the aging process, progressive reduction in skeletal muscle strength occurs. Reduced muscle strength, defined by maximum force, is a predictor for functional outcomes, independent of loss of muscle mass.[15,23] The age-related decline in muscle strength is more pronounced in the legs than in the arms.[18] Still, strength in older individuals is often assessed via handgrip strength, a measurement limited by the presence of common arthritis-like conditions,[4] and poorly able to detect changes in strength over time.[21] Therefore, assessment of lower extremity strength in older adults is

From †Division of Human Nutrition, Wageningen University, Wageningen, the Netherlands, and ‡Department of Human Biology, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, the Netherlands

*Address correspondence to Pol Grootswagers, MSc, Division of Human Nutrition, Wageningen University, PO Box 17, Wageningen, 6700 AA, the Netherlands (email: pol.grootswagers@wur.nl).

The authors report no potential conflicts of interest in the development and publication of this article.

The original project was funded by TI Food and Nutrition, a public-private partnership on precompetitive research in food and nutrition. The public partners are responsible for the study design, data collection and analysis, decision to publish, and preparation of the manuscript. The private partners have contributed to the project through regular discussion.

DOI: 10.1177/19417381211063847

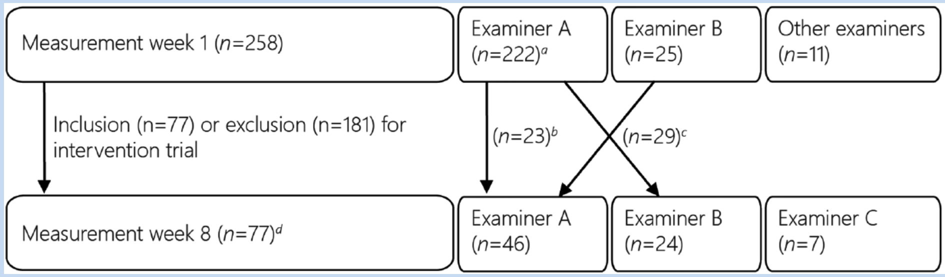

Figure 1. Flowchart of study design and origin of subsamples. Subsamples used for analysis: [a]same-day intrarater; [b]8-week interval intrarater; [c]interrater; [d]relative validity.

increasingly recognized as important.[11] However, traditional testing of lower extremity strength requires costly equipment (eg, Biodex system) and experienced staff (eg, physical therapists). Isometric leg strength testing using a handheld dynamometer (HHD) may offer an alternative as it is less costly and can be applied in various clinical settings.

Several studies have evaluated the accuracy of the HHD to assess leg strength in older people.[1,2,5,9,12,16,24] However, these studies only used a limited sample of n ≤ 25 subjects,[2,5,12] or did not compare measurements against a gold standard method.[1,9,16,24] Therefore, the purpose of this study was to assess the validity and reliability of HHD measurements in a large population of older adults. Validity was compared against the Biodex system 4 dynamometer. Reliability was not only assessed by same-day retests but also by retests 8 weeks apart to understand how larger intervals between measurements (without intervention) affect reliability.

METHODS

Study Population

This study presents an analysis of screening and baseline data collected for an intervention trial, registered at ClinicalTrials.gov as NCT02349282. In that trial,[22] isometric knee flexion and extension strengths were measured by HHD in 500 community-dwelling older adults, during eligibility screening. Eligibility screening was attended by older adults (≥65 years) who estimated their strength to be below average. Of these participants, 32.4% (n = 84) had a hand-grip strength below the cut-off level for frailty.[6] For activity and gait speed, respectively 7.3% (n = 19) and 3.5% (n = 9) scored below the frailty cut-off level. Criteria for subsequent inclusion in the intervention were age ≥65 years, body mass index of 18.5 to 30 kg/m^2, vitamin D status 20 to 50 nmol/L, physically frail or prefrail based on criteria of Fried et al,[6] free of medical conditions or medication use interfering with vitamin D metabolism, and consuming a maximum of 21 alcoholic units per week.

We used data of all participants that were eligible for the intervention trial (n = 77), plus data of the largest subsample of participants who were measured by 1 examiner during the screening session (n = 222, of which n = 41 are also included in the n = 77), resulting in a total of n = 258 participants. The 77 participants eligible for the intervention trial attended a second measurement day, where isometric knee flexion and extension strengths were measured by HHD and Biodex. Only data obtained before the interventions started have been used for this study.

Between December 2014 and June 2015, measurements were carried out at Wageningen University and Hospital Gelderse Vallei in Ede, the Netherlands. The Medical Ethics Committee of Wageningen UR approved the study protocol and all participants gave their written informed consent.

Examiners

Measurements by 3 examiners were used for the analyses. Examiners A and B were both males, examiner C was female. None of the examiners had prior experience with HHD measurements but were trained to perform the standardized operation procedure before the first measurements. Examiners trained with each other to ensure that a balanced position was found, and that all elements of the measurement, such as position and means of encouragement, were similar between examiners.

Figure 1 presents an overview of the measurements that were used for the analyses. Examiner A measured knee extension and flexion strengths in 222 participants on measurement day 1. These measurements were used to calculate same-day intrarater reliability. Of these 222 participants, 23 were measured by the same examiner on measurement day 2, and this subsample was used to calculate Δ8-week intrarater reliability. The calculations of the interrater reliability were based on a sample of 29 participants who were measured by examiner A on day 1 and examiner B on day 2, or vice versa. Finally, relative validity was assessed by using the measurements of all 77 participants who attended measurement day 2. These measurements were carried out by examiners A, B, and C.

Measurements

Handheld Dynamometer

Peak isometric knee flexion and extension forces were used as the strength measures. Force was assessed in newton (N), using the MicroFET 2 handheld dynamometer (Hoggan Health). The subjects were asked to sit upright on an examination table with their knees at a 90° angle. Peak force was measured 3 times per leg with 5 seconds of muscle contraction and 60 seconds of rest between repetitions. A "make test" was used, as literature shows that this test results in greater reliability than a "break test."[8,20] Subjects were instructed to gradually increase their force when the examiner counted "3-2-1, go" to a peak force until the examiner instructed to stop. All participants started with 1 rehearsal measurement to ensure full understanding of instructions. Consistent verbal encouragement was provided by the examiner during the 5 seconds of muscle contraction. Stabilization of examiner and HHD position was ensured by measuring while seated against a wall and by offering counter force with both hands (a visual of the setup can be found in the Supplemental Material available in the online version of this article). On measurement day 2, tests were performed on the Biodex chair instead of an examination table. Besides this difference, other conditions were equal, with participants being seated without the use of chair straps, at an equal seat-to-floor distance, and at an equal distance of the wall that was used to support the examiner.

Biodex Dynamometer

In a subgroup of 77 subjects, peak isometric knee flexion and extension torque were assessed in newton meter (N·m) using the Biodex System 4 (Biodex Medical Systems). Subjects performed the test in a sitting position while holding the handgrips. Chair straps were used to stabilize the subjects to prevent accessory movements. The dynamometer was positioned with the lever arm immediately adjacent to the participant's lower leg with the axis of the dynamometer aligned with the lateral condyle of the knee and flexed at a 60° angle. Peak torque of flexion and extension was measured 3 times per leg with 5 seconds of muscle contraction and 60 seconds of rest between repetitions. Consistent verbal encouragement and visual feedback on the computer monitor were given to the subjects by the examiners.

Anthropometrics

Body weight was measured using a calibrated analogue scale (SECA 716), without wearing heavy clothing. Weight was reported to the nearest 0.5 kg. Total height was measured to the nearest 0.1 cm using a stadiometer (SECA 213). Lever length was measured in the validity subsample, using a measuring tape, as the distance in meters between the lateral condyle of the femur and the lateral malleolus of the right leg.

Statistical Methods

Participants' characteristics are reported as mean with standard deviation (SD), or percentages of the total study group. Validity of HHD was assessed by comparison of readings with Biodex. Pearson correlation coefficients were used to test for validity of peak readings, and intraclass correlation coefficients (ICCs) were calculated to assess absolute validity between multiple measurements. Bland-Altman plots were created to assess the agreement between HHD and Biodex and to inspect possible strength-related bias. The significance of the difference between HHD and Biodex measurements was assessed by paired-sample *t*-tests, and between-examiner differences were inspected. To compare HHD and Biodex measurements, HHD force values were multiplied by lever length to convert into torque values. The peak torque of the 3 attempts was used for all analyses except for the same-day intrarater reliability assessment. To assess the intra- and interrater reliabilities of the HHD method, ICCs(2,1) and standard error of measurement (SEM) were calculated. ICCs were calculated via a 2-way random effect model, and the SEM was expressed as the product of the SD of the first HHD measurements and the square root of (1 − ICC). When ICCs were less than 0.5, reliability was considered "poor," between 0.5 and 0.75 "moderate," between 0.75 and 0.9 "good," and greater than 0.9 "excellent." SEMs are indicative of an expected variability in repeated measures. Data analyses were performed with the statistical program SPSS, Version 23 (IBM Corp). Graphs were created using GraphPad Prism 5 (GraphPad Software Inc).

RESULTS

Patients' characteristics are presented in Table 1. The mean ± SD age of the total study population was 73.0 ± 5.9 years and 58.3% were men. Mean body mass index was 27.3 ± 3.6 kg/m^2. The mean peak knee flexion and extension forces, measured by HHD were 205 ± 70 and 345 ± 108 N, respectively. The mean lever length in the group that attended the second measurement day was 0.42 ± 0.03 m, and mean peak torques as measured by Biodex were 66 ± 21 N·m for knee flexion and 130 ± 41 N·m for knee extension.

Validity

The HHD and Biodex measurements were highly correlated (Table 2, Figure 2). The Bland-Altman plots in Figure 3 show the difference between the Biodex and HHD outcomes. Overall, the HHD measurements significantly overestimated torque as compared with the Biodex (Table 2), but Bland-Altman plots revealed no strength-related bias. Differences between HHD and Biodex measurements were more pronounced in the right leg and in the knee flexion measurements (Table 2). The magnitude of torque overestimation by HHD differed between examiners.

HHD measurements of examiner A (in n = 45) overestimated torque on average by 9.2% (95% CI 3.4%-15.0%, $P = 0.002$), and HHD measurements of examiner B (in n = 24) overestimated torque on average by 18.5% (95% CI 12.7%-24.3%, $P < 0.001$). This difference in overestimation was not statistically significant ($P = 0.18$). The HHD measurements of examiner C (in n = 7) underestimated Biodex measurements by 6.1% (95% CI −36.2% to 23.9%, $P = 0.64$), which was significantly different from examiner B ($P = 0.01$), but not from examiner A ($P = 0.16$).

Table 1. Participants' characteristics

	Total Population (n = 259)	Validity Subsample (n = 77)
Demographics		
Gender, male, % (n)	58.3 (151)	55.1 (43)
Age, y, mean ± SD	73.0 ± 5.9	74.1 ± 6.3
Body mass index, kg/m^2, mean ± SD	27.3 ± 3.6	27.9 ± 3.6
Lever length, m, mean ± SD	—	0.42 ± 0.03
Biodex torque, N·m (right leg), mean ± SD		
Peak knee flexion	—	66 ± 21
Peak knee extension	—	130 ± 41
MicroFET force, N (right leg), mean ± SD		
Peak knee flexion	205 ± 70	181 ± 56
Peak knee extension	355 ± 108	334 ± 109

Table 2. Means, correlation, and difference between torque measures obtained by Biodex and MicroFET in older adults (n = 77)

	Biodex, Mean (SD)	MicroFET, Mean (SD)	ICC (95% CI)	Pearson's *r*	Difference (95% CI)
Peak knee flexion, N·m					
Right	65.6 (21.2)	75.5 (25.2)	0.79 (0.57-0.89)	0.73*	9.9 (6.0-13.9)*
Left	63.5 (21.3)	70.2 (24.5)	0.88 (0.77-0.93)	0.82*	6.6 (3.5-9.8)*
Peak knee extension, N·m					
Right	130.3 (41.2)	140.0 (50.0)	0.85 (0.76-0.91)	0.77*	9.7 (2.4-17.0)*
Left	130.1 (45.1)	135.8 (54.1)	0.94 (0.90-0.96)	0.90*	5.7 (0.25-11.2)*

ICC, intraclass correlation coefficient.
*$P < 0.05$.

Reliability

The intrarater reliability of the HHD measurements performed by one examiner is presented in Table 3. Same-day test-retest resulted in ICCs between 0.97 and 0.98. Peak force measures performed 54 ± 9 days apart resulted in somewhat lower ICCs, with 0.90 for the flexion measurements and 0.96 for the extension measurements. Table 3 also shows the interrater reliability of the HHD measurements. On average, examiner B tended to measure greater force output than examiner A, with a statistically significant difference observed in left knee extension (24 N, $P = 0.03$). The ICCs between the tests performed by the examiners were higher for the extension measures (ICC 0.93 and 0.95) compared with the flexion measures (ICC 0.83 and 0.89).

DISCUSSION

In this study, we tested the relative validity of knee flexion and extension torque measured by HHD in a convenient and replicable test setup as compared with Biodex dynamometry in a large sample size. We also assessed the intra- and interrater reliabilities of isometric knee flexion and extension force assessments by HHD.

We observed strong correlations between HHD and Biodex torque measurements. However, HHD significantly overestimated torque, which was most pronounced for isometric knee flexion measurements. This overestimation is unexpected, since the greater measurement angle of the leg (90° in HHD vs 60° in Biodex), and the absence of trunk stabilization straps

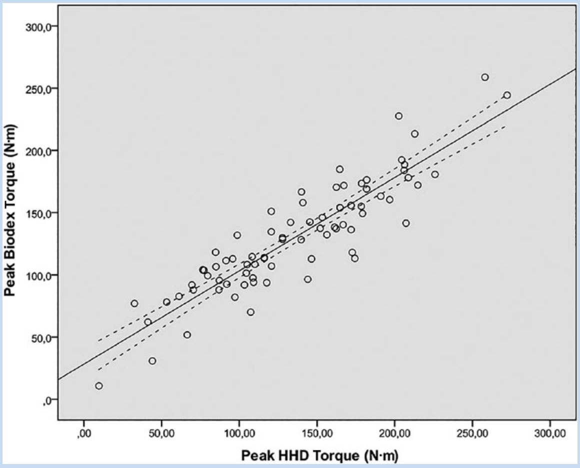

Figure 2. Correlation between Biodex and handheld dynamometer (HHD) measurements of left knee extension peak torque ($r = 0.90$, n = 77). Line of best fit (Biodex = 28.3 + 0.75 × HHD) with 95% confidence interval.

during HHD measurements, have both been reported to decrease peak torque.[7,17] Previously performed studies comparing Biodex and HHD have indeed found lower values for HHD measurement.[2,12] These studies used HHD devices (Lafayette Instrument Inc) that were a different manufacturer than ours (MicroFET, Hoggan Health), which may account for a difference. However, in young adults, a comparison between these 2 devices did not suggest that such an interdevice difference exists.[14] Alternatively, the peak force of our participants was more than 1.5 times higher compared with their participants, which potentially may have played a role in our different findings. The overestimation of torque measured by HHD was systematic, and equal over the range of measures levels of torque. This suggests that calibration of HHD readings to Biodex readings should be possible, and that caution is warranted when HHD readings are being compared with cut-off values based on Biodex measurements.

The overestimation of torque measured by HHD was more explicit for knee flexion than for knee extension measurements. Knee flexion measurements also scored worse than knee extension on interrater reliability, and on intrarater reliability when measured 8 weeks apart. This is in line with the study of Martins et al,[13] which found higher validity and intrarater validity for knee extension than for knee flexion in a young population. It is yet unclear which mechanisms account for the difference between flexion and extension. Speculatively, fatigue-related differences do not play a role as intrarater reliability differences are only apparent when measured 8 weeks apart, and they are not with same-day measurements. That suggests that output consistency plays a role,

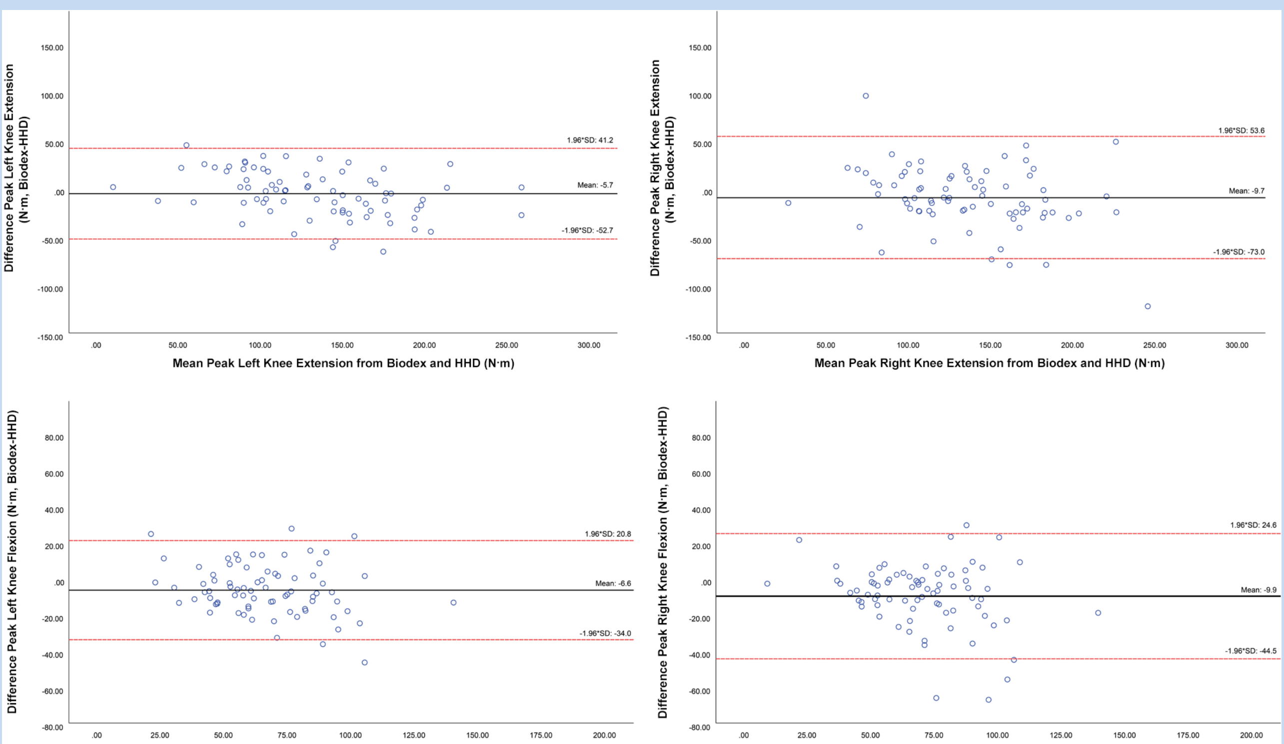

Figure 3. Bland-Altman plots of Biodex versus MicroFET measurements (n = 77). Note that the scales are standardized across both legs for flexion and extension assessments. HHD, handheld dynamometer.

Table 3. Intrarater reliability for force measures obtained by MicroFET (same-day test-retest period, n = 222, 8-week test-retest period, n = 23) and interrater reliability for force measures obtained by MicroFET in older adults (2 examiners, n = 29).

Same-Day Intrarater	Trial 1, Mean ± SD	Trial 2, Mean ± SD	Trial 3, Mean ± SD	ICC (95% CI)	SEM
Knee flexion, N					
Right	194 ± 70	200 ± 67	201 ± 70	0.98 (0.98-0.99)	10
Left	179 ± 68	183 ± 65	186 ± 64	0.98 (0.97-0.98)	10
Knee extension, N					
Right	347 ± 105	343 ± 104	344 ± 105	0.98 (0.97-0.98)	15
Left	315 ± 106	314 ± 101	314 ± 102	0.97 (0.97-0.98)	18
Δ8-Week Intrarater	**Trial Week 0, Mean ± SD**	**Trial Week 8, Mean ± SD**		**ICC (95% CI)**	**SEM**
Peak knee flexion, N					
Right	183 ± 53	173 ± 46		0.90 (0.78-0.96)	22
Left	167 ± 55	161 ± 46		0.90 (0.76-0.96)	22
Peak knee extension, N					
Right	330 ± 117	332 ± 102		0.96 (0.91-0.98)	21
Left	297 ± 120	312 ± 113		0.96 (0.90-0.98)	21
Δ8-Week Interrater	**Examiner A, Mean ± SD**	**Examiner B, Mean ± SD**	**Difference (95% CI)**	**ICC (95% CI)**	**SEM**
Peak knee flexion, N					
Right	199 ± 64	198 ± 50	–1 (–18.3 to 15.6)	0.83 (0.63-0.92)	29
Left	179 ± 65	189 ± 53	11 (–3.0 to 25.2)	0.89 (0.76-0.95)	23
Peak knee extension, N					
Right	331 ± 111	342 ± 114	11 (–10.5 to 33.0)	0.93 (0.85-0.97)	28
Left	317 ± 131	341 ± 136	24 (2.1-45.6)*	0.95 (0.87-0.98)	24

ICC, intraclass correlation coefficient; SEM, standard error of measurement.
*$P < 0.05$.

which may be influenced by neurological activation of the muscle group. It is important to note that although knee flexion assessment performs worse than knee extension assessment, good validity (ICC 0.79-0.88), good interrater reliability (ICC 0.83-0.89), and excellent intrarater reliability (ICC 0.90-0.98) were observed. The SEMs were around 5% of the mean for same-day intrarater tests, without clear differences between flexion and extension. For 8-week intra- and inter-rater, the SEMs were around 13% of the mean for flexion, and 7% for extension. These values are comparable with the better-performing studies reported in the review by Chamorro et al.[3]

Despite the systematic overestimation, the HHD measurements can still be used to rank participants on their strength level, for instance, in the case of a cross-sectional study, because of the good to excellent reliability (depending on inter- or intrarater). Also, because of the systematic nature of the overestimation, it should not bias within-person strength changes over time, especially when the assessment is performed by the same examiner. After all, we found excellent intrarater reliability of HHD measurements with an 8-week break in between (ICCs 0.90-0.96), even though the test conditions were slightly different. This is a strength of our study, as this has not often been investigated before. In a similar population, Schaubert and Bohannon[19] also found excellent intrarater reliability (ICCs 0.92-0.93) for HHD measurements separated by a 6-week period. Intrarater reliability of HHD measurements on the same day was even higher (ICCs 0.97-0.98). These same-day values are comparable with those found by Wang et al[24] (0.98-0.99) but higher than values reported by Katoh and Isozaki[9] (0.88-0.91). Katoh and Isozaki[9] did not include a familiarization trial and performed only 2 HHD measurements per leg. The lack of a familiarization trial forms a likely explanation of lower intrarater reliability and indicates the importance of including such a familiarization trial for HHD measurements.

We observed good to excellent interrater reliability for the assessment of knee flexion and extension force. Interrater reliability was calculated from peak measurements that were 8 weeks apart, which likely results in an underestimation of interrater reliability measured with shorter between-measurement intervals. Despite the good to excellent interrater reliability, we did find differences between examiners. Examiner B measured higher forces with HHD than examiner A, which reached significance for left knee extension force. Compared with Biodex readings, torque overestimation was twice as high for HHD measurements performed by examiner B compared with those performed by examiner A. Both examiners were similarly trained, used the same technique, were similarly experienced with HHD measurements, were both male, and were of similar age. The difference in measured HHD force between these 2 examiners might be due to factors beyond these similarities, such as strength of the examiner.[10]

Significant limitations should be considered when interpreting the results. First, HHD testing positions on measurement days 1 and 2 were slightly different. Likely, this led to lower reliability. Second, despite allowing short periods of rest between tests, fatigue may have occurred on the measurement days, resulting in lower readings over time, and thus lower reliability. Third, the strength of the examiners is likely related to the extent of overestimation. Fourth, examiners were not blinded to test results, which may have introduced bias. Finally, knee angles for HHD and Biodex were different (90° vs 60°), which might have had an effect on validity results.

In conclusion, HHD performed by well-trained examiners in a stable position represents a reliable and valid alternative but is not equivalent to isokinetic testing to assess lower extremity strength in an older population. Comparing HHD readings with isokinetic testing cut-offs or normative values should be done with caution because of systematic overestimation.

REFERENCES

1. Abizanda P, Navarro JL, Garcia-Tomas MI, Lopez-Jimenez E, Martinez-Sanchez E, Paterna G. Validity and usefulness of hand-held dynamometry for measuring muscle strength in community-dwelling older persons. *Arch Gerontol Geriatr.* 2012;54:21-27.
2. Arnold CM, Warkentin KD, Chilibeck PD, Magnus CR. The reliability and validity of handheld dynamometry for the measurement of lower-extremity muscle strength in older adults. *J Strength Cond Res.* 2010;24:815-824.
3. Chamorro C, Armijo-Olivo S, De la Fuente C, Fuentes J, Javier Chirosa L. Absolute reliability and concurrent validity of hand held dynamometry and isokinetic dynamometry in the hip, knee and ankle joint: systematic review and meta-analysis. *Open Med (Warsaw, Poland).* 2017;12:359-375.
4. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Writing Group for the European Working Group on Sarcopenia in Older People 2 (EWGSOP2), the Extended Group for EWGSOP2. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing.* 2019;48:16-31.
5. Dunn JC, Iversen MD. Interrater reliability of knee muscle forces obtained by hand-held dynamometer from elderly subjects with degenerative back pain. *J Geriatric Phys Ther.* 2003;26:23-29.
6. Fried LP, Tangen CM, Walston J, et al. Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci.* 2001;56:M146-M156.
7. Hart DL, Stobbe TJ, Till CW, Plummer RW. Effect of trunk stabilization on quadriceps femoris muscle torque. *Phys Ther.* 1984;64:1375-1380.
8. Jeon I-C. Comparisons of test-retest reliability of strength measurement of gluteus medius strength between break and make test in subjects with pelvic drop. *J Korean Phys Ther.* 2019;31:147-150.
9. Katoh M, Isozaki K. Reliability of isometric knee extension muscle strength measurements of healthy elderly subjects made with a hand-held dynamometer and a belt. *J Phys Ther Sci.* 2014;26:1855-1859.
10. 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.
11. Manini TM, Clark BC. Dynapenia and aging: an update. *J Gerontol A Biol Sci Med Sci.* 2012;67:28-40.
12. Martin HJ, Yule V, Syddall HE, Dennison EM, Cooper C, Aihie Sayer A. Is hand-held dynamometry useful for the measurement of quadriceps strength in older people? A comparison with the gold standard Bodex dynamometry. *Gerontology.* 2006;52:154-159.
13. Martins J, da Silva JR, da Silva MRB, Bevilaqua-Grossi D. Reliability and validity of the belt-stabilized handheld dynamometer in hip- and knee-strength tests. *J Athl Train.* 2017;52:809-819.
14. Mentiplay BF, Perraton LG, Bower KJ, et al. Assessment of lower limb muscle strength and power using hand-held and fixed dynamometry: a reliability and validity study. *PLoS One.* 2015;10:e0140822.
15. Newman AB, Kupelian V, Visser M, et al. Strength, but not muscle mass, is associated with mortality in the health, aging and body composition study cohort. *J Gerontol A Biol Sci Med Sci.* 2006;61:72-77.
16. Ottenbacher KJ, Branch LG, Ray L, Gonzales VA, Peek MK, Hinman MR. The reliability of upper- and lower-extremity strength testing in a community survey of older adults. *Arch Phys Med Rehabil.* 2002;83:1423-1427.
17. Pincivero DM, Salfetnikov Y, Campy RM, Coelho AJ. Angle- and gender-specific quadriceps femoris muscle recruitment and knee extensor torque. *J Biomech.* 2004;37:1689-1697.
18. Samuel D, Wilson K, Martin HJ, Allen R, Sayer AA, Stokes M. Age-associated changes in hand grip and quadriceps muscle strength ratios in healthy adults. *Aging Clin Exp Res.* 2012;24:245-250.
19. Schaubert KL, Bohannon RW. Reliability and validity of three strength measures obtained from community-dwelling elderly persons. *J Strength Cond Res.* 2005;19:717-720.
20. Stratford PW, Balsor BE. A comparison of make and break tests using a hand-held dynamometer and the Kin-Com. *J Orthop Sports Phys Ther.* 1994;19:28-32.
21. Tieland M, Verdijk LB, de Groot LC, van Loon LJ. Handgrip strength does not represent an appropriate measure to evaluate changes in muscle strength during an exercise intervention program in frail older people. *Int J Sport Nutr Exerc Metab.* 2015;25:27-36.
22. Vaes AMM, Tieland M, Toussaint N, et al. Cholecalciferol or 25-hydroxycholecalciferol supplementation does not affect muscle strength and physical performance in prefrail and frail older adults. *J Nutr.* 2018;148:712-720.
23. Visser M, Goodpaster BH, Kritchevsky SB, et al. Muscle mass, muscle strength, and muscle fat infiltration as predictors of incident mobility limitations in well-functioning older persons. *J Gerontol A Biol Sci Med Sci.* 2005;60:324-333.
24. Wang CY, Olson SL, Protas EJ. Test-retest strength reliability: hand-held dynamometry in community-dwelling elderly fallers. *Arch Phys Med Rehabil.* 2002;83:811-815.