Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Test-retest and inter-rater reliability of a field-based test battery to assess lower limb isometric strength and inter-limb asymmetry in healthy adult endurance runners of various training levels

JOACHIM D'HONDT
Joachim.dhondt@vub.be
Vrije Universiteit Brussel
LAURENT CHAPELLE
Vrije Universiteit Brussel
CHRIS BISHOP
Middlesex University
DIRK AERENHOUTS
Vrije Universiteit Brussel
KEVIN DE PAUW
Vrije Universiteit Brussel
PETER CLARYS
Vrije Universiteit Brussel
EVA D'Hondt
Vrije Universiteit Brussel

Research Article

Keywords: variability, handheld dynamometry, symmetry, athlete monitoring, endurance running, imbalance

Posted Date: April 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9354122/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Objectives

This study aimed to evaluate the test-retest and inter-rater reliability of lower-limb isometric strength assessments using handheld dynamometry, and to examine the magnitude of inter-limb strength asymmetries in healthy adult endurance runners.

Methods

In total, 53 recreational runners of various training levels completed field-based tests measuring isometric hip adduction, hip abduction, knee extension, knee flexion, ankle dorsiflexion, and ankle plantar flexion strength across three sessions. Reliability was assessed using the intraclass correlation coefficient (ICC), standard error of measurement (SEM), standard error of prediction and coefficient of variation (CV). Systematic bias was assessed via Wilcoxon signed-rank tests.

Results

Test-retest reliability of the different tests ranged from moderate to good (ICC = 0.60–0.84), while inter-rater reliability ranged from moderate to excellent (ICC = 0.60–0.91). Variability was relatively low, with CV values ≤ 5.9%. These measures were generally comparable between subgroups. No significant differences were found across test sessions, except for left hip abduction ($p = 0.022$), right ankle dorsiflexion ($p = 0.007$) and right ankle plantar flexion ($p = 0.019$). Inter-rater comparisons showed no significant differences, except for right knee flexion ($p = 0.048$) and left ankle dorsiflexion ($p = 0.006$). Inter-limb asymmetry magnitudes ranged from 8.0% to 19.1% across tests.

Conclusion

Acceptable reliability for assessing group-level isometric lower-limb strength in distance runners was demonstrated. However, given the moderate to good test-retest ICCs and measurement variability, practitioners should exercise caution when using this test battery to monitor individual-level asymmetry magnitudes over time.

Trial registration:

ClinicalTrials.gov (NCT06808399). Registered on 02 April 2025.

Introduction

Endurance running performance is influenced by various physiological and biomechanical factors [1]. Running economy reflects the energy cost of running at a given velocity and has been shown to improve significantly with strength training, which can also effectively enhance time trial performance in middle- and long-endurance runners [2]. These improvements in running performance are likely attributed to increased muscle-tendon unit stiffness, improved neuromuscular function and mechanical efficiency as well as optimized muscle recruitment patterns, following explosive or heavy-

weight strength training [3]. Given its important role in terms of running performance, strength measures are commonly employed among both elite and recreational endurance runners to monitor their athletic performance and guide training approaches [2, 4].

Over the past decade, there has been increasing interest in understanding the relationship between side-to-side strength differences and sports performance [5]. While the evidence available is rather inconclusive, a recent systematic review shows that a greater magnitude of lower inter-limb asymmetry may be negatively associated with running performance [4]. For instance, Blagrove et al. [6] demonstrated that a more pronounced hip abduction strength asymmetry was strongly associated with impaired running economy (i.e., average energy cost of running across all speeds below the lactate turn point and scaled for body mass) in female endurance runners ($r = 0.85$). In contrast, no significant association was observed between the same measures in male runners ($r = -0.19$).

Recent studies emphasize the highly variable nature of inter-limb asymmetries, particularly at the lower limb level. More specifically, lower inter-limb asymmetry has been shown to be highly individual-specific, while also being dependent on the specific task and timing of the assessment [7–9]. Due to the inconsistent and context-dependent nature of lower inter-limb asymmetries, a recent systematic review by D'Hondt and Chapelle [4] stressed the importance of employing robust and reliable measurements to minimize testing errors and accurately detect true (changes in) inter-limb asymmetry. In support of this view, Bishop et al. [10] noted that when the variability of a measurement tool exceeds the magnitude of the expected change, its value for individual-level monitoring becomes limited. This highlights the critical importance of selecting reliable assessments to ensure meaningful interpretation of inter-limb asymmetry data.

To date, several methods are available for assessing muscle strength, with isokinetic dynamometry often being regarded as the gold standard for measuring this outcome whilst performing a specific movement [11]. Although isokinetic dynamometry is widely acknowledged for its excellent validity and reliability [12], it also requires expensive, laboratory-based, non-portable equipment, which makes it impractical for field use. To overcome these limitations, hand-held dynamometry (HHD) was developed as a low-cost, portable and convenient alternative for field-based strength assessments. Although previous research has reported substantial validity and reliability of using HHD as compared to isokinetic testing, discrepancies in results exist [11]. This is possibly related to differences in resistance support generated by the examiner, the lack of stabilization of the participant, and/or the specifics of the device being used [13].

Among the various HHD tools available, the "MusTec HD" is developed for measuring isometric muscle strength. However, to date, the test-retest (i.e., consistency of a measure over time) and inter-rater reliability (i.e., agreement between two or more raters) of this specific "MusTec HD" device has not been evaluated. While test-retest reliability has previously been examined using the "microFET2" HHD and a gauge tension dynamometer in healthy adult endurance runners (with ICCs ranging between 0.93 and 0.99), no studies to date have assessed the inter-rater reliability of using HHD for assessing muscle strength (and its magnitude of asymmetry) in this particular population [6, 14].

Given the widespread use of HHD in both laboratory and field-based settings, the primary aim of this study was to assess both test-retest reliability and inter-rater reliability of different lower limb isometric strength tests (i.e., hip adduction, hip abduction, knee extension, knee flexion, ankle dorsiflexion, ankle plantar flexion) using the "MusTec HD" dynamometer in endurance runners of

various training levels. Secondly, we examined the magnitude of (lower) inter-limb strength asymmetries within our population of healthy endurance runners. We hypothesized that the strength tests would demonstrate acceptable reliability and that runners would display small but meaningful inter-limb asymmetries.

Methods

Participants

In total, 27 male (age = 35.5 ± 9.7 years, body mass index (BMI) = 23.4 ± 3.6 kg/m^2), and 26 female (age = 33.4 ± 8.9 years, BMI = 22.3 ± 2.7 kg/m^2) endurance runners from various training levels were recruited using convenience and snowball sampling techniques. All participants' eligibility was assessed using a pre-screening online questionnaire. They were required to be free from medical conditions, injuries, or surgeries involving the lower limbs, hips, trunk, or back for at least six months prior to data collection. Participants with a history of any medical condition potentially affecting musculoskeletal or neurological functioning were excluded from the study. Study participants were also prohibited from taking any medication that might affect their performance during the tests.

Measures

Basic anthropometry

During the T_f, participants' stature (SECA 213 Portable Stadiometer, SECA, Germany) and body weight (Radwag WLT60/10/X/3, AllScales Europe, Veen, The Netherlands) were measured to the nearest 0.1 cm and 0.002 kg, respectively, according to the International Society for the Advancement of Kinanthropometry guidelines [15].

Functional test battery

Before the start of the strength tests during each of the three sessions (T_f, T_1 and T_2), a 10-minute standardized warm-up was implemented according to the RAMP method (i.e., Raise, Activate, Mobilize and Potentiate), including light jogging, body weight strength exercises (e.g., unilateral squat) and some jumps [16].

Muscle strength was examined using tests requiring the participants to exert maximal force against a stationary HHD (MusTec HD, MusTec, Almere, Netherlands), being secured with a tension strap. Joint positions for each test were visually aligned with the prescribed starting angles. The HHD recorded data at a sampling rate of 62.5 Hz, and results were expressed in kilograms (kg). All strength tests (N = 6, see Fig. 1) included one practice trial and three actual test trials per body side during each testing session. Measurements were stopped after reaching peak force. The best score of three attempts was used for analysis. Every test started with the right leg, followed by the left leg. At least 60 seconds between trials (on the same or opposite side of the body) was provided to allow sufficient recovery. All participants were verbally encouraged to deliver maximal force during the testing.

Hip adduction (see Fig. 1, panel A) and *hip abduction* (see Fig. 1, panel B) were measured in side lying position with the feet and ankles surpassing the treatment table's edge. The hips were strapped

down to isolate the target movement and minimize hip movement. Prior to test execution, participants were instructed to adopt a starting position in which the hip and knee of the resting leg were flexed to ~ 45° and ~ 90°, respectively. The HHD was attached to a tension strap (i.e., superior to the lateral malleolus for hip abduction and 5 cm superior to the medial malleolus for hip adduction) that was anchored to the ground. Participants were instructed to push their extended leg vertically upwards whilst maintaining the knee and hip in a fixed position. Specifically for hip abduction, the raters ensured that the leg remained fully horizontal during the test.

Knee extension (see Fig. 1, panel C) was measured in seated position with hip flexion of ~ 90° and knee flexion of ~ 90° over the edge of a treatment table with the strap of the HHD attached to one straight leg of the table. The HHD was positioned proximal to the lateral malleolus. To isolate the intended movement, the upper legs were strapped down to the treatment table and the arms of the participants were resting across their chest. They were also instructed to extend their knee as hard as possible.

Knee flexion (see Fig. 1, panel D) was measured in prone position with the pelvis fixed on the treatment table. The HHD was placed just superior of the calcaneus. Again, the tension strap of the HHD was attached on one straight leg of the treatment table. The participants were instructed to flex their knee as hard as possible.

Ankle dorsiflexion (see Fig. 1, panel E) was conducted barefoot in supine position with no hip and knee flexion. The start position of the ankle was plantar flexion mid-range. The HHD was positioned on the dorsal surface proximal of the metatarsophalangeal joint and attached to a tension strap that was anchored to the leg of a nearby standing table. The participants were instructed to perform ankle dorsiflexion as hard as possible.

Ankle plantar flexion (see Fig. 1, panel F) was performed barefoot by sitting on a chair with bent knees and the feet placed on the ground. The HHD was placed proximal to the medial and lateral condyle of the femur and fixed with a tension strap around the plantar surface of the foot proximal to the metatarsal head. Knee flexion was determined by vertically aligning both reference points. The participants were instructed to push their foot as hard as possible towards plantar flexion.

Design and Procedures

A repeated measures study design was implemented with participants being measured in one familiarization session (T_f) to minimize potential learning effects, followed by two data collection sessions (T_1 and T_2) to evaluate test-retest reliability of the strength assessments. All testing sessions were separated by two weeks, as is generally advised for health-related test-retest assessments [17]. Each participant was scheduled at the same time of day across all three test sessions. The data collection was carried out by the principal investigator (JD) and 12 Master's students in Movement and Sport Sciences or Physiotherapy and Rehabilitation Sciences, all of whom received extensive hands-on training prior to testing, including a theoretical session where all study procedures were reviewed in detail and at least two comprehensive practical session during which all tests were carried out on practice participants. At T_2, the tests were independently conducted by two raters (T_{2a} and T_{2b}) to assess inter-rater reliability. As such, the first measurement (i.e., which served as the reference measurement) was performed by the principal investigator (JD, T_{2a}) whilst the second

measurement was performed by one of the trained Master's students (T_{2b}). At least 60 minutes between the same strength tests was provided for recovery. All participants were instructed not to participate in vigorous physical activity for 24 hours prior to each data collection session, refrain from alcohol consumption for 12 hours, abstain from caffeine on the day of testing, and avoid diuretics for seven days before. They were also asked to maintain their usual training routines between sessions.

Statistical Analysis

All statistical analyses were conducted using SPSS version 29 (SPSS Inc., Chicago, IL, USA). Test-retest and inter-rater reliability were assessed using intraclass correlation coefficients (ICCs [3, 1]) with 95% confidence intervals, based on a two-way mixed model and consistency type. In addition, the standard error of measurement (SEM), standard error of prediction (SEP), and coefficient of variation (CV) were calculated for the entire study sample of endurance runners. ICC values were classified as follows: < 0.50 = poor, 0.50–0.74 = moderate, 0.75–0.90 = good, > 0.90 = excellent [18]. SEM, which was calculated as $SD \times \sqrt{1 - ICC}$ reflected the expected variability in scores around a participant's true score [19]. The SEP, being determined as $SD \times \sqrt{1 - ICC^2}$, represented the variability in an individual's score when performing the test a second time, considering their performance on the initial test [19]. The CV was calculated as $\frac{SD}{mean} \times 100$ [20]. Shapiro-Wilk tests were used to verify the normality of the data. Given that the raw performance values were non-normally distributed, Wilcoxon signed-rank tests were used to determine differences between test sessions T_1 and T_2 as well as differences between two raters on T_2 (i.e., T_{2a} and T_{2b}). Effect sizes for these Wilcoxon tests were computed as $r = \frac{Z}{\sqrt{N}}$ and interpreted as follows: 0.10 to 0.30 = small, 0.30 to 0.50 = medium, and ≥ 0.50 = large [21]. The magnitude of inter-limb asymmetry was calculated using the percentage difference method: ((dominant value – non-dominant value) / dominant value) × 100 [22].

Results

The total sample included 53 healthy adult endurance runners. However, due to unforeseen technical issues (i.e., a malfunction of the HHD caused by excessive forces (> 130 kg) during the initial ankle plantar flexion test), the inter-rater reliability could only be assessed in 49 participants (92.5%). In addition to equipment failure, the testing protocol for ankle plantar flexion was modified partway through the study resulting in 34 measured participants (64.2%). Initially, this particular strength test was performed seated with the leg fully extended, using the examiner for resistance support. This posture inadvertently introduced force, potentially compromising the specificity of the measurement. Consequently, the participants with collected ankle plantar flexion data based on this initial testing protocol were excluded (n = 19, 35.8%) to reflect the reliability of the adjusted testing position in the remaining participants.

Test-retest and inter-rater reliability results as well as lower inter-limb asymmetry magnitudes of the endurance runners are presented in Table 1 and Table 2, respectively. Across the 6 (functional) strength tests in our field-based test battery, the mean test-retest reliability ICC values ranged from moderate to good (ICC = 0.60–0.84), while the mean inter-rater reliability ICC values ranged from moderate to excellent (ICC = 0.60–0.91). In addition, SEM values ranged from 4.9 to 18.8 kg for test-retest reliability and from 4.1 to 26.2 kg for inter-rater reliability. Similarly, SEP values ranged from 6.6 to 25.1 kg for test-retest and from 5.6 to 33.3 kg for inter-rater assessments. CV values ranged

from 0.1% to 5.9% for test-retest reliability and from 0.8% to 5.3% for inter-rater reliability. The magnitude of lower inter-limb asymmetry observed across T_1 and T_{2a} test sessions varied between 8.0% and 19.0%.

Wilcoxon signed-rank test results indicated no significant differences between T_1 and T_{2a} assessments for hip adduction (both legs), hip abduction (right leg), knee extension and flexion (both legs), ankle dorsiflexion (left leg), and ankle plantar flexion (left leg), with p-values ranging from 0.101 to 0.929 (see Table 3). However, significant differences between T_1 and T_{2a} test results were found for hip abduction (left leg), ankle dorsiflexion (right leg), and ankle plantar flexion (right leg), with p-values ranging between 0.007 and 0.022. Additionally, no significant differences were observed between raters for most strength tests included in the presented battery, including hip adduction and abduction (both legs), knee extension (both legs), knee flexion (left leg), ankle dorsiflexion (right leg), and ankle plantar flexion (both legs), with p-values ranging between 0.055 and 0.929. In contrast, knee flexion of the right leg ($p = 0.048$) and ankle dorsiflexion of the left leg ($p = 0.006$) showed significant differences between raters (see Table 3).

Table 1
Test-retest reliability measures and inter-limb asymmetry magnitudes of all isometric strength tests in distance runners

Strength measure	N	Mean ± SD T_1 (kg)	Asymmetry (%)	Mean ± SD T_{2a} (kg)	Asymmetry (%)	ICC (95% CI)	SEM (kg)	SEP (kg)	CV (%)
Hip adduction, R	53	42.8 ± 13.9	12.8 ± 9.6	44.8 ± 15.7	11.8 ± 9.8	0.81 (0.70, 0.89)	6.5	8.7	3.2
Hip adduction, L	53	42.6 ± 14.0		44.2 ± 14.7		0.71 (0.55, 0.82)	7.7	10.1	2.5
Hip abduction, R	53	31.0 ± 9.6	17.5 ± 13.2	29.3 ± 8.4	19.1 ± 12.0	0.60 (0.40, 0.75)	5.7	7.2	4.0
Hip abduction, L	53	31.6 ± 10.5		29.0 ± 9.5		0.70 (0.54, 0.82)	5.3	6.9	5.1
Knee extension, R	53	92.8 ± 28.6	10.1 ± 10.5	92.9 ± 28.8	9.4 ± 7.1	0.84 (0.74, 0.90)	11.4	15.5	0.1
Knee extension, L	53	88.9 ± 27.3		88.1 ± 26.4		0.78 (0.64, 0.87)	12.5	16.7	0.6
Knee flexion, R	53	54.2 ± 16.6	13.1 ± 11.2	55.5 ± 18.7	12.8 ± 8.8	0.67 (0.49, 0.80)	9.5	12.3	1.6
Knee flexion, L	53	50.6 ± 15.3		51.8 ± 15.0		0.77 (0.63, 0.86)	7.3	9.6	1.7
Ankle dorsiflexion, R	53	39.0 ± 12.2	12.6 ± 11.9	40.1 ± 11.7	12.9 ± 10.3	0.83 (0.72, 0.90)	4.9	6.6	2.1
Ankle dorsiflexion, L	53	38.5 ± 11.2		41.7 ± 11.0		0.75 (0.61, 0.85)	5.6	7.4	5.1

N = number of participants assessed, SD = standard deviation, T_1 = first test session, T_2 = second test session, first test trial, ICC = intraclass correlation coefficient, CI = confidence interval, SEM = standard error of measurement, SEP = standard error of prediction, CV = coefficient of variation, R = right, L = left

Strength measure	N	Mean ± SD T_1 (kg)	Asymmetry (%)	Mean ± SD T_{2a} (kg)	Asymmetry (%)	ICC (95% CI)	SEM (kg)	SEP (kg)	CV (%)
Ankle plantar flexion, R	34	127.8 ± 43.6	11.9 ± 11.3	134.4 ± 40.6	8.0 ± 6.8	0.84 (0.70, 0.92)	16.8	22.7	3.6
Ankle plantar flexion, L	34	120.5 ± 40.5		130.9 ± 37.9		0.77 (0.58, 0.88)	18.8	25.1	5.9

N = number of participants assessed, SD = standard deviation, T_1 = first test session, T_2 = second test session, first test trial, ICC = intraclass correlation coefficient, CI = confidence interval, SEM = standard error of measurement, SEP = standard error of prediction, CV = coefficient of variation, R = right, L = left

Table 2
Inter-rater reliability measures and inter-limb asymmetry magnitudes of all isometric strength tests in distance runners

Strength measure	N	Mean ± SD T_{2a} (kg)	Asymmetry (%)	Mean ± SD T_{2b} (kg)	Asymmetry (%)	ICC (95% CI)	SEM (kg)	SEP (kg)	CV (%)
Hip adduction, R	49	44.8 ± 15.7	11.5 ± 9.5	43.7 ± 13.0	13.3 ± 9.1	0.86 (0.76, 0.92)	5.4	7.4	1.7
Hip adduction, L	49	44.2 ± 14.7		42.9 ± 14.5		0.84 (0.73, 0.91)	5.8	7.9	2.1
Hip abduction, R	49	29.3 ± 8.4	10.9 ± 9.0	27.3 ± 9.8	13.6 ± 12.4	0.65 (0.45, 0.79)	5.4	6.9	5.1
Hip abduction, L	49	29.0 ± 9.5		27.8 ± 9.3		0.74 (0.58, 0.84)	4.8	6.3	3.0
Knee extension, R	49	92.9 ± 28.8	9.0 ± 7.0	91.8 ± 30.6	12.2 ± 9.2	0.84 (0.73, 0.90)	11.8	16.0	0.8
Knee extension, L	49	88.1 ± 26.4		84.9 ± 27.6		0.91 (0.84, 0.95)	8.1	11.1	2.7
Knee flexion, R	49	55.5 ± 18.7	12.5 ± 8.9	51.5 ± 14.2	11.7 ± 8.7	0.75 (0.60, 0.85)	7.8	10.3	5.3
Knee flexion, L	49	51.8 ± 15.0		49.0 ± 13.7		0.75 (0.59, 0.85)	7.2	9.5	4.0
Ankle dorsiflexion, R	49	40.1 ± 11.7	12.2 ± 10.2	38.4 ± 10.9	12.4 ± 11.3	0.87 (0.78, 0.93)	4.1	5.6	3.1
Ankle dorsiflexion, L	49	41.7 ± 11.0		38.7 ± 10.8		0.83 (0.72, 0.90)	4.5	6.1	5.3

N = number of participants assessed, SD = standard deviation, T_{2a} = Second test session conducted by the principal investigator (JD), T_{2b} = Second test session conducted by a trained Master's student, ICC = intraclass correlation coefficient, CI = confidence interval, SEM = standard error of measurement, SEP = standard error of prediction, CV = coefficient of variation, R = right, L = left

Strength measure	N	Mean ± SD T_{2a} (kg)	Asymmetry (%)	Mean ± SD T_{2b} (kg)	Asymmetry (%)	ICC (95% CI)	SEM (kg)	SEP (kg)	CV (%)
Ankle plantar flexion, R	34	134.4 ± 40.6	8.0 ± 6.8	139.6 ± 44.7	14.7 ± 11.3	0.62 (0.35, 0.79)	26.2	33.3	2.7
Ankle plantar flexion, L	34	130.9 ± 37.9		139.9 ± 42.5		0.60 (0.34, 0.78)	25.4	32.2	4.7

N = number of participants assessed, SD = standard deviation, T_{2a} = Second test session conducted by the principal investigator (JD), T_{2b} = Second test session conducted by a trained Master's student, ICC = intraclass correlation coefficient, CI = confidence interval, SEM = standard error of measurement, SEP = standard error of prediction, CV = coefficient of variation, R = right, L = left

Table 3
Test–retest and inter-rater reliability of all isometric strength tests in distance runners, based on Wilcoxon signed-rank test comparisons in distance runners

	Test-retest reliability					Inter-rater reliability				
	N	p-value	Z-value	ES	ES levels	N	p-value	Z-value	ES	ES levels
Hip adduction, R	53	0.253	-1.142	-0.157	Small	49	0.647	-0.458	-0.065	Small
Hip adduction, L	53	0.339	-0.956	-0.131	Small	49	0.842	-0.199	-0.028	Small
Hip abduction, R	53	0.101	-1.639	-0.225	Small	49	0.055	-1.920	-0.274	Small
Hip abduction, L	53	**0.022**	-2.289	-0.314	Medium	49	0.479	-0.708	-0.101	Small
Knee extension, R	53	0.929	-0.089	-0.012	Small	49	0.929	-0.089	-0.013	Small
Knee extension, L	53	0.153	-1.430	-0.196	Small	49	0.153	-1.430	-0.204	Small
Knee flexion, R	53	0.390	-0.859	-0.118	Small	49	**0.048**	-1.975	-0.282	Small
Knee flexion, L	53	0.541	-0.611	-0.084	Small	49	0.303	-1.031	-0.147	Small
Ankle dorsiflexion, R	53	**0.007**	-2.714	-0.373	Medium	49	0.236	-1.184	-0.169	Small
Ankle dorsiflexion, L	53	0.462	-0.736	-0.101	Small	49	**0.006**	-2.746	-0.392	Medium
Ankle plantar flexion, R	34	**0.019**	-2.351	-0.403	Medium	34	0.369	-0.898	-0.128	Small
Ankle plantar flexion, L	34	0.103	-1.633	-0.280	Small	34	0.132	-1.505	-0.215	Small

N = number of participants assessed, ES = effect size, R = right, L = left, note: significant differences ($p < 0.05$) between test occasions are bolded

Discussion

This study aimed to evaluate the test-retest and inter-rater reliability of the “MusTec HD” as a HHD tool for assessing isometric strength during hip adduction, hip abduction, knee extension, knee flexion, ankle dorsiflexion and ankle plantar flexion among healthy adult endurance runners of various training levels. Additionally, we examined the magnitude of these study participants’ lower inter-limb strength asymmetries.

In our overall study sample of endurance runners, the field-based strength test performances assessed with the “MusTec HD” demonstrated moderate to good test-retest reliability and moderate to excellent inter-rater reliability as indicated by the respective mean ICC values. No significant differences were observed between test (T_1) and retest (T_{2a}) sessions, except for hip abduction of the left leg and ankle dorsiflexion of the right leg. This suggests minimal systematic bias between test occasions. Similarly, inter-rater comparisons revealed no significant differences in test outcomes, with the exception of knee flexion of the right leg and ankle dorsiflexion of the left leg. SEM values ranged from 4.1 to 26.2 kg, and SEP values ranged from 5.6 to 6.6 kg for both test–retest and inter-rater analyses. Finally, all CV values were below 6% for both test-retest and inter-rater reliability, suggesting that asymmetry values below this threshold should be interpreted cautiously, although this may vary depending on the specific test.

Compared to previous research also involving endurance runners, the present study yielded lower ICC values for test-retest reliability. For example, Blagrove et al. [6] reported excellent test-retest reliability for isometric hip abduction strength (ICC: right leg = 0.97, left leg = 0.98) and knee flexion (ICC: right leg = 0.97, left leg = 0.99) in a sample of 12 adolescent competitive endurance runners using a slightly different test setup with a gauge tension dynamometer. In contrast, our study found moderate test-retest reliability for hip abduction strength (ICC: right leg = 0.60, left leg = 0.70) and good test-retest reliability for knee flexion strength (ICC: right leg = 0.67, left leg = 0.77) in the total sample of endurance runners. Furthermore, Jackson et al. [14] reported higher test-retest ICCs (range: 0.93–0.98) across isometric hip adduction, hip abduction, knee flexion, and ankle plantar flexion strength tests in 75 adult recreational endurance runners using a MicroFet 2 dynamometer. Despite the differences in ICC values, the SEM values were more comparable between studies. Our SEM values ranged from 5.5 to 17.2 kg in the total sample, while those reported by Jackson et al. [14] ranged from 5.3 to 18.8 kg, suggesting a similar level of expected score variability around each participant’s true value.

A possible explanation for the lower ICCs in our study may lie in methodological differences. While we secured the HHD using a strap, Jackson, Cheng [14] employed a PVC pipe for the stabilization of the leg. Another key distinction is the timing of the test-retest sessions. Jackson et al. [14] conducted intra-session re-testing (i.e., within a single test session) and Blagrove et al. [6] used a 48-hour interval, whereas our study employed a minimum interval of two weeks between sessions. The two-week interval used in the present study was considered a suitable compromise to minimize recall bias while avoiding unintended clinical changes [17]. However, this extended duration possibly increased the likelihood of natural fluctuations in performance, which may have contributed to lower but, yet arguably ecologically valid reliability estimates.

In terms of inter-rater reliability, only moderate inter-rater reliability was found for isometric hip abduction of the right leg and ankle plantar flexion strength in both legs. These findings suggest that, for these specific assessments, using the same examiner is recommended to enhance measurement consistency. Hip abduction, as assessed in this study, probably presented the greatest degree of freedom among the different strength tests, potentially contributing to an increased variability. While the top leg was visually aligned to a horizontal position, even small deviations in hip angle may have affected force output and, therefore, reliability of the associated test result [23]. Additionally, during the ankle plantar flexion assessment, the HHD required manual stabilization by the examiner, which

may have introduced inconsistencies due to differences in force application or positioning across raters [13].

The observed isometric strength asymmetry between the left and right leg ranged from 8.0% to 19.1% across tests, with hip adduction displaying the highest magnitude of inter-limb asymmetry. All asymmetry values exceeded CV%, indicating that these differences represent meaningful asymmetries rather than measurement error. The magnitudes observed are comparable to those reported in previous studies involving endurance runners [6, 24, 25]. Although endurance running is often considered a symmetrical sport, the present results indicate that a certain magnitude of inter-limb strength asymmetry is common. While evidence regarding the relationship between inter-limb asymmetry and running performance yields mixed findings [4], prior research has suggested that asymmetries in isometric hip abduction strength exceeding 9% may be negatively associated with running economy in female endurance runners [6]. Therefore, future research should aim to confirm whether such asymmetries are indeed detrimental to endurance running performance, and practitioners may consider implementing targeted strength training interventions to reduce excessive inter-limb imbalances [10].

Although higher magnitudes of inter-limb asymmetry could potentially impair running performance, it is important to note that the CV values in the present study ranged from 0.1% to 5.9%. These levels of measurement error may obscure real changes in asymmetry over time when the (lower) inter-limb asymmetry magnitudes are low. As Bishop et al. [10] emphasized in their review, when the variability of a measurement tool exceeds the magnitude of the expected change, its usefulness for individual monitoring is limited. Therefore, practitioners should be cautious when interpreting low (≤ 6%) individual isometric strength asymmetry magnitudes, as they may fall within the range of measurement error.

Despite the large sample size and comprehensive test battery, this study has several limitations. First, due to failure of equipment, inter-rater reliability could only be assessed in 49 of the 53 participants. Additionally, necessary adjustments to the ankle plantar flexion protocol partway through the study just allowed us to evaluate the test-retest and inter-rater reliability in 34 participants. Second, the use of multiple examiners, as well as their inconsistent participation across the different test sessions, for the test-retest reliability may have introduced additional variability. Given that inter-rater reliability was only moderate to good, this variability could have negatively influenced the consistency of the test-retest results. Future studies should therefore consider evaluating the intra-rater reliability as well as the validity against established gold standard tools of these assessments with the “MusTec HD”. Third, joint angles were visually confirmed rather than objectively measured, which may have increased measurement error. However, as all limbs were stabilized against a standardized support with minimal freedom of movement, we expect that this verification method was appropriate.

In summary, the “MusTec HD” showed moderate to good test-retest reliability and moderate to excellent inter-rater reliability across all field-based isometric strength tests in healthy adult endurance runners of various training levels. While group-level reliability was generally acceptable across tests, its use for detecting small changes in individual-level strength asymmetries should be approached with caution due to the high measurement variability relative to potential low asymmetry magnitudes (≤ 6%). More research on this topic is thus warranted to further refine testing protocols and explore methods for enhancing reliability of strength assessments in applied athletic settings.

Abbreviations

HHD
hand-held dynamometry
BMI
body mass index
ICC
intraclass correlation coefficient
SEM
standard error of measurement
SEP
standard error of prediction
CV
coefficient of variation

Declarations

Ethics approval and consent to participate

After being fully informed regarding all study details, every participant provided their written informed consent. Our study protocol followed the principles outlined in the Declaration of Helsinki and was approved by the medical ethics committee of Vrije Universiteit Brussel and UZ Brussel (B.U.N. 1432023000042) and has been registered online on ClinicalTrials.gov (ID: NCT06808399).

Consent for publication

Not applicable

Competing interests

The authors declare no competing interests.

Funding

The authors declare that no funding was received to support the conduct of this study.

Author Contribution

JD, CB, ED, LC, DA, PC and KDP contributed to the conceptualization of the study. JD was responsible for data collection. JD prepared the original manuscript draft, while CB, ED, LC, DA, PC and KDP provided critical review and revisions. All authors read and approved the final version of the manuscript.

Acknowledgments

The authors warmly thank the participants of this study.

Data Availability

The data of the present study is available in Vrije Universiteit Brussel (VUB) Institutional Repository with the Persistent Identifiers VUB/MOVE/1/000013 and VUB/MOVE/1/000015.

References

1. Bassett DJr, Howley ET. Limiting factors for maximum oxygen uptake and determinants of endurance performance. Med Sci Sports Exerc. 2000;32(1):70–84.
2. Blagrove RC, Howatson G, Hayes PR. Effects of Strength Training on the Physiological Determinants of Middle- and Long-Distance Running Performance: A Systematic Review. Sports Med. 2018;48(5):1117–49.
3. Denadai BS, de Aguiar RA, de Lima LCR, Greco CC, Caputo F. Explosive Training and Heavy Weight Training are Effective for Improving Running Economy in Endurance Athletes: A Systematic Review and Meta-Analysis. Sports Med. 2017;47(3):545–54.
4. D'Hondt J, Chapelle L, Bishop C, Aerenhouts D, De Pauw K, Clarys P et al. Association Between Inter-Limb Asymmetry and Determinants of Middle- and Long-distance Running Performance in Healthy Populations: A Systematic Review. Sports Med-Open. 2024;10(1).
5. Bishop C, Turner A, Read P. Effects of inter-limb asymmetries on physical and sports performance: a systematic review. J Sport Sci. 2018;36(10):1135–44.
6. Blagrove RC, Bishop C, Howatson G, Hayes PR. Inter-limb strength asymmetry in adolescent distance runners: Test-retest reliability and relationships with performance and running economy. J Sport Sci. 2021;39(3):312–21.
7. Chapelle L, Bishop C, D'Hondt J, D'Hondt E, Clarys P. Morphological and functional asymmetry in elite youth tennis players compared to sex- and age-matched controls. J Sport Sci. 2022;40(14):1618–28.
8. D'Hondt J, Chapelle L. Change of Direction Asymmetry in Youth Elite Tennis Players: A Longitudinal Study. Int J Sports Med. 2024;45(6):436–42.
9. D'Hondt J, Vanhoudt M, Maluta D, Chapelle L, Aerenhouts D. Does a muscle fatigue-inducing protocol alters the magnitude of jump inter-limb asymmetry in healthy adolescents? J Hum Kinetics. 2025.
10. Bishop C, de Keijzer KL, Turner AN, Beato M. Measuring Interlimb Asymmetry for Strength and Power: A Brief Review of Assessment Methods, Data Analysis, Current Evidence, and Practical Recommendations. J Strength Cond Res. 2023;37(3):745–50.
11. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held Dynamometry Correlation With the Gold Standard Isokinetic Dynamometry: A Systematic Review. Pm&R. 2011;3(5):472–9.
12. Drouin JM, Valovich-mcLeod TC, Shultz SJ, Gansneder BM, Perrin DH. Reliability and validity of the Biodex system 3 pro isokinetic dynamometer velocity, torque and position measurements. Eur J Appl Physiol. 2004;91(1):22–9.
13. Kolber MJ, Cleland JA. Strength testing using hand-held dynamometry Physical therapy reviews. 2005;10:99–112.

14. Jackson SM, Cheng MS, Smith AR, Kolber MJ. Intrarater reliability of hand held dynamometry in measuring lower extremity isometric strength using a portable stabilization device. Musculoskel Sci Prac. 2017;27:137–41.
15. Marfell-Jones MJSAD, JH dR. International standards for anthropometric assessment. International Society for the Advancement of Kinanthropometry; 2012.
16. Jeffreys I. Warm-up revisited: The ramp method of optimizing warm-ups. Prof Strength Cond. 2007:6–12.
17. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RE. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol. 2003;56(8):730–5.
18. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.
19. McManus IC. The misinterpretation of the standard error of measurement in medical education: A primer on the problems, pitfalls and peculiarities of the three different standard errors of measurement. Med Teach. 2012;34(7):569–76.
20. Cormack SJ, Newton RU, McGuigan MR, Doyle TLA. Reliability of Measures Obtained During Single and Repeated Countermovement Jumps. Int J Sport Physiol. 2008;3(2):131–44.
21. Fritz CO, Morris PE, Richler JJ. Effect Size Estimates: Current Use, Calculations, and Interpretation. J Exp Psychol Gen. 2012;141(1):2–18.
22. Bishop C, Read P, Chavda S, Turner A. Asymmetries of the Lower Limb: The Calculation Conundrum in Strength Training and Conditioning. Strength Cond J. 2016;38(6):27–32.
23. Savelberg HHCM, Meijer K. The effect of age and joint angle on the proportionality of extensor and flexor strength at the knee joint. J Gerontol a-Biol. 2004;59(11):1120–8.
24. Tabor P, Mastalerz A, Iwańska D, Grabowska O. Asymmetry indices in female runners as predictors of running velocity. Pol J Sport Tourism. 2019;26(3):3–8.
25. Vagenas G, Hoshizaki B. Functional Asymmetries and Lateral Dominance in the Lower-Limbs of Distance Runners. Int J Sport Biomech. 1991;7(4):311–29.

Figures



Figure 1

Assessment of isometric strength upon performing various lower limb strength tests: hip adduction (A), hip abduction (B), knee extension (C), knee flexion (D), ankle dorsiflexion (E), and ankle plantar flexion (F).