2017

Reliability and Minimum Detectable Change for Common Clinical Physical Function Tests in Sarcopenic Men and Women

Nathaniel D.M. Jenkins
Oklahoma State University, nathaniel.jenkins@okstate.edu

Joel T. Cramer
University of Nebraska-Lincoln, jcramer@unl.edu

Published in *Journal of the American Geriatrics Society* (2017), 8 pp. doi: 10.1111/jgs.14769

Reliability and Minimum Detectable Change for Common Clinical Physical Function Tests in Sarcopenic Men and Women

Nathaniel D.M. Jenkins, PhD,[1] and Joel T. Cramer, PhD[2]

1 Applied Neuromuscular Physiology Laboratory, Oklahoma State University, Stillwater, Oklahoma

2 Neuromuscular Research and Imaging Laboratory, University of Nebraska-Lincoln, Lincoln, Nebraska

Corresponding author — Joel T. Cramer, Associate Professor, Department of Nutrition and Health Sciences, University of Nebraska-Lincoln, 211 Ruth Leverton Hall, Lincoln, NE 68583, email jcramer@unl.edu

Objectives: To determine the test–retest reliability and minimum detectable change scores for seven common clinical measurements of muscle strength and physical function in a multiethnic sample of sarcopenic, malnourished men and women.

Design: Each participant visited the laboratory seven times over 25 to 26 weeks. Reliability was assessed for each measurement from Familiarization 1 to Familiarization 2 (R1), Familiarization 2 to baseline testing (R2), Familiarization 3 to 12-week testing (R3), and Familiarization 4 to 24-week testing (R4).

Setting: Data were collected during a clinical trial at 23 sites in the United States, Belgium, Italy, Mexico, Poland, Spain, Switzerland, and the United Kingdom.

Participants: Sarcopenic, malnourished, older adults (N = 257; n = 98 men aged 76.8 ± 6.3, n = 159 women aged 75.9 ± 6.6).

Measurements: During each visit, participants completed the Short Physical Performance Battery (SPBB) and isometric handgrip and isokinetic leg extensor and flexor strength testing at a slow (1.05 rad/s) and fast (3.15 rad/s) velocity.

Results: Handgrip strength, gait speed, SPPB score, and isokinetic leg extension and flexion peak torque (PT) had intraclass correlation coefficients (ICCs) that were significantly greater than 0 (all ≥ 0.59) at R1, R2, R3, and R4, although most of these variables demonstrated systematic increases at R1, and several exhibited systematic variability beyond the baseline testing session.

Conclusion: The ICCs and standard errors of the measurement (SEMs) generally improved with familiarization, which emphasizes the need for at least one familiarization trial for these measurements in sarcopenic, malnourished older adults. A three tier-approach to interpreting the clinical importance of statistically significant results that includes null hypothesis testing, examination and interpretation of the effect magnitude, and comparison of individual changes with the SEM and minimum detectable change of the measurements used is recommended.

Keywords: sarcopenia; aging; muscle strength; physical function; sensitivity

Aging is associated with a progressive loss of skeletal muscle mass and strength known as sarcopenia. Sarcopenia has multiple contributing factors, including chronic disease, a sedentary lifestyle, and malnutrition, and is associated with deleterious consequences such as impaired mobility, impaired ability to perform activities of daily living, greater risk of falls and fractures, loss of independence, and greater risk of death.[1] In 2010, it was estimated that sarcopenia affected more than 50 million people and would affect more than 200 million older adults through 2050.1 Consequently, the identification and treatment of sarcopenia has become an emphasis of investigators and clinicians worldwide, and identifying reliable measurements with which to identify and track sarcopenia is of great importance.

In 2009, a group of representatives from the European Geriatric Medicine Society, the European Society for Clinical Nutrition and Metabolism, the International Association of Gerontology and Geriatrics—European Region, and the International Association of Nutrition and Aging formed the European Working Group on Sarcopenia in Older People (EWGSOP) to develop a clinical definition of and diagnostic criteria for sarcopenia.[1] In its report, the EWGSOP developed primary outcome domains for interventions in sarcopenic older adults that included muscle strength and physical performance.[1] The EWGSOP identified handgrip, leg extensor, and leg flexor strength as three primary muscle strength measurements and the Short Physical Performance Battery (SPPB) and usual gait speed as two primary tests of physical function for use in studies on sarcopenia,[1] but they recognized that information on the feasibility of muscle strength testing was needed from a wider range of ages and

ethnicities in older adults.[1] Furthermore, although studies have identified meaningful changes in gait speed and SPPB score in older adults with mild to moderate mobility impairment and older sedentary adults,[2,3] no previous studies have examined the reliability or of meaningful changes in sarcopenic, malnourished older adults. Therefore, the purpose of this study was to determine the test–retest reliability and minimum detectable change (MDC) scores for seven common clinical measurements of muscle strength and physical function in a multiethnic sample of sarcopenic, malnourished older men and women.

Methods

Participants

Two hundred fifty-seven men (n = 98; mean age 76.8 ± 6.3, height 171.5 ± 7.4 cm, weight 77.8 ± 13.9 kg) and women (n = 159; mean age 75.9 ± 6.6, height 156.8 ± 7.7 cm, weight 65.1 ± 12.7 kg) completed this study. All participants were aged 65–90 and were ambulatory and had a Subjective Global Assessment of Nutritional Status score of B or C,[4] a gait speed slower than 0.8 m/s or low handgrip strength (<20 kg for women, <30 kg for men), and a low skeletal muscle mass index according to whole-body dual x-ray absorptiometry (Class 1 or Class 2 sarcopenia),1 which classified them as sarcopenic, malnourished older adults. Further details regarding inclusion and exclusion criteria have been published elsewhere.[5]

The institutional review board or ethics committee at each of the study sites approved this study. All participants provided written informed consent before data collection, and a physician's clearance was required to verify all inclusion and exclusion criteria before enrollment.

Experimental Design

Each participant visited the laboratory on seven occasions over 25–26 weeks (Figure S1). At each visit, participants completed handgrip strength testing, the SPPB, and isokinetic leg extension and flexion testing. Reliability was assessed in men and women for each measurement from Familiarization 1 to Familiarization 2 (R1), Familiarization 2 to baseline testing (R2), Familiarization 3 to 12-week testing (R3), and Familiarization 4 to 24-week testing (R4).

These data were collected during a clinical trial (ClinicalTrials.gov Identifier: NCT01191125) from February 2011 to September 2012 at 23 sites in the United States, Belgium, Italy, Mexico, Poland, Spain, Switzerland, and the United Kingdom. Before any testing, all study staff were trained in a webinar and in person. This clinical trial involved two intervention arms comparing an experimental oral nutritional supplement with a control oral nutritional supplement, but the aim of this substudy was to examine the test–retest reliability at each discrete testing phase (R1, R2, R3, R4), and the effects of the intervention arms were not examined, so the intervention did not affect the reliability of these tests reported herein. Baseline isokinetic muscle strength data from this clinical trial have been previously published.[5]

Isometric Handgrip Strength Testing

Maximal isometric handgrip strength was measured using a calibrated hand-held dynamometer (Jamar, Lafayette Instrument, Lafayette, IN). Participants completed three trials during which they were asked to squeeze the dynamometer handle as forcefully as possible for 3 to 5 seconds. Each trial was recorded in kilograms, and the average of three trials was used as the representative score.

Short Physical Performance Battery

The SPPB has been described previously.[6] Each of its components (balance tests, gait speed) were scored on a scale from 0 to 4, and a composite SPPB score was created from all three tests and used for analyses. Individual gait speed times during the 4-m course were converted to m/s and analyzed separately from the composite SPPB score.

Isokinetic Leg Extension Strength Testing

For all isokinetic testing, participants were seated on a calibrated isokinetic dynamometer. They completed four submaximal warm-up leg extension and leg flexion contractions at 25%, 50%, 75%, and 100% of their perceived effort at 1.05 rad/s and 3.14 rad/s. After the warm-ups, each participant completed three maximal leg extension and flexion contractions as quickly and explosively as possible at each velocity. The dynamometer software calculated peak torque (PT; Nm), which was the peak of the isokinetic torque-angle curve for each repetition. The repetition yielding the highest PT value at each velocity for each participant was used for further analysis.[5]

Statistical Analyses

The dependent variables used for analysis were handgrip strength (kg), gait speed (m/s), SPPB score, isokinetic leg extension PT at 1.05 rad/s and 3.14 rad/s (Nm), and isokinetic leg flexion PT at 1.05 rad/s and 3.14 rad/s (Nm).

Before statistical analyses were conducted, the data were checked for data entry errors. Errors or nonphysiological values were found for up to eight subjects, such as values that were outside of the equipment's reporting capabilities. In each instance, all data for that participant were removed for that variable at the corresponding time point. Tables 1–3 indicate where data entry errors were removed (indicated by a ^ in the sample size column).

Test–retest reliability was quantified for each of the dependent variables between visits at R1, R2, R3, and R4. Two-way analyses of variance (ANOVAs) were used to compare the means between visits for systematic variability. Cohen's d effect sizes were calculated for the difference in means between visits. The likelihood that the magnitude of the effect was meaningful was determined using magnitude-based inferences using a smallest important effect value of 0.2.[7] Test–retest reliability was calculated by determining the intraclass correlation coefficient (ICC) using model "2, k."[8,9] The 95% confidence interval (CI) for each $ICC_{2,k}$ was also calculated.[8,10,11] For measures of absolute reliability, the standard error of the measurement (SEM) was calculated as

Table 1. Means, Standard Deviations and Reliability Statistics for Handgrip Strength, Gait Speed, and Short Physical Performance Battery (SPPB) at Familiarization 1 and Familiarization 2, Familiarization 2 and Baseline Testing, Familiarization 3 and 12-Week Testing, and Familiarization 4 and 24-Week Testing

	n	1 Mean±Standard Deviation	2 Mean±Standard Deviation	Change, %	*P*-Value	ES (%)	Intraclass Correlation Coefficient (95% Confidence Interval)	Standard Error of the Measurement	Coefficient of Variation, %	Minimum Detectable Change
Handgrip strength, kg										
R1: Familiarization 1–familiarization 2										
Men	97^	27.84 ± 8.86	28.43 ± 9.04	2.10	.02[a]	0.24 (34.7)	0.96 (0.94–0.97)	1.72	6.11	4.76
Women	159	16.60 ± 5.76	16.87 ± 5.76	1.66	.12	0.12	0.93 (0.90–0.95)	1.56	9.30	4.31
R2: Familiarization 2–baseline										
Men	98	28.37 ± 9.01	28.01 ± 9.08	-1.27	.10	-0.17	0.97 (0.96–0.98)	1.51	5.35	4.18
Women	159	16.87 ± 5.76	16.93 ± 5.83	0.36	.62	0.04	0.96 (0.95–0.97)	1.11	6.57	3.08
R3: Familiarization 3–12 weeks										
Men	98	29.76 ± 9.27	29.62 ± 8.83	-0.46	.63	-0.05	0.95 (0.93–0.97)	1.98	6.68	5.50
Women	159	17.60 ± 6.10	17.63 ± 6.08	0.16	.83	0.02	0.96 (0.95–0.97)	1.22	6.91	3.37
R4: Familiarization 4–24 weeks										
Men	98	30.42 ± 9.09	30.61 ± 8.98	0.61	.48	0.07	0.96 (0.94–0.97)	1.83	6.01	5.08
Women	159	18.17 ± 5.82	18.27 ± 5.72	0.53	.37	0.07	0.97 (0.96–0.98)	0.96	5.28	2.67
Gait speed, m/s										
R1: Familiarization 1–familiarization 2										
Men	98	0.77 ± 0.18	0.78 ± 0.18	1.43	.30	0.09	0.82 (0.75–0.88)	0.08	9.97	0.21
Women	159	0.72 ± 0.18	0.74 ± 0.18	3.33	<.01[a]	0.22 (38.0)	0.86 (0.81–0.90)	0.06	8.64	0.18
R2: Familiarization 2–baseline										
Men	98	0.78 ± 0.18	0.79 ± 0.18	1.15	.40	0.09	0.82 (0.74–0.88)	0.08	9.84	0.21
Women	159	0.74 ± 0.18	0.76 ± 0.19	2.28	.04[a]	0.19 (54.3)	0.84 (0.78–0.88)	0.07	9.39	0.20
R3: Familiarization 3–12 weeks										
Men	98	0.83 ± 0.18	0.86 ± 0.18	3.62	<.01[a]	0.40 (1.1)	0.90 (0.83–0.94)	0.05	6.29	0.15
Women	159	0.78 ± 0.19	0.79 ± 0.20	1.55	.04[a]	0.14 (80.7)	0.93 (0.91–0.95)	0.05	6.56	0.14
R4: Familiarization 4–24 weeks										
Men	98	0.87 ± 0.19	0.89 ± 0.20	2.29	.02[a]	0.26 (28.4)	0.92 (0.87–0.94)	0.05	6.22	0.15
Women	159	0.80 ± 0.20	0.81 ± 0.20	1.00	.18	0.1	0.92 (0.90–0.94)	0.06	6.94	0.16
SPPB										
R1: Familiarization 1–familiarization 2										
Men	98	8.49 ± 2.01	8.80 ± 2.15	3.60	<.01[a]	0.28 (22.4)	0.84 (0.77–0.89)	0.80	9.28	2.22
Women	159	7.76 ± 2.45	8.08 ± 2.38	4.22	<.01[a]	0.28 (17.0)	0.88 (0.83–0.91)	0.82	10.37	2.28
R2: Familiarization 2–baseline										
Men	98	8.80 ± 2.15	8.99 ± 2.22	2.21	.11	0.16 0.85	(0.79–0.90)	0.84	9.39	2.32
Women	159	8.08 ± 2.38	8.30 ± 2.46	2.72	.02[a]	0.18 (60.0)	0.87 (0.83–0.90)	0.86	10.49	2.38
R3: Familiarization 3–12 weeks										
Men	98	9.39 ± 2.11	9.56 ± 2.17	1.84	.03[a]	0.21 (45.9)	0.93 (0.89–0.95)	0.57	5.97	1.57
Women	159	8.40 ± 2.26	8.52 ± 2.29	1.50	.14	0.11	0.89 0.85–0.92)	0.76	9.01	2.11
R4: Familiarization 4–24 weeks										
Men	98	9.89 ± 2.08	9.85 ± 2.21	–0.41	.65	-0.05	0.92 (0.88–0.94)	0.63	6.34	1.73
Women	159	8.75 ± 2.37	9.05 ± 2.16	3.45	<.01[a]	0.27 (19.3)	0.87 (0.82–0.91)	0.80	8.98	2.21

[a]Systematic variability.
The Cohen's d effect size (ES) and chance that the value of the statistic is trivial (%) for the significant systematic changes are also provided.

Table 2. Means, Standard Deviations and Reliability Statistics for Isokinetic Leg Extension and Flexion Peak Torque (PT) at 1.05 rad/s at Familiarization 1 and Familiarization 2, Familiarization 2 and Baseline Testing, Familiarization 3 and 12-Week Testing, and Familiarization 4 and 24-Week Testing

	n	1	2	Change, %	*P*-Value	ES (%)	Intraclass Correlation Coefficient (95% Confidence Interval)	Standard Error of the Measurement	Coefficient of Variation, %	Minimum Detectable Change
		Mean±Standard Deviation								
Isokinetic leg extension PT at 1.05 rad/s, Nm										
R1: Familiarization 1–familiarization 2										
Men	95^	82.92 ± 38.48	88.43 ± 40.03	6.64	<.01a	0.32 (11.8)	0.90 (0.84–0.93)	12.12	14.15	33.59
Women	159	51.41 ± 22.6	53.81 ± 22.30	4.68	<.01a	0.25 (26.5)	0.90 (0.86–0.93)	6.94	13.19	19.23
R2: Familiarization 2–baseline										
Men	96^	88.84 ± 40.02	86.19 ± 38.59	–2.98	.03a	–0.23 (38.8)	0.95 (0.93–0.97)	8.39	9.59	23.26
Women	159	53.81 ± 22.30	54.13 ± 20.96	0.59	.65	0.04	0.92 (0.89–0.94)	6.14	11.37	17.01
R3: Familiarization 3–12 weeks										
Men	98	93.40 ± 37.79	92.73 ± 37.75	±0.71	.58	-0.06	0.95 (0.93–0.97)	8.44	9.07	23.39
Women	158^ 5	7.17 ± 23.52	55.84 ± 20.98	2.33	.23	-0.10	0.80 (0.74–0.85)	9.89	17.50	27.41
R4: Familiarization 4–24 weeks										
Men	98	94.00 ± 36.61	91.93 ± 35.26	±2.21	.06	-0.19	0.95 (0.93–0.97)	7.74	8.07	21.46
Women	159	57.11 ± 20.77	56.37 ± 20.08	±1.28	.21	-0.10	0.93 (0.91–0.95)	5.22	9.20	14.48
Isokinetic leg flexion PT at 1.05 rad/s, Nm										
R1: Familiarization 1–familiarization 2										
Men	95^	42.11 ± 23.19	46.47 ± 23.22	10.34	<.01a	0.34 (8.3)	0.84 (0.75–0.89)	8.96	20.22	24.83
Women	159	27.35 ± 11.81	29.23 ± 10.89	6.84	<.01a	0.26 (23.6)	0.78 (0.71–0.84)	5.18	18.32	14.36
R2: Familiarization 2–baseline										
Men	96^	46.58 ± 23.	12 49.73 ± 22.88	6.77	<.01a	0.31 (14.1)	0.90 (0.84–0.93)	7.19	14.94	19.94
Women	159	29.23 ± 10.89	30.75 ± 11.46	5.21	.02a	0.19 (55.1)	0.74 (0.66–0.81)	5.59	18.65	15.50
R3: Familiarization 3–12 weeks										
Men	98	50.75 ± 20.62	51.84 ± 20.62	2.14	.17	0.14	0.93 (0.90–0.95)	5.45	10.62	15.10
Women	158^	31.64 ± 11.61	31.79 ± 10.67	0.49	.80	0.02	0.78 (0.71–0.83)	5.29	16.70	14.68
R4: Familiarization 4–24 weeks										
Men	98	52.86 ± 21.49	52.56 ± 22.24	-0.58	.75	-0.03	0.90 (0.86–0.94)	6.80	12.90	18.85
Women	159	32.06 ± 11.13	33.05 ± 11.49	3.07	.04a	0.16 (69.5)	0.85 (0.80–0.89)	4.29	13.19	11.91

[a]Systematic variability.

The Cohen's d effect size (ES) and chance that the value of the statistic is trivial (%) for the significant systematic changes are also provided.

Table 3. Means, Standard Deviations and Reliability Statistics for Isokinetic Leg Extension and Flexion Peak Torque (PT) at 3.14 rad/s at Familiarization 1 and Familiarization 2, Familiarization 2 and Baseline Testing, Familiarization 3 and 12-Week Testing, and Familiarization 4 and 24-Week Testing

	n	1	2	Change, %	*P*-Value	ES (%)	Intraclass Correlation Coefficient (95% Confidence Interval)	Standard Error of the Measurement	Coefficient of Variation, %	Minimum Detectable Change
		Mean±Standard Deviation								
Isokinetic leg extension PT at 3.14 rad/s, Nm										
R1: Familiarization 1–familiarization 2										
Men	94^	50.57 ± 25.84	53.17 ± 27.19	5.15	.03a	0.23 (38.7)	0.90 (0.86–0.94)	8.11	15.63	22.47
Women	159	30.57 ± 15.07	31.90 ± 15.89	4.35	.01a	0.20 (50.0)	0.91 (0.87–0.93)	4.70	15.04	13.02
R2: Familiarization 2–baseline										
Men	96^	53.13 ± 26.90	52.70 ± 26.77	-0.81	.68	0.04	0.93 (0.90–0.95)	7.12	13.45	19.72
Women	159	31.90 ± 15.89	31.87 ± 15.16	±0.09	.96	±0.01	0.91 (0.79–0.93)	4.70	14.74	13.02
R3: Familiarization 3–12 weeks										
Men	98	56.27 ± 27.53	55.65 ± 27.81	-1.10	.42	±0.08	0.96 (0.95–0.98)	5.27	9.42	14.61
Women	157^	34.24 ± 16.10	33.80 ± 16.33	-1.27	.48	±0.06	0.89 (0.85–0.92)	5.49	16.15	15.23
R4: Familiarization 4–24 weeks										
Men	98	57.62 ± 26.37	56.56 ± 25.64	-1.85	.17	±0.14	0.96 (0.94–0.97)	5.42	9.50	15.03
Women	159	34.46 ± 15.35	33.68 ± 14.93	-2.29	.10	±0.13	0.92 (0.90–0.94)	4.19	12.30	11.61
Isokinetic leg flexion PT at 3.14 rad/s, Nm										
R1: Familiarization 1–familiarization 2										
Men	95^	28.21 ± 16.26	31.07 ± 16.61	10.14	<.01[a]	0.31 (23.1)	0.83 (0.74–0.89)	6.59	22.25	18.28
Women	159	17.40 ± 8.55	17.67 ± 8.29	1.57	.65	0.04 0.59	(0.48–0.69)	5.38	30.67	14.90
R2: Familiarization 2–baseline										
Men	97^	30.74 ± 16.60	30.80 ± 16.80	0.21	.94	0.01	0.87 (0.81–0.91)	6.09	19.79	16.88
Women	159	17.67 ± 8.29	18.55 ± 8.97	4.96	.08	0.14	0.74 (0.66–0.80)	4.39	24.25	12.17
R3: Familiarization 3–12 weeks										
Men	98	31.89 ± 16.98	32.95 ± 16.52	3.34	.21	0.13	0.87 (0.82–0.91)	5.94	18.32	16.46
Women	158^	20.11 ± 10.44	19.36 ± 9.05	-3.71	.23	0.10	0.69 (0.60–0.76)	5.44	27.58	15.09
R4: Familiarization 4–24 weeks										
Men	98	32.84 ± 18.32	32.07 ± 17.68	-2.34	.29	0.11	0.92 (0.88–0.95)	5.10	15.71	14.13
Women	159	20.47 ± 9.51	20.24 ± 9.53	-1.12	.65	0.04	0.78 (0.71–0.84)	4.47	21.94	12.38

a. Systematic variability.

The Cohen's d effect size (ES) and chance that the value of the statistic is trivial (%) for the significant systematic changes are also provided.

the square root of the mean square error from the ANOVA table.[9,12] The coefficient of variation (CV) was calculated as a normalized measure of the SEM.[12] The MDC was computed by calculating a 95% interval about the SEM using the equation described by Weir.[9]

All data were analyzed using SPSS version 22 (IBM, Armonk, NY) and a custom-written spreadsheet (Excel, Microsoft, Redmond, WA). A type I error rate was set a priori at 5%.

Results

The means, standard deviations, and test–retest reliability statistics for the dependent variables at R1, R2, R3, and R4 are provided in Tables 1–3, Table S1, and Figures S2 and S3. The ICCs for each dependent variable at R1, R2, R3, and R4 were all greater than 0 ($P \leq .05$).

Discussion

This is the first study to quantify the test–retest reliability of commonly used clinical assessments of physical function in sarcopenic, malnourished older men and women over 24 weeks. The results indicate that handgrip strength, gait speed, SPPB score, and isokinetic leg extension and flexion PT had ICCs that were greater than 0 (all ≥ 0.59) at all four of the test–retest time points, although most of these variables demonstrated systematic increases at R1, and several of the variables, such as gait speed, SPPB score, and leg flexion PT at 1.05 rad/s, exhibited systematic variability beyond the baseline testing session. Finally, ranges are provided for SEM and MDC scores (Table S1) for each of the tests that may be used to evaluate individual responses to interventions.

All of the variables in the present study demonstrated relative reliability (ICCs) greater than 0 (Figure S2) at each of the time points assessed, although there were several variables for which the ICCs improved with familiarization. For example, the ICCs for gait speed, SPPB score, and isokinetic leg extension and flexion at 1.05 rad/s improved from R1 and/or R2 to R3 and/or R4 in men. Similarly, the ICCs for gait speed, handgrip strength, and isokinetic leg flexion at 3.14 rad/s improved from R1 and/ or R2 to R3 and/ or R4 in women, indicating that familiarization may increase the relative reliability of these variables in sarcopenic, malnourished older men and women.

The ICCs for isokinetic leg flexion in women were generally lower than for isokinetic leg extension (Figure S2), and this difference was more evident at 3.14 rad/s than 1.05 rad/s. A previous study[5] demonstrated that the variability in PT over three isokinetic repetitions was greater in sarcopenic, malnourished elderly individuals than in those who were healthy, greater for leg flexion than extension, and greater at 3.14 rad/s than 1.05 rad/s. It also found a significant negative relationship between intraindividual variability and strength.[5] Therefore, it is possible that the lower ICCs for leg flexion PT in the sarcopenic, malnourished women in the present study was a function of lower leg flexion strength. It has been suggested that the inability of older adults to express strength reliably may be a direct consequence of detrimental age-related changes in neuromuscular function,[13,14] and low leg flexion strength has been associated with greater risk of falls in older adults.[15–17] Thus, although isokinetic leg flexion strength tests may exhibit lower reliability in sarcopenic, malnourished older adults, increasing leg flexion strength may be a way to enhance physical function in this population.

With the exception of handgrip strength in women, gait speed in men, and isokinetic leg flexion PT at 3.14 rad/s in men, there were systematic increases for each variable from Familiarization 1 to 2 (at R1). Furthermore, gait speed, SPPB score, and leg extension and flexion PT at 1.05 rad/s exhibited systematic variability beyond R1, although as previously described, "(s)upraoptimal sample sizes can produce. . . statistically significant effects that are likely to be clinically useless."[7] Therefore, because of the large sample sizes in this study, magnitude-based inferences were used to measure the clinical meaningfulness of these changes objectively.[7] The magnitude-based inferences indicated that the significant changes ($P \leq .05$) in SPPB score at R2 (% chance of indicated outcome (likely or trivial) = 60.0%; outcome = probably trivial), gait speed at R2 (54.3%; probably trivial) and R3 (80.7%; likely trivial), and isokinetic leg flexion PT at 1.05 rad/s at R2 (55.1%; probably trivial) and R4 (69.5%; probably trivial) for women were trivial (Tables 1–3). The meaningful systematic changes at R1 further establish the importance of familiarization when using these variables in sarcopenic, malnourished older adults. Finally, the meaningful, systematic changes beyond R1 for variables such as gait speed in men and SPPB score in men and women demonstrate that significant changes in large-scale clinical trials should be interpreted with caution and emphasize the need for a metric that provides a standard for researchers and clinicians to evaluate meaningful changes. In addition, because the SPBB is a composite of several tests, researchers may choose to report the results of the individual tests and the composite SPBB score.

Previous authors have used or suggested the use of the SEM and MDC as indices for investigators seeking to understand the importance of an individual's change based on the precision of the measurement.[2,18] The SEM provides a quantification of the trial-to-trial noise of a test and has been described as an index that "links the reliability of the measurement. . . to the standard deviation of the population."[9,18] The MDC is calculated as a 95% confidence interval (CI) about the SEM and has been described as "the difference needed between separate measures on a subject for the difference in the measures to be considered real."[9] In other words, investigators can be 95% confident that an individual's change exceeds the measurement error if the change is equal to or greater than the MDC, although the 95% CI used to calculate the MDC may be too conservative in a clinical setting. For example, a skeletal muscle mass index 1 standard deviation unit below the mean in young adults (a value below the 68% confidence interval) has been used to define Class I sarcopenia, and a value 2 standard deviation units below

the mean (a value below the 95% CI) has been used to define Class II sarcopenia in older adults.[1,19] Consequently, previous investigators have used the SEM as an indicator of clinically meaningful change for individuals on a test.[2,18] Therefore, in addition to reporting the results of null hypothesis testing, investigators may wish to report the proportion of subjects with improvements that exceeded the SEM as a moderate change and the MDC as a large change. Accordingly, a range of SEM and MDC values have been provided for each of the seven variables examined in this study (Table S1) that can be used to assess changes after an intervention on a subject-by-subject basis.

Much like the ICCs, a qualitative examination suggested that the SEM and MDC tended to decrease with greater familiarization (Tables 1–3). In addition, MDC for isokinetic leg flexion PT in men and women were large, especially at 3.14 rad/s (Table S1), suggesting that the absolute reliability of these measurements may be lower than that of the other measurements examined in this study. Therefore, combined with the low ICCs, isokinetic leg flexion may be unreliable in sarcopenic, malnourished older adults.

Overall, the results of the present study indicated that handgrip strength, gait speed, SPPB score, and isokinetic leg extension and flexion PT had relative reliability scores that were greater than 0 at R1, R2, R3, and R4, although there were systematic increases from Familiarization 1 to Familiarization 2 and the ICCs and SEMs generally improved with familiarization for most of the variables examined (Figure S4). Together, these data emphasize the need for at least one familiarization trial for these measurements in sarcopenic, malnourished older adults. These data also stress the importance of control groups in interventions that use these measurements as outcome variables. Finally, in an attempt to interpret the clinical importance of statistically significant results in clinical trials, a three-tiered approach that includes null hypothesis testing, examination and interpretation of the effect magnitude, and comparison of individual changes relative to the SEM and MDC of the measurements used is recommended.

Acknowledgments

The authors would like to thank Dr. Vikkie Mustad for her thoughtful discussions and input.

Conflict of Interest: When this manuscript was submitted, Dr. Nathaniel D.M. Jenkins had disclosed a financial interest as a paid consultant with General Nutrition Centers, Inc. (GNC) to present at a GNC-sponsored symposium. Dr. Joel T. Cramer had disclosed a significant financial interest in Abbott Nutrition as a paid consultant whose responsibilities were to assist with clinical trial design, analysis, interpretation of data, and training key study personnel at all clinical trial sites. In accordance with its Conflict of Interest policy, the University of Nebraska-Lincoln's Conflict of Interest Review Committee has determined that this must be disclosed.

Financial Disclosure: This clinical trial was funded in part by Abbott Nutrition (clinicaltrials.gov identifier: NCT01191125). Dr. Joel T. Cramer is supported by the U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch project (Accession Number: 1009500).

Author Contributions: NDMJ: primary manuscript writer. JTC: primary manuscript reviewer. NDMJ, JTC: data analysis and interpretation, approval of final manuscript for submission.

Sponsor's Role: The study sponsor did not contribute to analysis or interpretation of this data, nor preparation of this manuscript.

References

1. Cruz-Jentoft AJ, Baeyens JP, Bauer JM et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing 2010;39:412–423.
2. Perera S, Mody SH, Woodman RC et al. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006;54:743–749.
3. Kwon S, Perera S, Pahor M et al. What is a meaningful change in physical performance? Findings from a clinical trial in older adults (the LIFE-P study). J Nutr Health Aging 2009;13:538–544.
4. Detsky AS, McLaughlin JR, Baker JP et al. What is subjective global assessment of nutritional status? JPEN J Parenter Enteral Nutr 1987;11:8–13.
5. Cramer JT, Jenkins ND, Mustad VA et al. Isokinetic dynamometry in healthy versus sarcopenic and malnourished elderly: Beyond simple measurements of muscle strength. J Appl Gerontol. 2015.
6. Guralnik JM, Simonsick EM, Ferrucci L et al. A short physical performance battery assessing lower extremity function: Association with self reported disability and prediction of mortality and nursing home admission. J Gerontol 1994;49:M85–M94.
7. Hopkins WG. A spreadsheet for deriving a confidence interval, mechanistic inference and clinical inference from a P value. Sportscience 2007;11:16– 20.
8. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–428.
9. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231–240.
10. Vincent WJ, Weir JP. Statistics in Kinesiology, 4th Ed. Champaign, IL: Human Kinetics, 2012.
11. Jenkins ND, Buckner SL, Cochrane KC et al. Age-related differences in rates of torque development and rise in EMG are eliminated by normalization. Exp Gerontol 2014;57C:18–28.
12. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30:1–15.
13. Manini TM, Hong SL, Clark BC. Aging and muscle: A neuron's perspective. Curr Opin Clin Nutr Metab Care 2013;16:21–26.

14. Jenkins ND, Buckner SL, Bergstrom HC et al. Reliability and relationships among handgrip strength, leg extensor strength and power, and balance in older men. Exp Gerontol. 2014;58:47–50.
15. Bento PC, Pereira G, Ugrinowitsch C et al. Peak torque and rate of torque development in elderly with and without fall history. Clin Biomech 2010;25:450–454.
16. Palmer TB, Thiele RM, Thompson BJ. Age-related differences in maximal and rapid torque characteristics of the hip extensors and dynamic postural balance in healthy young and old females. J Strength Cond Res 2016. May 21. [Epub ahead of print]
17. Ding L, Yang F. Muscle weakness is related to slip-initiated falls among community-dwelling older adults. J Biomech 2016;49:238–243.
18. de Vet HC, Terwee CB, Ostelo RW et al. Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 2006;4:54.
19. Janssen I, Heymsfield SB, Ross R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. J Am Geriatr Soc 2002;50:889–896.

Supporting Information (following)

Figure S1. Experimental timeline.

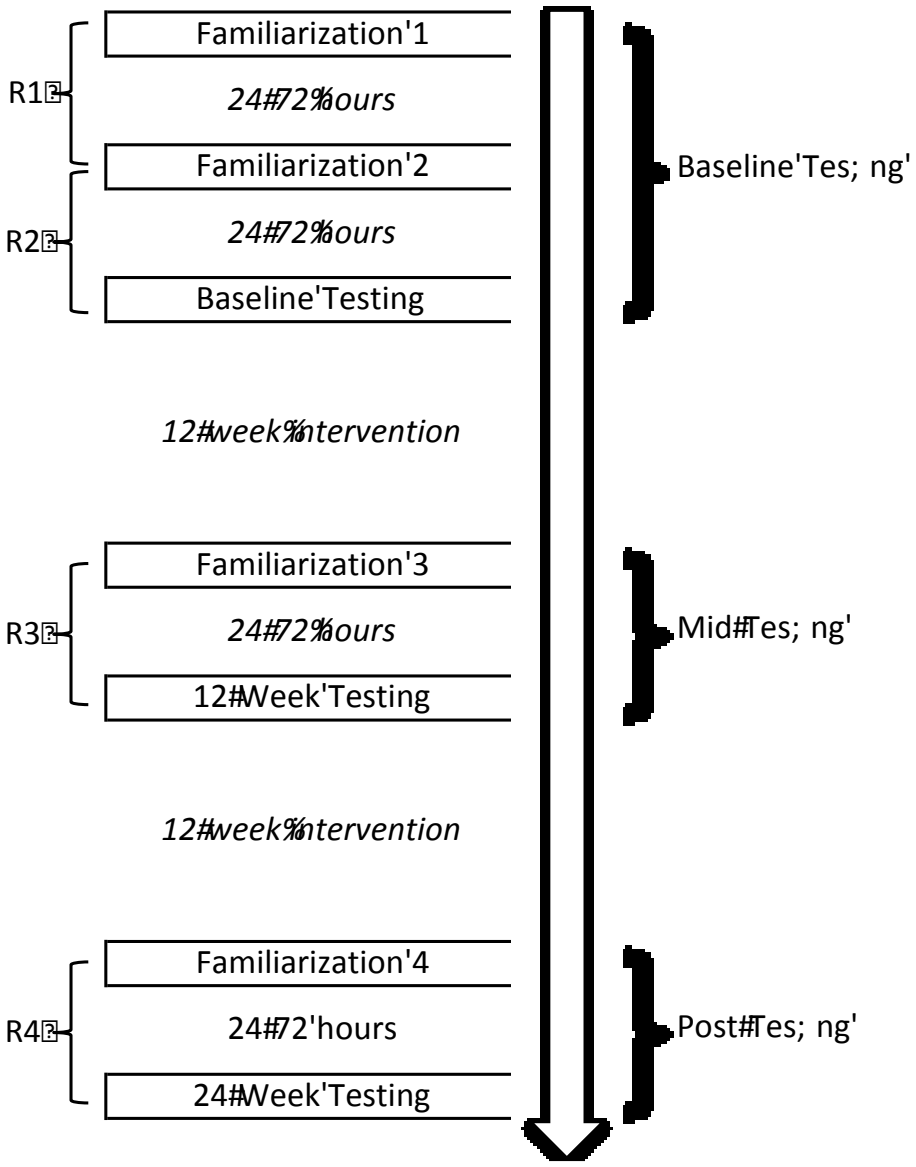
Figure S2. Intraclass correlation coefficients with 95% confidence intervals for each of the dependent variables in men (A, B, C) and women (D, E, F) at Familiarization 1 and Familiarization 2 (R1), Familiarization 2 and Baseline Testing (R2), Familiarization 3 and 12-Week Testing (R3), and Familiarization 4 and 24-Week Testing (R4).

Figure S3. Minimum detectable change (MDC; solid dots) and percentage change (% Change; open squares) for each of the dependent variables in men (A, B, C) and women (D, E, F) at Familiarization 1 and Familiarization 2 (R1), Familiarization 2 and Baseline Testing (R2), Familiarization 3 and 12-Week Testing (R3), and Familiarization 4 and 24-Week Testing (R4).

Figure S4. Intraclass correlation coefficients and coefficients of variation for each dependent variables in men (A and C, respectively) and women (B and D, respectively) at Familiarization 1 and Familiarization 2 (R1), Familiarization 2 and Baseline Testing (R2), Familiarization 3 and 12-Week Testing (R3), and Familiarization 4 and 24-Week Testing (R4).

Table S1. Absolute and Relative (Percentage of Grand Mean) Standard Error of the Measurement (SEM) and Minimum Detectable Change (MDC) Range

Figure S1. Experimental timeline. Handgrip strength, isokinetic leg extension strength, and the Short Physical Performance Battery were assessed at Familiarization 1, Familiarization 2, baseline testing, Familiarization 3, 12-week testing, Familiarization 4, and 24-week testing.



Supplementary Figure S2. Intraclass correlation coefficients with 95% confidence intervals for each of the dependent variables in men (A, B, C) and women (D, E, F) at R1, R2, R3, and R4.

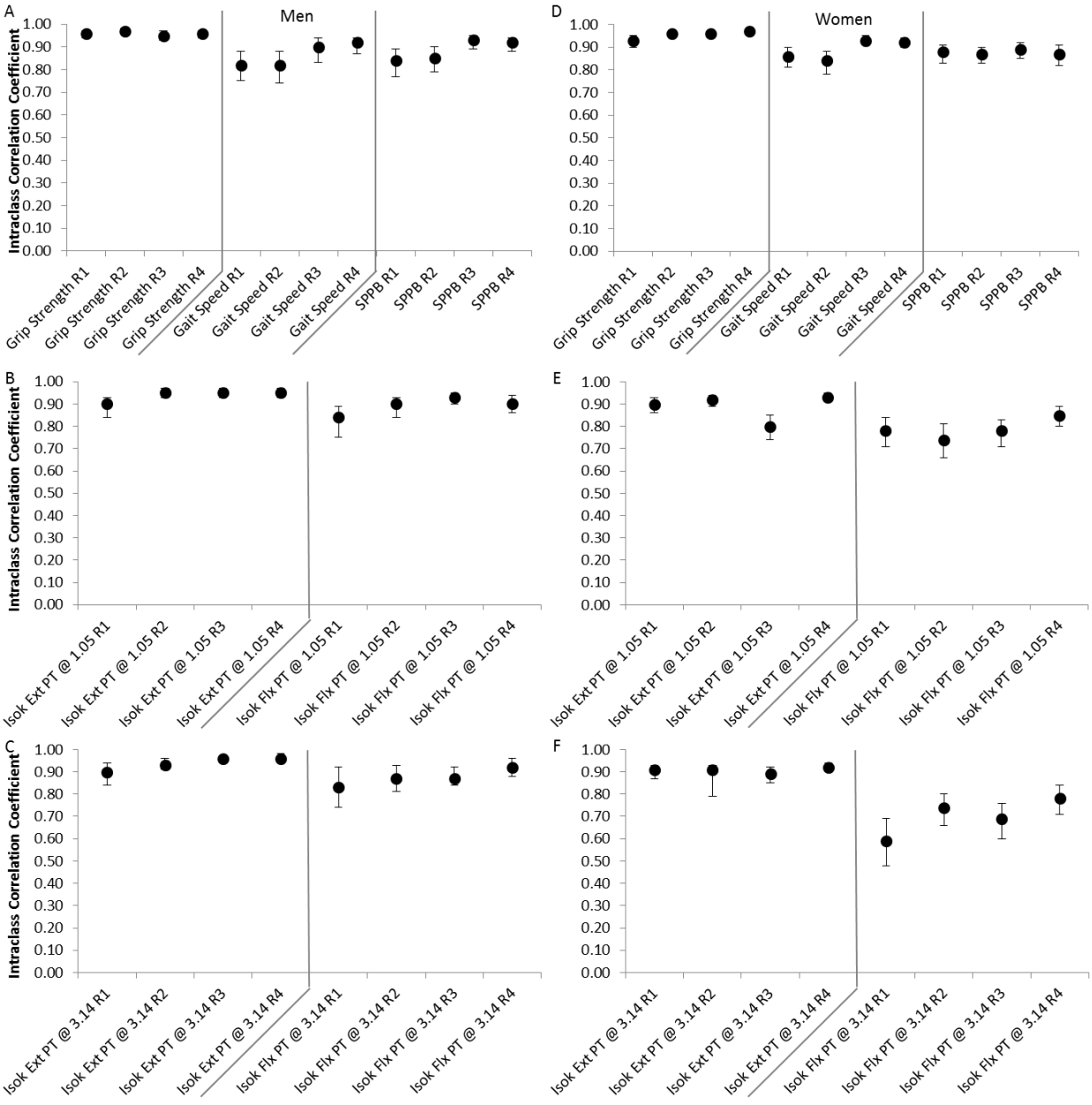
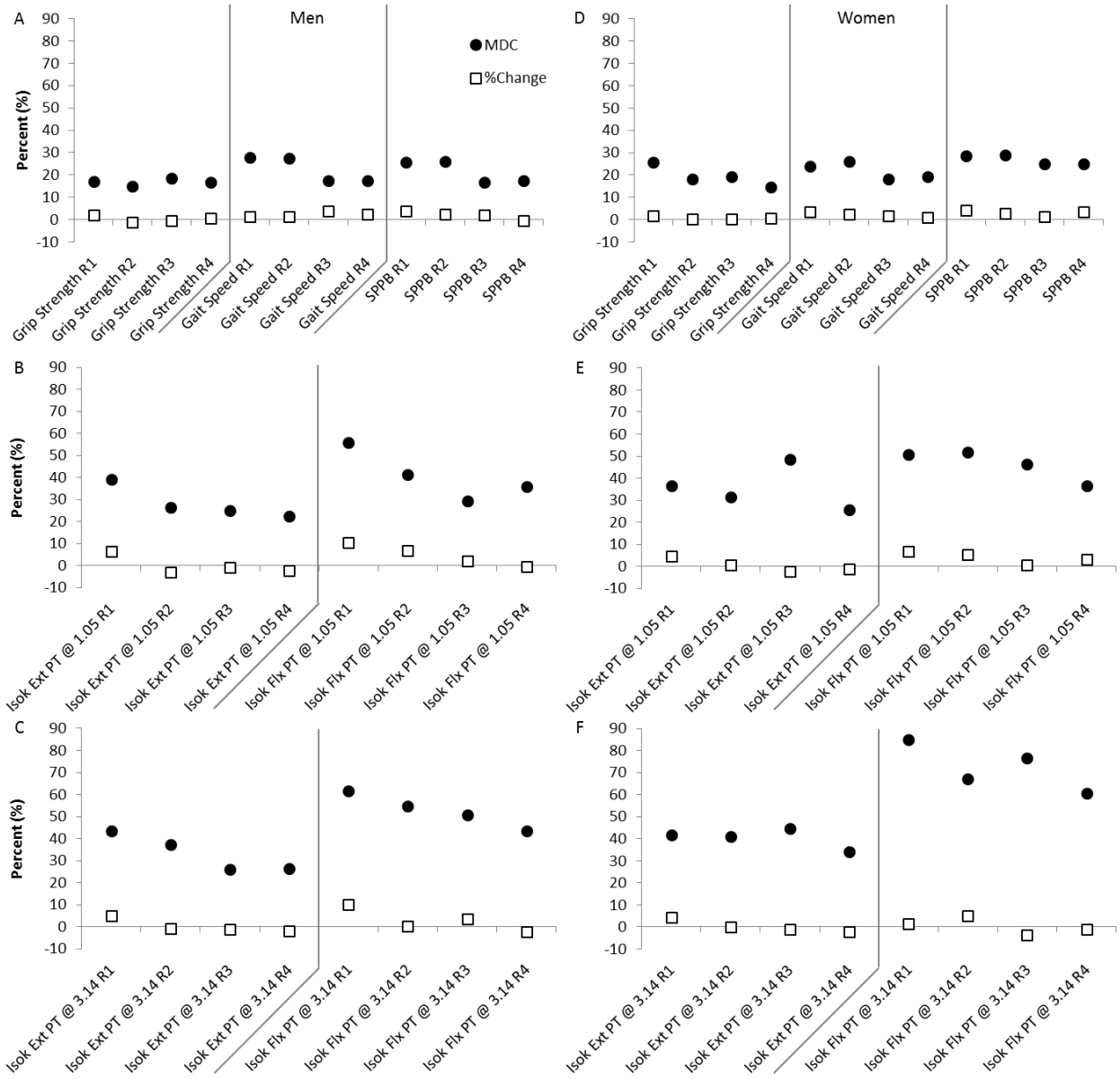


Figure S3. Minimum detectable change (MDC; solid dots) and percentage change (%Change; open squares) for each of the dependent variables in men (A, B, C) and women (D, E, F) at R1, R2, R3, and R4.



Supplementary Figure S4. Intraclass correlation coefficients and coefficients of variation for each dependent variables in men (A and C, respectively) and women (B and D, respectively) at R1, R2, R3, and R4.

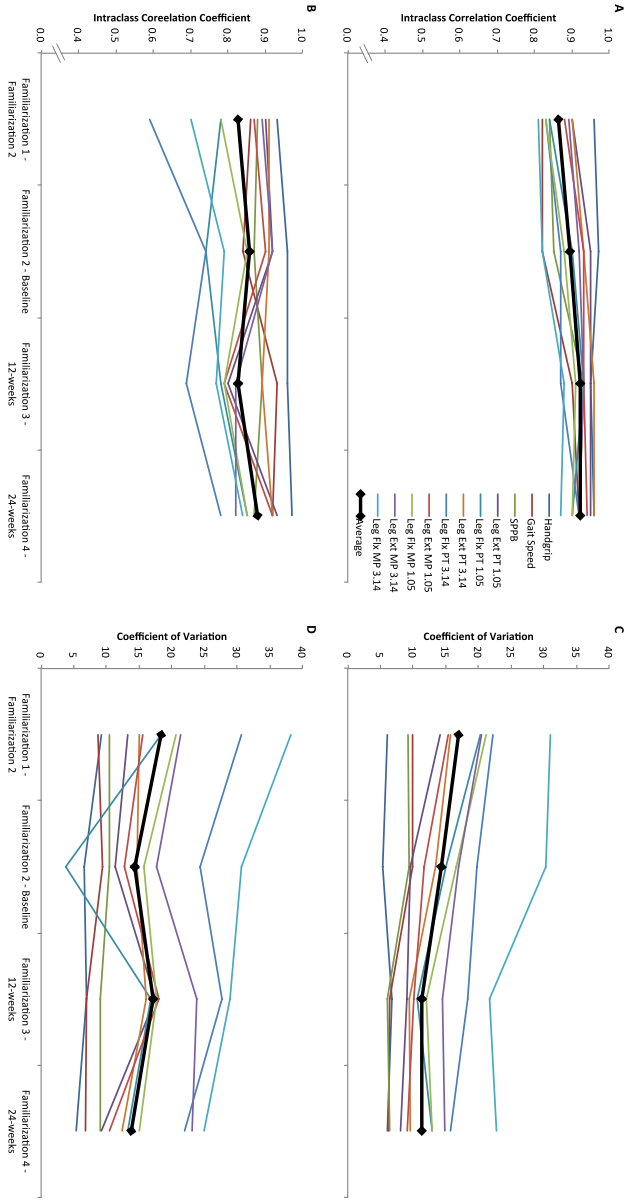


Table S1. Absolute and Relative (Percentage of Grand Mean) Standard Error of the Measurement (SEM) and Minimum Detectable Change (MDC) Range

Variable	Absolute		Relative (%; CV)	
	SEM	MDC	SEM	MDC
Men				
Grip strength, kg	1.51–1.98	4.18–5.50	5.35–6.68	14.83–18.52
Gait speed, m/s	0.05–0.08	0.15–0.21	6.22–9.97	17.05–27.10
SPPB score	0.57–0.84	1.57–2.32	5.97–9.39	16.57–26.08
Isokinetic leg extension PT 1.05 rad/s, Nm	7.74–12.12	21.46–33.59	8.07–14.15	23.08–39.21
Isokinetic leg flexion PT 1.05 rad/s, Nm	5.45–8.96	15.10–24.83	10.62–20.22	34.09–48.41
Isokinetic leg extension PT 3.14 rad/s, Nm	5.27–8.11	14.61–22.47	9.42–15.63	26.11–43.32
Isokinetic leg flexion PT 3.14 rad/s, Nm	5.10–6.59	14.13–18.28	15.71–22.25	43.54–61.67
Women				
Grip strength, kg	0.96–1.56	2.67–4.31	5.28–9.30	14.65–25.75
Gait speed, m/s	0.05–0.07	0.14–0.20	6.56–9.39	17.83–26.67
SPPB score	0.76–0.86	2.21–2.38	9.01–10.49	26.12–29.06
Isokinetic leg extension PT 1.05 rad/s, Nm	5.22–9.89	14.48–27.41	9.20–17.50	25.52–48.51
Isokinetic leg flexion PT 1.05 rad/s, Nm	4.29–5.59	11.91–15.50	13.19–18.65	36.58–51.68
Isokinetic leg extension PT 3.14 rad/s, Nm	4.19–5.49	11.61–15.23	12.30–16.15	34.08–44.77
Isokinetic leg flexion PT 3.14 rad/s, Nm	4.39–5.44	12.17–15.09	24.25–27.58	67.20–76.46

SPPB=Short Physical Performance Battery; PT=peak torque.