



Absolute reliability of measurements of muscle strength and physical performance measures in older people with high functional capacities

Yoshikazu Suzuki¹ · Naoto Kamide^{2,3} · Yui Kitai² · Masataka Ando⁴ · Haruhiko Sato^{2,3} · Shiba Yoshitaka² · Miki Sakamoto²

Received: 27 March 2019 / Accepted: 28 June 2019 / Published online: 5 July 2019
© European Geriatric Medicine Society 2019

Key summary points

Aim The aim of this study was to clarify the absolute reliability of muscle strength and physical performance measures in older people.

Findings Only the five-times chair stand test was found to have a significant systematic error. The minimum detectable changes to the measurement value of grip strength, 5-m walking time, and the timed up and go test were all < 10%, whereas that of knee extension strength was 12%.

Message Changes beyond the minimum detectable change are necessary to detect true changes in muscle strength and physical performance tests over time in community-dwelling older people.

Abstract

Purpose The purpose of this study was to clarify the absolute reliability of muscle strength and physical performance measures in older people.

Methods The participants were 718 community-dwelling older people who were living independently. Muscle strength and physical performance tests were administered twice for all participants by the same rater. Grip and knee extension strength during isometric contractions were used as muscle strength tests, and the five-times chair stand test (FCST), 5-m walking time at comfortable pace, and the timed up and go test (TUG) as physical performance tests. Bland–Altman analysis was performed to determine the systematic errors for each muscle strength and physical performance test, and the amount of errors was estimated using the minimum detectable change (MDC). Further, %MDC was calculated by dividing the MDC with the mean of two measurement values for each test.

Results No systematic errors were found in any of the muscle strength or physical performance tests, except for the FCST, for which MDC and %MDC could not be calculated. The %MDC of grip strength, 5-m walking time, and the TUG were estimated at < 10%, whereas that of knee extension strength was 12%. Age and gender were not found to affect any systematic errors or MDC and %MDC.

Conclusion Grip strength, 5-m walking time, and the TUG, which all have excellent reliability, were suggested to be appropriate indexes as outcome measures of muscle strength and physical performance in community-dwelling older people.

Keywords Minimum detectable change · Muscle strength · Older people · Physical performance · Reliability

✉ Naoto Kamide
naokami@kitasato-u.ac.jp

¹ Department of Rehabilitation, Kitasato University East Hospital, Sagamihara, Japan

² School of Allied Health Sciences, Kitasato University, 1-15-1 Kitasato, Minami-ku, Sagamihara, Kanagawa 252-0373, Japan

³ Graduate School of Medical Sciences, Kitasato University, Sagamihara, Japan

⁴ Graduate School of Medicine, Hirosaki University, Hirosaki, Japan

Introduction

Generally, in older people, adverse health-related outcomes such as frailty, sarcopenia, falling, and fractures are well known to increase with age [1, 2]. These adverse health-related outcomes are critical risk factors in terms of causing functional disabilities in older people who live independently in their community. Muscle strength and physical performance tests are commonly used in clinical and community settings to evaluate the risk of frailty, falling, and sarcopenia [3–5]. Physical exercise interventions are carried out to help prevent frailty, falling, and sarcopenia, and muscle strength and physical performance tests are used to verify their effectiveness [6]. However, estimating measurement errors for muscle strength and physical performance tests in community-dwelling older people is indispensable for accurate evaluation. On the other hand, measurement errors for muscle strength and physical performance tests for older people have not been clearly defined because previous studies have been insufficient in terms of having relatively small sample sizes and participants that did not include older people with high functional capacities [7, 8]. These issues must be addressed to clarify the absolute reliability of muscle strength and physical performance tests and establish accurate evaluations. Verifying the absolute reliability and determining the reasonable measurement errors in muscle strength and physical performance tests could, therefore, contribute to the establishment of judgment criteria for the effectiveness of physical exercise interventions and the risk of functional decline. Especially, the determination of judgment criteria could be a useful index for the evaluation of intervention effects and the risk of functional decline at the individual rather than the population level.

The purpose of the present study was to clarify the absolute reliability of muscle strength and physical performance tests for community-dwelling older people with high functional capacities.

Materials and methods

Participants

We analyzed data obtained from 718 community-dwelling older volunteers recruited from Sagami city in Kanagawa Prefecture, Japan, using advertisements in newspapers and community newsletters. The inclusion criteria were as follows: age 65 years and older, and able to perform activities of daily living (ADL) independently. The participants who could perform ADL independently were defined as those who lacked certification of care level by long-term care insurance (LTCI) system, and this was confirmed by

interviews with experienced researchers at recruitment. The LTCI system is a public insurance system established by the Japanese government that covers all Japanese individuals aged 65 years and older. An individual's certified care level is determined according to uniform criteria under the LTCI, and is assessed by a trained investigator and a primarily physician. A certified care level means that an older person requires support to perform ADL [9]. Further, participants suspected of having dementia based on these interviews were excluded.

This study was approved by the Institutional Review Board of the School of Allied Health Sciences at Kitasato University (approval number 2016-G021B), and written informed consent was obtained from all participants.

Muscle strength and physical performance tests

We conducted two different isometric muscle strength tests and four different physical performance tests, as described below. For the isometric muscle strength tests, we administered a grip strength test and a knee extension strength test. For the physical performance tests, we administered the five-times chair stand test (FCST) and the timed up and go test (TUG), and measured 5-m walking time at a comfortable pace. These tests have been shown to be useful to predict adverse health-related events such as falling and functional disabilities in community-dwelling older people. To verify the test–retest reliability and the measurement errors for all tests, all measurements were carried out twice by the same experienced physical therapist.

1. Grip strength

Grip strength, which is a simple and inexpensive isometric muscle strength test, is commonly used as a diagnosis criterion in sarcopenia and frailty [5, 10, 11]. Further, a meta-analysis found grip strength to be a predictable index for adverse outcomes in older people [12]. In this study, maximum voluntary grip strength was measured using a Smedley-type dynamometer (T.K.K.5401, TAKEI Scientific Instruments Co., Ltd., Niigata, Japan). All measurements were performed on the dominant hand with the elbow joint extended at the side of the trunk while the participant was in a standing position. The measurement precision was 0.1 kgf.

2. Knee extension strength

Knee extension strength is clinically used to assess isometric muscle strength for the quadriceps muscle and has been reported to be associated with mortality and functional decline in older people [13, 14]. In this study, maximum voluntary knee extension strength was measured in the right

leg using a handheld dynamometer (μ -Tas F-1; Anima Inc., Tokyo, Japan). All measurements were performed while the participants were in a seated position with the knee joint in 90° flexion. The sensor of the handheld dynamometer was secured on the lower part of the crus (top of the medial and lateral malleolus) using a belt. The measurement precision was 0.1 kgf.

3. Five-times chair stand test [15]

The FCST is clinically used to assess lower extremity function and has been reported to be associated with knee extension strength [16]. Recently, the FCST has also been recommended as a clinical parameter of muscle strength for the identification of sarcopenia [5]. For the FCST, the participants sat on a standard armless chair (height 42 cm) with their feet apart at shoulder width and their arms crossed in front of their chest. They were then instructed to stand up from the chair and sit back down five times as quickly as possible. The time required to complete the task was measured using a digital stopwatch (ALBA W072; Seiko Watch Corporation, Tokyo, Japan). The measurement precision was 0.01 s.

4. 5-m walking time

Walking time at a comfortable pace is simple test for an assessment of mobility and walking ability, and is commonly used as a diagnosis criterion in sarcopenia and frailty [5, 10, 11]. Further, walking time, compared with muscle strength and other physical performance tests such as balance and lower extremity function, has been suggested to be powerful predictor of impaired ADL in older people [17]. In this study, the total length of the walkway was set at 9 m, including acceleration and deceleration zones at the start and end of the walkway. All participants were instructed to walk on the walkway at their usual walking pace without any assistance. The time required to walk the 5-m length in the middle of the walkway was measured using a digital stopwatch (ALBA W072; Seiko Watch Corporation). The measurement precision was 0.01 s.

5. Timed up and go test [18]

The TUG, which was developed as a test of functional mobility for older people, is also used as a screening tool for the assessment of frailty [11] and fall risk in older people [19]. For the TUG, the participants were instructed to rise from a standard armless chair (height 42 cm), walk 3 m, turn around, walk back to the chair, and sit back down as quickly as possible. The time required to complete this task was measured using a digital stopwatch (ALBA W072; Seiko Watch Corporation). The measurement precision was 0.01 s.

Basic characteristics

The participants' age, height, body weight, body mass index, and Tokyo Metropolitan Institute of Gerontology Index of Competence (TMIG-IC) scores were recorded. The TMIG-IC, which is used to assess functional capacities higher than ADL in community-dwelling older people, consists of the following three sub-items: instrumental ADL (IADL), intellectual activity, and social roles [20]. TMIG-IC total scores range from 0 to 13 (the sub-scores for IADL, intellectual activity, and social roles range from 0 to 5, 0 to 4, 0 to 4, respectively), with higher scores indicating greater functional capacity.

Statistical analysis

With respect to the reliability of the isometric muscle strength and physical performance tests, relative and absolute reliability were verified based on two measurement values for each test. Intraclass correlation coefficients (ICCs) were calculated to investigate relative reliability [21]. For the purposes of the present study, the following ICC criteria were used: 0.90–0.99 indicated high reliability, 0.80–0.89 indicated good reliability, 0.70–0.79 indicated fair reliability, and < 0.69 indicated poor reliability [22]. Further, the detection of systematic errors and the calculation of the margin of error were carried out to investigate absolute reliability. For systematic errors, Bland–Altman analysis was performed. In addition, the presence or absence of fixed and proportional biases were investigated [23]. The presence of a fixed bias was determined using a one-sample *t* test [23] and effect size (Cohen's *d*) [24], and that of a proportional bias was detected based on the relationship between the differences and means of two measurement values using Pearson's correlation coefficient [25]. The presence of a fixed bias was defined as a probability of < 0.05 by the one-sample *t* test and an effect size (*d*) of > 0.5. The presence of a proportional bias was defined as a Pearson's correlation coefficient (*r*) of > 0.3. If no fixed or proportional biases were found, further analysis was performed to determine the margin of error. These criteria were used in accordance with a previous report [24].

Regarding the margin of error, the minimal detectable change (MDC) was calculated using the following formula: $1.96 \times \sqrt{2} \times SD \times \sqrt{(1 - ICC)}$ (SD: standard deviation) [26]. In addition, %MDC was calculated by dividing the MDC with the mean of two measurement values for each test. A %MDC of < 30% was considered acceptable, and a %MDC of < 10% was considered excellent [27].

Statistical analysis was performed using the R programming language and environment (R version 3.2.2) [28].

Results

The participants' demographic characteristics are presented in Table 1. The mean TMIG-IC total scores were 11.8 ± 1.6

Table 1 Characteristics of the participants in this study

Characteristics	Men (mean \pm SD)	Women (mean \pm SD)
<i>n</i>	197	521
Age (years)	73.4 \pm 5.3	71.2 \pm 4.5
Height (cm)	166.4 \pm 5.3	153.2 \pm 5.5
Body weight (kg)	63.1 \pm 8.0	51.9 \pm 7.6
Body mass index (kg/m ²)	22.8 \pm 2.5	22.1 \pm 3.1
TMIG-IC (/13 points)	11.8 \pm 1.6	12.5 \pm 1.2
IADL sub-score (5 points)	4.8 \pm 0.6	5.0 \pm 0.2

TMIG-IC Tokyo Metropolitan Institute of Gerontology Index of Competence, IADL instrumental activities of daily living

in men and 12.5 ± 1.2 in women, and 37.6% of men and 46.6% of women showed the best functional capacity (i.e., highest possible TMIG-IC score). Further, 81.3% of men and 97.3% of women had the highest possible IADL sub-score (five points) and could perform IADL independently.

To assess systematic bias, Bland–Altman analysis was conducted. Bland–Altman plots of the muscle strength and physical performance tests are shown in Fig. 1. The presence or absence of fixed and proportional biases is shown in Table 2. Fixed and proportional biases were found in the FCST, but not in any of the other tests. The ICCs and margins of error for the muscle strength and physical performance tests are shown in Table 3. Furthermore, we stratified the data by gender and age (< 75 and \geq 75 years) and analyzed the presence or absence of fixed and proportional biases. As a result, no fixed or proportional biases were found, except for the FCST in both gender and age groups. The ICCs of the total sample for grip and knee extension strength indicated high reliability, while those for the 5-m

Fig. 1 Bland–Altman plot representing the differences between repeated measurements of muscle strength and physical performance. Dotted lines indicate the mean difference between repeated measures. Bold lines indicate the 95% limit of agreement

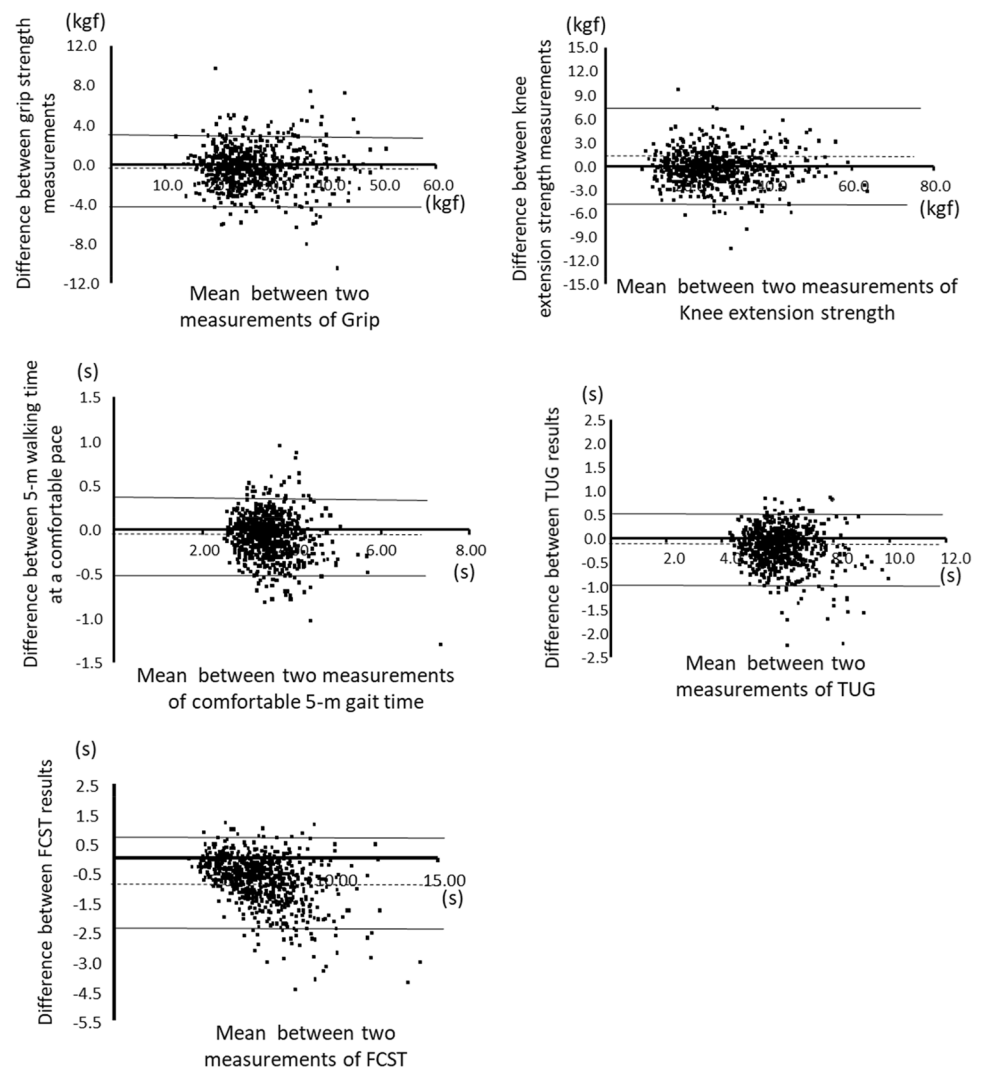


Table 2 Systematic errors for the muscle strength and physical performance tests

	First measurement	Second measurement	Mean difference	Fixed bias		Proportional bias	
	Mean \pm SD (min:max)	Mean \pm SD (min:max)	Mean \pm SD	<i>p</i> value	ES	<i>p</i> value	<i>r</i>
Grip strength (kgf)	26.3 \pm 6.8 (10.6:50.1)	26.0 \pm 6.8 (12.2:51.6)	-0.34 \pm 2.01	<0.05	0.1	0.60	0.0
Knee extension strength (kgf)	26.2 \pm 9.1 (7.8:63.3)	28.1 \pm 9.2 (8.8:65.8)	1.21 \pm 3.15	<0.05	0.2	0.28	0.0
FCST (s)	7.2 \pm 1.8 (3.5:15.9)	6.4 \pm 1.4 (3.5:12.4)	-0.76 \pm 0.81	<0.05	0.5 ^a	<0.05	0.4 ^a
5-m walking time (s)	3.6 \pm 0.6 (2.5:8.0)	3.5 \pm 0.5 (2.4:6.7)	-0.08 \pm 0.25	<0.05	0.2	<0.05	0.2
TUG (s)	6.0 \pm 1.0 (-4.0:11.2)	5.8 \pm 0.9 (4.0:9.3)	-0.19 \pm 0.40	<0.05	0.2	<0.05	0.2

FCST five-times chair stand test, TUG timed up and go test, *r* Pearson's correlation coefficient

^aPresence of fixed or proportional bias

Table 3 Relative and absolute reliability for the muscle strength and performance tests

	ICC (95%CI)	MDC	%MDC
Grip strength (kgf)			
Total (<i>n</i> = 718)	0.96 (0.95–0.96)	1.18	0.05
Men (<i>n</i> = 197)	0.91 (0.88–0.93)	2.2	0.06
Women (<i>n</i> = 521)	0.89 (0.87–0.90)	1.62	0.07
< 75 years (<i>n</i> = 506)	0.95 (0.95–0.96)	1.2	0.05
\geq 75 years (<i>n</i> = 212)	0.96 (0.95–0.97)	1.1	0.04
Knee extension strength (kgf)			
Total (<i>n</i> = 718)	0.90 (0.88–0.91)	3.29	0.12
Men (<i>n</i> = 197)	0.87 (0.84–0.90)	2.55	0.07
Women (<i>n</i> = 520)	0.88 (0.86–0.90)	2.98	0.12
< 75 years (<i>n</i> = 506)	0.90 (0.88–0.91)	2.81	0.10
\geq 75 years (<i>n</i> = 212)	0.88 (0.84–0.90)	2.91	0.12
FCST (s)			
Total (<i>n</i> = 715)	0.77 (0.74–0.80)	NA	NA
Men (<i>n</i> = 197)	0.78 (0.72–0.83)	NA	NA
Women (<i>n</i> = 518)	0.77 (0.73–0.80)	NA	NA
< 75 years (<i>n</i> = 505)	0.76 (0.72–0.79)	NA	NA
\geq 75 years (<i>n</i> = 210)	0.76 (0.69–0.81)	NA	NA
5-m walking time (s)			
Total (<i>n</i> = 717)	0.88 (0.87–0.90)	0.23	0.07
Men (<i>n</i> = 197)	0.88 (0.85–0.91)	0.23	0.07
Women (<i>n</i> = 520)	0.88 (0.86–0.90)	0.23	0.07
< 75 years (<i>n</i> = 505)	0.88 (0.86–0.90)	0.23	0.06
\geq 75 years (<i>n</i> = 212)	0.88 (0.84–0.91)	0.26	0.07
TUG (s)			
Total (<i>n</i> = 714)	0.88 (0.87–0.90)	0.37	0.06
Men (<i>n</i> = 197)	0.88 (0.84–0.90)	0.44	0.08
Women (<i>n</i> = 517)	0.89 (0.87–0.90)	0.35	0.06
< 75 years (<i>n</i> = 505)	0.85 (0.82–0.87)	0.43	0.08
\geq 75 years (<i>n</i> = 209)	0.90 (0.87–0.92)	0.36	0.06

FCST five-times chair stand test, TUG timed up and go test, ICC intraclass correlation coefficient, MDC minimal detectable change with a 90% confidence level, %MDC MDC divided with the mean of the first and second measurement values for each test, NA not applicable

walking time and TUG indicated good reliability. The MDC and %MDC of the FCST are not shown because of the presence of fixed and proposal biases. The %MDCs of grip strength, 5-m walking time, and the TUG were all < 10%, and the %MDCs of the sub-groups stratified by gender and age were also all < 10% (Table 3). By contrast, that of knee extension strength was 12%, and the %MDCs of the sub-groups were also \geq 10%, except for that in men (Table 3).

Discussion

In present study, we analyzed data obtained from 718 community-dwelling older people who could perform ADL independently and had high functional capacity, and clarified the absolute reliability of muscle strength and physical performance tests that are widely used in the clinical setting. The absolute reliability of the muscle strength and physical performance tests for older people with high functional capacities is essential so that healthcare professionals can evaluate precisely the risk of adverse health-related events such as frailty and falling. The MDCs and %MDCs for the absolute reliability of such tests determined in the present study could be a useful index for muscle strength and physical performance tests in community-dwelling older people.

With respect to muscle strength and physical performance tests, relative reliability was investigated using ICCs, which are commonly used to determine test-retest reliability [21], and systematic errors were investigated using Bland-Altman analysis. The ICCs for all tests were found to be better than a fair level [22]. Furthermore, no systematic errors were found in any tests except for the FCST. These findings did not differ as a result of the analysis stratified by gender and age. Regarding the systematic bias of the FCST, both fixed and proportional biases were found; the first measurement value was larger than the second, and the error between the first and second measurement values increased in accordance with increments in the mean values. That is, better performances

tended to be seen for the FCST in the second trial, and this tendency was prominent in individuals who had a worse performance in the first trial; this finding appeared to be the result of a learning effect [29]. However, the extent of the learning effect could not be clarified from the data, and we are not aware of any method to offset this type of learning effect. Further study will be necessary to address this issue. Consequently, we could not estimate the MDC or clarify the margin of error for the FCST in this study. In a recent consensus published by the European Working Group on Sarcopenia in Older People regarding the diagnosis of sarcopenia, 15 s or more in the FCST was defined as low muscle strength [5]. Although the FCST is known to be a useful indicator of low muscle strength, careful attention is necessary when assessing changes over time sake for critical systematic error in measurement.

Regarding the margin of error, the %MDCs of grip strength, 5-m walking time, and the TUG were considered excellent. On the other hand, knee extension strength was only considered acceptable. The detectability of grip strength, 5-m walking time, and the TUG were higher than that for knee extension strength. To detect changes in function and performance over time in older people, grip strength, 5-m walking time, and the TUG are, therefore, considered to be useful.

MDC and %MDC were investigated by previous studies in older people or those with a condition such as Parkinson's disease. With respect to muscle strength, the MDC for grip strength has been reported to be 5.2 kgf in patients who have undergone cardiac rehabilitation [30]. Furthermore, the %MDC of knee extension strength has been reported to be 50% in older nursing home residents [7]. Regarding physical performance tests, the %MDC of gait speed has been reported to be about 20% in older people who can walk independently; however, that study utilized a relatively small sample ($n=52$). Furthermore, the %MDC of gait speed in people recovering from stroke has been reported to be about 60% [31]. In addition, the MDCs of the TUG have been reported to be 3.5 s in people with Parkinson's disease [32] and 4.1 s in people with Alzheimer's disease [33]. Thus, the MDCs and %MDCs reported in previous studies were larger than those in the present study, which included community-dwelling older people with high functional capacities. These findings suggest that the absolute reliability of muscle strength and physical performance tests are greatly influenced by factors such as age, functional level, and disease. Therefore, the absolute reliability of muscle strength and physical performance tests appears to be a useful index for community-dwelling older people with high functional capacities.

This study had several limitations. First, grip strength was measured using a Smedley-type dynamometer, which is used widely and clinically in Japan [34]. However, a

previous study reported that grip strength measurement values differed between Jamar- and Smedley-type dynamometers [35]. Therefore, whether MDC and %MDC for grip strength estimated using a Smedley-type dynamometer are identical to those assessed by a Jamar-type dynamometer remains unclear. Second, the measurement of knee extension strength was performed without body stabilization of the abdomen and pelvis because it was simpler and more clinically practical. Previous research has suggested that body stabilization in the measurement of knee extension strength affects muscle strength [36]. In this study, a large measurement error was found in knee extension strength compared with grip strength; however, no mention was made of measurement errors for knee extension strength with body stabilization. Therefore, the measurement errors for knee extension strength estimated in this study may be limited to measurement values taken without body stabilization. Third, all measurements were carried out in two trials in this study, and absolute reliability was verified based on these data. Therefore, the absolute reliability established by this study may be limited to tests in which measurements were carried out twice. The number of trials in this study were set considering previous studies and clinical applications. For instance, many studies set the number of trials from one to three for measurements of grip strength and the TUG [34, 37]. Furthermore, applying additional trials may be difficult in clinical settings because of patient fatigue and burden, and the restricted amount of time available for treatment and assessment. Therefore, the absolute reliability established by this study may be limited to tests in which measurements are carried out twice.

In conclusion, the present study clarified the absolute reliability of muscle strength and physical performance tests commonly carried out in the clinical setting, which could therefore serve as reference values for measurement errors to detect changes in muscle strength and physical performance over time at the individual level in community-dwelling older people with high functional capacities. Because of critical systematic errors, the FCST was suggested to be an inappropriate physical performance test as an outcome measure to detect changes over time in lower extremity function. On the other hand, the measurement errors for grip strength, 5-m walking time, and the TUG were found to be < 10%. Thus, these tests were suggested to be a useful index to detect changes over time in muscle strength and physical performance tests in community-dwelling older people. Clinically, the findings suggest that changes of $\geq 10\%$ detected in these tests may be interpreted as true changes in muscle strength or physical performance at the individual level. However, knee extension strength showed relatively large measurement errors, thereby suggesting that the sensitivity for detection of changes over time might be more limited than grip strength, 5-m walking time, and the TUG.

Funding This study was funded by JSPS KAKENHI (Grant number 16K21348). The funder played no role in the design, methods, recruitment, data collection, analysis, or preparation of the paper.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval All procedures in this study that involved human participants were performed in accordance with the ethical standards of the Institutional Review Board of the School of Allied Health Sciences at Kitasato University [approval number 2016-G021B] and of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all participants.

References

- Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D (2004) Multimorbidity and quality of life in primary care: a systematic review. *Health Qual Life Outcomes* 2:51
- Metsemakers JF, Hoppener P, Knottnerus JA, Kocken RJ, Limonard CB (1992) Computerized health information in The Netherlands: a registration network of family practices. *Br J Gen Pract* 42:102–106
- Shumway-Cook A, Silver IF, LeMier M, York S, Cummings P, Koepsell TD (2007) Effectiveness of a community-based multifactorial intervention on falls and fall risk factors in community-living older adults: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* 62:1420–1427
- Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E et al (2017) Determining risk of falls in community dwelling older adults: a systematic review and meta-analysis using posttest probability. *J Geriatr Phys Ther* 40:1–36
- Cruz-Jentoft AJ, Bahat G, Bauer J, Boirie Y, Bruyere O, Cederholm T et al (2019) Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* 48:16–31
- Cadore EL, Rodriguez-Manas L, Sinclair A, Izquierdo M (2013) Effects of different exercise interventions on risk of falls, gait ability, and balance in physically frail older adults: a systematic review. *Rejuvenation Res* 16:105–114
- Buckinx F, Croisier JL, Reginster JY, Dardenne N, Beaudart C, Slomian J et al (2017) Reliability of muscle strength measures obtained with a hand-held dynamometer in an elderly population. *Clin Physiol Funct Imaging* 37:332–340
- Mangione KK, Craik RL, McCormick AA, Blevins HL, White MB, Sullivan-Marx EM et al (2010) Detectable changes in physical performance measures in elderly African Americans. *Phys Ther* 90:921–927
- Tsutsui T, Muramatsu N (2005) Care-needs certification in the long-term care insurance system in Japan. *J Am Geriatr Sci* 53:522–527
- Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J et al (2001) Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 56:M146–M156
- Dent E, Lien C, Lim WS, Wong WC, Wong CH, Ng TP et al (2017) The Asia-Pacific clinical practice guidelines for the management of frailty. *J Am Med Dir Assoc* 18:564–575
- Rijk JM, Roos PR, Deckx L, van den Akker M, Buntinx F (2016) Prognostic value of handgrip strength in people aged 60 years and older: a systematic review and meta-analysis. *Geriatr Gerontol Int* 16:5–20
- Newman AB, Kupelian V, Visser M, Simonsick EM, Goodpaster BH, Kritchevsky SB et al (2006) 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* 61:72–77
- Kojima N, Kim H, Saito K, Yoshida H, Yoshida Y, Hirano H et al (2014) Association of knee-extension strength with instrumental activities of daily living in community-dwelling older adults. *Geriatr Gerontol Int* 14:674–680
- Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB (1994) A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 49:M85–M94
- Corrigan D, Bohannon RW (2001) Relationship between knee extension force and stand-up performance in community-dwelling elderly women. *Arch Phys Med Rehabil* 82:1666–1672
- Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP (2011) Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC Geriatr* 11:33. <https://doi.org/10.1186/1471-2318-11-33>
- Podsiadlo D, Richardson S (1991) The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 39:142–148
- Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K et al (2013) Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *J Am Geriatr Soc* 61:202–208
- Koyano W, Shibata H, Nakazato K, Haga H, Suyama Y (1991) Measurement of competence: reliability and validity of the TMIG Index of Competence. *Arch Gerontol Geriatr* 13:103–116
- Lexell JE, Downham DY (2005) How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil* 84:719–723
- Currier DP (1990) Elements of research in physical therapy, 3rd edn. Williams and Wilkins, Baltimore
- Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160
- Cohen J (1992) A power primer. *Psychol Bull* 112:155–159
- Ludbrook J (2002) Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* 29:527–536
- Weir JP (2005) Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19:231–240
- Smidt N, van der Windt DA, Assendelft WJ, Mourits AJ, Deville WL, de Winter AF et al (2002) Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil* 83:1145–1150
- R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>. Accessed 4 Sep 2015
- Cani KC, Silva IJCS, Karloh M, Gulart AA, Matte DL, Mayer AF (2018) Reliability of the five-repetition sit-to-stand test in patients with chronic obstructive pulmonary disease on domiciliary oxygen therapy. *Physiother Theory Pract* 1:1–7
- Puthoff ML, Saskowski D (2013) Reliability and responsiveness of gait speed, five times sit to stand, and hand grip strength for patients in cardiac rehabilitation. *Cardiopulm Phys Ther J* 24:31–37

31. Fulk GD, Echternach JL (2008) Test–retest reliability and minimal detectable change of gait speed in individuals undergoing rehabilitation after stroke. *J Neurol Phys Ther* 32:8–13
32. Huang SL, Hsieh CL, Wu RM, Tai CH, Lin CH, Lu WS (2011) Minimal detectable change of the timed “up & go” test and the dynamic gait index in people with Parkinson disease. *Phys Ther* 91:114–121
33. Ries JD, Echternach JL, Nof L, Gagnon Blodgett M (2009) Test–retest reliability and minimal detectable change scores for the timed “up & go” test, the six-minute walk test, and gait speed in people with Alzheimer disease. *Phys Ther* 89:569–579
34. Kamide N, Kamiya R, Nakazono T, Ando M (2015) Reference values for hand grip strength in Japanese community-dwelling elderly: a meta-analysis. *Environ Health Prev Med* 20:441–446
35. Kim M, Shinkai S (2017) Prevalence of muscle weakness based on different diagnostic criteria in community-dwelling older adults: a comparison of grip strength dynamometers. *Geriatr Gerontol Int* 17:2089–2095
36. Ushiyama N, Kurobe Y, Momose K (2017) Validity of maximal isometric knee extension strength measurements obtained via belt-stabilized hand-held dynamometry in healthy adults. *J Phys Ther Sci* 29:1987–1992
37. Bohannon RW (2006) Reference values for the timed up and go test: a descriptive meta-analysis. *J Geriatr Phys Ther* 29:64–68

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.