Irish Journal of Medical Science (1971 -)
https://doi.org/10.1007/s11845-020-02492-0

ORIGINAL ARTICLE

Test-retest reliability of the functional reach test and the hand grip strength test in older adults using nursing home services

Soraia Ferreira[1,2] · Armando Raimundo[1,2] · José Marmeleira[1,2]

Received: 9 November 2020 / Accepted: 17 December 2020

Abstract

Objective This study aimed to determine the absolute and relative test-retest reliability of the functional reach test (FRT) and the handgrip strength test (HGST) in older adults using nursing homes.

Methods Participants (≥ 65 years old), living in nursing homes or using their day care services, were distributed into a group without cognitive impairment (GWCI, $n = 43$) and a group with mild cognitive impairment (GCI; $n = 22$). A 1-week test-retest was performed for the FRT and the HGST. Relative reliability was measured by the intraclass correlation coefficient ($ICC_{3.1}$), and absolute reliability by the standard error of measurement (SEM), minimal detectable change (MDC_{95}), and Bland-Altman plots.

Results The ICC showed high reliability for the FRT (GWCI, ICC = 0.83; GCI, ICC = 0.87) and the HGST (ICC ≥ 0.95 in both hands and participant groups). The absolute reliability was good: FRT, SEM = 2.96/2.29, MDC_{95} = 8.20/6.35 for the GWCI and the GCI, respectively; HGST dominant hand SEM = 1.26/0.82, MDC_{95} = 3.50/2.29, and HGST non-dominant hand SEM = 1.05/0.80, MDC_{95} = 2.90/2.21, for the GWCI and the GCI, respectively. Bland-Altman showed that there was not a systematic bias for the tests in both groups.

Discussion Findings show that the FRT and the HGST are reliable, have acceptable measurement error, and may be used for research and clinical purposes to assess functional balance and strength of the hands in older adults using nursing homes.

Keywords Balance · Elderly · Handgrip strength · Nursing homes · Reliability

Introduction

Estimations for the European Union suggest that the number of institutionalized older people will triple by 2060, reaching approximately 8.3 million [1]. In general, the level of frailty of people entering nursing homes is high, resulting from several factors, including disability, cognitive impairment, and acute illness [2–4]. This reality is a major challenge for health care personnel working in institutionalized settings.

In parallel with institutionalization, there are day care centers where the elderly stay during the day (they return to their homes to sleep at the end of the day). A recent review of the literature indicated that the main reasons for attending a day care center are social isolation, loss of mobility, emotional problems, loneliness, feeling “stuck” at home, and looking for something to do [5]. Thus, as people living in long care facilities (nursing homes), frequently, day care attendees also have deficits in their health and functional status. One of the purposes of day care centers is to provide interventions focused on older attendees’ mental health, physical function, social life, and quality of life [5]. In Portugal, day care centers are usually a valence of nursing homes, and at some point in time, attenders start to live in the nursing homes.

For planning appropriate interventions to stimulate people using nursing home services (day care and residence), it is necessary to measure their level of functioning. Consequently, it is important to select the best field tests to be used by health care personnel. Unfortunately, few studies have examined the psychometric properties of physical tests that measure the functional status of older adults using nursing home services [6–9]. Despite the marked growth over the

✉ Soraia Ferreira
sdpf@uevora.pt

1 Departamento de Desporto e Saúde, Escola de Ciências e Tecnologia, Universidade de Évora, Rua de Reguengos de Monsaraz, n° 14, 7005-399 Évora, Portugal

2 Comprehensive Health Research Centre (CHRC), Lisboa, Portugal

Published online: 21 January 2021

years in the number of people using nursing home services, still, most studies examined the reliability of physical performance tests only in community-dwelling older adults [10–14].

In the present study, we focus on two well-known field tests of physical fitness, the functional reach test (FRT) and the handgrip strength test (HGST), for which there is limited evidence on their psychometric properties in institutionalized settings. The FRT is a clinical measure of balance that measures the maximal distance one can reach forward beyond the length of one arm while maintaining a fixed base of support in the standing position [15]. This test could be very useful in institutionalized settings, as it is well known that balance is fundamental for decreasing the risk of falls and for the performance of ADLs in older adults [16, 17]. The HGST is a measure of the hand strength and forearm muscles and has been shown to provide important information related to frailty and the physical status of elderly individuals [18].

A several of studies showed that both the FRT and the HGST are reliable in different groups of community-dwelling older adults (FRT, stroke, frailty, and community-dwelling elderly; HGST, Parkinson’s disease, and unilateral thumb carpometacarpal osteoarthritis) [11, 12, 15, 19–22]. In contrast, information on the reliability of both the FRT and the HGST is scarce regarding older people using nursing home services. We found just two studies that focused the FRT in nursing homes residents [8, 23], reporting discrepant levels of test-retest reliability, and one study for the HGST performed in institutionalized people, which reported good test-retest reliability [9]. More investigations are needed to confirm these results.

As previously referred, there is limited information on the psychometric properties of physical performance tests in institutionalized older adults. Therefore, our goal is to examine the absolute and relative test-retest reliability of two of the most popular physical performance tests used with community-dwelling older adults—the FRT and the HGST, in a sample of older adults using nursing home services. This is of high relevance as both balance and strength are fundamental abilities for the functional capacity of the growing number of older people using nursing home services, and therefore, they should be measured regularly.

Methods

Participants

The volunteers in this study were older adults who were using the services of five nursing homes (day care center or living in) in the region of Évora (Portugal) and were selected as part of a convenience sample. The health care personnel in the nursing homes helped to identify potential participants according to the following inclusion criteria: living in a nursing home or using a day care center, being aged 65 years or older, and being capable of walking without the assistance of another person.

The total sample included sixty-five older adults, which were distributed in one group with mild cognitive impairment (GCI, $n = 22$; 17 women) and another group without cognitive impairment (GWCI, $n = 43$; 28 women), considering the scores on the Portuguese version of the Mini-Mental State Examination with cut-offs of ≤ 27 points or persons with > 11 years of school education, ≤ 22 for persons ranging from 1 to 11 years of school education, and ≤ 15 points for illiterate persons [24].

Table 1 shows the general characteristics of the participants by group. The sample was composed of old and very old people (85.4 ± 6.0 years), and in general, the participants had a low educational level (~ 40% illiterate). Regarding BMI, according to the World Health Organization criteria, 41.5% of the participants were overweight, and 30.8% were obese. Most participants (73.6%) resided in nursing homes, and the remaining participants used day care services and stayed overnight in their homes. There were no significant differences between the participants who lived in the nursing homes and the participants who used the day care services of the nursing homes regarding the age, BMI, years of education, and scores on the MMSE.

All participants or their legal representatives were informed about the objectives of the study and provided informed consent prior to participation. The study was approved by the University of Évora ethics committee and conducted in accordance with the Declaration of Helsinki.

Procedure

In this study, the handgrip strength test and the functional reach test were performed two times 1 week apart

Table 1 Descriptive characteristics of the participants

	GWCI (*n* = 43)	GCI (*n* = 22)
Age (years)	84.5 (6.5)	87.1 (5.0)
Height (cm)	152 (9.5)	150 (8.0)
Weigh (kg)	64.3 (12.5)	61.3 (11.2)
BMI (kg/m^2)	28 (5.2)	26.9 (3.2)
Education (years)	2.2 (2)	2.8 (3.4)
MMSE (points)	24.5 (3.9)	14.4 (6.1)
Day care services *n* (%)	13 (30.2%)	5 (22.7%)
Female *n* (%)	28 (65.1%)	17 (77.3%)

Data are expressed as mean ± standard deviation, except where *n* (%)

GWCI, group without cognitive impairment; *GCI*, group with cognitive impairment; *BMI*, body mass index; *MMSE*, mini-mental state examination to evaluate the inter-session test-retest reliability [23]. The same kinesiologist (with experience in assessing physical performance in older adults) collected the data at both time points. The tests were administered to each individual separately at the nursing homes in a quiet room.

Instruments

The strength of the hand and forearm muscles was measured (in kilograms) using a handgrip dynamometer (Baseline Smedley, Model 12-0286, White Plains, NY, USA). The handle of the handgrip was adjusted for each individual according to the instructions in the manual. The test was performed with the elbow flexed at 90° and in the sitting position [9]. The participant was instructed to grip the device with maximum strength for 3 s. Following a demonstration of the protocol by the researcher, the participant performed one practice trial (when necessary, more trials were performed to ensure that they understood the protocol) for each hand and then performed three test trials for each hand (starting with the dominant hand), with a 1-min rest period between trials. The mean of the three trials was used for data analysis. It has been reported that the HGST shows predictive validity in identifying persons at risk of mobility limitation [25], premature mortality, and health limitations [26].

Balance was assessed with the FRT [15]. The participants stood without any type of physical support, and their dominant arm was positioned close to the wall (without touching). They were asked to raise their dominant arm with a closed fist to approximately 90° of shoulder flexion, and the instructor recorded the position of the 3rd metacarpal head in this initial position (the participant held a pencil, which served as a reference). The participants were then instructed to keep their feet flat on the floor and to reach (while maintaining a horizontal position of the arm, and without taking a step) as far as possible and to stay in that position for 3 s. The result was the difference (in cm) between this position and the initial position. Following the demonstration of the protocol by the researcher, each participant performed the test 2 times for familiarization (some participants had difficulty understanding the instructions, and in such cases, it was necessary to perform more practice trials). Afterwards, each participant performed the test 3 times, and the mean of three trials was used for data analysis. In the original study on the FRT and other subsequent studies, it was reported that the FRT has predictive validity for the occurrence of falls in older adults [27, 28]. More recent studies (systematic reviews) do not support the use of the FRT as a single measure to predict the risk of falls of older adults [29, 30].

Data analysis

Relative and absolute reliability was assessed for each test using a test-retest design. Relative reliability, which refers to the degree to which individuals in a sample maintain their position across repeated measurements [31], was determined by two-way mixed-model intraclass correlation coefficients (ICC) (3,1) with absolute agreement [32]. In the present study, the following cut-off points were used: ICC < 0.5 indicates poor reliability; 0.5 to 0.74 indicates moderate reliability; 0.75 to 0.9 indicates good reliability; and > 0.9 indicates excellent reliability [33, 34].

Absolute reliability, which refers to the degree to which repeated measurements vary for individuals (the less they vary, the higher the reliability) [31], was calculated by the standard error of measurement (SEM), the minimal detectable change (MDC) and Bland-Altman 95% limits of agreement (LoA). The SEM reflects the standard deviation of measurement errors [31], and lower values indicate high reliability [35]. The SEM was calculated using the formula $SEM = SD\sqrt{(1 - ICC)}$, where the SD is the mean of the standard deviations of the two trials [36]. The MDC is based on the SEM and is defined as the minimal changes in the scores [37]. The MDC was calculated at the 95% level of confidence using the formula $MDC95 = 1.96 \times \sqrt{2} \times SEM$ [38]. The MDC indicates the smallest within-person change in a score that can be considered as a "real" change, above the measurement error of an individual [39]. The Bland-Altman plots provide a graphical presentation of the differences between two tests plotted against the mean difference of the two tests, allowing a visual assessment of the scoring distribution and potential measurement bias [40]. The 95% limits of agreement (LoA) were estimated as the mean difference ± 1.96 × SD of the difference. The number of participants was calculated according to Walter, Eliasziw, and Donner (1998), considering $\alpha = 0.05$ and $\beta = 0.20$, in which the desired ICC was 0.8, with a CI of 0.60 [41].

Additionally, to detect possible systematic bias between the duplicate tests, paired sample *t* tests were computed for the test results in the two sessions. The level of significance was established to be $p < 0.05$. The data were analyzed using SPSS 24.0 for Windows (SPSS Inc., Chicago, IL).

Results

Sixty-five older adults using nursing home services participated in this study. In the GWCI, two participants did not perform the second evaluation of the FRT and the HGST due to health problems, and two participants chose not to participate in the second day of assessments of HGST. In the GCI, four and three participants decided not to participate in the second day of assessment of the FRT and the HGST, respectively. A single rater conducted all assessments.

Table 2 shows the relative and absolute reliability values of the two physical performance tests for both groups of participants. The relative reliability was good for the FRT (GWCI, ICC = 0.83; GCI, ICC = 0.87) and excellent for the HGST (dominant and non-dominant hand) in both groups (ICC = 0.95 to 0.98). The SEM and MDC_{95} were higher for the FRT than for the HGST in both groups. Thus, the highest SEM value (as a percentage) was 13.5% on the FRT and 6.7% on the HGST; the highest MDC_{95} values were 37.4% and 18.5% on the FRT and on the HGST, respectively. The GCI had smaller values for both the SEM and MDC_{95} than did the GWCI.

The mean scores obtained in sessions one and two were very similar for all tests and for both groups, and the paired sample *t* test confirmed that they were not significantly different. Hence, in the GWCI, the mean difference in the scores between the two test sessions for the FRT, HGST with the dominant hand, and HGST with the non-dominant hand was − 0.77 cm (3.2%; $t = -1.16$, $p = 0.25$), 0.02 kg (0.09%; $t = 0.06$, $p = 0.96$), and − 0.33 kg (1.8%; $t = -1.43$, $p = 0.16$), respectively. In the GCI, the mean difference in the score was 0.11 cm (0.5%; $t = 0.13$, $p = 0.90$) for the FRT, − 0.32 kg (1.2%; $t = -1.17$, $p = 0.26$) for the HGST with the dominant hand, and − 0.05 kg (0.32%; $t = -0.17$, $p = 0.87$) for the HGST with the non-dominant hand.

The Bland-Altman plots (Figs. 1 and 2) show the 95% limits of agreement for the tests. It is possible to observe that in all plots, most of the values were within 95% of LoA, indicating a normal distribution of the differences between the first and second sessions. Hence, there was no evidence of increasing variability with an increase in the mean, suggesting good agreement in the participants' performance between sessions.

Discussion

In the present investigation, the relative and absolute reliability (test-retest method) of the FRT and the HGST were studied in older adults with and without cognitive impairment, living in or using the day care services of nursing homes.

In the present study, the ICC for the FRT was good for both groups (GWCI, ICC = 0.83; GCI, ICC = 0.87). This finding is in line with findings from previous studies that reported high values of relative reliability (ICC 0.87 to 0.98) for the FRT in different groups of older adults, including those with stroke, frailty, and community-dwelling elderly individuals [15, 21–23]. We found only two studies that focused on the FRT reliability in participants living in nursing homes. The most recent study showed results similar to ours, reporting an ICC of 0.85 [8], but the study of Fox reported a low ICC of 0.38 [23]. Despite the difference in the ICC values between both studies, one should note that the sample of the study by Fox et al. was composed per older adults with dementia [23] while the participants in the study by Galhardas et al. had a normal cognitive status [8]. Although in the present research we included a group of older adults with cognitive impairment (for which the FRT showed to have high relative reliability), comparisons should be performed with caution since we only used the MMSE as a screening tool for cognitive functioning, and, therefore, it was not possible to determine whether participants (or how many participants) have dementia.

The FRT performance was better in the GWCI than in the GCI. This result was expected, as, in general, motor performance is negatively affected by decrements in cognitive functioning [42]. Curiously, the SEM and MDC_{95} were lower in the GCI (SEM = 2.29; MDC_{95} = 6.35) than in the GWCI (SEM = 2.96; MDC_{95} = 8.20), but the magnitude of the differences was relatively low. These small differences may be related to the sample size being smaller in the GCI than in the

Table 2 Intraclass correlation coefficient, standard error of measurement for repeated measures, and minimal detectable change scores at the 95% confidence interval for the tests

Test item	*n*	Mean (SD) Test	Mean (SD) Retest	Difference	ICC (95%)	SEM	MDC_{95}
GWCI							
FRT (cm)	41	21.6(7.6)	22.3 (6.7)	− 0.7	0.83 (0.70–0.90)	2.96	8.20
HGST dominant hand (kg)	39	18.9 (5.9)	18.9(5.8)	0.01	0.95 (0.91–0.98)	1.26	3.50
HGST non-dominant hand (kg)	39	17.1 (5.6)	17.4 (5.4)	− 0.3	0.96 (0.93–0.98)	1.05	2.90
GCI							
FRT (cm)	18	20.4 (6.6)	20.3 (6.2)	0.1	0.87 (0.68–0.95)	2.29	6.35
HGST dominant hand (kg)	19	16.7 (5.4)	16.9 (5.2)	− 0.2	0.97 (0.94–0.99)	0.82	2.29
HGST non-dominant hand (kg)	19	14.8 (5.3)	14.8 (5.2)	− 0.05	0.98 (0.94–0.99)	0.80	2.21

GWCI, group without cognitive impairment; *GCI*, group with cognitive impairment; *FRT*, functional reach test; *HGST*, handgrip strength test; *SD*, standard deviation; *ICC*, intraclass correlation coefficient; *SEM*, standard error of measurement; *MDC*, minimal detectable change

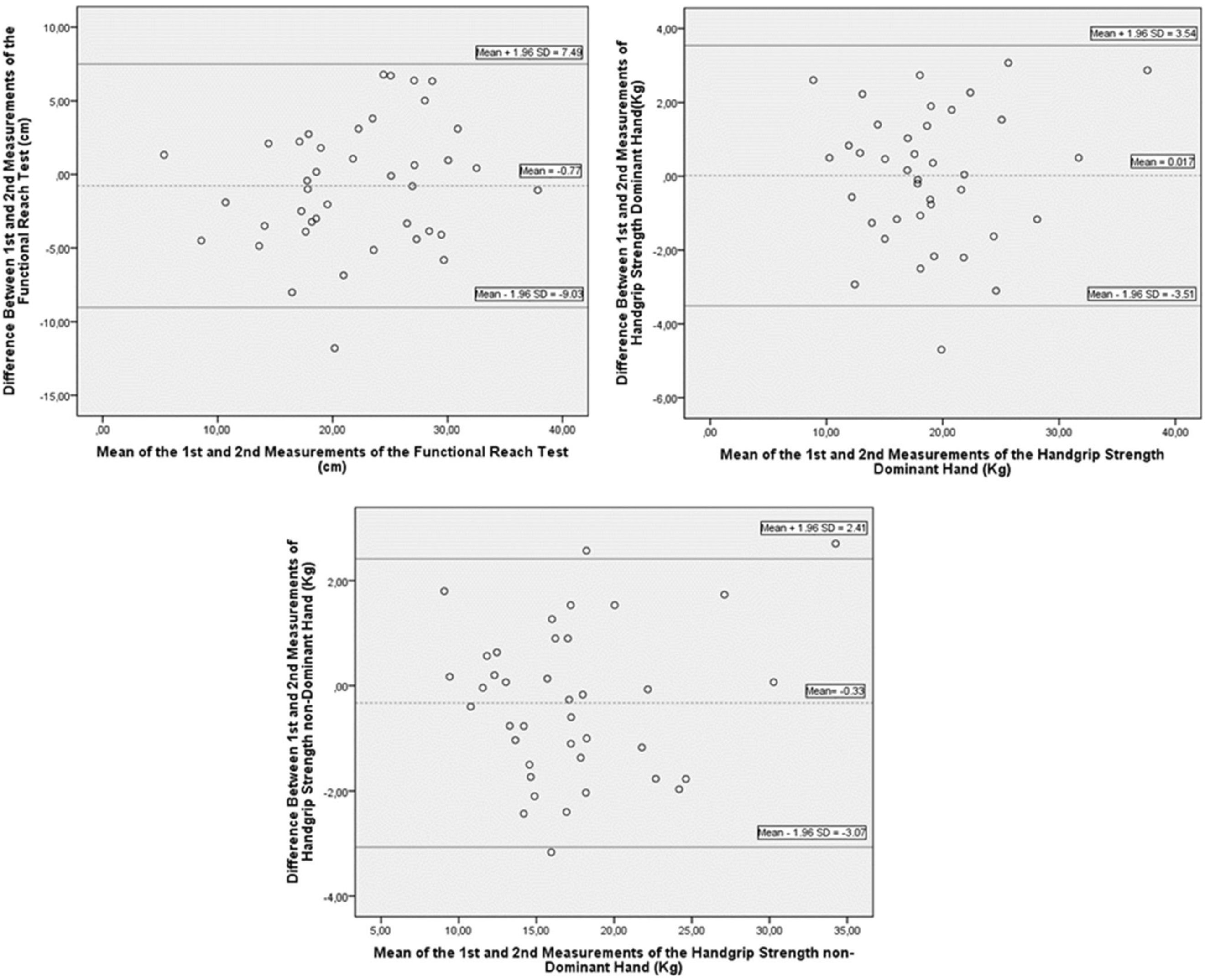

Fig. 1 Bland-Altman plots for the group without cognitive impairment (GWCI). The middle lines represent the mean difference between test-retest. The upper and lower lines represent the 95% limits of agreement (LoA)

GWCI. Some previous studies have reported the SEM and the MDC for the FRT. Regarding the two studies [8, 23] that investigated the FRT in nursing home residents, only the later reported results on the test-retest absolute reliability. In this case, the authors indicated better scores than in the present study (SEM = 1.5, and MDC_{95} = 4.0 cm) in participants without cognitive impairment. Another study reported worse scores than ours for the SEM (4.56 cm) and the MDC_{95} (12.64 cm) in community-dwelling older adults with mild to moderate Alzheimer's disease [43]. We also found a study with adults > 50 years old with osteoarthritis, which indicated an SEM of 3.43 cm and an MDC of 8.0 cm (in this case, the MDC was calculated for the 90% level of confidence). We highlight that the MDC_{95} is clinically relevant because it helps to identify real changes beyond measurement errors. Our data show that for the GWCI and the GCI, changes by 8.2 cm and 6.4 cm, respectively, are necessary for considering an intervention to be clinically relevant in participants using nursing home services or that the scores truly changed over time in this group. Regarding the FRT, the Bland-Altman plots and the paired sample *t* test showed that there was no systematic bias (the mean difference was close to zero) between the performance on the tests administered 1 week apart. The LoA for the FRT was relatively wide, and the visual inspection of the Bland-Altman plots confirmed the existence of some variability in the differences between the first and second measurements over the increments of the participants' mean performance.

Due to the reduced physical fitness and the general frailty of older adults living in nursing homes [17], and due to its practicality in clinical settings [29], the FRT is an appealing test for assessing balance in elderly individuals. Interestingly, we have not found in the literature any study on the validity of the FRT alone for the prediction of falls in older adults living in nursing homes. One should note, however, that as recent reviews indicate that the FRT might not be a valid instrument

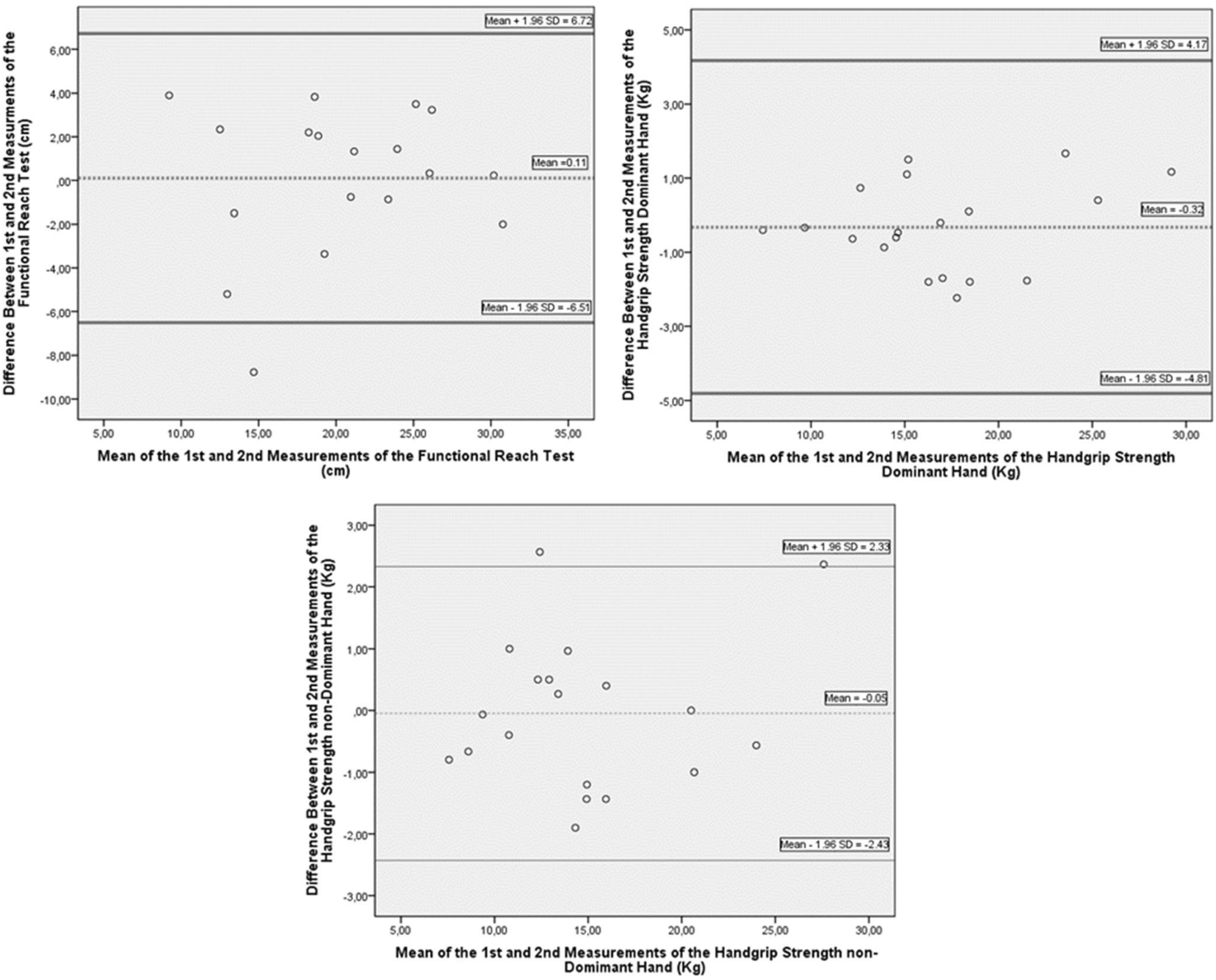

Fig. 2 Bland-Altman plots for the group with cognitive impairment (GCI). The middle lines represent the mean difference between test-retest. The upper and lower lines represent the 95% limits of agreement (LoA)

for predicting the occurrence of falls in older adults living in the community [29, 30], other measures (e.g., Berg Balance Scale or the Timed Up and Go test) may be necessary for that purpose in older adults living in institutionalized settings.

The HGST has been used as an indicator of overall muscle strength and function, and is associated with mobility and activities of daily living [44]. In the current study, the ICC for the HGST was excellent for both hands in the two groups (dominant hand, ICC = 0.95 for the GWCI and ICC = 0.97 for the GCI; non-dominant hand, ICC = 0.96 for the GWCI and ICC = 0.98 for the GCI). These results are similar to those of a previous study with nursing home residents and day care centers that reported excellent values for ICC (> 0.91) [9]. Furthermore, other investigations also reported excellent ICC (0.91 to 0.98) in different groups of community-dwelling elderly individuals, including Parkinson's disease, dementia, and unilateral thumb carpometacarpal osteoarthritis [11, 12, 19, 20, 23].

The GCI showed better performance than did the GWCI in the HGST. For the HGST, the SEM and MDC_{95} were relatively lower in the GCI (SEM = 0.82, MDC_{95} = 2.29, dominant hand; SEM = 0.80, MDC_{95} = 2.21, non-dominant hand) than in the GWCI (SEM = 1.26, MDC_{95} = 3.50, dominant hand; SEM = 1.05, MDC_{95} = 2.90, non-dominant hand). A previous study with older adults in nursing homes and day care center showed worse results than our study (right hand, SEM = 1.70, MDC_{95} = 4.71; left hand, SEM = 1.73, MDC_{95} = 4.80) [9]. In contrast, the SEM and MDC_{95} values were slightly higher than those reported in previous studies with older adults, namely, in those with Parkinson's disease (SEM = 0.05 for both hands) [11] and osteoarthritis (SEM = 0.61 for the affected right hand, SEM = 0.54 for the contralateral left hand) [12]. Curiously, a study with community-dwelling older adults reported SEM values that were much higher (SEM = 15.8, right hand; SEM = 21.3, left hand) [19]. Nevertheless, one should note that in this study, the HGST measures were obtained over a 12-week interval. In contrast, a recent study [20] reported lower values of the MDC_{95} (1.18) for the HGST in community-dwelling elderly individuals.

The Bland-Altman plots showed that there was no systematic bias in the HGST. The mean values were close to zero in all plots, and the LoA values were relatively small. Interestingly, the results of the Bland-Altman plots for the HGST in the GCI were close to those reported in a previous study [23] with older adults with dementia living in residential aged care facilities. Fox et al. (2014) also reported means close to zero (0.17 and − 0.57 for the right and left hand, respectively) and a small LoA (− 4.55 to 4.88 and − 3.93 to 2.80 for the right and left hand, respectively) [23].

Limitations and strengths of the study

This study has a number of limitations. The GCI was formed according to the MMSE scores, and in the absence of a more comprehensive cognitive evaluation, it was not possible to diagnose cases of mild cognitive impairment or dementia. Moreover, the sample size of the GCI was inferior to the optimal sample size (39 participants) previously calculated, limiting the generalizability of the findings. Nevertheless, considering the currently limited information on the reliability of physical tests for older adults with cognitive impairment, we choose to include the GCI in the analysis. Finally, we did not evaluate the functional capacity index of the participants or their level of fragility.

Conversely, there are some important strengths. The present study provides evidence that two of the most popular physical performance tests for older adults are reliable for older adults using nursing home services. This adds to the literature, as most related studies have focused on community-dwelling older adults. Furthermore, the present investigation also included people with cognitive impairment, a group for which there is a lack of available information regarding assessment instruments.

Conclusion

This study showed that both the FRT and the HGST have high reliability and have acceptable measurement error in older adults with and without cognitive impairment using nursing home services. Thus, these tests could be valuable clinical tools for assessing the balance and strength of the hands of older adults in institutionalized contexts.

Funding S.F. thanks the Foundation for Science and Technology (FCT), Portugal, for the Ph.D. grant SFRH/BD/141448/2018.

This study was supported by Horizon - Portugal – Alentejo 2020 (ALT20-03-0145-FEDER-000007) – ESACA.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bettio F, Verashchagina A (2012) Long-term care for the elderly. Provisions and providers in 33 European countries. Luxembourg, European Commission, 178 p
2. Miller E, Weissert W (2000) Predicting elderly people's risk for nursing home placement, hospitalization, functional impairment, and mortality: a synthesis. Med Care Res Rev. [cited 2019 Jun 11];57(3):259–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10981186
3. Luppa M, Luck T, Weyerer S et al (2010) Prediction of institutionalization in the elderly. A systematic review. Age Ageing. [cited 2019 Jun 11];39(1):31–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19934075
4. Bell C, LaCroix A, Desai M et al (2015) Factors associated with nursing home admission after stroke in older women. J Stroke Cerebrovasc Dis. [cited 2019 Jun 11];24(10):2329–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26169547
5. Orellana K, Manthorpe J, Tinker A (2020) Day centres for older people - attender characteristics, access routes and outcomes of regular attendance: findings of exploratory mixed methods case study research. BMC Geriatr. [cited 2020 Jul 10];20(1). Available from: https://pubmed.ncbi.nlm.nih.gov/32366223/
6. Conradsson M, Lundin-Olsson L, Lindelöf N et al (2007) Berg Balance Scale: intrarater test-retest reliability among older people dependent in activities of daily living and living in residential care facilities. Phys Ther. [cited 2019 Jun 16];87(9):1155–63. Available from: https://academic.oup.com/ptj/ptj/article/2742401/Berg
7. Tappen R, Roach K, Buchner D et al (1997) Reliability of physical performance measures in nursing home residents with Alzheimer's disease. Journals Gerontol - Ser A Biol Sci Med Sci 52(1)
8. Galhardas L, Raimundo A, Marmeleira J (2020) Test-retest reliability of upper-limb proprioception and balance tests in older nursing home residents. Arch Gerontol Geriatr 89(April)
9. Silva AG, Cerqueira M, Raquel Santos A et al (2019) Inter-rater reliability, standard error of measurement and minimal detectable change of the 12-item WHODAS 2.0 and four performance tests in institutionalized ambulatory older adults. Disabil Rehabil 41(3):366–73. Available from: https://doi.org/10.1080/09638288.2017.1393112
10. Yang L, He C, Pang MYC (2016) Reliability and validity of dual-task mobility assessments in people with chronic stroke. Tremblay F, editor. PLoS One. [cited 2019 Jun 3];11(1):e0147833. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26808662
11. Villafañe J, Valdes K, Buraschi R et al (2016) Reliability of the handgrip strength test in elderly subjects with Parkinson disease. HAND. [cited 2019 Jun 3];11(1):54–8. Available from: http://journals.sagepub.com/doi/10.1177/1558944715614852
12. Villafañe J, Valdes K, Vanti C et al (2015) Reliability of handgrip strength test in elderly subjects with unilateral thumb carpometacarpal osteoarthritis. HAND. [cited 2019 May 30];10(2):205–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26034431
13. Hesseberg K, Bentzen H, Bergland A (2015) Reliability of the senior fitness test in community-dwelling older people with cognitive impairment. Physiother Res Int [cited 2019 Jun 16];20(1):37–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24925585
14. Blankevoort C, van Heuvelen M, Scherder E (2013) Reliability of six physical performance tests in older people with dementia. Phys

Ther. [cited 2019 Jun 16];93(1):69–78. Available from: https://academic.oup.com/ptj/article-lookup/doi/10.2522/ptj.20110164
15. Duncan P, Weiner D, Chandler J et al (1990) Functional reach: a new clinical measure of balance. J Gerontol. [cited 2019 May 30];45(6):M192-7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2229941
16. Angın E, Can F, İyigün G et al (2016) Does balance influence daily living activities and quality of life in community-dwelling older people? Physiotherapy. [cited 2019 Aug 16];102:e227–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0031940616303546
17. Marmeleira J, Ferreira S, Raimundo A (2017) Physical activity and physical fitness of nursing home residents with cognitive impairment: a pilot study. Exp Gerontol. 100:63–69
18. Stessman J, Rottenberg Y, Fischer M et al (2017) Handgrip strength in old and very old adults: mood, cognition, function, and mortality. J Am Geriatr Soc. [cited 2019 Aug 16];65(3):526–32. Available from: http://doi.wiley.com/10.1111/jgs.14509
19. Bohannon R, Schaubert K (2005) Test–retest reliability of grip-strength measures obtained over a 12-week interval from community-dwelling elders. J Hand Ther. [cited 2019 Jun 3];18(4):426–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16271690
20. Suzuki Y, Kamide N, Kitai Y et al (2019) Absolute reliability of measurements of muscle strength and physical performance measures in older people with high functional capacities. Eur Geriatr Med. [cited 2019 Jul 25];1–8. Available from: http://link.springer.com/10.1007/s41999-019-00218-9
21. Thomas J, Lane J (2005) A pilot study to explore the predictive validity of 4 measures of falls risk in frail elderly patients. Arch Phys Med Rehabil. [cited 2019 May 30];86(8):1636–40. Available from: https://www.sciencedirect.com/science/article/pii/S0003999305003102?via%3Dihub
22. Merchán-Baeza J, González-Sánchez M, Cuesta-Vargas A (2014) Reliability in the parameterization of the functional reach test in elderly stroke patients: a pilot study. Biomed Res Int. [cited 2019 May 30];2014:637671. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24868537
23. Fox B, Henwood T, Neville C, Keogh J (2014) Relative and absolute reliability of functional performance measures for adults with dementia living in residential aged care. Int Psychogeriatrics. 26(10):1659–1667
24. Guerreiro M, Silva A, Botelho M et al (1994) Adaptação à população portuguesa da tradução do Mini Mental State Examination (MMSE). Rev Port Neurol. [cited 2019 Jul 14];3(suppl):9–10. Available from: https://www.scienceopen.com/document?vid=97d4329c-9dba-49fe-9df1-db1215f3d4c5
25. Sallinen J, Stenholm S, Rantanen T et al (2010) Hand-grip strength cut points to screen older persons at risk for mobility limitation. J Am Geriatr Soc. [cited 2019 Oct 15];58(9):1721–6. Available from: http://doi.wiley.com/10.1111/j.1532-5415.2010.03035.x
26. Bohannon R (2008) Hand-grip dynamometry predicts future outcomes in aging adults. J Geriatr Phys Ther. 31(1):3–10
27. Duncan P, Studenski S, Chandler J et al (1992) Functional reach: predictive validity in a sample of elderly male veterans. J Gerontol. [cited 2019 Oct 15];47(3):M93-8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1573190
28. Sugihara T, Mishima S, Tanaka M et al (2006) Physical ability estimation and falling in the elderly. J Phys Ther Sci. [cited 2019 Oct 15];18(2):137–41. Available from: http://joi.jlc.jst.go.jp/JST.JSTAGE/jpts/18.137?from=CrossRef
29. Rosa MV, Perracini MR, Ricci NA (2019 Mar) Usefulness, assessment and normative data of the Functional Reach Test in older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr. [cited 2019 Oct 15];81:149–70. Available from: https://www.sciencedirect.com/science/article/pii/S0167494318302322?via%3Dihub
30. Lusardi MM, Fritz S, Middleton A et al (2017) Determining risk of falls in community dwelling older adults. J Geriatr Phys Ther. [cited 2019 Oct 15];40(1):1–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27537070
31. Bruton A, Conway J, Holgate S (2000) Reliability: What is it, and how is it measured? Physiotherapy. 1 [cited 2019 May 22];86(2):94–9. Available from: https://www.sciencedirect.com/science/article/pii/S0031940605612114
32. Weir J (2005) Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. [cited 2019 May 22];19(1):231–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15705040
33. Koo T, Li M (2016) A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. [cited 2019 May 22];15(2):155–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27330520
34. Portney L, Watkins M (2015) Foundations of clinical research: applications to practice, 3rd edn. Pennsylvania, F.A. Davis Company 912 p
35. Šerbetar I (2015) Establishing some measures of absolute and relative reliability of a motor tests. Croat J Educ. 17(1):37–48
36. Hopkins W (2000) Measures of reliability in sports medicine and science. Sport Med. [cited 2019 May 22];30(1):1–15. Available from: http://link.springer.com/10.2165/00007256-200030010-00001
37. Charter R (1996) Revisiting the standard errors of measurement, estimate, and prediction and their application to test scores. Percept Mot Skills. [cited 2019 May 22];82(3_suppl):1139–44. Available from: http://journals.sagepub.com/doi/10.2466/pms.1996.82.3c.1139
38. de Vet H, Terwee C, Mokkink L, Knol D (2011) Measurement in medicine [Internet]. Cambridge: Cambridge University Press; [cited 2019 May 22]. Available from: http://ebooks.cambridge.org/ref/id/CBO9780511996214
39. Terwee C, Bot S, de Boer M et al (2007) Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. [cited 2019 Aug 16];60(1):34–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17161752
40. Bland J, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet (London, England). [cited 2019 May 22];1(8476):307–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2868172
41. Walter SD, Eliasziw M, Donner A (1998) Sample size and optimal designs for reliability studies. Stat Med. 17(1):101–110
42. Yang M, Guo Y, Gong J et al (2018) Relationships between functional fitness and cognitive impairment in Chinese community-dwelling older adults: a cross-sectional study. BMJ Open. [cited 2019 Jul 25];8:20695. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5961618/pdf/bmjopen-2017-020695.pdf
43. Muir-Hunter S, Graham L, Montero-Odasso M (2015) Reliability of the Berg Balance Scale as a clinical measure of balance in community-dwelling older adults with mild to moderate Alzheimer disease: a pilot study. Physiother Canada. Aug [cited 2019 Jul 25];67(3):255–62. Available from: https://utpjournals.press/doi/10.3138/ptc.2014-32
44. Kozakai R (2017) Grip strength and healthy aging. J Phys Fit Sport Med. [cited 2019 Jul 25];6(3):145–9. Available from: https://www.jstage.jst.go.jp/article/jpfsm/6/3/6_145/_article

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