ACRM
AMERICAN CONGRESS OF
REHABILITATION MEDICINE

Archives of Rehabilitation Research and Clinical Translation

Archives of Rehabilitation Research and Clinical Translation 2021;3:100146

Available online at www.sciencedirect.com

OPEN ACCESS

ARCHIVES of Rehabilitation Research & Clinical Translation

An OPEN ACCESS JOURNAL serving the rehabilitation community

AN OFFICIAL JOURNAL OF ACRM

Original Research

Reliability of Virtual Physical Performance Assessments in Veterans During the COVID-19 Pandemic

Elisa F. Ogawa, PhD [a,b], Rebekah Harris, PT, DPT, PhD [a], Alyssa B. Dufour, PhD [c,d], Miriam C. Morey, PhD [e,f], Jonathan Bean, MD, MPH [a,b,g]

[a] *Geriatric Research, Education, and Clinical Center/Veterans Affairs Boston Healthcare System, Boston, MA*
[b] *Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA*
[c] *Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA*
[d] *Department of Medicine, Harvard Medical School, Boston MA*
[e] *Geriatric Research, Education, and Clinical Center/Veterans Affairs Healthcare System, Durham, NC*
[f] *Department of Medicine, Duke University Center for Aging/Claude D. Pepper Older Americans Independence Center, Durham, NC*
[g] *Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Boston, MA*

KEYWORDS
COVID-19;
Physical functional performance;
Rehabilitation;
Telemedicine

Abstract *Objective:* To determine the reliability of 3 physical performance tests performed via a telehealth visit (30-s arm curls test, 30-s chair stand test, 2-min step test) among community-dwelling older veterans.

Design: Cross sectional study.

Setting: Virtual.

Participants: Veterans (N=55; mean age 75y) who enrolled in Gerofit, a virtual group exercise program.

Interventions: Not applicable.

Main Outcome Measures: Participants were tested by 2 different assessors at 1 time point. The intraclass correlation coefficient (ICC) with 95% confidence intervals and Bland-Altman plots were used as measures of reliability. To assess generalizability, ICCs were further evaluated by health conditions (type 2 diabetes, arthritis, obesity, depression).

List of abbreviations: COVID-19, coronavirus disease 2019; ICC, intraclass correlation coefficient; SFT, Senior Fitness Test; VA, Veterans Affairs.

Certain Gerofit sites were supported by the VHA Office of Geriatrics and Extended Care and the Office of Rural Health. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.
Disclosures: none.
Cite this article as: Arch Rehabil Res Clin Transl. 2021;3:100146

https://doi.org/10.1016/j.arrct.2021.100146

Results: Assessments were conducted among 55 participants. The ICC was above 0.98 for all 3 tests across health conditions and Bland-Altman plots indicated that there were no significant systematic errors in the measurement.

Conclusions: The virtual physical performance measures appear to have high reliability and the findings are generalizable across health conditions among veterans. Thus, they are reliable for evaluating physical performance in older veterans in virtual settings.

Maintenance of health and independence is important for all older adults, especially for the rapidly growing number of aging veterans.[1] Assessment of physical function and especially mobility is widely advocated as a means to monitor risk for adverse health outcomes as well as patient prioritized goals of care.[2,3] The Senior Fitness Test (SFT) is an established, well-validated, and reliable physical performance test commonly used in geriatric and rehabilitative care.[4,5] Furthermore, SFT is correlated with usual gait speed[6] and falls risks.[7] Veterans Affairs (VA) has placed a high priority on developing modes of telehealth, but the reliability of physical performance tests in the virtual setting is unknown. The coronavirus disease 2019 (COVID-19) pandemic rapidly transformed the US health care system with a substantial increase in virtual visits and assessments.[8] In response to the pandemic, the VA Gerofit program transformed its face-to-face group exercise program into telehealth-delivered classes.[9] Gerofit is a group-based exercise program that promotes health and wellness for older veterans. Gerofit was declared a VA best practice, with 17 VA medical centers having implemented it.[10] Since the beginning of the pandemic, Gerofit has served over 240 veterans at 13 VA medical centers. Before the COVID-19 pandemic, face-to-face Gerofit physical performance assessments included the SFT.[11] These assessments were done to help guide personalized exercise prescription and monitor progression.[4,5,10,12] Virtual Gerofit physical performance tests were chosen based on their ability to be safely performed at home with minimal space and equipment required.

As virtual physical performance assessment became a necessity, the need to evaluate psychometric properties of previously used face-to-face instruments in a virtual setting became a priority. Reliability is one of the measurement properties referring to the degree to which the results from the measurement are stable and consistent.[13] Reliability plays an important role in ensuring the quality of the results from the instrument in research, clinical practice, and health assessment.[14] Thus, the purpose of this study was to examine the reliability of the virtual physical performance assessments conducted during the Gerofit assessments.

Methods

To track outcomes, veterans who enroll in Gerofit undergo a physical and questionnaire assessment at baseline, 3, 6, and 12 months and then annually thereafter. This cross-sectional study was conducted from May 2020 to February 2021 during the COVID-19 pandemic. Gerofit is a clinical program; thus, participation is voluntary and written consent is not required. The Durham VA Health Care System maintains an annually reviewed and approved institutional review board protocol for retrospective analyses of program outcomes for all participating Gerofit sites. This analysis included 55 Gerofit participants from the VA Boston Healthcare System who completed the virtual physical performance assessment at their appropriate time point.

Referrals to Gerofit are generated by their medical providers. Exclusion criteria for Gerofit include an inability to perform activities of daily living, cognitive impairment, unstable angina pectoris, proliferative diabetic retinopathy, oxygen dependence, incontinence, open wounds, volatile behavior, inability to be effective in a group setting, active substance abuse, and homelessness.[10]

The virtual physical performance assessments were assessed by 2 testers, a physical therapist and an exercise physiologist, at the same point in time using VA Video Connect[a] or Zoom Video Conference.[b] Before the virtual assessment, the assessors obtained the veteran's name, location, phone number, emergency contact, and medical history during the medical chart review. Veteran's age, body mass index, sex, and race were recorded from the computerized patient record system. During the virtual assessment, global health, self-reported physical activity, pain, fear of falling, and self-reported conditions were also recorded. Veterans reported average pain using a 0-10 scale, with 0 indicating no pain and 10 indicating worst imaginable pain in the past 7 days.

Physical performance measures

Fully remote, virtual physical performance assessment included (1) 30-second arm curls to measure upper extremity function; (2) 30-second chair stands to measure lower extremity function; and (3) 2-minute step test to measure cardiorespiratory function from the SFT.[15] All veterans completed a 5-minute warmup and all test directions were standardized. Adaptations and methods for the performance tests to a virtual setting were developed by Durham Gerofit program.[9] Considering the possible video and audio delay, testers started the timer once they observed the initiation of the movement by the veteran. For the arm curl test, veterans were instructed to do as many curls as they possibly could using available weights or household items (eg, dumbbell, water jug) in 30 seconds. Veterans self-reported the weight that they were using. For the chair stand test, veterans were instructed to sit in the middle of any available chair, arms across their chest, and stand all the way up and down as fast and as many times as they could for 30 seconds. The number of repetitions was recorded for arm curls and chair stands. As part of the 30-second chair stands, time to complete the initial 5 chair stands was also recorded. For the 2-minute step test, veterans were instructed to march in place, bringing their knee halfway up between knee and hip as many times as they could in 2 minutes. The number of steps completed during the 2 minutes was recorded.

The video camera angle was adjusted for each test so that the testers were able to observe veterans' full range of motion (eg, arms for arm curl, chair and torso for chair stand, and legs for 2-minute step test). Additional verbal cues and demonstrations were provided by the tester who was leading the assessment when veterans were not performing the tests correctly (eg, not sitting or standing up all the way).

Statistical analysis

Sample characteristics are presented as means and SDs for continuous variables and frequencies and percentages for categorical variables. Student *t* test for paired sample was used to test the difference between the 2 testers. The reliability was assessed using the intraclass correlation coefficient (ICC) with 95% confidence intervals and Bland-Altman plots. ICC values range from 0-1, where 1 corresponds to perfect agreement. An ICC ≥0.80 was considered high, 0.60-0.79 moderate, and <0.60 poor relative reliability.[16] ICCs were calculated using a 2-way random effects model. As a secondary analysis, we examined the ICC based on presence or absence of specific health conditions (type 2 diabetes, arthritis, obesity body mass index≥30, and depression) that might influence performance. Sensitivity analysis was conducted by removing tests that were noted as having technical difficulties (eg, frozen screen). Data were analyzed using Stata 15.1[c] with 2-sided tests at an α=.05 significance level.

Results

Sample characteristics for the total sample (N=55) are presented in table 1. Participants were primarily male (n=47, 85.5%) and non-Hispanic White (n=48, 87.3%), with an average age of 75 years. Three-fourths of the veterans reported a diagnosis of arthritis, and one-third reported fear of falling.

The 2 testers assessed 60 virtual physical performance assessments, with 5 participants repeating their assessments (eg, baseline, 3mo). Technology and/or the internet connection were limiting factors during the assessments. Of the 60 visits, 3 visits were complicated by technology problems (5%). We experienced multiple rescheduling or canceling of the virtual performance assessments because of technical difficulties.

The means of the physical performance tests for each tester are presented in table 2 along with the results from the reliability analysis. The ICC reflects high reliability for all tests (ICC>0.99). Furthermore, reliability remained high across health conditions (ICC>0.98) (table 3). A visual inspection of the Bland-Altman plots with 95% limits of agreement between testers revealed no significant proportional bias (fig 1). The mean differences between testers were 0.15 repetitions for arm curls, 0.14 repetitions for chair stands, 0.14 steps for 2-minute step test, and 0.12 seconds for 5 chair stands. The largest difference between the 2 testers for each test was 5 reps, 2 reps, 9 steps, and 3.08 seconds, respectively. This included assessments both with and without technical difficulties. Removal of performance scores with noted technical difficulties did not materially alter the findings.

Table 1 Baseline characteristics of Gerofit virtual assessment participants (N=55)

Characteristics	Mean ± SD
Age (y)	74.6±8.1
BMI (kg/m^2)	29.4±5.8
Moderate-intensity aerobic activity/wk (min)	84.1±96.7
Moderate-intensity strength training/wk (min)	21.5±30.0
Pain rating (0-10 scale)	3.7±1.0
	n (%)
Sex (female)	8 (14.6)
Race (non-Hispanic White)	48 (87.27)
Fear of falling (yes)	17 (30.9)
Self-report health rating (excellent/very good)	34 (61.82)
Self-report diagnosis of diabetes	38 (69.09)
Self-report diagnosis of arthritis	42 (76.4)
Self-report diagnosis of neuropathy	11 (20.0)
Self-report diagnosis of depression	36 (65.5)

NOTE. Five of the 55 study participants completed multiple virtual assessments.
Abbreviation: BMI, body mass index.

Discussion

This study provides important psychometric information regarding the reliability of face-to-face physical performance assessment conducted within virtual settings among veterans. Our findings suggest that virtual physical performance assessments arm curl test, chair stand test, and 2-minute step test from the SFT are reliable (ICC>0.99), and there was no significant systematic error in the measurement among these older veterans.

The high ICC values observed from this study correspond well with the results presented for face-to-face performance assessments of community-dwelling older adults[5] and clinical populations.[17,18] For example, among community-dwelling older adults aged 60-94 years, ICC values for SFT were between 0.80 and 0.98.[5] Other studies among those with cognitive impairment and type 2 diabetes reported ICC values ≥0.92.[17,18] Unfortunately,

Table 2 Reliability for physical performance measurements (n=60)

Measurements	Tester 1 Mean ± SD	Tester 2 Mean ± SD	Difference Mean ± SD	ICC (95% CI)
30-second arm curls (reps)	20.62±6.22	20.47±6.40	0.27±1.11	0.992 (0.986-0.995)
30-second chair stand (reps)	11.49±3.53	11.38±3.42	0.14±0.72	0.989 (0.981-0.994)
2-minute step test (steps)	84.29±34.18	84.16±33.99	0.14±1.99	0.999 (0.999-0.999)
5 chair stands (s)	12.74±3.57	12.86±3.52	−0.12±0.68	0.990 (0.684-0.995)

NOTE. Five of the 55 study participants completed multiple virtual assessments.
Abbreviations: CI, confidence interval; ICC, interclass correlation coefficient.

Table 3 Reliability for physical performance measurements by comorbidities (n=60)

Comorbidities	30-Second Arm Curls	30-Second Chair Stand	2-Minute Step Test	5 Chair Stand
Diabetes				
Yes	0.987 (0.975-0.994)	0.988 (0.976-0.994)	0.999 (0.993-0.999)	0.987 (0.974-0.994)
No	0.996 (0.989-0.999)	0.989 (0.970-0.996)	0.998 (0.994-0.999)	0.995 (0.987-0.998)
Arthritis				
Yes	0.989 (0.979-0.994)	0.984 (0.969-0.992)	0.998 (0.997-0.999)	0.989 (0.979-0.994)
No	0.998 (0.992-0.999)	0.995 (0.983-0.998)	0.996 (0.999-0.999)	0.995 (0.984-0.999)
Obesity				
Yes	0.996 (0.993-0.998)	0.998 (0.995-0.999)	0.999 (0.998-0.999)	0.997 (0.994-0.999)
No	0.986 (0.968-0.994)	0.973 (0.938-0.988)	0.999 (0.998-0.999)	0.985 (0.967-0.994)
Depression				
Yes	0.992 (0.983-0.996)	0.989 (0.977-0.994)	0.999 (0.998-0.999)	0.988 (0.975-0.994)
No	0.993 (0.980-0.997)	0.985 (0.961-0.994)	0.999 (0.997-0.999)	0.993 (0.984-0.998)

NOTE. Five of the 55 participants completed multiple virtual assessments. Values are ICCs with 95% confidence intervals.

no prior data on reliability of the 2-minute step test exist within similar populations.

Face-to-face SFT was developed to assess older adults' fitness levels to estimate the fitness level needed to remain independent in their later life.[4,5] It has been used widely in both research and clinical settings.[4] Our study suggests virtual SFT is also reliable supporting its uses among veterans undergoing virtual care. Considering that Gerofit is serving over 240 veterans and monitoring progression over time using the assessments evaluated in this study, the findings support the clinical utility of these tests.

Technology problems are always a concern in telehealth.[19] Although we experienced multiple cancelations and rescheduling of the assessments owing to technology and/or the internet, we experienced relatively few connection delays (5%) where 1 of the testers was not able to observe the performance tests in real time. As a result, several of the largest differences between testers can be explained by technical difficulties. However, the exclusion of these performance scores did not materially alter the findings. It is important to note that in addition to technical difficulties, we experienced several challenges in participants' home environments that made it challenging to complete the performance assessment (eg, poor lighting and poor camera angle). Furthermore, there are known characteristics associated with lower rates of telehealth utilization, including advanced age, rural residency, lower socioeconomic status, and racial and ethnic background.[20] Thus, future studies should examine the reliability of virtual physical performance assessments in diverse population to generalize the findings.

Study limitations

There are limitations to our study that need to be considered. Firstly, the protocols of the virtual performance tests were modified from the face-to-face SFT. For example, veterans used available weights and chairs, for which we did not know the exact weight or height to complete the assessments. These modifications are inevitable for virtual assessments because veterans generally do not always have the identical equipment used in a normal face-to-face assessment. Secondly, we examined the interrater reliability at a single time point and not prospectively. This study was designed purposely to assess reliability in the framework of the virtual Gerofit assessments. Thus, we were unable to conduct test-retest reliability.

Despite these limitations, several strengths of this study are noteworthy. Although there are new technological developments in assessing physical performance virtually using mobile health,[21-23] these have been done in controlled laboratory settings. The strength of our study is in the setting, where we evaluated the reliability of real-life virtual

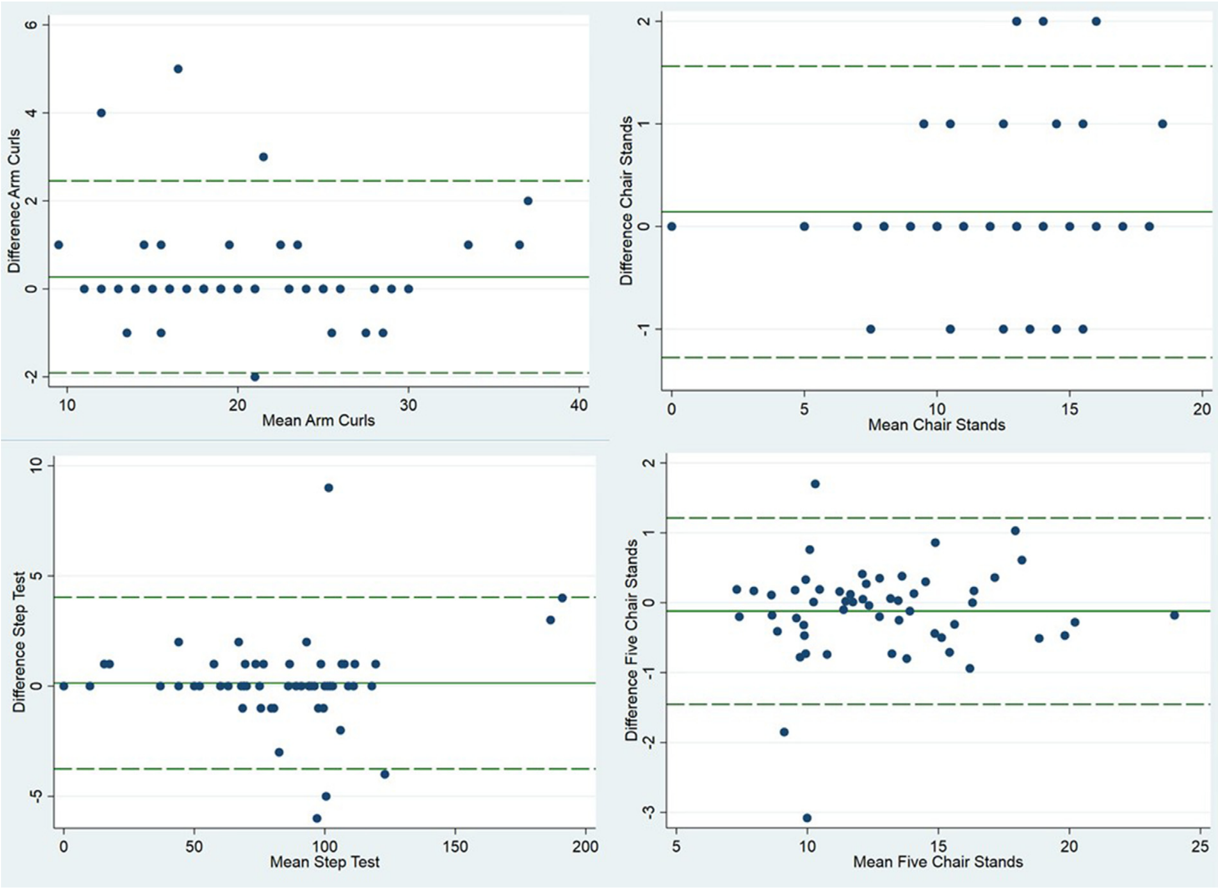

Fig 1 Bland-Altman plots for arm curls, chair stands, 2-minute step test, and 5 chair stands (N=60).

assessment (eg, veterans joining from their home). Our findings support the usage of performance tests in telehealth virtual settings, which we will continue to use post COVID-19 pandemic.[24] Thus, this study provides supporting evidence that virtual performance assessments can be used in clinical and research environments.

Conclusions

The virtual physical performance measures appear to have high reliability across health conditions among veterans. Thus, they are reliable for evaluating physical performance in older veterans in virtual settings.

Suppliers

a. VA Video Connect; VA Mobile Health.
b. Zoom Video Conference; Zoom Video Communications.
c. Stata 15.1; StataCorp.

Corresponding author

Elisa F. Ogawa, PhD, Geriatric Research, Education, and Clinical Center/Veterans Affairs Boston Healthcare System, 150 S Huntington Ave, Boston, MA 02130. *E-mail address:* elisa.ogawa@va.gov.

References

1. Pruchno R. Veterans aging. Gerontologist 2016;56:1-4.
2. Bean JF, Orkaby AR, Driver JA. Geriatric rehabilitation should not be an oxymoron: a path forward. Arch Phys Med Rehabil 2019;100:995-1000.
3. Studenski S, Perera S, Wallace D, et al. Physical performance measures in the clinical setting. J Am Geriatr Soc 2003;51:314-22.
4. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist 2013;53:255-67.
5. Rikli R, Jones CJ. Senior fitness test manual. Champaign: Human Kinetics 2001.
6. Wu T, Zhao Y. Associations between functional fitness and walking speed in older adults. Geriatr Nurs 2021;42:540-3.
7. Toraman A, Yildirim NU. The falling risk and physical fitness in older people. Arch Gerontol Geriatr 2010;51:222-6.
8. Mann DM, Chen J, Chunara R, et al. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020;27:1132-5.
9. Jennings SC, Manning KM, Bettger JP, et al. Rapid transition to telehealth group exercise and functional assessments in response to COVID-19. Gerontol Geriatr Med 2020;6:2333721420980313.
10. Morey MC, Lee CC, Castle S, et al. Should structured exercise be promoted as a model of care? Dissemination of the Department

of Veterans Affairs Gerofit Program. J Am Geriatr Soc 2018;66:1009–16.

11. Serra MC, Addison O, Giffuni J, et al. Physical function does not predict care assessment need score in older veterans. J Appl Gerontol 2019;38:412–23.
12. Morey MC, Crowley GM, Robbins MS, et al. The Gerofit Program: a VA innovation. South Med J 1994;87:S83–7.
13. Committee on Psychological Testing IVT. for Social Security Administration Disability Determinations. Board on the Health of Select Populations. Institute of Medicine. Overview of psychological testing. Psychological testing in the service of disability determination. WashingtonDC: National Academies Press; 2015.
14. Souza AC, Alexandre NMC, Guirardello EB. Psychometric properties in instruments evaluation of reliability and validity. Epidemiol Serv Saude 2017;26:649–59.
15. Rikli RE, Jones CJ. Functional fitness normative scores for community-residing older adults, ages 60-94. J Aging Phys Act 1999;7:162.
16. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231–40.
17. Hesseberg K, Bentzen H, Bergland A. Reliability of the Senior Fitness Test in community-dwelling older people with cognitive impairment. Physiother Res Int 2015;20:37–44.
18. Alfonso-Rosa RM, Del Pozo-Cruz B, Del Pozo-Cruz J, et al. Test-retest reliability and minimal detectable change scores for fitness assessment in older adults with type 2 diabetes. Rehabil Nurs 2014;39:260–8.
19. Lopez AM, Lam K, Thota R. Barriers and facilitators to telemedicine: can you hear me now? Am Soc Clin Oncol Educ Book 2021;41:25–36.
20. Hsiao V, Chandereng T, Lankton RL, et al. Disparities in telemedicine access: a cross-sectional study of a newly established infrastructure during the COVID-19 pandemic. Appl Clin Inform 2021;12:445–58.
21. Urena R, Chiclana F, Gonzalez-Alvarez A, et al. m-SFT: a novel mobile health system to assess the elderly physical condition. Sensors (Basel) 2020;20:1462.
22. Mellone S, Tacconi C, Chiari L. Validity of a smartphone-based instrumented Timed Up and Go. Gait Posture 2012;36:163–5.
23. Banos O, Moral-Munoz JA, Diaz-Reyes I, et al. mDurance: a novel mobile health system to support trunk endurance assessment. Sensors (Basel) 2015;15:13159–83.
24. Kichloo A, Albosta M, Dettloff K, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020;8:e000530.