J Musculoskelet Neuronal Interact 2019; 19(2):215-219

Journal of Musculoskeletal and Neuronal Interactions JMNI

Original Article

Determine real change in performance during single session of functional testing

Ibrahim Mustafa Altubasi

Department of Physiotherapy, School of Rehabilitation Sciences, The University of Jordan

Abstract

Objective: Performance-based measures of function are used in both clinical and research settings. During functional testing procedures, participants' performance change from one trial to another. Therefore, there is a need to know how much change in performance is considered a real change from one trial to another during the same testing visit. **Methods:** The current study has a cross-sectional design. A sample of 51 healthy subjects participated in this investigation. Participants attended 1 testing visit during which all subjects completed the functional tests twice. The functional tests used in the current study were: timed up and go, 10-meter walking time, 5 times chair rise, and climbing stairs. **Results:** Intra-class correlation coefficients were high on all functional tests for the whole sample and for different subgroups of age and gender. There were small standard error of measurement (SEM) and MDC values for the whole sample on all functional performance tests. Age groups showed different SEM and MDC on the repeated performance of tests. However, gender didn't have an effect on SEM and MDC. **Conclusions:** The current study provides data that can guide clinicians in determining whether trial-to-trial changes in performance would be real changes or just measurement error.

Keywords: Intra-Session Reliability, Agreement, Standard Error of Measurement, Minimal Detectable Change, Functional Performance

Introduction

Performance-based measures of function have been widely used in both clinical and research settings. These measures include but are not limited to gait speed, stair climbing, get up and go, and chair rising. Additionally, the reliability and responsiveness for many functional performance tests have been investigated in the literature in order to determine their ability to detect clinical changes over different time points.

During any functional testing procedure, researchers ask participants to perform the functional tests twice or more in order to better estimate their performance. Also, clinicians use functional tests as outcome measures in their daily practice and they ask their patients to perform the test many times during the same visit to estimate patients' performance. However, participants' or patients' performance change from one trial to another. Therefore, how many times the researcher asks the participates or patients to perform the test in order to accurately measure their level of function remains an issue that needs to be explored. Consequently, there is a need to know how much change in performance is considered a real change from one trial to another during the same testing visit

The purposes of this study were first to determine intra-session within rater relative and absolute reliability as well as minimal detectable change (MDC) in performance for 4 physical performance tests. The second purpose is to explore age and gender influences on the MDC for the 4 physical performance tests.

Materials and methods

A convenient sample of 51 healthy subjects was recruited in this investigation. Participants were recruited if they were 20 years old or older and could walk independently without using assistive devices. Participants were excluded from participation if they had any cardiovascular, neuromuscular,

The author has no conflict of interest.

Corresponding author: Ibrahim Mustafa Altubasi, PT, PhD, Department of Physiotherapy, School of Rehabilitation Sciences, The University of Jordan, Queen Rania St, Amman 11942, Jordan
E-mail: i.tubasi@ju.edu.jo

Edited by: G. Lyritis
Accepted 28 February 2019

or respiratory diseases, inflammatory arthritis, muscular disorders or were participating in a regular exercise program.

Recruitment of participants was carried out between March 2014 and October 2015 through disseminating the study information by the word of mouth among individuals in an academic institution. All study procedures were approved by The University of Jordan Hospital Institutional Review Board, and all subjects provided informed consent before participation.

The current study has a cross-sectional design. Participants in this investigation attended 1 testing visit at the School of Rehabilitation Sciences at the University of Jordan. For the within rater reliability aim in this study, a physiotherapist measured the time for all participants to completed the functional performance tests twice with up to 5 minute-rest between testing trials. Participants were asked to perform 1 practice trial of each functional performance test to get familiar with the tests before data collection. The functional tests were administered in the following order: timed up and go, 10-meter walking time, 5 times chair rise, and climbing stairs. All functional tests were timed by using the stopwatch application embedded within a smartphone.

Timed up and go test

Standardized performance of the timed up and go test was instructed to the participants as described by Yuksel et al.[1]. Participants were seated on a standard height chair with armrests. When instructed to start the test, participants stood up and walked 3 meters, then turned around, walked back to the chair and sat down. Participants were instructed to perform the test safely as quickly as possible.

Ten-meter walking time test

Two cones were placed on a floor of a corridor representing the 10 meter distance. Participants were asked to stand 2 meters before one of the cones to account for acceleration. Then they were instructed to walk at their regular speed and not to stop until passing by the other cone to account for deceleration. The time required to cross the 2 cones was recorded.

Five times chair rise test

Participants were instructed to stand up from a standard height chair as quickly as they safely could 5 times while keeping their arms crossed on their chests. The time to complete the 5 stand ups was recorded.

Climbing stairs test

Participants were instructed to climb a flight of stairs consisting of 9 steps as quickly as possible. Step height is 16 cm and step depth is 30 m. The time to climb the stairs was recorded.

Statistical procedures

All statistical procedures were performed on SPSS version 20 (IBM, SPSS Statistics 2009). Age variable was categorized into the following groups: young (between 20-39 years), middle-aged (40-59 years) and old (60 years and above).

Intra-class correlation coefficients (ICC) (2-way random model 2, 1) type consistency, were used to determine test-retest reliability. ICC values were calculated for the whole cohort then they were calculated for each age group. Next ICC values were calculated for females and males. Values of ICC were considered good to excellent if more than 0.75, however for the clinical use of measurement it was suggested that the ICC should be more than .9[2,3].

The standard error of measurement (SEM) is considered an absolute measure of reliability, it is used to quantify the error in the measurements of all functional tests. SEM was calculated based on the following formula[4,5].

$$SEM = SD\ pooled\ X \sqrt{(1\text{-}ICC)}$$

where SD pooled is the pooled standard deviation from repeated measurements and ICC is the intra-class correlation coefficients for the intra-rater reliability.

Also, MDC at 95% confidence level (MDC 95%) for all tests were calculated to determine the smallest change that is considered true. MDC 95% was calculated by using the following formula[4,5].

$$MDC\ 95\% = SEM\ X\ 1.96\ X \sqrt{2}$$

The standard error of measurement and MDC 95% were calculated for each age group and also for males and females.

Results

Fifty-one healthy participants were recruited in the current study. The average age of the study sample was 51.24 years (range: 21-82 years). There were 19 females (mean age 47.21 range 29-77 years) and 32 males (mean age 53.63 range 21-82 years) in the current sample. The number of participants in each age group was 11 in the young (mean age 29.64 years), 24 in the middle age (mean age 48.67 years) and 16 in the older adults (mean age 69.94 years). Table 1 presents the results of the functional performance test during the first and second trials. By looking at the whole sample, the time to complete all functional performance tests was significantly shorter in the second trial compared to the first one except for the timed up and go test. Among different age groups, there were significant differences between trials of on timed up and go test among the middle age group and walking time and chair rise tests among the older adults group. For females, there were significant differences between trials on timed up and go and chair rise tests. However, for males, there was a significant difference between trials only on chair rise test.

Table 2 presents single measure ICC, SEM and $MDC_{95\%}$ for the whole sample. ICC ranged from .963-.986. Table 3 and 4 present single measure ICC, SEM and $MDC_{95\%}$

Table 1. Mean performance on the first and second trials for all tests.

	TUG Trial 1	TUG Trial 2	10-meter Walking trail 1	10-meter Walking trial 2	Climbing stairs trial 1	Climbing stairs trial 2	Chair rise trial 1	Chair rise trial 2
Whole group	6.60	6.52	7.81	7.69*	3.59	3.54*	11.16	10.85*
Young	5.61	5.58	7.16	7.08	2.90	2.82	9.83	9.71
Middle age	6.44	6.28*	7.61	7.59	3.47	3.43	10.70	10.46
Older adults	7.53	7.51	8.62	8.30*	4.26	4.20	12.77	12.22*
Females	6.97	6.75*	8.03	7.86	3.77	3.70	11.46	11.13*
Males	6.39	6.38	7.69	7.59	3.49	3.45	10.97	10.69*
** Indicate significant difference between trials p-value < .05. TUG: Timed up and Got Test.*								

Table 2. Reliability measures and minimal detectable change for the whole group.

	ICC (CI 95%)	SEM	$MDC_{95\%}$
TUG	.974 (.955-.985)	.229	.636
10-meter Walking	.972 (.952-.984)	.262	.725
Climbing stairs	.986 (.976-.992)	.114	.317
Chair rise	.963 (.937-.979)	.456	1.264

Table 3. Reliability measures and minimal detectable change for age groups.

	Young			Middle Age			Older adults		
	ICC (CI 95%)	SEM	$MDC_{95\%}$	ICC (CI 95%)	SEM	$MDC_{95\%}$	ICC (CI 95%)	SEM	$MDC_{95\%}$
TUG	.961 (.862-.989)	.147	.407	.898 (.778-.954)	.223	.617	.981 (.945-.993)	.273	.756
Walking Time	.961 (.862-.989)	.165	.457	.970 (.931-.987)	.217	.602	.976 (.930-.992)	.335	.929
Climbing Stairs	.939 (.790-.983)	.098	.273	.973 (.939-.988)	.116	.320	.989 (.969-.996)	.123	.342
Chair Rise	.811 (.440-.945)	.413	1.146	.948 (.884-.977)	.449	1.246	.975 (.929-.991)	.457	1.268

Table 4. Reliability measures and minimal detectable change for gender.

	Females			Males		
	ICC (CI 95%)	SEM	MDC 95%	ICC (CI 95%)	SEM	MDC 95%
TUG	.973 (.932-.990)	.208	.576	.977 (.954-.989)	.227	.630
Walking Time	.959 (.896-.984)	.256	.710	.977 (.953-.989	.263	.728
Climbing Stairs	.987 (.967-.995)	.121	.336	.985 (.970-.993)	.111	.308
Chair Rise	.956 (.890-.983)	.439	1.216	.966 (.932-.983)	.468	1.297

for different age groups and gender respectively. ICC for young ranged from .811-.961 with chair rise test showed smallest ICC value= .811. For the middle-age group, the ICC values ranged from .898-.973 and for the older adults group the ICC values ranged from .975-.989. When the participants were examined by gender, the ICC values ranged from .956-.987 for females and .966-.985 for males on all tests of physical function.

Discussion

The current study examined the intra-session within rater reliability, measurement error and minimal detectable change in the performance of physical functional tests among healthy participants. The 4 physical performance tests used in this study are simple and could be used in different clinical settings. These tests assess different aspects of physical function; timed up and go test assesses balance and postural control[6,7]. Stair climbing and chair rise tests assess lower extremities strength[7], and the self-based walking time test assesses general mobility and gait speed[7].

As seen in Table 1, all functional performance tests on the second trial were shorter in time compared to the first trial. Although not all of them were statistically significant, trends on each functional test suggested that a learning effect had occurred, making participants faster on the second trial compared to the first one. Accordingly, in order to better estimate participants' true performance we need to identify how much change is considered a real change in performance on repeated measurements

Examining the intra-session test-retest reliability through relative and absolute measures of reliability revealed consistent findings. ICC values for all functional performance tests were excellent and SEM values were small indicating reproducibility of results on the repeated measurement of physical function.

The ICC is a relative measure of reliability; it is a ratio of the total variability in the measurement (between-subject variability + measurement error) that is explained by between-subjects variability[8]. The high ICC values found in this study revealed the tests' ability to distinguish between subjects' performance beyond measurement error[8]. Accordingly, the test-retest reliability of the functional performance tests was high due to small measurement error on repeated measurements. ICC values were high on all functional performance tests for the whole sample as well as for different subgroups of age and gender. The smallest ICC value was found on chair rise test for the young adults' group (ICC= .811) and is probably due to the small variability between young adults on the performance of chair rise test (pooled SD= 0.95 s).

In order to determine the $MDC_{95\%}$ for the 4 physical performance tests and compare them between different age groups and gender, SEM values were calculated first for all physical performance tests. There were small SEM values for the whole sample on all functional performance tests (Table 2) indicating agreement between repeated measurement. However, different age groups showed different measurement errors in the repeated performance of functional tests. Young adults have smaller SEM values on all functional tests compared to middle age and older adults groups (Table 3). This finding indicates that young adults are more stable in their performance on repeated measurements. SEM values for older adults were the largest compared to the other age groups which indicate that older adults fluctuate more in their performance of functional tests. Gender didn't seem to have an effect on measurement errors; SEM values were similar between males and females (Table 4).

In order to understand the clinical importance of the measurement error from the repeated performance of functional tests, the $MDC_{95\%}$ was calculated for all functional performance tests to help clinician and investigators determine whether the observed change on repeated measurement is true change or just measurement error. The $MDC_{95\%}$ for all functional tests for the whole sample are presented in Table 2. Those values represent the smallest differences from repeated performance of tests that are considered a real change in performance. However, $MDC_{95\%}$ for functional tests seems to be influenced by the age groups since it depends on its calculation on the SEM which is affected by age groups. Therefore it is important to know the age of the subjects before deciding whether the changes in repeated functional performance are true changes or not. For example, if someone had a change score on repeated timed up and go test equals .636 s (which is the $MDC_{95\%}$ for the whole sample in the current study as presented in Table 3), this would be a true change for young and middle age groups but it would be considered a measurement error for older adults. The $MDC_{95\%}$ for females and males seem to be similar to each other (Table 4), therefore gender variable did not influence the $MDC_{95\%}$ values.

The results of this study provide a guideline for better estimation of patients' or participants' performance on the 4 functional performance tests during single evaluation visit. On repeated testing during the same visit, if the time on second testing trial is shorter than the time on the first trial, then calculate the difference between trials and according to the age of the participant check whether the difference is larger than the $MDC_{95\%}$ for the specific test of interest. If the difference is larger than the $MDC_{95\%}$, a learning effect may occur, therefore ask the participants to perform a third trial. A difference between the third and the second trials that is less than the $MDC_{95\%}$ necessitates stopping evaluation on that specific test and either of the last two trials can be used as an outcome. However, a difference larger than $MDC_{95\%}$ between third and second trial indicates the performance of a fourth trial. Continued learning effect between trials indicates repeating the test until the difference is less than $MDC_{95\%}$. However, if the performance time on the second trial is longer than the one on the first trial and the change is larger than $MDC_{95\%}$, a fatigue effect may occur. Thus, measuring clinician should then ask the participants to perform a third trial after extended rest time. A difference between the third and second trials that is less than $MDC_{95\%}$ requires stopping evaluation of that specific test and any of the last two trials can be used as the outcome.

The current study sample included healthy subjects as evident on their performance of functional tests. The older adults group in the current study is the slowest group in their functional performance, however their performance in the current study is considered better than older adults in other studies. For example, the older adults' performance on the timed up and go test in the current study was better than what Pondal et al.[9] found in their study for older adults' group. Furthermore, the scores of the current study sample on the functional performance tests are better than those found in the literature for age-matched participants with other health related problems. Dobson et al.[10] reported that participants with hip and knee OA had a noteworthy longer time to complete the stair climb test compared to our sample. Also, Iijima et al.[11] found that participants with mild knee OA had longer time on the stair climb test compared to our participants. Furthermore, SEM and MDC in the current study for functional performance tests were smaller than that found in literature for participants with different health problems[1,3,5,10-11]. Therefore, clinicians need to take into consideration the influence of other health related conditions on functional performance when determining whether a real change had occurred with repeated testing or just measurement error.

In conclusion, the 4 functional performance tests: timed up and go, 10-meter walking time, climbing stairs and 5 times chair rise tests are reliable and can detect real change in performance that could occur during the repeated performance in a single session of evaluation of the physical function. The current study provides data that can guide clinicians in determining whether trial to trial changes in performance could be real changes or just measurement error.

Acknowledgments

This study is funded by the Deanship of Academic Research at the University of Jordan.

References

1. Yuksel E, Kalkan S, Cekmece S, Unver B, Karatosun V. Assessing Minimal Detectable Changes and Test-Retest Reliability of the Timed Up and Go Test and the 2-Minute Walk Test in Patients With Total Knee Arthroplasty. The Journal of Arthroplasty 2017;32(2): 426–430.
2. Marques A, Cruz J, Quina S, Regêncio M, Jácome, C. Reliability, Agreement and Minimal Detectable Change of the Timed Up & Go and the 10-Meter Walk Tests in Older Patients with COPD. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2015; 13(3): 279–287.
3. Ortega-Pérez de Villar L, Martínez-Olmos FJ, Junqué-Jiménez A, Amer-Cuenca JJ, Martínez-Gramage J, Mercer T, Segura-Ortí E. Test-retest reliability and minimal detectable change scores for the short physical performance battery, one-legged standing test and timed up and go test in patients undergoing hemodialysis. PLOS ONE 2018;13(8): e0201035.
4. Saxer S, Speich R, Toigo M, Mueller SM, Ulrich Somaini S. Reliability of parameters during stair ascent measured with Leonardo Mechanograph(®) Stair A in healthy subjects. J Musculoskelet Neuronal Interact 2015;15(3):257-63.
5. Puthoff ML, Saskowski D. Reliability and responsiveness of gait speed, five times sit to stand, and hand grip strength for patients in cardiac rehabilitation. Cardiopulm Phys Ther J 2013;24(1):31-7.
6. Dawson N, Dzurino D, Karleskint M, Tucker J. Examining the reliability, correlation, and validity of commonly used assessment tools to measure balance. Health Science Reports 2018;1(12), e98.
7. Bennell K, Dobson F, Hinman R. Measures of physical performance assessments: Self-Paced Walk Test (SPWT), Stair Climb Test (SCT), Six-Minute Walk Test (6MWT), Chair Stand Test (CST), Timed Up & Go (TUG), Sock Test, Lift and Carry Test (LCT), and Car Task. Arthritis Care & Research 2011;63(S11), S350–S370.
8. Hernaez R. Reliability and agreement studies: a guide for clinical investigators. Gut 2015;64(7):1018–1027.
9. Pondal M, del Ser T. Normative Data and Determinants for the Timed "Up and Go" Test in a Population-Based Sample of Elderly Individuals Without Gait Disturbances. Journal of Geriatric Physical Therapy 2008;31(2):57–63.
10. Dobson F, Hinman RS, Hall M, Marshall CJ, Sayer T, Anderson C, Newcomb N, Stratford PW, Bennell KL. Reliability and measurement error of the Osteoarthritis Research Society International recommended performance-based tests of physical function in people with hip and knee osteoarthritis. Osteoarthritis and Cartilage 2017;25(11):1792-1796.
11. Iijima H, Shimoura K, Eguchi R, Aoyama T, Takahashi M. Concurrent Validity and Measurement Error of Stair Climb Test in People with Pre-radiographic to Mild Knee Osteoarthritis. Gait Posture 2019;68:335-339.