



ORIGINAL RESEARCH

Impaired performance of women with patellofemoral pain during functional tests



Camile Ludovico Zamboti^{a,b}, Carlos Augusto Marçal Camillo^{a,c},
Amanda Paula Ricardo Rodrigues da Cunha^{a,b}, Thaiuana Maia Ferreira^{a,b},
Christiane Souza Guerino Macedo^{a,b,*}

^a Laboratory of Human Movement, Department of Physical Therapy, Universidade Estadual de Londrina (UEL), Londrina, PR, Brazil

^b Education, Research and Extension Laboratory of Sports Physical Therapy, Department of Physical Therapy, Universidade Estadual de Londrina (UEL), Londrina, PR, Brazil

^c Department of Rehabilitation Sciences, Universidade Pitágoras UNOPAR, Londrina, PR, Brazil

Received 25 April 2019; received in revised form 5 May 2020; accepted 11 May 2020

Available online 22 May 2020

KEYWORDS

Knee;
Anterior knee pain;
Reproducibility of results;
Physical therapy;
Exercise test

Abstract

Objective: To compare the performance, reliability, and validity of functional tests between women with and without patellofemoral pain.

Methods: Twenty women with a diagnosis of patellofemoral pain between 18 and 40 years of age and 20 age-matched pain-free controls participated in the study. All participants performed a set of five function tests: sitting-rising test, sit-to-stand in 30 seconds, stair-climb test, stair descent test, and six-minute step test. To investigate reliability, participants were assessed on two different days, seven days apart, by two independent investigators blinded to the results of the other investigator. Validity was evaluated through associations with the results on the Anterior Knee Pain Scale.

Results: Performance in the tests was worse in women with patellofemoral pain than in the control group for the sit-to-stand in 30 seconds (mean difference [MD] 3.4reps; 95%CI: 0.4, 6.4), stair-climb test (MD: 0.36 s; 95%CI: 0.1, 0.63), and six-minute step test (MD: 45reps; 95%CI: 20, 70). No differences were observed for the sitting-rising and stair descent tests. All tests in both groups showed moderate to excellent intra- and inter-rater reliability (intraclass correlation coefficients: 0.61 to 0.91 and 0.72 to 0.96, respectively). Finally, only the results on the sit-to-stand in 30 seconds test correlated with the Anterior Knee Pain Scale ($r = 0.44$, $p = 0.047$) in the patellofemoral pain group.

Conclusion: Women with patellofemoral pain present lower performance on some functional tests. Functional tests are reliable in patients with patellofemoral pain, although they are not associated with the results on the Anterior Knee Pain Scale self-questionnaire.

© 2020 Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia. Published by Elsevier Editora Ltda. All rights reserved.

* Corresponding author at: Departamento de Fisioterapia, Universidade Estadual de Londrina, Avenida Robert Kock, nº 60, Vila Operária, CEP: 86038-350, Londrina, PR, Brazil.

E-mail: chmacedouel@yahoo.com.br (C.S. Macedo).

Introduction

Patellofemoral pain (PFP) is a condition characterized by insidious and diffuse pain in the anterior region of the knee,¹ which worsens during activities involving repetitive loading of the patellofemoral joint.^{2,3} PFP is more prevalent in women and impacts daily activities exacerbating pain during functional tasks such as sitting and getting up from a chair or when going up and down stairs.⁴ The clinical impact is that an estimated 75% of patients with PFP present a reduction in activities, especially sports and recreational activities, due to pain.⁵

Function can be assessed using self-reported measures and performance-based tests.⁶ The Anterior Knee Pain Scale (AKPS) is a commonly used, reliable, self-reported measurement of function.⁷ Performance and function tests aim to provide an indirect but objective and comparable estimate of patient tolerance to the physical demands of everyday life.⁸ However, it is essential that the results of functional tests are reliable and valid.⁹

Studies that used functional tests in individuals with PFP observed exacerbation of pain while performing the tests,^{4,10} with the anteromedial lunge, step down, and single leg press being moderately correlated with pain intensity on the visual analogue scale (VAS). Tests similar to activities of daily living that exacerbate pain in this population, such as going up and down stairs and sit to stand movements, may offer a more realistic assessment of functional limitations.⁴ Additionally, specific functional tests for PFP should include overload with movements throughout knee flexion, including varied movements requiring dynamic muscular control.¹⁰

Although Loudon et al.¹⁰ reported reliable functional tests associated with knee pain, these tests were not sensitive to differences in function among people with PFP. Therefore, the literature lacks evidence for valid, reliable, and sensitive tests of functional status and changes in patients with PFP. The aim of the present study was to compare the functional performance of women with or without PFP, as well as to evaluate the reliability and validity of functional tests that mimic movements exacerbating knee pain.

Methods

Study design

This cross-sectional study was conducted in accordance with resolution 466/2012 of the Brazilian National Health Council and was approved by the institutional review board (#1.757.451) of Universidade Estadual de Londrina (UEL), Londrina, PR, Brazil. All participants provided an informed consent form prior to participation in the study.

Sample characteristics

Female participants, between 18 and 40 years of age were screened for eligibility in a non-probabilistic design with a convenience sample. To be included in the study, they could not present any of the following conditions: history of traumatic knee injury; previous lower extremity surgery;

patellar dislocation; clinical evidence of meniscal injury; ligament instability or patellar tendinopathy; neurological, cardiovascular, or rheumatic diseases; diabetes; sensory change on the plantar surface of the foot; using medication and/or performing physical therapy sessions in the previous six months; being pregnant at the time of the study.¹ Additionally, participants in the PFP group needed to present the following features: (1) a clinical diagnosis of patellofemoral dysfunction by a physician; (2) anterior knee pain of four or more points on a 0–10 VAS for at least eight weeks before evaluation; (3) insidious onset of symptoms unrelated to trauma; (4) presence of knee pain in at least three of the following tasks: climbing/descending stairs, crouching, running, kneeling, or sitting for prolonged periods.¹¹ Participants in the control group (i.e. without PFP) were included if they did not present any complaints of pain or disability, in any joints of the body. Participants were excluded if they did not complete all tests or attend the retest session.

Procedures

During the initial visit, all participants were informed about the research and data collection procedures. In addition, on the same day, they completed a general questionnaire for sample characterization, the VAS for pain, and the AKPS (both translated and validated for Brazilian Portuguese).¹² Participants, were then asked to perform five functional tests in a random sequence: sitting–rising test (SRT), sit-to-stand in 30 s (STS30), stair-climbing test (SCT), stair descent test (SDT), and six-minute step test (6MST).

To assess inter-rater reliability, the set of 5 tests was performed twice and assessed by separate investigators. The sequence of tests and the order of the two investigators were chosen randomly on Day 1 (while one assessor performed the evaluation the other remained in a separate room to blind the assessment and result). Participants returned to the laboratory one week later to repeat the entire procedure (Day 2), with the same randomized sequence as the first visit to avoid differences due to fatigue. A total of three experienced investigators performed the data collection. To guarantee standardized operational procedures, all investigators were trained for the study protocol and participated in a pilot study of this protocol.

Functional tests

Sitting–rising test: The SRT assesses the ability to sit down and stand up from the floor with partial scores for each of the two actions. The participant is required to get up from the floor, and after standing up, return to the floor, trying to use minimal support.¹³ This test evaluates the quality of movements to sit and rise from the floor. The quality of movement is analysed and scored from zero to 11. For each external support necessary, for example if the participant need to put their hand on the floor to perform sitting, or if imbalance occurs during execution, the total score is reduced by a point and a half point, respectively. Higher values refer to better functional performance. Due to the lack of evidence describing the number of repetitions necessary to assure reliability of the test, on each visit the test was repeated three times

with 1-min rest between each execution. The mean of all attempts was used as the test result.

Sit-to-stand in 30 s: The STS30 evaluates the number of repetitions of sitting down and getting up from a 42 cm-chair in 30 s. The complete movement of sitting and getting up was counted as one repetition, where higher number of repetitions correspond to a better result. If participants did not fully extend their knees or sit on the chair, the repetition was not counted. On each visit, two executions of the test were performed with 2 min rest between them, as proposed in the literature.¹⁴ The mean of the attempts was used as the test result.

Stair-climbing test and stair descent test: These tests evaluate the time taken to go up and down a set of stairs, respectively. In the SCT, the volunteer climbs eight steps in the shortest time possible without external support, with lower values corresponding to greater agility and better performance. The same protocol was used for stair descent. On each visit, participants performed two trials of each test with a 2 min rest between attempts.¹⁵ The mean of the attempts for both SCT and SDT was used as the test result.

Six-minute step test: The 6MST measures the number of times the participant climbed and descended from a 20 cm high step in 6 min. The aim is to go up and down the step as fast as possible, with higher values corresponding to better performance. The number of repetitions was counted, considering one repetition when the participant went up and down the step. The test is similar to the six-minute walk test in terms of rhythm, timing, and encouragement. Participants performed the test twice on each day, one time with each assessor,¹⁶ and the best attempt for each assessor (done on different days) was used as the test result.

For safety reasons, heart rate (using a cardio-frequency metre) and the subjective sensation of dyspnoea and exertion (using the Borg scale) were monitored during all tests. The presence of syncope, vertigo, major tachycardia (100% of maximal heart rate), or cyanosis, were indications to discontinue testing; however, none of these events were observed in this study. To control fatigue and pain, the Borg scale and VAS were completed before and after each test, if there was increase in pain or fatigue participant waited before performing another trial or test.

Statistical analyses

Difference in the STS30 between the two groups was defined a priori as the primary outcome. Based on a pilot study, forty participants (2 groups of twenty) were needed to obtain 90% power, based on a difference of 5.16 repetitions in the STS30, assuming a standard deviation of 4.27 repetitions, and an alpha lower than 0.01.

Statistical analyses were performed using Statistical Package for the Social Sciences (SPSS®) 20.0. Data are described as mean \pm SD or median [interquartile range] according to data distribution. Data normality was assessed by the Shapiro–Wilk test. Characteristics of the participants and performance in the tests were compared between patients with PFP and the control group using unpaired t tests or Mann–Whitney tests, according to the data distribution. Descriptive data of the tests and the comparison between the two groups were done using the results of the first patients' assessment during the first day, with the

exception of 6MST which was the only one performed by two raters on both days, so the best attempt of the first patients' assessment on both days was used.

Intra-rater analysis was done using the two-way mixed effects, absolute agreement and multiple raters (i.e. all test–retest data were used). Interrater analysis, was done using the two-way random effects, absolute agreement and multiple raters (i.e. all the three raters were used in the analysis).^{17,18} Reliability was deemed moderate if values were between 0.50 and 0.75, good if between 0.76 and 0.90, and excellent if above 0.90.¹⁹ To observe inter-evaluator and execution differences, two-way ANOVAs were performed, and data with non-parametric distribution were normalized with log10. The Standard Error of Measurement (SEM) was calculated to establish the absolute reliability, by means of the following equation: $SEM = SD \sqrt{1 - ICC}$, in which SD is the standard deviation and ICC is the Intraclass correlation coefficient.²⁰

Although there is no gold standard to evaluate function in individuals with PFP, validity of the tests was assessed by determining the extent of association between scores on the tests and score on the AKPS, using the Spearman correlation coefficient.

In addition, Spearman correlation was used to determine the association between pain intensity (VAS) and performance for each test. The Mann–Whitney test was used to compare differences between pain scores between groups for each test. To compare pain between groups, the average of VAS in the end of all tests of the SRT, 30STS, SCT, and SDT and the reported VAS during the best attempt of the 6MST were used. Statistical significance was set at $p < 0.05$ for all analyses.

Results

Forty participants were evaluated in the study, 20 in the PFP group and 20 in the control group. Three raters performed all study assessments. The raters were all women, with mean age of 21.6 years with experience in the application of all tests performed. In total, 160 assessments were performed by the raters.

Characteristics of the participants are presented in Table 1. Participants in both groups were similar in age and height, but those in the control group, had a significantly higher body weight and body mass index (BMI), as well as total AKPS score. The PFP group demonstrated worse scores compared to the control group in three tests, STS30, SCT, and 6MST (Table 2).

Intra-rater reliability was deemed from moderate to excellent ($0.61 < ICC < 0.91$) and inter-rater reliability from moderate to excellent ($0.72 < ICC < 0.96$) in both groups. The two-way analysis of variance did not indicate significant differences neither between raters nor between tests executions and no interaction between raters and test were observed.

Correlation values between the score on the AKPS and performance on the functional tests were low and not significant ($r = -0.30$ to 0.23 ; $p > 0.05$ for all), with the exception of AKPS and STS30 which presented a moderate correlation ($r = 0.44$; $p = 0.04$, Fig. 1). Additionally, the relationship between pain scores on the VAS and performance on the

Table 1 Characteristics of the participants.

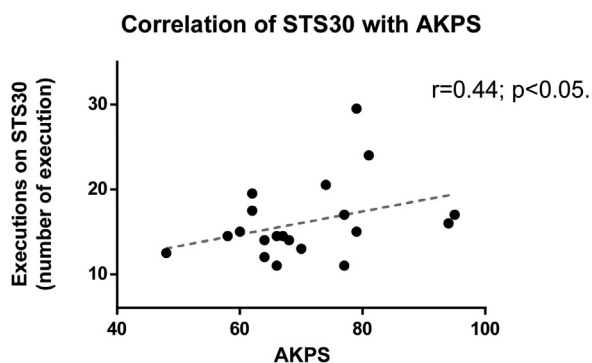
	PFP (n=20)	Control (n=20)	p value
Age (years)	25.60 ± 4.97	26.50 ± 4.13	0.53
Body weight (kg)	67.55 ± 12.40	59.15 ± 9.40	0.01
Height (cm)	164.00 ± 0.05	165.00 ± 0.08	0.76
BMI (kg/m ²)	25.05 ± 3.49	21.75 ± 2.77	0.002
Pain (0–10 VAS)	5.40 ± 1.30	–	–
Pain onset (months from initial evaluation)	70.35 ± 77.84	–	–
AKPS (0–100 points)	70.50 ± 11.61	96.7 ± 4.80	<0.001

Data presented as mean ± SD. SD, standard deviation; BMI, body mass index; VAS, visual analogue scale; AKPS, Anterior Knee Pain Scale; PFP, patellofemoral pain.

Table 2 Results on the functional performance tests.

Tests	PFP (n=20)	Control (n=20)	p value
SRT (total score)	10 [9–10]	11 [9–11]	0.08
STS30 (repetitions)	15 [13–17]	19.2 [15–23]	0.01
SCT (s)	2.4 [2.0–2.6]	2.0 [1.7–2.2]	0.003
SDT (s)	2.1 [1.8–2.2]	1.9 [1.7–2.0]	0.09
6MST (repetitions)	182 [163–210]	227 [197–261]	0.001

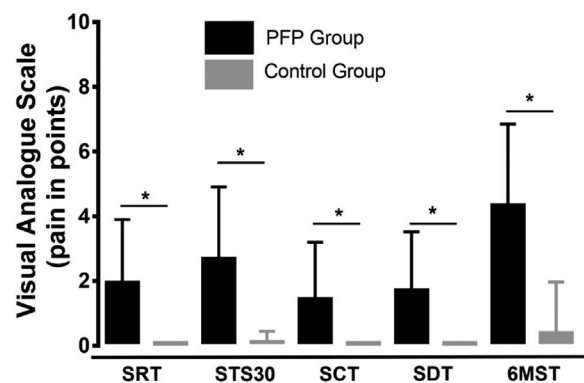
Data presented as median [IQR]. IQR, interquartile range; PFP, patellofemoral pain; SRT, sitting–rising test; STS30, sit-to-stand in 30 s; SCT, stair-climb test; SDT, stair descent test; 6MST, six-minute step test.

**Figure 1** Correlation of performance on STS30 with AKPS. STS30: sit-to-stand in 30 s; AKPS, anterior knee pain scale.

functional tests, were not significant ($r = -0.09$ to 0.15 ; $p > 0.05$ for all). Pain perception was significantly higher after each of the tests in the PFP group compared to the control group (Fig. 2).

Discussion

The hypothesis of this study was partially confirmed, based on the worse functional performance being found in the PFP group for the STS30, SCT, and 6MST, compared to the control group. The results of the present study differ from the study of Loudon et al.,¹⁰ which did not find differences in functional tests (i.e. leg press, balance and reach, antero-medial lunge, and bilateral squat) between individuals with PFP and a control group.¹⁰ A likely explanation for the differences may be the different characteristics of the tests used. In the present study we included tests that induce overload to the knees and that mimic biomechanical move-

**Figure 2** Pain perception at the end of the functional tests. PFP, patellofemoral pain; SRT, sitting–rising test; STS30: sit-to-stand in 30 s; SCT: stair-climb test; SDT: stair descent test; and 6MST: six-minute step test. * Statistically significant difference; $p < 0.05$; analysed by the Mann–Whitney test.

ments that trigger pain (e.g. sit and rise; climb and descend a stairway).¹⁰

The five tests used in this study reached moderate to excellent reliability in both groups. Consistent with the present study, Luque-Siles et al.²¹ demonstrated moderate to good intra-rater reliability in the STS30 (ICC = 0.7–0.8) in a healthy sample.²¹ Furthermore, the SCT and 6MST presented similar results to those found in the literature, with good intra- and inter-rater reliability in healthy individuals.^{16,22} Although measurement error (Table 3) in the functional tests were lower than the difference in both groups, in clinical practice values' dispersion must be observed to verify real clinical changes.

Table 3 Reliability of functional performance tests.

Tests		Intra-rater			Inter-raters		
		ICC	95% CI	SEM (units of measurement)	ICC	95% CI	SEM (units of measurement)
SRT (total score)	PFP	0.89	0.84, 0.92	0.35	0.77	0.56, 0.87	0.35
	Control	0.86	0.73, 0.92	0.29	0.76	0.54, 0.87	0.34
STS30 (repetitions)	PFP	0.91	0.84, 0.95	1.64	0.96	0.92, 0.98	0.45
	Control	0.93	0.67, 0.97	1.53	0.94	0.86, 0.97	0.96
SCT (s)	PFP	0.76	0.63, 0.85	0.17	0.76	0.55, 0.87	0.16
	Control	0.90	0.81, 0.95	0.12	0.90	0.81, 0.94	0.08
SDT (s)	PFP	0.61	0.40, 0.75	0.23	0.72	0.46, 0.86	0.14
	Control	0.85	0.70, 0.92	0.13	0.78	0.60, 0.88	0.11
6MST (repetitions)	PFP	0.91	0.80, 0.94	8.70	0.95	0.90, 0.97	6.93
	Control	0.89	0.80, 0.94	13.88	0.84	0.70, 0.91	16.58

Data presented as ICC, 95% CI and SEM in units of measurement of each test. ICC, intraclass correlation coefficient; 95% CI, 95% confidence interval. SEM, Standard Error of Measurement; SRT, sitting–rising test; STS30, sit-to-stand in 30 s; SCT, stair-climb test; SDT, stair descent test; 6MST, six-minute step test.

The validity of functional tests was assessed using AKPS as an anchor, as there are no other objective parameters for this measure, and it is a widely used and reliable scale of function in studies on PFP.¹² The results showed a moderate correlation between AKPS and performance in STS30 in the PFP group, although only weak correlation between self-report questionnaires and STS were found in healthy participants in the literature.²³ Considering the significant difference in the performance of the STS30 between groups and its moderate correlation with AKPS, it can be hypothesized that this is the best tool (objectively measured) to identify functional limitation in individuals with PFP.

Additionally, although the pain was stratified as low intensity in VAS after tests (Fig. 2), participants present functional limitation in tests, reinforcing the importance of evaluating functional performance. The results of the present study should be interpreted considering some potential limitations. Participants were asked to perform a large number of tests and repetitions, which may have interfered in the obtained results, although, to avoid this potential bias the randomization of test sequence was performed. Finally, this study did not aim to evaluate the responsiveness of the tests to a given intervention for women with PFP.

Conclusion

Women with PFP demonstrated worse performance in the activities of sitting and rising and going up and down stairs evaluated by different functional tests as compared to the control group. The tests presented good to excellent intra- and inter-rater reliability, with the exception of the SDT for the PFP group which demonstrated moderate intra-rater reliability. Although validity of the tests could not be established, the STS30 correlated moderately with function as assessed by the AKPS.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - financing code 001. We thank all participants who agreed to participate in this study.

References

1. Crossley KM, Stefanik JJ, Selfe J, et al. 2016 Patellofemoral pain consensus statement from the 4th International Patellofemoral Pain Research Retreat, Manchester. Part 1: Terminology, definitions, clinical examination, natural history, patellofemoral osteoarthritis and patient-reported outcome m. *Br J Sports Med.* 2016;50(14):839–843.
2. Silva DDO, Briani RV, Ferrari D, Pazzinatto MF. No son buenos indicadores de dolor y de limitaciones funcionales el ángulo Q y la pronación subastragalina en los sujetos con síndrome de dolor patelofemoral. *Fisiot Pesq.* 2015;22(2):169–175.
3. Briani RV, Pazzinatto MF, de Oliveira Silva D, de Azevedo FM. Different responses of pain in relation to distinct levels of physical activity in women with patellofemoral pain. *Brazilian J Phys Ther.* 2016;21(2):138–143.
4. Piazza L, Lisboa ACA, Costa Vda, et al. Sintomas e limitações funcionais de pacientes com síndrome da dor patelofemoral. *Rev Dor.* 2012;13(1):50–54.
5. Esculier JF, Roy JS, Bouyer LJ. Psychometric evidence of self-reported questionnaires for patellofemoral pain syndrome: a systematic review. *Disabil Rehabil.* 2013;35(26):2181–2190.
6. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthr Cartil.* 2013;21(8):1042–1052.
7. Terwee CB, Mokkink LB, Steultjens MP, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Rheumatology.* 2006;45(7):890–902.
8. Pessoa BV, Arcuri JF, Labadessa IG, Costa JNF, Sentanin AC, Di Lorenzo VAP. Validity of the six-minute step test of free cadence in patients with chronic obstructive pulmonary disease. *Brazilian J Phys Ther.* 2014;18(3):228–236.

9. Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. *J Appl Physiol*. 2008;105(3):977–987.
10. Loudon JK, Wiesner D, Goist-Foley HL, Asjes C, Loudon KL. Intrarater reliability of functional performance tests for subjects with patellofemoral pain syndrome. *J Athl Train*. 2002;37(3):256–261.
11. Baldon RM, Serrão FV, Scattone Silva R, Piva SR. Effects of functional stabilization training on pain, function, and lower extremity biomechanics in women with patellofemoral pain: a randomized clinical trial. *J Orthop Sport Phys Ther*. 2014;44(4), 240-A8.
12. da Cunha RA, Costa LOP, Hespanhol Junior LC, Pires RS, Kujala UM, Lopes AD. Translation, cross-cultural adaptation, and clinimetric testing of instruments used to assess patients with patellofemoral pain syndrome in the Brazilian population. *J Orthop Sports Phys Ther*. 2013;43(5):332–339.
13. Brito LBB, Ricardo DR, de Araújo DSMS, Ramos PS, Myers J, de Araújo CGS. Ability to sit and rise from the floor as a predictor of all-cause mortality. *Eur J Prev Cardiol*. 2012:1–7.
14. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community-residing older adults. *J Aging Phys Act*. 1999;7:129–161.
15. HYF T, ARA KCT, YHM C. Reliability of the Stair-Climb Test (SCT) of cardiorespiratory fitness. *Adv Exerc Sport Physiol*. 2004;10(3):77–83.
16. Arcury JF, Borghi-Silva A, Labadessa IG, Sentanin AC, Candolo C, Pires di Lorenzo VA. Validity and reliability of the 6-minute step test in healthy individuals: a Cross-sectional study. *Clin J Sport Med*. 2016;26(1):69–75.
17. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163.
18. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.
19. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. vol. 47. Pearson/Prentice Hall; 2002.
20. Matheson GJ. We need to talk about reliability: making better use of test–retest studies for study design and interpretation. *PeerJ*. 2019;2019(5):1–25.
21. Luque-Siles C, Gallego-izquierdo T, Jiménez-rejano JJ, Granados-de-la-orden S, Plaza-manzano G. Reliability and minimal detectable change of three functional tests: forward-lunge, step-up-over and sit-to-stand. *J Phys Ther Sci*. 2016;28:3384–3389.
22. Plat MCJ, Frings-Dresen MHW, Sluiter JK. Reproducibility and validity of the stair-climb test for fire fighters. *Int Arch Occup Environ Health*. 2010;83(7):725–731.
23. Baldwin JN, McKay MJ, Hiller CE, Moloney N, Nightingale EJ, Burns J. Relationship between physical performance and self-reported function in healthy individuals across the lifespan. *Musculoskelet Sci Pract*. 2017;30:10–17.