Reliability and Feasibility of Physical Fitness Tests in Female Fibromyalgia Patients

Abstract
▼
The aim of the present study was to determine the reliability and feasibility of physical fitness tests in female fibromyalgia patients. 100 female fibromyalgia patients (aged 50.6 ± 8.6 years) performed the following tests twice (7 days interval test–retest): chair sit and reach, back scratch, handgrip strength, arm curl, chair stand, 8 feet up and go, and 6-min walk. Significant differences between test and retest were found in the arm curl (mean difference: 1.25 ± 2.16 repetitions, Cohen d = 0.251), chair stand (0.99 ± 1.7 repetitions, Cohen d = 0.254) and 8 feet up and go (−0.38 ± 1.09 s, Cohen d = 0.111) tests. Intraclass correlation coefficients (ICC) range from 0.92 in the arm curl test to 0.96 in the back scratch test. The feasibility of the tests (patients able to complete the test) ranged from 89% in the arm curl test to 100% in the handgrip strength test. Therefore, the reliability and feasibility of the physical fitness tests examined is acceptable for female fibromyalgia patients.

Introduction
▼
People who report chronic pain generally suffer from limitations in their daily life compared to individuals not suffering from pain. These limitations impact physical mobility in general as well as activities such as lifting groceries or climbing stairs [6]. Fibromyalgia is recognized as a poly-symptomatic syndrome with pain as the predominant symptom [16]. Fibromyalgia patients show a reduced functional capacity [10, 17] compared to age-matched healthy peers [4], and similar to healthy older adults [10, 26]. Some physical fitness tests are able to discriminate between fibromyalgia and non-fibromyalgia females [4]. Associations have already been established between the performance on physical fitness tests and the patients’ subjective ratings of their activity limitations [22]. Moreover, an inverse association has been observed between pain and some physical fitness tests performance [10]. Those patients with higher performance in the 6-min walk test presented lower tenderness, perceived pain and fatigue, higher vitality and perceived functional capacity and better mental health than those patients with lower performance [11]. A better physical fitness performance has also been associated with higher cognitive function in fibromyalgia patients [14].
On the other hand, temporary work disability has been significantly correlated with poor physical fitness [29]. Consequently, the assessment of physical fitness in this population is important for understanding the relationship between interventions [24]. The Senior Fitness test battery and the handgrip strength test have been previously used in fibromyalgia patients [3–5, 10, 14, 32], but its reliability in these patients is unknown. Reliability is a characteristic that needs to be assured for any measurement tool [13]. A test is considered to be reliable when an individual obtains similar results when performing the test on 2 or more occasions under the same conditions and in close succession [20]. We studied the reliability of those physical fitness tests in female fibromyalgia patients.

Methods
▼
Participants
We contacted a local association of fibromyalgia patients from Granada (southern Spain). A total of 116 potentially eligible patients responded and gave their written informed consent after receiving detailed information about the aims
and study procedures in an informative session. Inclusion criteria were: (i) being female; (ii) being younger than 65; (iii) being diagnosed with fibromyalgia and meeting the American College of Rheumatology criteria: having widespread pain for more than 3 months and pain with 4 kg/cm² of pressure reported for 11 or more of 18 tender points [36]; (iv) absence of acute or terminal illness; (v) having had no severe trauma within the last 6 months, orthopedic or musculoskeletal limitations that precluded ambulation. The Spanish version of the revised Physical Activity Readiness Questionnaire was also administered to identify people at risk for adverse events while testing [12]. 6 patients were men, 3 were older than 65 years, 2 had orthopedic or musculoskeletal limitations and 5 patients did not carry out one of the 2 assessments. Finally, a total of 100 female patients (50.6 ± 8.6 years; range: 21–65 years) were included in the analysis. The research protocol was reviewed and approved by the Ethics Committee of the Hospital Virgen de las Nieves (Granada, Spain) and followed the ethical standards of the International Journal of Sports Medicine [19].

Body composition
Weight and height were measured following standard procedures with a scale (Inbody R20, Biospace, Seoul, Korea) and a stadiometer (Seca 780, Hamburg, Germany), respectively, and body mass index [weight/(kg)/height(\text{m}^2)] was calculated.

Physical fitness assessment
Patients’ physical fitness was assessed following the standardised Functional Senior Fitness Test Battery [28]. Additionally, the handgrip strength test was measured due to its potential to discriminate between presence/absence of fibromyalgia [5]. Although the Functional Senior Fitness Test battery was originally designed for elderly people, we felt that it could be appropriate for this population for 3 main reasons: i) the battery has shown no ceiling and floor effects, which is a relevant aspect for this study sample due to the high heterogeneity of fibromyalgia patients [35]; ii) fibromyalgia patients have similar physical fitness as elderly healthy people [10, 26] and iii) the use of this battery would allow comparisons with population-based samples of healthy adults or older adults. The physical fitness components, muscular strength, flexibility, motor agility/dynamic balance and aerobic capacity, were assessed using the tests detailed below:

Lower body muscular strength: The “chair stand test” involves counting the number of times within 30 s that an individual can rise to a full stand from a seated position with back straight and feet flat on the floor, without pushing off with the arms. The arms were crossed at the chest level. The patients performed one trial after familiarization [28].

Upper body muscular strength: The “arm curl test” involves determining a number of times a hand weight (2.3 kg) can be curled through a full range of motion in 30 s [28]. Patients performed the test with both hands and the average was used in the analyses. Additionally, “handgrip strength test” was measured using a digital dynamometer (TKK 5101 Grip-D; Takey, Tokyo, Japan). Optimal grip span was calculated using the formula suggested by Ruiz et al. [31]: \( y = x \times (5 + 1.5) \) in women, where “\( x \)” is hand size and “\( y \)” grip length. Each patient performed the test with the arm fully extended, forming a 30° angle in relation to the trunk. Patients performed the test twice (alternating hands), with a 1-min rest period between measures. The best value of 2 trials for each hand was chosen and the average of both hands was used in the analyses.

Lower body flexibility: In the “chair sit and reach test”, the patient, while seated with one leg extended, slowly bends forward sliding the hands down the extended leg in an attempt to touch (or pass) the toes. The number of centimeters short of reaching the toe (minus score) or reaching beyond it (plus score) was be recorded [28]. 2 trials with each leg were measured and the best value of each leg was registered. The average of both legs was used in the analyses.

Upper body flexibility: The “back scratch test”, a measure of overall shoulder range of motion, involves measuring the distance between (or overlap of) the middle fingers behind the back with a ruler [28]. Participants performed this test twice, alternating hands, and the best value of each hand was registered. The average of both hands was used in the analyses.

Motor agility/dynamic balance: The “8 feet up and go test” involves standing up from a chair, walking 8 feet to and around a cone, and returning to the chair in the shortest possible time [28]. The best time of 2 trials was recorded and used in the analyses.

Cardiorespiratory fitness: We used the “6-min walk test”. This test involves determining the maximum distance (meters) that can be walked in 6 min along a 45.7 meters rectangular course [21, 28]. All patients were asked to complete a visual analogue scale for pain at each day of assessment. It is a simple assessment tool consisting of a 10 cm line with 0 on one end, representing no pain, and 10 on the other, representing the worst pain ever experienced. Previous findings suggest that baseline pain acts as an occasion setter which determines the level of physical activity the patient is willing to perform, regardless of pain increase [15]. Additionally, the rate of perceived exertion (RPE) was monitoring using the Borg’s conventional (6–20 point) scale [8] after each test. It was therefore possible to analyze whether the baseline pain and perceived exertion were similar in both assessment appointments.

Feasibility of the functional fitness tests
We scored how many participants were able to complete the tests satisfactorily on both test and retest occasions.

Fibromyalgia severity and pain
To better describe the characteristic of the study sample, we assessed fibromyalgia severity and pain with the Fibromyalgia Impact Questionnaire (FIQ) [30]. We also measured the number of tender points [36] using a standard pressure algometer (FPK 20; Wagner Instruments, Greenwich, CT, USA).

Statistical analysis
Descriptive statistics are presented as mean ± standard deviation (SD). A one-way repeated measures analysis of variance (ANOVA) was used to compare mean differences between measurements (test-T1 vs. retest-T2). Additionally, we performed the same analysis after adjusting for change in pain (ANOVA, pain T2-pain T1). Cohen d was computed to quantify the magnitude of the difference between test and retest. The Wilcoxon signed-rank
test was selected to analyze the differences in RPE and baseline pain between the 2 appointments. Relative reliability of the battery was determined by Intraclass Correlation Coefficient (ICC, model 2.1) and was interpreted as follows: >0.75, 0.50–0.75 and <0.50 were considered as good, moderate and poor reliability scores, respectively [27]. The standard error of measurement (SEM) provides an absolute index of reliability and was calculated as: SEM = SD(1 − ICC) [34]. The minimal detectable change (MDC) represents the minimum detectable change necessary to exceed the measurement error of 2 repeated measures at a specified interval confident (CI) and was calculated for the 95 % CI as MDC = SEM × 1.96 × √2.

The agreement between the corresponding fitness variables obtained during the 2 successive measurements was also graphically examined by plotting the difference between each pair of measurements against their mean, according to the Bland-Altman approach [7]. The 95 % limits of agreement for all the physical fitness variables were calculated as the inter-trial mean difference ± 1.96 SD. The association between the difference and the magnitude of each fitness test (i.e., heteroscedasticity) was examined by regression analysis.

The feasibility of each test was calculated as the percentage of participants who were able to complete the test. All analyses were performed using the Statistical Package for Social Sciences (IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp). For all analyses, the significant level was set at P < 0.05.

### Results

Characteristics of the study participants are shown in Table 1. The mean number of tender points is 17.3 ± 1.6, and the mean FIQ total score is 69.2 ± 15.4. These data shows that the sample included in the present study is severely affected by the disease. Significant mean differences between test and retest were found in the arm curl, chair stand, and 8 feet up and go tests (Table 2). Pain levels differed between test and retest (5.8 ± 2.0 vs. 5.1 ± 2.4; P < 0.001). The results did not changes after adjusting for change in pain (pain T2-pain T1) (data not shown). There was good test-retest reliability in all tests (ICCs from 0.91 to 0.96) (Table 2). There were no differences in RPEs between test and retest in any tests, except for the arm curl test (14.7 ± 2.5 vs. 12.0 ± 2.6; respectively, P = 0.013). The mean RPE score for the other physical fitness tests studied in test and retest were: 12.3 ± 3.0 vs. 12.0 ± 2.4 for the chair sit and reach test; 12.7 ± 3.1 vs. 12.1 ± 2.6 for the back scratch test; 11.7 ± 2.8 vs. 11.5 ± 2.5 for the handgrip strength test; 13.3 ± 2.7 vs. 12.9 ± 2.3 for the chair stand test; 11.6 ± 3.0 vs. 11.0 ± 2.4 for the 8 feet up and go test and 13.7 ± 2.6 vs. 13.2 ± 2.6 for the 6-min walk test, respectively.

The Bland-Altman plots (Fig. 1, 2) graphically show the reliability patterns in terms of systematic errors (bias or mean inter-trial differences) and the limits of agreement for the physical fitness tests studied: chair sit and reach (−12.46, 12.12 cm), back scratch (−9.20, 11.26 cm), handgrip strength (−5.59, 5.93 kg), arm curl (−2.99, 5.47 rep), chair stand (−2.34, 4.32 rep), 8 feet up and go (−2.52, 1.76 s), and 6-min walk (−87.81, 95.83 m). The heteroscedasticity analysis showed a positive association between inter-trial mean differences and inter-trial average in back scratch, 8 feet up and go (R = 0.46 and R = 0.43, respectively, both P < 0.001) and chair stand (R = 0.2 P < 0.01) tests. All patients were able to perform the handgrip strength test, while 99% completed the chair sit and reach test, chair stand test and 8 feet up and go test, and 95% the back scratch test and 6-min walk test. The lowest feasibility was observed in the arm curl test (89%). The main problem in those patients whose did not complete the arm curl test or the back scratch test was that...
they felt more pain during the test. During the 6-min walk test the main problem was discomfort with the shoes.

**Discussion**

The purpose of the present study was to determine the reliability and feasibility of a battery of physical fitness tests in female fibromyalgia patients. The ICC values for all the physical fitness tests in the present study were high and represented a good reliability. The SEMs provide a low index of error. We used the SEM to estimate the MDC, providing clinically useful information in terms of defining the minimal amount of change needed to be considered “real” and that is not likely to be attributable to a chance variation in measurement [34]. All of the physical fitness tests were well tolerated by the patients. These results confirm that this battery of tests is successful in measuring performance across a wide range of ability levels. The ICC values observed in our female fibromyalgia patients are similar to those reported in the original study of the senior fitness test battery with older women [28].

While we observed significant differences between test and retest performance in the arm curl, chair stand and 8 feet up and go tests, the mean inter-trial difference for these measurements were low. Furthermore, the effect size of the mean differences, as measured by the Cohen’s d were small ( < 0.25). This suggests that, from a practical point of view, these tests can be used to evaluate physical fitness in female fibromyalgia patients. The better scores performed in the retest assessment may suggest learning effects (positive systematic bias), but others reasons, such as intensity of symptoms, may also have influenced the results. The intensity of the symptoms is not always constant across the day or between days, and the changes in one symptom can affect other symptoms [18, 25, 37]. The fatigue, the quality of rest the night before, or even the weather, could also affect the patients, and hence physical performance [9, 25, 33, 37]. Previous studies have observed an association between the performance in physical fitness tests and the level of pain [10, 11, 15, 22]. In the present study, however, the results remained unchanged after the adjustment for change in pain.

Results from the heteroscedasticity analysis and Bland-Altman plots indicate that the worse the performance in the back scratch, 8 feet up and go chair sit and reach tests, the worse the degree of the agreement. Mannerkorpi et al. [23] analyzed in 15 females fibromyalgia patients the reliability of a battery composed by the following tests: shoulder range of motion (flexion and abduction), hand-to-neck, hand-to-scapula and isometric endurance of the shoulder abductor muscles for the upper extremity, and chair test and 6-min walk for lower extremity. They did not find differences between the test and retest performance (interval of 2–3 days) and concluded, that all the tests, except the hand to-neck test, possessed acceptable reliability. 2 studies focused on the reliability of the 6-min walk test. The ICC obtained in the study by Pankoff et al. (0.91) was similar to ours (0.92). King et al. [21] repeated the 6-min walk test 5 times over 10 days in 12 patients. They reported a correlation of 0.733 for five 6-min walk tests over 10 days, and 0.885 for the final 4 walks.

3 recent studies have analyzed the reliability of strength tests. Adsuar et al. carried out 2 studies to analyze the reliability of isokinetic peak torque and work and isometric peak torque measurements for knee flexion and extension [1], and the rela-
bility of isokinetic strength measurements during concentric and eccentric actions of the shoulder muscles [2]. In the first study [1], which focused on the lower extremities, all peak torque measures had an ICC > 0.90 with the exception of eccentric flexion (0.83), and all work measures had an ICC > 0.85. In the second study [2], which focused on the upper extremities, they found that peak torque showed high reliability for the abduction (ICC= 0.88) and adduction (ICC= 0.89) phases in the concentric/eccentric test. However, in the concentric/concentric test, peak torque showed low reliability in the abduction phase (ICC=0.29) and good reliability in the adduction phase (ICC=0.92). Finally, Munguia et al. [24] analyzed the reliability of low-load endurance strength tests for upper and lower extremities. The tests showed a good reliability (ICC = 0.973–0.979) and the SEMs were between 1.44 and 1.66 repetitions.

The physical fitness assessment in daily clinical practice is limited by the availability of scientific instruments or time. In this sense, the proposed tests have a great potential in a clinical setting by the availability of scientific instruments or time. In this sense, the proposed tests have a great potential in a clinical setting for several reasons. First, the equipment needed to perform these tests is inexpensive. Second, these 7 tests can all be performed within 30 or 40 min. Third, the procedures for these tests are simple and do not require any special training.

The present study has limitations that need to be mentioned. The sample was of convenience and was only composed by women.

Conclusions

The physical fitness tests evaluated in the present study present good reliability for use in female fibromyalgia patients. They have shown a high feasibility and they are quick and easy to administer.

Acknowledgements

We would like to thank the AGRAFIM (association of fibromyalgia from Granada) members involved in the field work for their effort and great enthusiasm. This study was supported by the Spanish Ministry of Science and Innovation (I+D+I DEP2010-15639, grants: BES-2011-047133, RYC-2010-05957), the Consejería de Turismo, Comercio y Deporte (CTCD-201000019242-TRA) and the Spanish Ministry of Education (AP-2010-0963).

Conflict of interest: The authors have no conflict of interest to declare.

Affiliations
1 Department of Physical Education, Faculty of Education Sciences, University of Cadiz, Spain
2 Department of Physical Education and Sport, Faculty of Sport Sciences, University of Granada, Granada, Spain
3 Department of Physiology and Institute of Nutrition and Food Technology, Faculty of Pharmacy, University of Granada, Spain

References


34 Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005; 19: 231–240

