ANNE THERESE TVETER, PT, MSc[1] • **HANNE DAGFINRUD,** PT, PhD[1,2] • **TUVA MOSENG,** PT, MSc[1] • **INGER HOLM,** PT, PhD[1,3]

Measuring Health-Related Physical Fitness in Physiotherapy Practice: Reliability, Validity, and Feasibility of Clinical Field Tests and a Patient-Reported Measure

The consequences of physical inactivity are an important public health problem,[6] with reduced well-being for the individual and high societal costs.[29] The scientific evidence for numerous health benefits of exercise is indisputable, and engaging in physical activity can delay all-cause mortality in many patients.[18]

STUDY DESIGN: A cross-sectional study with a test-retest design.

OBJECTIVES: To assess measurement properties of the physical fitness questionnaire, the 6-minute walk test, the stair test, the hand-grip test, the 30-second sit-to-stand test, and the fingertip-to-floor test in patients with various musculoskeletal conditions (MSCs).

BACKGROUND: Patients suffering from MSCs tend to be more deconditioned and less physically active than healthy people. Physiotherapists should, therefore, focus on health-related physical fitness in addition to their patients' specific MSCs to offer optimal treatment. To enable good decision making, a core set of feasible measures with acceptable measurement properties is needed.

METHODS: Eighty-one patients with MSCs (57.6 ± 14.2 years of age) were recruited from outpatient physiotherapy clinics. Relative reliability was analyzed with intraclass correlation coefficient model 2,1, and absolute reliability with standard error of measurement and smallest detectable change. Construct validity was assessed with a priori hypotheses. Time spent and assistance needed to accomplish the measures were used to assess feasibility.

RESULTS: The 6-minute walk test, the hand-grip test, and the physical fitness questionnaire showed acceptable reliability (49 m, 4 kg, and 2 points, respectively) and construct validity. The stair test showed acceptable reliability (8 seconds) but not validity. The 30-second sit-to-stand test showed acceptable validity but not reliability (4 sit-to-stands), whereas the fingertip-to-floor test showed neither acceptable reliability (9 cm) nor validity.

CONCLUSION: The 6-minute walk test, the hand-grip test, and the physical fitness questionnaire can be recommended as a core set of reliable and valid measures to assess health-related physical fitness in patients with various MSCs. *J Orthop Sports Phys Ther 2014;44(3):206-216. Epub 22 January 2014. doi:10.2519/jospt.2014.5042*

KEY WORDS: *measurement properties, musculoskeletal conditions, performance-based tests*

People suffering from musculoskeletal conditions (MSCs) tend to be less physically active and more deconditioned than healthy controls.[16,23,43] Thus, physiotherapists treating patients with MSCs in outpatient clinics should include an individually tailored exercise program in the treatment plan for these patients. For optimal health benefits, it is important that exercise programs are in accordance with current recommendations. For older adults and adults with chronic conditions, the American College of Sports Medicine recommends a focus on physical activity that may improve health-related physical fitness.[38] Physical fitness is defined as the characteristics related to the ability to perform physical activity,[11] and the health-related components are cardiorespiratory endurance, muscle strength, flexibility, and body composition.[2,11]

Measuring and monitoring patients' health-related physical fitness are important for patients, clinicians, and health authorities. For this purpose, physiotherapists need measurement tools with acceptable measurement properties. To be clinically feasible, these tools must be readily available and easy to perform, require no or only portable equipment,

[1]Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway. [2]National Resource Center for Rehabilitation in Rheumatology, Diakonhjemmet Hospital, Oslo, Norway. [3]Department of Orthopaedic Surgery, Division of Surgery and Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. The study was approved by the Regional Ethical Committee in Norway. Grant support was provided by the Norwegian Fund for Postgraduate Training in Physiotherapy. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Anne Therese Tveter, PO Box 1089, Blindern, NO-0318 Oslo, Norway. E-mail: a.t.tveter@medisin.uio.no • Copyright ©2014 *Journal of Orthopaedic & Sports Physical Therapy*®

and be time efficient.[4] Clinical field tests to evaluate physical performance can be used for this purpose.[2] Though clinical field tests are less accurate and less specific than laboratory-based tests, they may be applicable for measuring the different aspects of health-related physical fitness. In addition, to enable good clinical decision making, it has been argued that information about both what patients can do and what they perceive they can do is needed,[4,48] emphasizing the use of patient-reported instruments as well.

Several studies have shown that clinical field tests and patient-reported measures can be used to reflect different aspects of health-related physical fitness. These tests include the 6-minute walk test (6MWT)[10,28] and the stair test (ST)[12,34] (reflecting cardiorespiratory endurance), the 30-second sit-to-stand test (30sSTS)[5,32] and the hand-grip test[9,42] (reflecting muscle strength), the fingertip-to-floor test (FTF)[41] (reflecting flexibility), and the physical fitness questionnaire[49] (capturing the patients' perspectives regarding their physical fitness level).

Validity and reliability of these instruments have been investigated, but estimates of reliability have only been reported in some specific subgroups of patients with MSCs[20,22,30,33,47,52] and validity has been evaluated mainly in patients with other diagnoses.[5,9,10,12,28,32,34,41,42] Patients seen by physiotherapists in outpatient clinics vary in terms of diagnosis and disease severity, and instruments for measuring physical fitness must, therefore, be applicable across diagnostic groups, age, gender, and physical fitness levels. The aim of this study was to examine reliability, validity, and feasibility of a core set of clinical field tests and a patient-reported measure reflecting health-related physical fitness in patients with a variety of MSCs.

METHODS

Design and Participants

This cross-sectional study with a test-retest design is part of FYSIOPRIM, a large-scale research program conducting methodological and clinical physiotherapy research in primary health care. The program comprises researchers from different research centers working in close collaboration with experienced physiotherapists from 6 different outpatient clinics in Norway.

Eighty-one patients between 20 and 90 years of age with MSCs were recruited from the outpatient physiotherapy clinics and tested on 2 occasions 1 week apart. The patient-reported measure was answered before the clinical field tests and laboratory-based tests were completed, and a standardized test protocol including verbal instructions was used at both test and retest times.

Selection of Measures

Relevant clinical field tests were identified through a literature search and thoroughly discussed in the group of researchers and experienced clinicians. Through an informal consensus process, researchers and clinicians decided on a core set of field tests reflecting the different aspects of health-related physical fitness. Further, a questionnaire asking for the patient's perceived physical fitness (aerobic capacity, muscle strength, and flexibility) was added to the core set. Well-established laboratory-based tests measuring muscle strength and aerobic capacity were selected for validation purposes.

All participants answered a set of sociodemographic questions about age, gender, employment status, occupation, main complaint, and duration of complaints. Body composition was measured by body weight and height and presented as body mass index (kg/m^2). Physical activity level was measured by the International Physical Activity Questionnaire short form (IPAQ short),[13] consisting of 7 questions asking about the time spent in vigorous- and moderate-intensity activities, in walking, and in sedentary activity. The results were transformed into metabolic equivalent task (minutes per week) scores and categorized into low, moderate, and high levels of participation in physical activity, according to the guidelines for the IPAQ short (www.ipaq.ki.se).

Patient-Reported Measures

A revised version of the physical fitness questionnaire[49] that included 4 items (aerobic fitness, muscle strength, flexibility, and balance) to be answered on a 9-point numeric rating scale, anchored with 1 as poor and 9 as good and with an illustration at either end, was used. In scoring the different items, the participants were instructed to compare themselves to others of the same age and gender, with a middle score of 5 constituting "normality." The balance item will not be investigated in this article, as balance is not a part of the health-related physical fitness construct.[11]

Clinical Field Tests

The 6MWT has been described as an inexpensive and simple walking test that can be used as a predictor of aerobic fitness.[10] The 6MWT was conducted according to the American Thoracic Society guidelines.[3] Participants were instructed to walk as fast as possible back and forth between 2 cones (18 m apart) on a flat, hard surface for 6 minutes, and the walking distance was measured in meters. After the test, the patient's heart rate was recorded with a heart-rate monitor (Polar Electro Oy, Kempele, Finland), and perceived exertion was measured with the Borg rating of perceived exertion (15-point scale, ranging from 6 to 20, with 6 as very, very light and 20 as very, very hard).[45]

The ST has been performed in several different ways[12,34] and is referred to as a test of submaximal aerobic capacity.[12] A revised version, in which participants were instructed to ascend and descend 50 standard-sized steps with 3 landings in between as quickly as possible, was used. Participants were allowed to run and could use the banisters if necessary, but were not allowed to skip any step. Results were measured in seconds and heart rate was recorded after the test (Polar Electro Oy).

TABLE 1 — A Priori Hypotheses for the Clinical Field Tests and the Physical Fitness Questionnaire

Hypothesis	Clinical Field Tests: 6MWT	ST	30sSTS	Hand	FTF	Physical Fitness Questionnaire: Aerobic Fitness	Muscle Strength	Flexibility
1. A correlation of 0.4 or greater between clinical field tests and laboratory-based tests, reflecting corresponding constructs	+	+	+	+	...	...	...	...
2. A correlation of 0.4 or less between clinical field tests and laboratory-based tests, reflecting noncorresponding constructs	–	–	–	+	+	...	...	...
3. A correlation of 0.4 or greater between self-assessed physical fitness dimensions and clinical field tests, reflecting corresponding constructs	...	...	...	...	...	+	+	+
4. A correlation of 0.4 or less between physical fitness dimensions and clinical field tests, reflecting noncorresponding constructs	...	...	...	...	...	–	+	–
5. A correlation of greater than 0.4 between heart rate and m (6MWT)/s (ST)	+	+	...	...	...	...	...	...
6. Significant difference between patients reporting moderate/low physical activity level and patients reporting high physical activity level	+	–	+	–	–	+	+	+
Hypothesis confirmed	3/4	2/4	2/3	2/3	1/2	2/3	3/3	2/3

Abbreviations: 30sSTS, 30-second sit-to-stand test; 6MWT, 6-minute walk test; FTF, fingertip-to-floor test; Hand, hand-grip test; ST, stair test.

The hand-grip test is a simple method of assessing muscle strength.[42] In the present study, grip strength was measured using the Baseline dynamometer (Fabrication Enterprises, White Plains, NY). The participants were seated with their arm alongside the trunk and elbow in 90° of flexion. The dynamometer has 5 handle positions, and the second position was used for all participants.[21] The mean value (in kilograms) of the 2 trials was used to analyze the right and left hands, with the dominant hand measured first.

The 30sSTS is described as a test measuring lower extremity strength.[5,32] The participants were instructed to complete as many full stands as possible during 30 seconds, starting from a seated position with arms folded across the chest.[32] A chair of standard height was used and the number of full stands was recorded.

The FTF test has been described as a measure of flexibility of the spine, pelvic girdle, and hamstrings.[41] The test was performed with participants standing on a stool. They were instructed to reach as far as possible with their fingertips toward the floor, with knees fully extended.[19] The results were measured in centimeters (negative values reflecting an inability to reach the stool and positive values reflecting the ability to reach beyond the level of the stool).

Laboratory-Based Tests

Isokinetic muscle strength of the quadriceps and hamstrings muscles was tested bilaterally using the CYBEX 6000 isokinetic dynamometer (Computer Sports Medicine Inc, Stoughton, MA). The test protocol consisted of 5 repetitions at a velocity of 60°/s (muscle strength) and 30 repetitions at a velocity of 240°/s (muscle endurance), with a 1-minute rest between tests.[24] A moderate to good test-retest correlation (r = 0.60-0.83) has been shown for both sides and speeds in sports students.[25] Only results from the right quadriceps muscle were reported.

The Åstrand submaximal cycle ergometer test was developed as a test to predict cardiorespiratory fitness by indirectly calculating maximal oxygen uptake ($VO2_{max}$). The 6-minute submaximal cycle ergometer test was performed at a standardized working rate, with a steady-state heart rate between 120 and 170 beats per minute required at the end of the work stage.[2] $VO2_{max}$ was estimated using Åstrand's nomogram, and results for patients older than 65 years were extended from the same formula. The test has been referred to as a reliable, valid, and feasible instrument for calculating $VO2_{max}$ in patients with nonspecific low back pain[23] and stroke.[35]

Statistical Analyses and Hypotheses

The data were analyzed using SPSS Statistics 20 (IBM Corporation, Armonk, NY). The results were presented as mean ± SD if normally distributed, or median and interquartile range (25th-75th quartile) if skewed. Predicted maximal heart rate was estimated using the formula 211 – 0.64 × age.[39]

Reliability A paired t test was used to assess the mean difference between test

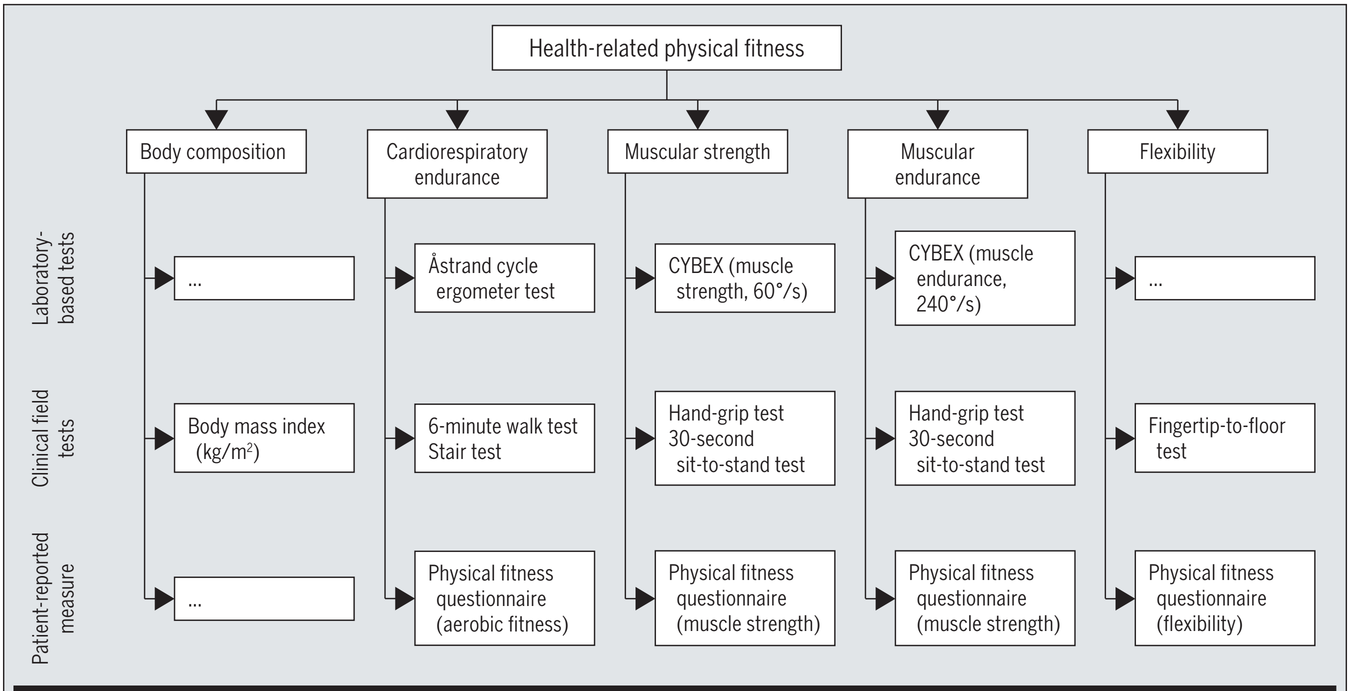

FIGURE 1. The health-related physical fitness construct.[2] Predefined hypotheses about the correspondence between the health-related physical fitness construct and the laboratory-based tests, clinical field tests, and the patient-reported measure.

and retest. An intraclass correlation coefficient ($ICC_{2,1}$) was used to assess relative reliability, and standard error of measurement (SEM) and smallest detectable change (SDC) were used to analyze absolute reliability (measurement error). $ICC_{2,1}$ and $SEM_{agreement}$ were used to account for the systematic difference between the first and second tests,[14] and a 90% confidence interval was used when reporting $SDC_{90\%}$.[20,33,46] SEM was estimated from a 2-way random analysis of variance and, based on this, the $SDC_{90\%}$ was estimated ($SDC_{90\%} = 1.64 \times \sqrt{2} \times SEM_{agreement}$).[14] $ICC_{2,1}$ of 0.70 or greater was considered acceptable.[50] The relationships between the mean scores of the 2 tests and the mean difference between test and retest are illustrated in scatter plots with 90% limits of agreement.[1] The mean of the test and retest (x-axis) was divided into quartiles, and the variance across the scale (y-axis) was examined using a 1-way analysis of variance[1] with a Brown-Forsythe test (heteroscedasticity assumed if $P<.05$).

Validity Construct validity was investigated with a priori hypotheses (**TABLE 1**). Hypotheses of convergent and divergent validity were formulated to address the associations between aspects of the health-related physical fitness and laboratory-based tests, clinical field tests, and the patient-reported measure (**FIGURE 1**). It was hypothesized that there would be a moderate correlation (0.4 or greater) between measures reflecting corresponding constructs (convergent validity) and a low correlation (0.4 or less) between measures reflecting noncorresponding constructs (divergent validity) (hypotheses 1-4). The responses on the physical fitness questionnaire were adjusted for age and gender, and for analytical purposes, and similar adjustments were made to the data from the clinical field tests. A sample of 370 healthy adults from 18 to 90 years of age was used to estimate the performance, adjusted for age and gender (data in progress). The difference between this estimated performance and the performance of study participants on the clinical field tests was used in the correlation analyses. The heart response is important in aerobic capacity testing, thus moderate correlations (0.4 or greater) were hypothesized between the heart rate and the distance walked in 6 minutes and the time used to complete the ST, respectively (hypothesis 5). Pearson or Spearman correlation coefficients were used, depending on the distribution of the scores.

The physical fitness level of an individual is primarily determined by physical activity patterns.[7] Therefore, it was hypothesized that patients reporting high levels of physical activity, according to the IPAQ categorization (www.ipaq.ki.se), would score themselves as significantly more fit on the patient-reported measure and perform significantly better on the field tests than those who reported moderate or low levels of physical activity (hypothesis 6). Values of $P<.05$ were considered to be statistically significant. The measures were considered to have acceptable construct validity if 2 of 3 (66%) of the hypotheses were confirmed.

Feasibility Feasibility is defined as the burden in time and resources required

TABLE 2 DEMOGRAPHIC DATA FOR PATIENTS WITH VARIOUS MUSCULOSKELETAL CONDITIONS (N = 81)*

Characteristic	Value
Gender (female), n (%)	58 (72)
Age, y	57.6 ± 14.2
Height, cm	169.1 ± 9.5
Weight, kg	75.3 ± 17.3
Body mass index, kg/m^2	26.2 ± 4.5
Main problem, n (%)	
Upper extremities/neck	9 (11)
Back/pelvic	11 (14)
Lower extremities	36 (44)
Multiple sites (upper/lower extremities/back)	15 (19)
Systemic (RA, ankylosing spondylitis, fibromyalgia)	10 (12)
IPAQ short (n = 75), n (%)	
Low physical activity level	13 (17)
Moderate physical activity level	24 (32)
High physical activity level	38 (51)

Abbreviations: IPAQ, International Physical Activity Questionnaire; RA, rheumatoid arthritis.
**Values are mean ± SD unless otherwise indicated.*

TABLE 3 RESULTS OF THE CLINICAL FIELD TESTS, THE LABORATORY-BASED TESTS, AND THE PHYSICAL FITNESS QUESTIONNAIRE AT INITIAL TEST*

Test	Value
Clinical field tests	
6-minute walk test, m†	569 ± 94
Stair test, s†	53.1 ± 16.5
30-second sit-to-stand test, n†	15 ± 4
Hand-grip test right, kg†	29.2 ± 10.2
Hand-grip test left, kg†	28.9 ± 10.0
Fingertip-to-floor test, cm‡	–4.6 ± 14.3
Laboratory-based tests	
Åstrand, mL/min/kg§	30.4 ± 10.1
Maximal strength quadriceps right leg (work per repetition), 60°/s‖¶	87 (64-113)
Endurance strength quadriceps right leg (total work), 240°/s‖¶	1222 (934-1633)
Physical fitness questionnaire†	
Aerobic fitness	5.0 ± 2.0
Muscle strength	5.0 ± 1.9
Flexibility	4.7 ± 2.1

**Values are mean ± SD unless otherwise indicated.*
†n = 81.
‡n = 63.
§n = 72.
‖n = 80.
¶Values are median and interquartile range (25th-75th quartile).

by the physiotherapist to collect the data.[17] Feasibility was measured in time (minutes) used to complete the different measures (standardized instructions included), as well as the percentage of participants needing assistance to answer the physical fitness questionnaire. With no existing criteria, we considered the measures to be feasible if the total time did not exceed a standard physiotherapy session (30 minutes) and fewer than 1 of 3 patients (33%) needed assistance in completing the patient-reported measure.

All participants were included in the analyses of validity and feasibility (n = 81). The different analyses have some missing values, due either to missing registration (FTF; n = 18) or equipment failure (Åstrand cycle ergometer test; n = 9). The IPAQ short responses from 6 patients were removed due to guideline exclusion criteria (www.ipaq.ki.se).

The participants rated their change in physical fitness from test to retest on a 7-point global rating scale, ranging from much worse to much better. While test-retest reliability preferably should be assessed in a stable population,[15] only participants rating themselves as unchanged were included in the analyses of reliability (n = 55).

Ethical Considerations

The participants gave their written consent before participation. Ethical approval was given by the Regional Ethical Committee in Norway. The requirements of the tests in this study were considered activities of daily living that were not of any danger to the participants.

RESULTS

Demographic data are presented in **TABLE 2**. The majority of the participants were women, with a mean age of 57.6 ± 14.2 years. The median duration of the MSCs was 10 years, ranging from 16 weeks to 55 years. Most participants localized their complaints to the lower extremities (44%). According to the IPAQ criteria, 49% of the patients reported

TABLE 4	Reliability and Measurement Error*					
	Test	Retest	Difference	$ICC_{2,1}$†	$SEM_{agreement}$	$SDC_{90\%}$
Clinical field tests						
6-minute walk test, m	556.1 ± 95.0	574.8 ± 93.0	18.7 ± 23.7	0.95 (0.80, 0.98)	21.2	49.2
Stair test, s	54.8 ± 16.5	51.7 ± 15.4	–3.1 ± 3.9	0.95 (0.81, 0.98)	3.5	8.1
Hand-grip test right, kg	29.3 ± 10.9	29.4 ± 10.9	0.1 ± 2.6	0.97 (0.95, 0.98)	1.8	4.2
Hand-grip test left, kg	28.8 ± 10.7	29.3 ± 11.6	0.5 ± 2.3	0.98 (0.96, 0.99)	1.6	3.7
30-second sit-to-stand test, n	14.7 ± 4.2	16.4 ± 4.8	1.7 ± 1.7	0.87 (0.38, 0.96)	1.7	3.9
Fingertip-to-floor test, ±cm‡	–4.4 ± 13.8	–4.5 ± 15.9	–0.2 ± 5.3	0.94 (0.89, 0.97)	3.7	8.6
Physical fitness questionnaire						
Aerobic capacity	4.9 ± 2.0	5.0 ± 1.8	0.1 ± 1.0	0.86 (0.78, 0.92)	0.7	1.6
Strength	5.0 ± 1.9	5.1 ± 1.7	0.1 ± 1.0	0.85 (0.75, 0.91)	0.7	1.6
Flexibility	4.7 ± 2.0	4.9 ± 2.0	0.3 ± 1.3	0.78 (0.65, 0.87)	0.9	2.1

Abbreviations: ICC, intraclass correlation coefficient; SDC, smallest detectable change; SEM, standard error of measurement.
**Values are mean ± SD unless otherwise indicated. n = 55.*
†Values in parentheses are 95% confidence interval.
‡n = 43.

having a low or moderate physical activity level, whereas 51% reported a high physical activity level (**TABLE 2**). No differences in age, gender, height, and weight were found between these 2 groups (moderate/low versus high activity level). The results of the laboratory-based tests, clinical field tests, and the patient-reported measure for all participants (initial test) are presented in **TABLE 3**.

Reliability

Relative and absolute reliability values calculated for the unchanged patient group are shown in **TABLE 4**. All clinical field tests showed acceptable relative reliability ($ICC_{2,1}$>0.7). All tests, except for the FTF and the hand-grip test, showed a significant improvement (P<.001) from test to retest.

The scatter plot of the 30sSTS test (**FIGURE 2C**) shows an increase in measurement error with increasing number of sit-to-stands (Brown-Forsythe test, P = .02). In contrast, the variation in the mean difference between test and retest of the other measures was assumed to be equal across the scale (Brown-Forsythe test, P = .08-.87). The absolute reliability for the FTF test was almost 9 cm (**TABLE 4**), but 1 participant accounted for almost half of the measurement error in this instrument. With the data from that participant removed from the analysis, the $SDC_{90\%}$ was reduced to 5.3 cm.

The physical fitness questionnaire showed acceptable relative reliability ($ICC_{2,1}$>0.7). A 2-point change for all items was required to be 90% certain that a true change occurred (**TABLE 4**). No significant improvement from test to retest was found for any of the items.

Validity

Mean ± SD heart rate was 129 ± 24 and 139 ± 20 beats per minute after performing the 6MWT and the ST, respectively, which constitutes a percentage of predicted maximal heart frequency of 74% ± 13% and 79% ± 10%. The Borg rating of perceived exertion for the 6MWT was 12 ± 2.

The 6MWT, the ST, the 30sSTS, and the hand-grip test all showed moderate correlation (greater than 0.40) with corresponding laboratory-based tests; however, the 6MWT and the ST also showed moderate correlations (greater than 0.40) with noncorresponding laboratory-based tests (**TABLE 5**). The hand-grip test was highly associated with muscle strength in the quadriceps, explaining between 50% and 58% of the variance. Both the 6MWT and the ST showed correlations greater than or equal to 0.57 with heart rate after the tests (**TABLE 5**). Additional analyses showed that the 6MWT and the ST explained 17% and 19%, respectively, of the variance in estimated VO_{2max} (Åstrand test). When adjusting for heart rate, the explained variance of VO_{2max} increased to 35% for the 6MWT but was not influenced for the ST. The only significant differences in performance between patients with low to moderate and patients with high physical activity levels were found for the 6MWT and the 30sSTS test (**TABLE 6**).

All items of the physical fitness questionnaire showed moderate correlations (greater than 0.40) (**TABLE 5**) with at least 1 of the corresponding clinical field tests (**FIGURE 1**). All items also showed significant difference in scoring between patients reporting high physical activity levels and patients reporting moderate to low physical activity levels (**TABLE 6**).

The results of the a priori hypotheses are shown in **TABLE 1**, with the 6MWT, the 30sSTS, the hand-grip test, and the physical fitness questionnaire showing acceptable construct validity.

Feasibility

The 6MWT was the most time-consuming test (about 8 minutes, instructions

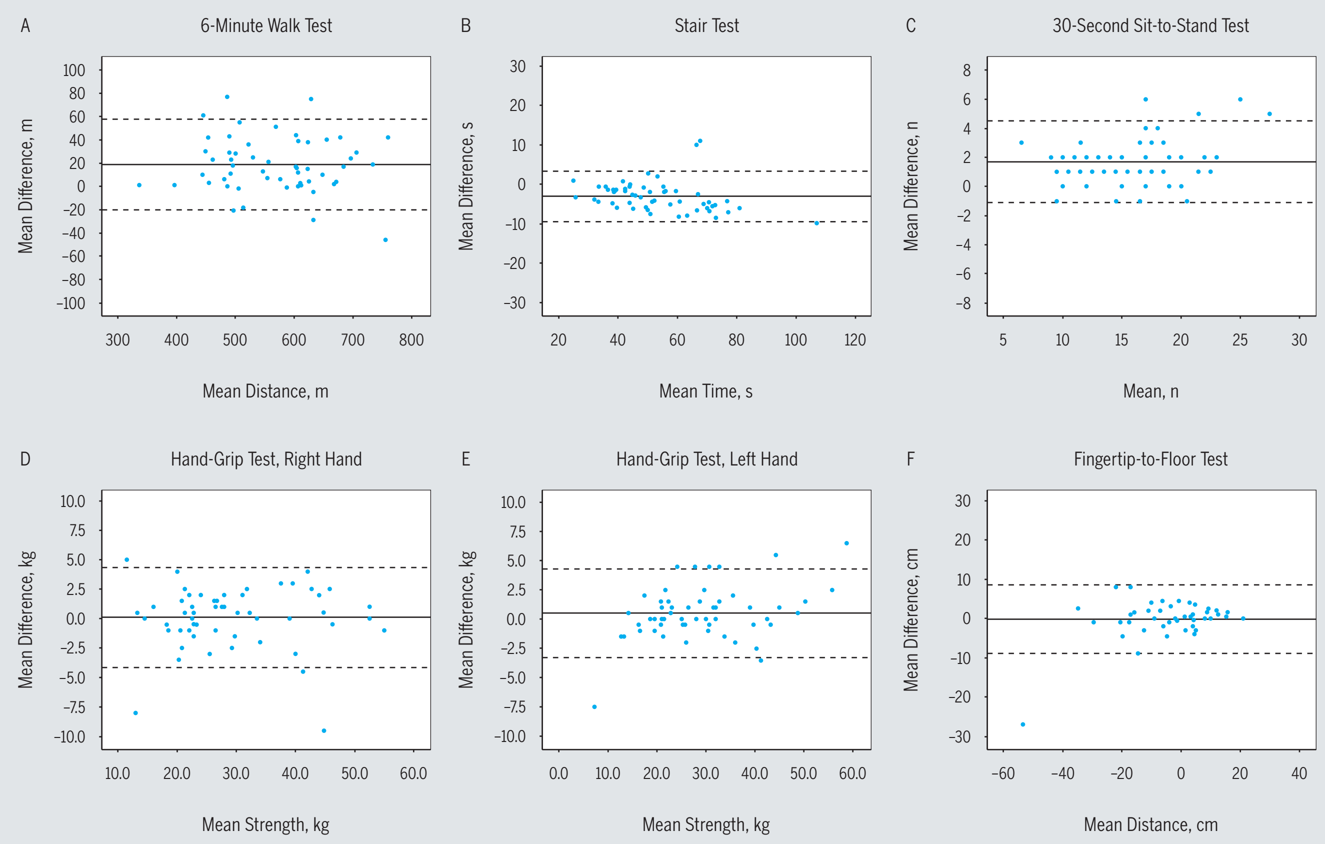

FIGURE 2. Clinical field tests. The scatter plots show the mean scores [(test + retest)/2] on the *x*-axis, correlated with the mean difference between scores (retest/test) on the *y*-axis, presented for the 6-minute walk test (A), the stair test (B), the 30-second sit-to-stand test (C), the hand-grip test for right (D) and left (E) hands, and the fingertip-to-floor test (F).

included), whereas all other clinical field tests took less than 3 minutes to complete. The physical fitness questionnaire took about a minute to complete, with 19% of the participants needing assistance at the initial test and 7% at retest. Altogether, the clinical field tests and the physical fitness questionnaire took about 15 minutes to complete at the initial test, whereas the time was reduced to 13 minutes at retest.

DISCUSSION

THE RESULTS OF THIS STUDY SHOWED that the 6MWT, the hand-grip test, and the physical fitness questionnaire displayed acceptable validity and reliability in patients with various MSCs. The 30sSTS test only displayed acceptable validity, the ST only showed acceptable reliability, and the FTF test showed neither acceptable reliability nor validity. All measures showed acceptable feasibility for use in a clinical practice.

The importance of using standardized measures in physiotherapy practice has been recognized; however, in busy physiotherapy clinics, measurement tools have to be easy to administer and not too time consuming.[31] Gold standard tests based on sophisticated and expensive equipment are not viable, but simple clinical field tests may be used as a substitute for these tests.[2] However, a larger degree of inaccuracy must be expected with the use of clinical field tests. The findings of the present study indicate that the 6MWT, the hand-grip test, and the physical fitness questionnaire are reliable and valid measures that should be included in a core set evaluating the different aspects of health-related physical fitness in patients with various MSCs.

The absolute reliability, presented as measurement error and reported in the actual scale unit, is more clinically useful than the relative reliability. The 6MWT, the ST, the hand-grip test, and the physical fitness questionnaire displayed measurement errors of 49 m, 8 seconds, 4 kg, and 2 points, respectively, indicating that improvements or deteriorations exceeding these values are needed to distinguish a true change from the measurement error.

The 30sSTS and the FTF, however, showed measurement errors of 4 STSs and an almost 9-cm fingertip-to-floor distance. A measurement error of 4 STSs is higher than that presented in previous studies,[20,40,52] probably due to younger

TABLE 5 Correlation Between Clinical Field Tests and Laboratory-Based Tests, and the Physical Fitness Questionnaire and the Clinical Field Tests*

	Laboratory-Based Tests			Clinical Field Tests						
	Åstrand Cycle Ergometer†	CYBEX, 60°/s	CYBEX, 240°/s	6MWT†	ST	30sSTS†	HandR†	HandL	FTF	Heart Rate†
Clinical field tests										
6MWT	0.44	0.55	0.61	...	...	...	...	...	...	0.64
ST	−0.41	−0.53	−0.58	...	...	...	...	...	...	−0.57
30sSTS	0.43	0.41	0.43	...	...	...	...	...	...	...
HandR	0.12	0.76	0.72	...	...	...	...	...	...	...
HandL	0.12	0.70	0.65	...	...	...	...	...	...	...
FTF	0.16	0.03	0.00	...	...	...	...	...	...	...
Physical fitness questionnaire										
Aerobic fitness	...	...	...	0.59	−0.43	0.56	0.10	0.28	0.18	...
Muscle strength	...	...	...	0.38	−0.33	0.45	0.22	0.33	0.15	...
Flexibility	...	...	...	0.52	−0.41	0.23	0.24	0.36	0.42	...

Abbreviations: 30sSTS, 30-second sit-to-stand test; 6MWT, 6-minute walk test; CYBEX, CYBEX 6000 isokinetic dynamometer (Computer Sports Medicine Inc, Stoughton, MA); FTF, fingertip-to-floor test; HandL, hand-grip test left; HandR, hand-grip test right; ST, stair test.
Analyzed with the Spearman correlation coefficient (or Pearson if normally distributed). A priori corresponding constructs (FIGURE 1**) are marked in gray (n = 59-81).*
†Pearson correlation coefficient.

TABLE 6 Known-Group Validity*

	Low/Moderate Physical Activity Level (n = 37)	High Physical Activity Level (n = 38)	Mean Difference†	*P* Value
Clinical field tests				
6-minute walk test, m	549.3 ± 89.4	593.6 ± 95.7	44.3 (1.7, 87.0)	.04
Stair test, s	54.5 ± 14.4	51.5 ± 18.8	−3.1 (−10.8, 4.6)	.43
30-second sit-to-stand test, n	13.9 ± 4.3	16.2 ± 3.7	2.3 (0.5, 4.2)	.01
Hand grip right, kg	27.6 ± 10.2	31.8 ± 10.4	4.2 (−0.5, 9.0)	.08
Hand grip left, kg	27.4 ± 9.5	31.3 ± 10.6	3.9 (−0.8, 8.5)	.10
Fingertip-to-floor test, cm	−5.6 ± 16.7	−2.3 ± 11.6	3.4 (−3.5, 10.2)	.33
Physical fitness questionnaire				
Aerobic fitness	4.1 ± 1.6	5.8 ± 2.0	1.7 (0.9, 2.5)	>.001
Muscle strength	4.4 ± 1.8	5.7 ± 1.8	1.3 (0.5, 2.1)	.003
Flexibility	4.2 ± 2.0	5.3 ± 2.1	1.2 (0.2, 2.1)	.01

**Values are mean ± SD unless otherwise indicated.*
†The mean difference between patients reporting low/moderate physical activity level and patients reporting high physical activity level (according to the International Physical Activity Questionnaire guidelines) on the clinical field tests and the physical fitness questionnaire (n = 75). Values in parentheses are 95% confidence interval.

participants in our study, resulting in a higher mean number of STSs. Accordingly, we would expect a decreased measurement error in participants with fewer STSs, which was confirmed in our analyses. With higher accuracy in patients with lower capacity, this test may be most suitable for older adults.

Neither absolute reliability nor construct validity of the FTF test was found to be acceptable in the present study, indicating that the test has poor methodological qualities for use in patients with varying MSCs. However, the test has been reported to be appropriate in patients with more specific MSCs, that is, patients with back pain.[47]

The 6MWT, the hand-grip test, and

the 30sSTS and the physical fitness questionnaire all showed acceptable construct validity, confirming at least 2 of 3 of the a priori hypotheses. Our findings showed that the 6MWT could serve as an acceptable measure of cardiorespiratory endurance, but the test also correlated moderately with muscle strength and muscle endurance, underlining the typical characteristics of field tests, namely, reflecting a combination of several physical dimensions. However, reporting heart rate after the 6MWT is recommended to enhance the interpretation of the test as a measure of cardiorespiratory endurance.

The hand-grip test showed high correlation, whereas the 30sSTS only showed moderate correlation, with the laboratory-based test measuring muscle strength/endurance. A possible explanation may be that the 30sSTS is a multidimensional measure possibly involving other aspects than just muscle strength or muscle endurance.[36] In line with previous studies, we found a high correlation between grip strength and quadriceps strength, indicating that grip strength might be used as a measure of total body strength.[8] Given the higher accuracy of the hand-grip test compared to the 30sSTS test, the hand-grip test might be a more suitable measure of general muscle strength in patients with various MSCs.

The ST was previously found to be a measure of both aerobic capacity[12] and muscle strength[44]; therefore, the a priori hypotheses of divergent and convergent validity in our study might have been too ambitious. Although not fulfilling the acceptable number of hypotheses, the ST may still be a functional test measuring several aspects of physical fitness in patients with various MSCs.

Moderate correlations were found between the different items in the physical fitness questionnaire and the corresponding clinical field tests (adjusted for age and gender) (**FIGURE 1**). The only exception was the hand-grip test, possibly indicating that perceived strength for most people is more associated with lower-limb strength than upper-limb strength. The different items of the physical fitness questionnaire all showed acceptable construct validity, and the questionnaire may be a feasible method for measuring and monitoring the different aspects of health-related physical fitness in patients with MSCs.

The mean time to complete all measures included in this study was approximately 15 minutes, including standardized instructions but not the rest periods between tests. Taking into account that, for most patients, only a selection of tests will be used, the time consumed must be considered acceptable. Although some assistance was needed for completion of the physical fitness questionnaire, this was within acceptable limits.

Although the laboratory-based tests are well known and extensively used, some limitation in using them should be mentioned. The analyses might have been hampered by the uncertainty regarding estimation of $VO2_{max}$ using a submaximal test as well as correlating walking or running tests to a cycle ergometer test. In addition, the measurement properties depend on the population in question, and, even if the CYBEX and the Åstrand tests are valid for specific diagnoses, the applicability for patients with various types of MSCs is unclear.

A surprising finding in this study was that 32% of the patients reported change in their physical fitness at just 1 week, indicating that MSCs are constantly fluctuating conditions influencing the patients' perception of their fitness level. However, because a reliability study assesses the degree of measurement error in a measure and not changes in the construct, only patients with stable conditions were included in the analyses of reliability.[15,37] Nevertheless, patients reporting to be unchanged from one week to the next still showed a significant change in performance for most of the clinical field tests, indicating the presence of a learning effect in the tests.[27] The measurement error found in the present study, therefore, accounts for a combination of the measurement error in the tests and the learning effect of such tests.

Men and women were not equally represented in this study, and this could threaten the generalizability of the findings. However, MSCs are shown to be more prevalent in women,[26] and a previous study found that 70% of the patients in outpatient physiotherapy clinics were women.[51] Thus, the results of the present study are probably representative of the patients seeking treatment for MSCs in outpatient physiotherapy clinics.

CONCLUSION

THE ST, THE 30sSTS TEST, AND THE FTF test showed varying measurement properties in this heterogeneous group of patients, and further studies are needed to examine the usefulness of these measures for subgroups of patients. In contrast, the 6MWT, the hand-grip test, and the physical fitness questionnaire showed acceptable validity and reliability and are recommended to constitute a core set of measures assessing health-related physical fitness in patients with varying MSCs. ◉

KEY POINTS

FINDINGS: The 6MWT, the hand-grip test, and the physical fitness questionnaire showed acceptable methodological qualities for use in physiotherapy practice in patients with various MSCs.

IMPLICATIONS: The present study shows that the 6MWT, the hand-grip test, and the physical fitness questionnaire can be included in a core set of measures evaluating different aspects of health-related physical fitness in patients with various MSCs.

CAUTION: Clinical field tests and patient-reported measures are less accurate than laboratory-based tests but can be used in clinical practice.

ACKNOWLEDGEMENTS: *We want to thank the patients participating in the study and the physiotherapists recruiting them.*

REFERENCES

1. Altman DG. *Practical Statistics for Medical Research*. London, UK: Chapman & Hall; 1991.
2. American College of Sports Medicine. *ACSM's Health-Related Physical Fitness Assessment Manual*. 3rd ed. Baltimore, MD: Lippincott Williams & Wilkins; 2010.
3. American Thoracic Society. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med*. 2002;166:111-117. http://dx.doi.org/10.1164/ajrccm.166.1.at1102
4. 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 Res (Hoboken)*. 2011;63 suppl 11:S350-S370. http://dx.doi.org/10.1002/acr.20538
5. Benton MJ, Alexander JL. Validation of functional fitness tests as surrogates for strength measurement in frail, older adults with chronic obstructive pulmonary disease. *Am J Phys Med Rehabil*. 2009;88:579-583; quiz 584-586, 590. http://dx.doi.org/10.1097/PHM.0b013e3181aa2ff8
6. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. *Br J Sports Med*. 2009;43:1-2.
7. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc*. 2001;33:S379-S399; discussion S419-S420.
8. Bohannon RW. Are hand-grip and knee extension strength reflective of a common construct? *Percept Mot Skills*. 2012;114:514-518.
9. Bohannon RW. Dynamometer measurements of grip and knee extension strength: are they indicative of overall limb and trunk muscle strength? *Percept Mot Skills*. 2009;108:339-342.
10. Burr JF, Bredin SS, Faktor MD, Warburton DE. The 6-minute walk test as a predictor of objectively measured aerobic fitness in healthy working-aged adults. *Phys Sportsmed*. 2011;39:133-139. http://dx.doi.org/10.3810/psm.2011.05.1904
11. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985;100:126-131.
12. Cataneo DC, Cataneo AJ. Accuracy of the stair climbing test using maximal oxygen uptake as the gold standard. *J Bras Pneumol*. 2007;33:128-133.
13. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35:1381-1395. http://dx.doi.org/10.1249/01.MSS.0000078924.61453.FB
14. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033-1039. http://dx.doi.org/10.1016/j.jclinepi.2005.10.015
15. de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine*. Cambridge, UK: Cambridge University Press; 2011.
16. Farr JN, Going SB, Lohman TG, et al. Physical activity levels in patients with early knee osteoarthritis measured by accelerometry. *Arthritis Rheum*. 2008;59:1229-1236. http://dx.doi.org/10.1002/art.24007
17. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2:i-iv, 1-74.
18. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43:1334-1359. http://dx.doi.org/10.1249/MSS.0b013e318213fefb
19. Gauvin MG, Riddle DL, Rothstein JM. Reliability of clinical measurements of forward bending using the modified fingertip-to-floor method. *Phys Ther*. 1990;70:443-447.
20. Gill S, McBurney H. Reliability of performance-based measures in people awaiting joint replacement surgery of the hip or knee. *Physiother Res Int*. 2008;13:141-152. http://dx.doi.org/10.1002/pri.411
21. Hamilton A, Balnave R, Adams R. Grip strength testing reliability. *J Hand Ther*. 1994;7:163-170.
22. Haywood KL, Garratt AM, Jordan K, Dziedzic K, Dawes PT. Spinal mobility in ankylosing spondylitis: reliability, validity and responsiveness. *Rheumatology (Oxford)*. 2004;43:750-757. http://dx.doi.org/10.1093/rheumatology/keh169
23. Hodselmans AP, Dijkstra PU, Geertzen JH, van der Schans CP. Nonspecific chronic low back pain patients are deconditioned and have an increased body fat percentage. *Int J Rehabil Res*. 2010;33:268-270. http://dx.doi.org/10.1097/MRR.0b013e328335213f
24. Holm I, Fosdahl MA, Friis A, Risberg MA, Myklebust G, Steen H. Effect of neuromuscular training on proprioception, balance, muscle strength, and lower limb function in female team handball players. *Clin J Sport Med*. 2004;14:88-94.
25. Holm I, Ludvigsen P, Steen H. Isokinetic hamstrings/quadriceps ratios: normal values and reproducibility in sport students. *Isokinet Exerc Sci*. 1994;4:141-145. http://dx.doi.org/10.3233/IES-1994-4403
26. Holth HS, Werpen HK, Zwart JA, Hagen K. Physical inactivity is associated with chronic musculoskeletal complaints 11 years later: results from the Nord-Trøndelag Health Study. *BMC Musculoskelet Disord*. 2008;9:159. http://dx.doi.org/10.1186/1471-2474-9-159
27. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30:1-15.
28. Hovington CL, Nadeau S, Leroux A. Comparison of walking parameters and cardiorespiratory changes during the 6-minute walk test in healthy sexagenarians and septuagenarians. *Gerontology*. 2009;55:694-701. http://dx.doi.org/10.1159/000240015
29. Ihlebaek C, Laerum E. [Hits most, costs most and gets least]. *Tidsskr Nor Laegeforen*. 2010;130:2106. http://dx.doi.org/10.4045/tidsskr.10.1035
30. Jakobsen TL, Kehlet H, Bandholm T. Reliability of the 6-min walk test after total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. 2013;21:2625-2628. http://dx.doi.org/10.1007/s00167-012-2054-y
31. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. *Phys Ther*. 2009;89:125-135. http://dx.doi.org/10.2522/ptj.20080234
32. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Res Q Exerc Sport*. 1999;70:113-119.
33. Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: a longitudinal study evaluating outcome following total hip and knee arthroplasty. *BMC Musculoskelet Disord*. 2005;6:3. http://dx.doi.org/10.1186/1471-2474-6-3
34. Koegelenberg CF, Diacon AH, Irani S, Bolliger CT. Stair climbing in the functional assessment of lung resection candidates. *Respiration*. 2008;75:374-379. http://dx.doi.org/10.1159/000116873
35. Lennon OC, Denis RS, Grace N, Blake C. Feasibility, criterion validity and retest reliability of exercise testing using the Astrand-Rhyming test protocol with an adaptive ergometer in stroke patients. *Disabil Rehabil*. 2012;34:1149-1156. http://dx.doi.org/10.3109/09638288.2011.635748
36. McCarthy EK, Horvat MA, Holtsberg PA, Wisenbaker JM. Repeated chair stands as a measure of lower limb strength in sexagenarian women. *J Gerontol A Biol Sci Med Sci*. 2004;59:1207-1212.
37. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745. http://dx.doi.org/10.1016/j.jclinepi.2010.02.006
38. Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc*. 2007;39:1435-1445. http://dx.doi.org/10.1249/mss.0b013e3180616aa2
39. Nes BM, Janszky I, Wisløff U, Støylen A, Karlsen T. Age-predicted maximal heart rate in healthy subjects: the HUNT Fitness Study. *Scand J Med Sci Sports*. 2013;23:697-704. http://dx.doi.org/10.1111/j.1600-0838.2012.01445.x
40. Overend T, Anderson C, Sawant A, Perryman B,

Locking-Cusolito H. Relative and absolute reliability of physical function measures in people with end-stage renal disease. *Physiother Can*. 2010;62:122-128. http://dx.doi.org/10.3138/physio.62.2.122

41. Perret C, Poiraudeau S, Fermanian J, Colau MM, Benhamou MA, Revel M. Validity, reliability, and responsiveness of the fingertip-to-floor test. *Arch Phys Med Rehabil*. 2001;82:1566-1570. http://dx.doi.org/10.1053/apmr.2001.26064

42. Roberts HC, Denison HJ, Martin HJ, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. *Age Ageing*. 2011;40:423-429. http://dx.doi.org/10.1093/ageing/afr051

43. Ryan CG, Grant PM, Dall PM, Gray H, Newton M, Granat MH. Individuals with chronic low back pain have a lower level, and an altered pattern, of physical activity compared with matched controls: an observational study. *Aust J Physiother*. 2009;55:53-58.

44. Salem GJ, Wang MY, Young JT, Marion M, Greendale GA. Knee strength and lower- and higher-intensity functional performance in older adults. *Med Sci Sports Exerc*. 2000;32:1679-1684.

45. Scherr J, Wolfarth B, Christle JW, Pressler A, Wagenpfeil S, Halle M. Associations between Borg's rating of perceived exertion and physiological measures of exercise intensity. *Eur J Appl Physiol*. 2013;113:147-155. http://dx.doi.org/10.1007/s00421-012-2421-x

46. Segura-Ortí E, Martínez-Olmos FJ. Test-retest reliability and minimal detectable change scores for sit-to-stand-to-sit tests, the six-minute walk test, the one-leg heel-rise test, and handgrip strength in people undergoing hemodialysis. *Phys Ther*. 2011;91:1244-1252. http://dx.doi.org/10.2522/ptj.20100141

47. Strand LI, Anderson B, Lygren H, Skouen JS, Ostelo R, Magnussen LH. Responsiveness to change of 10 physical tests used for patients with back pain. *Phys Ther*. 2011;91:404-415. http://dx.doi.org/10.2522/ptj.20100016

48. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol*. 2006;59:160-167. http://dx.doi.org/10.1016/j.jclinepi.2005.07.012

49. Strøyer J, Essendrop M, Jensen LD, Warming S, Avlund K, Schibye B. Validity and reliability of self-assessed physical fitness using visual analogue scales. *Percept Mot Skills*. 2007;104:519-533.

50. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42. http://dx.doi.org/10.1016/j.jclinepi.2006.03.012

51. Vasseljen O, Hansen AE. Pasienter i privat praksis – hvem er de og hva lider de av? *Fysioterapeuten*. 2002;5:13-18.

52. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. *J Orthop Sports Phys Ther*. 2011;41:319-327. http://dx.doi.org/10.2519/jospt.2011.3515

@ MORE INFORMATION
WWW.JOSPT.ORG

GO GREEN By Opting Out of the Print *Journal*

JOSPT subscribers and APTA members of the Orthopaedic and Sports Physical Therapy Sections can **help the environment by "opting out"** of receiving *JOSPT* in print each month as follows. If you are:

- **A *JOSPT* subscriber:** Email your request to **jospt@jospt.org** or call the *JOSPT* office toll-free at **1-877-766-3450** and provide your name and subscriber number.

- **APTA Orthopaedic or Sports Section member:** Go to http://www.apta.org/, log in, and select **My Profile**. Next click on **Email Management/GoGreen**. Toward the bottom of the list, you will find the **Publications** options and may opt out of receiving the print *JOSPT*. **Please save this preference.**

Subscribers and members alike will continue to have access to *JOSPT* online and can retrieve current and archived issues anytime and anywhere you have Internet access.