OXFORD

THE GERONTOLOGICAL SOCIETY OF AMERICA®

Research Practice

The Stair Climb Power Test as an Efficacy Outcome in Randomized Trials of Function Promoting Therapies in Older Men

Thiago Gagliano-Jucá, MD, PhD,[1,†,◉] Zhuoying Li, MS,[1,†] Karol M. Pencina, PhD,[1] Tinna Traustadóttir, PhD,[2] Thomas G. Travison, PhD,[1,3,◉] Linda Woodhouse, PT, PhD,[4,◉] Shehzad Basaria, MD,[1] Panayiotis D. Tsitouras, MD,[5] S. Mitchell Harman, MD,[6] Shalender Bhasin, MBBS,[1] and Thomas W. Storer, PhD[1,*]

[1]Research Program in Men's Health, Aging, and Metabolism; Boston Claude D. Pepper Older Americans Independence Center; Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. [2]Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona. [3]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Roslindale, Massachusetts. [4]Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada. [5]Health Sciences Center, University of Oklahoma, Oklahoma. [6]Phoenix VA Health Care System, Arizona.

*Address correspondence to: Thomas W. Storer, PhD, Laboratory of Exercise Physiology and Physical Performance Research Program on Men's Health, Aging, and Metabolism, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue, BLI-542 Boston, MA 02115. E-mail: tstorer@bwh.harvard.edu

†These authors contributed equally to this work.

Abstract

Background: Standardization of performance-based physical function measures that are reliable and responsive to intervention is necessary for efficacy trials of function promoting anabolic therapies (FPTs). Herein, we describe a standardized method of measuring stair climbing power (SCP) and evaluate its ability to assess improvements in physical function in response to an FPT (testosterone) compared to gait speed.
Methods: We used a 12-step SCP test with and without carrying a load (loaded, LSCP or unloaded, USCP) in two testosterone trials in older men. SCP was determined from mass, total step-rise, and time of ascent measured with an electronic timing system. Associations between SCP and leg press performance (strength and power), testosterone levels, and gait speed were assessed. Test–retest reliability was evaluated using interclass correlation and Bland–Altman analyses.
Results: Baseline SCP was negatively associated with age and positively with leg strength and power and gait speed. Both tests of SCP were safe and showed excellent reliability (intra-class correlation 0.91–0.97 in both cohorts). Changes in testosterone concentrations were associated with changes in USCP and LSCP, but not gait speed in mobility-limited men. Changes in leg press performance were associated with SCP in both trials.
Conclusions: Both USCP and LSCP are safe and have high test–retest reliability. Compared to gait speed, SCP is associated more robustly with leg press performance and is sensitive to testosterone therapy. The LSCP might be a more responsive outcome than gait speed to evaluate the efficacy of FPT in randomized trials.

Keywords: Physical Function, Testosterone, Muscle Power, Anabolic Intervention.

The past decade has witnessed remarkable advances in the development of novel function promoting anabolic therapies (FPTs) for the prevention and treatment of aging-associated sarcopenia and functional limitations. Testosterone, selective androgen receptor modulators, myostatin antagonists, growth hormone secretagogues, and insulin-like growth factor-mimetics have been studied in this con-

text (1). Demonstrating improvements in performance-based measures of physical function is important in establishing the efficacy of FPTs. Therefore, careful selection of measures of physical function that are patient-important, utilize the appropriate domains of muscle performance, and are responsive to the intervention is critically important in efficacy trials of FPTs.

Several performance-based measures of physical function have been used in randomized trials of FPTs including chair stands, the timed get-up-and-go, and gait speed over varying distances and under varying conditions (2). Composite tests that assess more than one functional measure, such as the physical performance test (3) and the short physical performance battery (4), have also been used. Most of these measures, including walking speed, have shown inconsistent or only modest responsiveness to the administration of FPTs in randomized trials in healthy older adults or in older adults with pre-frailty or mild-to-moderate degrees of functional limitation. In part, this may be due to a ceiling effect where there is little improvement in performance of a task even though a person's physical ability to do that task increases substantially after treatment. Other tests may exhibit a floor effect such that very substantial improvements in ability may be needed before any change in test performance above the performance floor are observed.

The ability to climb stairs is an important functional attribute, especially in older adults. Difficulty climbing stairs is one of the two functional limitations most frequently reported by older adults with mobility limitation (5–7) and was among the top five tasks that older adults rated as being most difficult (5). Difficulty in climbing stairs has also been reported as a marker of functional decline that can lead to loss of independence (8). Although stair climbing requires contributions from musculoskeletal, cardiopulmonary, and somatosensory systems, it is most reliant on leg muscle power, the latter expressed as the rate of muscle force generation (9–12). The stair climbing power (SCP) test is being used increasingly as a performance-based measure of physical function in efficacy trials of FPTs (13–17). However, previous studies that have used the stair climb test have had substantial variation in how the test is performed, including the number of steps, step height, ascents with or without descents, timing, and the criterion variable (eg, time of ascent and/or descent, or power). Because of its association with lower extremity muscle power, its potential for a higher ceiling, and importance in activities of daily living, we posited that the SCP test could provide a meaningful assessment of the efficacy of FPTs. Here, we describe a standardized approach to performing the SCP test, its performance characteristics and response to the administration of a FPT in the setting of randomized trials. The standardization of the SCP test could reduce variation across trials and enhance its usefulness as a performance-based measure of physical function.

Methods

Subjects

The data for the SCP test were collected in two randomized controlled trials in older men investigating the efficacy of a transdermal testosterone gel in improving lean mass, muscle performance, and physical function; the Testosterone in Older Men with Mobility Limitation (TOM) (13,18) (N = 146), and Testosterone's Effects on Atherosclerosis Progression in Aging Men (TEAAM) trials (14,19) (N = 253). The participants in the TOM trial were at least 65 years of age with low to low-normal serum total testosterone (100–350 ng/dL) or low free testosterone (<50 pg/mL). These subjects were mobility-limited defined by self-reported difficulty in walking two

blocks or climbing stairs and an short physical performance battery total score between 4 and 9 indicating mild-to-moderate mobility limitations (20). The TEAAM study included men at least 60 years of age who had low to low-normal serum total testosterone (100–400 ng/dL) or low free testosterone (<50 pg/mL) and were relatively healthy. All muscle performance variables assessed in the TEAAM trial were collected at baseline and at 6 months. Exclusion criteria for both studies included any contraindication to testosterone administration or other orthopedic, cardiac, cognitive, or neurological condition that would prohibit functional assessment. Fasting serum total testosterone concentrations were measured using an immunoassay with a sensitivity of 10 ng/dL in both trials (21). Subjects provided informed consent. The studies were approved by the institutional review boards at the participating sites.

Interventions

Details of the TOM and TEAAM trials have been described (18,19). Participants in the TOM trial were randomized to apply 10 g of either placebo or 1% testosterone transdermal gel daily for 6 months. Men in the TEAAM trial were randomized to either 7.5 g of 1% testosterone gel or placebo gel daily for 3 years. Two weeks after randomization, doses were adjusted in both trials to maintain on-treatment testosterone levels in the mid-normal range for young eugonadal men (~500 ng/dL). To maintain blinding, an unblinded study staff adjusted the dose of participants in the placebo group simultaneously.

Assessment of SCP

Subject preparation

Participants wore comfortable clothes and athletic shoes, or shoes with non-slip soles. A checklist was used to determine if the subject had any contraindications for testing, including severe limitations in lower body mobility, balance problems, major visual impairment, or degree of cardiopulmonary disease deemed to require exclusion from the test. Body mass was measured with shoes and clothing. The subjects rested for 5 minutes before the test began. Blood pressure and heart rate were measured at the end of the rest period.

Equipment

We used a well-lit 12-step interior staircase with handrails for the stair climb task. The rise (height) of each step was measured to the nearest millimeter and summed to obtain the total rise for the 12 steps. Time of ascent was measured with an electronic timing system consisting of a digital timer and switch mats with non-skid surfaces (Lafayette Instruments, Lafayette, IN) placed 15 cm from the base of the steps and on the 12th step). Orange cones were placed on the 12th step to mark the number of steps over which subjects should maintain their effort.

Procedure

Subjects were instructed to climb the stairs one step at a time, as fast as safely possible without running, skipping steps, or assistance. The subjects were not allowed to use the handrail except as needed for balance. The examiner demonstrated the test and used scripted instructions. Subjects were positioned one stride length behind the first switch mat and instructed to start on the command "go." Time started when the subject stepped on the first switch mat. Examiners followed the subjects during the climb to provide assistance if needed but without pacing the subject. After the ascent, subjects descended the steps with examiner assistance, were seated, and provided with 1.5 minutes of rest before the second trial.

Stopping criteria for the test
The test was stopped if the participant reported chest discomfort, significant shortness of breath, feeling faint or lightheaded, or if the subject did not reach the 12th step within 2 minutes. If the subject discontinued the test before the 12th step, total time for ascent was recorded as 2:00 minutes.

The Stair Climb Power Test

The SCP test was administered under two conditions: climbing the stairs normally ("unloaded," USCP) and while carrying a load ("loaded," LSCP). We hypothesized that the loaded condition would raise the ceiling of the test and increase the likelihood of detecting a response to the intervention compared to the unloaded test. Participants performed the LSCP test while carrying 20% of body mass. This mass was in the form of weight plates divided evenly among two canvas tote bags with one tote bag carried in each hand. The same absolute mass was used in subsequent tests. Prior to the test, each subject was allowed to lift the bags, take a few steps, and then agree or disagree with the question "do you feel it is safe for you to climb the steps caring these weights?" The test was not administered to those who disagreed.

Criterion Measure

Total ascent time was registered and power was calculated from the subjects' body mass (kg) plus the load carried, total rise (m) of the steps, the acceleration of gravity, and time of ascent [equation (1)].

$$SCP = \frac{(\text{body mass} + \text{load carried (kg)}) \times 9.81 \, \text{m/s}^2 \times \text{total rise (m)}}{\text{time (s)}} \tag{1}$$

Other Measures

Additional muscle performance tests were evaluated as outcomes in the original studies, but the relationship between their changes and changes in SCP were assessed for this report. Leg press strength was measured with a seated leg press machine (A420; Keiser Sport, Inc., Fresno, CA) using pneumatic resistance and the one-repetition maximum (1-RM) method (22). Leg press power was assessed using the same leg press machine with loads of 50%, 60%, and 70% of the 1-RM. The highest peak power was used in the analysis. All study sites used the same make and model leg press machine. Gait speed was measured during unloaded and loaded (20% body mass) 40 m walk tests using a switch mat and infrared timing system (22). Subjects were instructed to walk as fast as safely possible without running, and were allowed to use assistive devices (eg, canes and walkers) as needed. Two trials were given for each condition with 1.5 minutes rest between trials (22).

Data Analyses

Feasibility of the statistical assumptions was explored graphically and quantitatively. Assessment of potential outlying values was examined using outer Tukey's fences criteria. In addition, measures of physical function that were not biologically plausible were also considered outliers. Baseline associations between LSCP and USCP with demographic data, total testosterone and performance measures were assessed using linear regression models. Associations between changes in SCP and changes in leg press strength, leg press power, and gait speed were also examined using linear regression models; these analyses were conducted in testosterone groups only using the best trial of SCP and gait speed for each condition (loaded or unloaded).

In the absence of linear relationships between measures, nonlinear associations were investigated using restricted cubic splines (23). Estimates, corresponding 95% confidence intervals (CIs), R-squared metrics, and *p* values are presented for each association. Reliability of the selected measures was determined by examining the agreement between Trial 1 and Trial 2 using intra-class correlation (ICC) and Bland–Altman analysis (24). For ICCs, a two-way mixed-model for repeated measures was used. In the Bland–Altman method, the difference between Trials 1 and 2 (Trial 2 subtracted from Trial 1) was plotted against the mean of the two measurements.

All analyses were conducted separately for each study using SAS v.9.3 (SAS Institute, Cary, NC) and R software version 3.2.4. Statistical tests were two-sided with Type I error alpha of 0.05.

Results

Baseline Characteristics of the Participants

The baseline characteristics of the participants in the two trials included in the present study, including those randomized to placebo and testosterone groups, are summarized in Supplementary Table 1. The mean ± SD age of the TOM trial participants was 73.0 ± 5.2 years, body mass of 89.5 ± 13.8 kg, and body mass index of 29.9 ± 4.2 kg/m^2, and total testosterone was 244.9 ± 62.1 ng/dL. TEAAM trial participants were 67.0 ± 5.0 years of age, had body mass of 86.6 ± 11.0 kg, with body mass index of 28.2 ± 3.0 kg/m^2, and total testosterone was 312.2 ± 63.9 ng/dL. No subject exceeded 36 seconds to perform a SCP test. Community-dwelling older men (TEAAM trial) exhibited higher USCP and higher LSCP than older men with mobility limitations (TOM trial; Supplementary Table 1).

Distribution of SCP in Community-Dwelling Older Men and Older Men With Mobility Limitations

Both USCP and LSCP were negatively associated with age in community-dwelling and mobility-limited older men (all *p* values <.001; Figure 1; Table 1). Baseline total testosterone in TEAAM trial, had a weak positive association with USCP (*p* = .007; R^2 = .029), but not with LSCP (*p* = .220; R^2 = .007). Baseline total testosterone in TOM trial participants was not associated with USCP or LSCP (Figures 1; Table 1). Baseline body mass and body mass index were positively associated with USCP and LSCP in both cohorts (Table 1; Supplementary Figure 1). Baseline USCP and LSCP were strongly and positively associated with leg press 1-RM and peak power (Figure 1; Table 1), and unloaded and loaded gait speed in both cohorts (all *p* values <.001; Figure 2; Table 1).

Changes in Men Treated With Testosterone
TOM trial

In the TOM trial, serum total testosterone concentrations increased more in men treated with testosterone (mean increase at 6 months of 358.6 ng/dL; 95% CI = 241.5 to 475.8 ng/dL) than in those in the placebo arm (64.0 ng/dL; 95% CI = 24.0 to 104.0 ng/dL). In men treated with testosterone, age-adjusted changes in leg press 1-RM were associated with changes in USCP [similarly to unadjusted analyses previously reported (13)] but not with changes in LSCP; changes in leg press power were associated with LSCP (Figure 3; Table 2). Also, changes in leg press strength and power were not associated with changes in gait speed. Changes in testosterone levels in the testosterone arm were associated with changes in both USCP and LSCP but not with changes in gait speed (Figure 4). Changes in gait speed were not associated with changes in USCP or LSCP (Figure 3).
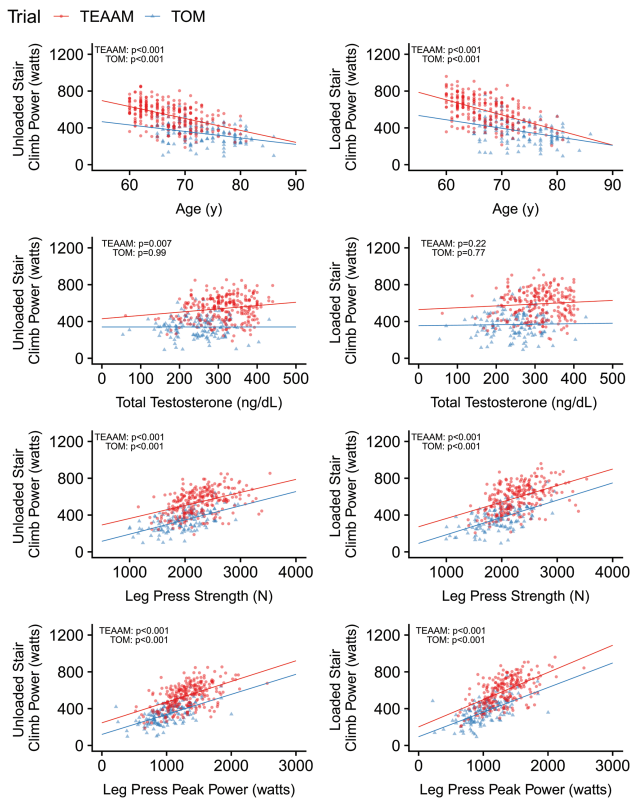
Figure 1. Baseline associations between stair climbing power and subject characteristics in both trials.

TEAAM trial

In the TEAAM trial, mean on-treatment increase in serum total testosterone concentration at 6 months was 311.0 ng/dL (95% CI = 239.6 to 382.4 ng/dL) in men randomized to testosterone, while mean level in the placebo arm did not change significantly (52.0 ng/dL; 95% CI = 28.6 to 75.4 ng/dL). In men randomized to testosterone, changes in USCP and LSCP, but not gait speed, were associated with changes in leg press 1-RM and leg press power (Figure 3; Table 2). Contrasting with TOM trial findings, changes in testosterone levels were not associated with performance in SCP tests or gait speed (Figure 4), and changes in gait speed were not associated with changes in USCP or LSCP (Figure 3).

Test–Retest Reliability

The test–retest measures of the USCP and the LSCP tests demonstrated excellent agreement; at baseline, in the TOM trial, ICC was 0.93 (95% CI = 0.90 to 0.95) and 0.97 (95% CI = 0.95 to 0.98), for USCP and LSCP, respectively. The Bland–Altman analysis revealed significant bias between Trial 1 and Trial 2 of −11.53 W (95% CI = −17.81 to −5.24 W) for USCP and −10.32 W (95% CI = −14.71 to −5.92) for LSCP measures (Supplementary Figure 2A) at baseline. ICC values similar to those seen at baseline were found at 6 months for both groups (Supplementary Table 2), and bias between Trials 1 and 2 in the testosterone group was nonsignificant (1.28 W; 95% CI = −10.09 to 12.64 W) for USCP and for LSCP (−5.39 W; 95% CI = −15.42 to 4.65 W) (Supplementary Figure 2B); bias between Trials 1 and 2 in the placebo group was also nonsignificant at 6 months for USCP (−3.82 W; 95% CI = −17.56 to 9.92 W) and for LSCP (−7.71 W; 95% CI = −17.15 to 1.74 W) (Supplementary Figure 2C).

For the TEAAM trial, at baseline ICC was 0.91 (95% CI = 0.87 to 0.94) and 0.95 (95% CI = 0.93 to 0.97), for USCP and LSCP, respectively. The Bland–Altman analysis also revealed significant bias for USCP (−29.98 W; 95% CI = −37.63 to −22.33 W) and for LSCP −31.84 W (95% CI = −37.32 to −26.35) between Trials 1 and 2 at baseline (Supplementary Figure 3A). As observed in the TOM trial, ICC at 6 months was similar to baseline, regardless of treatment group (Supplementary Table 2). Bias between Trials 1 and 2 in the testosterone group at 6 months was −29.50 W (95% CI = −39.50 to −19.50 W) for USCP and −22.44 W (95% CI = −32.82 to −12.07 W) for LSCP (Supplementary Figure 3B); bias between Trials 1 and 2 in the placebo group was −29.15 W (95% CI = −39.30 to −18.99 W) for USCP and −31.81 W (95% CI = −44.09 to −19.54 W) for LSCP (Supplementary Figure 3C).

Safety

No adverse events were recorded in nearly 2,500 USCP and LSCP tests in these two trials.

Discussion

We describe here a standardized method for the SCP test that can be used as a performance-based measure of physical function in randomized trials of FPTs. Implementation of a standardized approach to the SCP test should allow a more accurate evaluation of efficacy across different trial sites and studies, as well as the execution of appropriate meta-analytical analyses of findings in future trials of FPT. The SCP test is safe and has excellent test–retest reliability in community-dwelling and mobility-limited older men. Importantly, the SCP test was consistently responsive to changes in muscle strength and muscle power associated with testosterone treatment in both trials. Other randomized trials of promyogenic drugs, such as testosterone and selective androgen receptor modulators, have also reported that SCP is substantially more responsive to change during administration of promyogenic drugs than gait speed. Therefore, the SCP test could serve as a reliable, convenient, and safe performance-based measure of physical function in efficacy trials of promyogenic drugs that act primarily by increasing muscle mass and muscle strength.

We show that increasing serum total testosterone in older men with low or low-normal baseline serum testosterone levels who are mobility-limited significantly associates with change in SCP but not gait speed. However, neither changes in USCP, LSCP, nor gait speed were associated with changes in total testosterone in older community-dwelling healthy men. Changes in SCP were well associated with changes in more proximal measures of muscle performance (eg, leg muscle strength and power) in men treated with testosterone in both trials, whereas gait speed was not. The significant associations between SCP and leg press power add to previous reports of the validity of the SCP test (11,25).

The study also confirms the excellent reliability of the SCP we previously noted in small samples of young and older men (22). Agreement by ICCs was excellent (0.942–992) but performance in Test 2 was marginally better than in Test 1. The ICCs for the LSCP test were also greater than 0.940, bias was similar, and the LSCP test better discriminated between young men, healthy older men, and mobility-limited older men (22). In the present study, ICCs in both cohorts were high and some bias toward higher SCP in the first trial was observed at baseline in both studies, as well as at 6 months in the TEAAM trial. The reasons for these contrasting findings in bias between our present and past studies are not clear but might be related

Table 1. Baseline Associations Between Stair Climb Power and Gait Speed With Subject Characteristics

| | Unloaded Stair Climb Power | | | | | | | | Loaded Stair Climb Power | | | | | | | |
| | TOM | | | | TEAAM | | | | TOM | | | | TEAAM | | | |
	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²
Age (years)	143	−7.1 (−10.3 to −3.8)	<.001	0.118	253	−13.0 (−15.9 to −10.1)	<.001	0.238	136	−9.3 (−13.3 to −5.2)	<.001	0.135	229	−16.4 (−19.9 to −12.9)	<.001	0.275
Body mass (kg)	143	2.6 (1.4 to 3.8)	<.001	0.111	253	4.3 (2.9 to 5.7)	<.001	0.125	136	2.7 (1.1 to 4.3)	.001	0.078	229	3.8 (2.1 to 5.5)	<.001	0.078
BMI (kg/m²)	143	6.5 (2.4 to 10.6)	.002	0.064	253	10.3 (4.9 to 15.8)	<.001	0.053	136	6.6 (1.4 to 11.8)	.013	0.045	229	9.4 (2.9 to 16.0)	.005	0.035
Total testosterone (ng/dL)	143	0.0003 (−0.3, 0.3)	.99	<0.001	252	0.36 (0.10 to 0.61)	.007	0.029	136	0.05 (−0.30 to 0.41)	.77	<0.001	229	0.20 (−0.12 to 0.52)	.22	0.007
Leg press strength (N)	144	0.15 (0.12 to 0.19)	<.001	0.338	249	0.14 (0.10 to 0.18)	<.001	0.189	138	0.19 (0.14 to 0.23)	<.001	0.343	225	0.18 (0.14 to 0.22)	<.001	0.225
Leg press peak power (W)	143	0.22 (0.17 to 0.26)	<.001	0.375	247	0.22 (0.18 to 0.27)	<.001	0.276	137	0.27 (0.21 to 0.32)	<.001	0.390	223	0.30 (0.24 to 0.35)	<.001	0.347
Unloaded gait speed (m/s)	136	171.1 (133.9 to 208.3)	<.001	0.382	85	165.7 (122.3 to 209.1)	<.001	0.410	130	207.0 (157.2 to 256.8)	<.001	0.346	84	206.1 (156.8 to 255.4)	<.001	0.457
Loaded gait speed (m/s)	136	160.0 (120.8 to 198.7)	<.001	0.330	81	164.5 (123.8 to 205.2)	<.001	0.451	134	203.3 (157.0 to 249.5)	<.001	0.364	81	208.9 (165.4 to 252.4)	<.001	0.536
Unloaded Gait Speed / Loaded Gait Speed																
Total testosterone (ng/dL)	133	0.001 (−0.0001 to 0.002)	.08	0.015	96	0.001 (−0.0004 to 0.002)	.17	0.020	133	0.0004 (−0.001 to 0.002)	.40	0.005	82	0.0003 (−0.001 to 0.002)	.66	0.002
Leg press strength (N)	134	0.0004 (0.0002 to 0.0005)	<.001	0.133	87	0.0004 (0.0002 to 0.0006)	<.001	0.200	135	0.0003 (0.0002 to 0.0005)	<.001	0.127	82	0.0004 (0.0003 to 0.0006)	<.001	0.226
Leg press peak power (W)	133	0.001 (0.0005 to 0.001)	<.001	0.233	85	0.001 (0.0005 to 0.001)	<.001	0.315	135	0.001 (0.0004 to 0.001)	<.001	0.193	81	0.001 (0.0005 to 0.001)	<.001	0.365

Note: BMI = body mass index; CI = confidence interval; TEAAM = Testosterone's Effects on Atherosclerosis Progression in Aging Men; TOM = Testosterone in Older Men with Mobility Limitation.
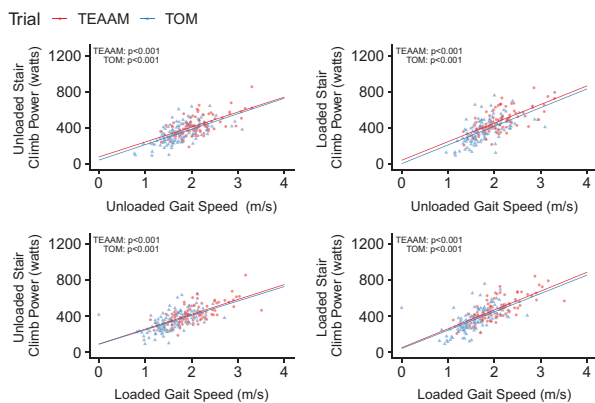
Figure 2. Baseline associations between stair climbing power and gait speed in both trials.

to the smaller sample size of our previous report. It is also possible that 1.5 minutes rest between trials is insufficient to allow for full recovery in some men. Nonetheless, using the best of the two trials should minimize the impact of bias to evaluations. Future studies should evaluate the impact of longer rest periods between trials.

Gait speed has been used widely as an efficacy outcome in randomized trials of FPTs. However, the changes in gait speed in most trials of testosterone, selective androgen receptor modulators, and myostatin antagonists have been small and not statistically significant despite significant gains in muscle mass and strength (13,14,26–28). In the present analyses, changes in gait speed in testosterone-treated men were not associated with changes in leg press strength, leg power, or testosterone levels. In contrast, SCP tests were responsive to testosterone administration, and the improvements in both USCP and LSCP were associated with changes in testosterone levels (in the TOM trial), as well as leg press strength and leg power. Our analyses suggest that gait speed and SCP assess different attributes of muscle performance and physical function and respond differently to promyogenic agents in older men. These findings support the use of the SCP test as a more sensitive measure of lower extremity function that is better aligned with lower extremity muscle strength and power and is more responsive than gait speed to interventions that increase muscle strength. The ability to climb stairs is important during daily activities as well as in emergency situations. Difficulty in climbing stairs is a common complaint among older adults with mobility limitation (5–7). Therefore, the SCP test may be particularly useful as a performance-based patient-important measure of physical function in efficacy trials of FPTs.

Because of the short time it takes to climb 12 steps, a high level of precision in timing is necessary in reducing test–retest variability. Timing precision is probably increased by the utilization of switch mats to trigger the starting and stopping of electronic timers. In addition, the test's reliability can be optimized by performing two trials and using the better of the two trials as the criterion measure. Manual measurement of stair climb time with a stopwatch may be affected by delays related to simple reaction time (29) from the examiner (activating/stopping the device), as well as chronostasis (distortion of time perception in visual recognition of movement (30) by the examiner). Nonetheless, one might still apply hand-held devices if one is willing to accept these potential sources of error.

Both USCP and LSCP tests are safe; we performed more than 2,500 stair climb tests without an adverse event even in mobility-limited older men. We attribute this to careful screening, instruction,
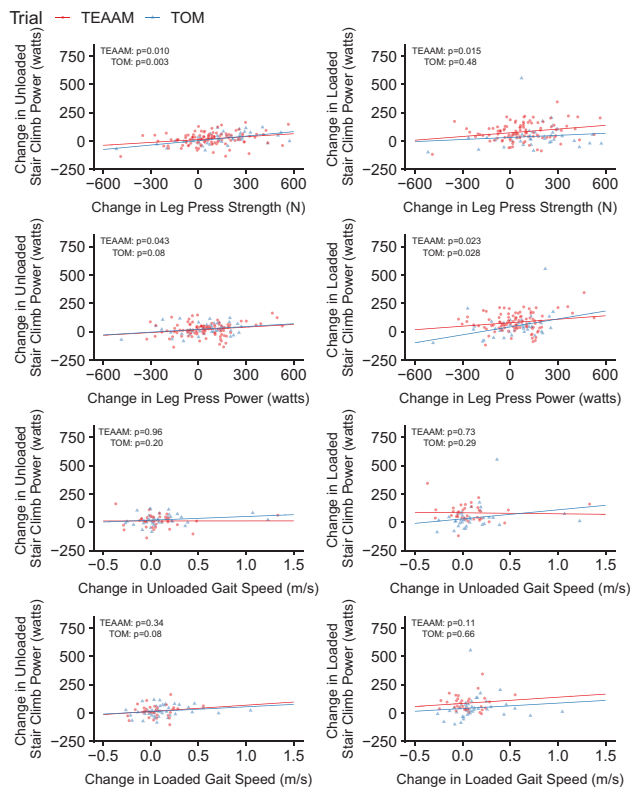

Figure 3. Associations between changes from baseline at 6 months in stair climbing power and changes in muscle strength, muscle power, and gait speed in both trials.

and careful monitoring during the climbs. In addition, verifying if subjects felt comfortable climbing the stairs while carrying the tots before the test took place might have played a role in the overall safety of the LSCP test.

Knowledge of the minimally important difference is valuable for interpreting the clinical meaningfulness of intervention efficacy. We previously reported minimally important differences of 38.9 and 60.7 W for USCP and LSCP, respectively, for the same sample of mobility-limited older men included is this report (13). These minimally important differences were computed using an anchor-based method that used a Likert scale that compared global ratings of perceived change in stair climbing ability. Because minimally important differences are specific to the context and the study population, similar data should be derived in other populations using the SCP test so as to guide sample size estimates and to understand the clinical meaningfulness of treatment effects in pivotal trials. New technologies such as wearable sensors to measure acceleration, force, and vertical distance for automated calculation of SCP could reduce staff burden and enhance precision but need validation.

Limitations

This report has some limitations. These studies enrolled either healthy older men or older men with mobility limitation with low or low-normal testosterone levels; generalizability to women, more disabled populations or to younger adults should be investigated. The tests may not be possible or safe in older adults with severe mobility limitation or balance problems. Many participants did not have follow-up measures of some physical function tests, which decreased the power of some analyses; with our limited power, statistically

Table 2. Association of Changes in Different Outcomes With Changes in Physical Function Tests

| | Change of Unloaded Stair Climb Power | | | | | | | | Change of Loaded Stair Climb Power | | | | | | | |
| | TOM | | | | TEAAM | | | | TOM | | | | TEAAM | | | |
	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²
Change of total testosterone (ng/dL)	48	0.04 (0.01 to 0.08)	.012	0.136	96	0.005 (−0.03 to 0.04)	.73	0.004	44	0.14 (0.07 to 0.20)	<.001	0.287	88	0.02 (−0.03 to 0.06)	.46	0.009
Change of leg press strength (N)	48	0.09 (0.03 to 0.15)	.003	0.175	108	0.08 (0.02 to 0.13)	.010	0.070	42	0.05 (−0.09 to 0.18)	.48	0.017	97	0.10 (0.02 to 0.18)	.015	0.066
Change of leg press peak power (W)	45	0.09 (−0.01 to 0.18)	.08	0.067	103	0.07 (0.002 to 0.14)	.043	0.047	40	0.23 (0.03 to 0.44)	.028	0.129	92	0.11 (0.02 to 0.21)	.023	0.056
Change of unloaded gait speed (m/s)	39	57.4 (−54.0 to 168.8)	.30	0.022	36	−54.7 (−169.9 to 60.6)	.34	0.033	35	91.6 (−53.1 to 236.3)	.21	0.045	35	1.1 (−128.8 to 131.0)	.99	<0.001
Change of loaded gait speed (m/s)	44	88.6 (3.3 to 174.0)	.04	0.062	33	55.8 (−61.7 to 173.3)	.34	0.029	41	146.2 (30.1 to 262.3)	.02	0.100	33	100.7 (−24.3 to 225.6)	.11	0.077

| | Change of Unloaded Gait Speed | | | | | | | | Change of Loaded Gait Speed | | | | | | | |
| | TOM | | | | TEAAM | | | | TOM | | | | TEAAM | | | |
	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²	N	Estimate (95% CI)	p value	R²
Change of total testosterone (ng/dL)	37	0.000002 (−0.0001 to 0.0001)	.97	0.0003	32	−0.00003 (−0.0003 to 0.0002)	.81	0.002	42	0.00006 (−0.0001 to 0.0002)	.35	0.002	26	−0.00002 (−0.0003 to 0.0003)	.91	0.0005
Change of leg press strength (N)	37	0.0001 (−0.0001 to 0.0003)	.49	0.040	37	0.0001 (−0.0003 to 0.0005)	.76	0.002	41	0.0002 (−0.0001 to 0.0004)	.13	0.096	34	0.0004 (−0.00005 to 0.0008)	.08	0.078
Change of leg press peak power (W)	34	0.0006 (0.0003 to 0.001)	.001	0.351	35	0.0001 (−0.0003 to 0.001)	.68	0.004	39	0.0003 (0.00002 to 0.0006)	.035	0.153	32	0.0003 (−0.0001 to 0.001)	.10	0.092

Notes: Data displayed are mean estimated changes from baseline and their 95% CI for muscle performance tests at the end of the intervention period (controlling for age at baseline) in each study. *p* values and R² for each association with changes in physical function tests are given. CI = confidence interval; TEAAM = Testosterone's Effects on Atherosclerosis Progression in Aging Men; TOM = Testosterone in Older Men with Mobility Limitation.
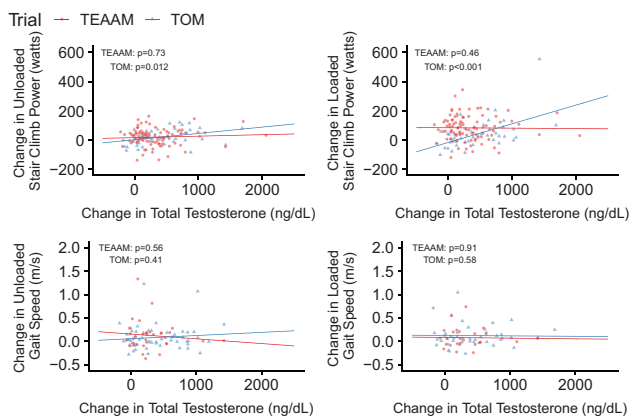
Figure 4. Associations between change from baseline at 6 months in total testosterone with changes in stair climb power and gait speed in both trials.

significant associations demonstrate sufficiently strong relationships between dependent and independent variables and, potentially, a ceiling effect. However, if a true ceiling effect is present in the population, it cannot be determined with our data. Both trials used testosterone as the promyogenic drug; the experience with the SCP test in trials of other FPTs is limited but might be similarly reliable for other therapies and types of interventions (eg, resistance training). We used a 12-step staircase in our studies, but this might not be available at all institutions. However, it is possible that a staircase with eight steps could provide a reasonable alternative; this remains to be demonstrated.

Conclusions

The SCP test is a safe and useful performance-based measure of physical function which can be standardized to achieve high levels of precision and test–retest reliability. Because of the importance of stair climbing in daily activities, the relation of SCP to lower extremity muscle strength and muscle power, and its relatively higher ceiling, the SCP test may serve as a useful performance-based measure of physical function in efficacy trials of promyogenic FPT, which act by increasing muscle strength.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Funding

Conflict of Interest

L.W. has served as a consultant to Eli Lilly and Scholar Rock. S.Ba. has served as a consultant to AbbVie Pharmaceuticals and Regeneron. S.Bh. has received research grant support from AbbVie Pharmaceuticals, Transition Therapeutics, Takeda Pharmaceuticals, and Eli Lilly for investigator-initiated research unrelated to this study. S.Bh. has served as a consultant to AbbVie, Regeneron, Novartis, and Eli Lilly. S.Bh. has a financial interest in Function

Promoting Therapies, LLC, a company aiming to develop innovative solutions that enhance precision and accuracy in clinical decision making and facilitate personalized therapeutic choices in reproductive health. T.W.S. has served as a consultant to Regeneron Pharmaceuticals and ScholarRock. Other authors report no conflicts.

References

1. Dennison EM, Sayer AA, Cooper C. Epidemiology of sarcopenia and insight into possible therapeutic targets. *Nat Rev Rheumatol*. 2017;13:340–347. doi:10.1038/nrrheum.2017.60
2. 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. doi:10.1002/acr.20538
3. Reuben DB, Siu AL. An objective measure of physical function of elderly outpatients. The Physical Performance Test. *J Am Geriatr Soc*. 1990;38:1105–1112. doi:10.1111/j.1532-5415.1990.tb01373.x
4. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci*. 2000;55:M221–M231. doi:10.1093/gerona/55.4.m221
5. Startzell JK, Owens DA, Mulfinger LM, Cavanagh PR. Stair negotiation in older people: a review. *J Am Geriatr Soc*. 2000;48:567–580. doi:10.1111/j.1532-5415.2000.tb05006.x
6. Hamel KA, Cavanagh PR. Stair performance in people aged 75 and older. *J Am Geriatr Soc*. 2004;52:563–567. doi:10.1111/j.1532-5415.2004.52162.x
7. Oh-Park M, Wang C, Verghese J. Stair negotiation time in community-dwelling older adults: normative values and association with functional decline. *Arch Phys Med Rehabil*. 2011;92:2006–2011. doi:10.1016/j.apmr.2011.07.193
8. Verghese J, Wang C, Xue X, Holtzer R. Self-reported difficulty in climbing up or down stairs in nondisabled elderly. *Arch Phys Med Rehabil*. 2008;89:100–104. doi:10.1016/j.apmr.2007.08.129
9. Bassey EJ, Fiatarone MA, O'Neill EF, Kelly M, Evans WJ, Lipsitz LA. Leg extensor power and functional performance in very old men and women. *Clin Sci (Lond)*. 1992;82:321–327. doi:10.1042/cs0820321
10. Bean JF, Leveille SG, Kiely DK, Bandinelli S, Guralnik JM, Ferrucci L. A comparison of leg power and leg strength within the InCHIANTI study: which influences mobility more? *J Gerontol A Biol Sci Med Sci*. 2003;58:728–733. doi:10.1093/gerona/58.8.m728
11. Bean JF, Kiely DK, LaRose S, Alian J, Frontera WR. Is stair climb power a clinically relevant measure of leg power impairments in at-risk older adults? *Arch Phys Med Rehabil*. 2007;88:604–609. doi:10.1016/j.apmr.2007.02.004
12. Reid KF, Fielding RA. Skeletal muscle power: a critical determinant of physical functioning in older adults. *Exerc Sport Sci Rev*. 2012;40:4–12. doi:10.1097/JES.0b013e31823b5f13
13. Travison TG, Basaria S, Storer TW, et al. Clinical meaningfulness of the changes in muscle performance and physical function associated with testosterone administration in older men with mobility limitation. *J Gerontol A Biol Sci Med Sci*. 2011;66:1090–1099. doi:10.1093/gerona/glr100
14. Storer TW, Basaria S, Traustadottir T, et al. Effects of testosterone supplementation for 3 years on muscle performance and physical function in older men. *J Clin Endocrinol Metab*. 2017;102:583–593. doi:10.1210/jc.2016-2771
15. Hanson ED, Srivatsan SR, Agrawal S, Menon KS, Delmonico MJ, Wang MQ, et al. Effects of strength training on physical function: influence of power, strength, and body composition. *J Strength Cond Res*.. 2009;23:2627–2637. doi:10.1519/JSC.0b013e3181b2297b
16. Dobs AS, Boccia RV, Croot CC, et al. Effects of enobosarm on muscle wasting and physical function in patients with cancer: a double-blind, randomised controlled phase 2 trial. *Lancet Oncol*. 2013;14:335–345. doi:10.1016/S1470-2045(13)70055-X
17. Dalton JT, Barnette KG, Bohl CE, et al. The selective androgen receptor modulator GTx-024 (enobosarm) improves lean body mass

and physical function in healthy elderly men and postmenopausal women: results of a double-blind, placebo-controlled phase II trial. *J Cachexia Sarcopenia Muscle*. 2011;2:153–161. doi:10.1007/s13539-011-0034-6

18. Basaria S, Coviello AD, Travison TG, et al. Adverse events associated with testosterone administration. *N Engl J Med*. 2010;363:109–122. doi:10.1056/NEJMoa1000485

19. Basaria S, Harman SM, Travison TG, et al. Effects of testosterone administration for 3 years on subclinical atherosclerosis progression in older men with low or low-normal testosterone levels: a randomized clinical trial. *JAMA*. 2015;314:570–581. doi:10.1001/jama.2015.8881

20. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med*. 1995;332:556–561. doi:10.1056/NEJM199503023320902

21. Salameh WA, Redor-Goldman MM, Clarke NJ, Reitz RE, Caulfield MP. Validation of a total testosterone assay using high-turbulence liquid chromatography tandem mass spectrometry: total and free testosterone reference ranges. *Steroids*. 2010;75:169–175. doi:10.1016/j.steroids.2009.11.004

22. LeBrasseur NK, Bhasin S, Miciek R, Storer TW. Tests of muscle strength and physical function: reliability and discrimination of performance in younger and older men and older men with mobility limitations. *J Am Geriatr Soc*. 2008;56:2118–2123. doi:10.1111/j.1532-5415.2008.01953.x

23. Harrell FE. *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag. 2010.

24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310. doi:10.1016/S0140-6736(86)90837-8

25. Ni M, Brown LG, Lawler D, Bean JF. Reliability, validity, and minimal detectable change of four-step stair climb power test in community-dwelling older adults. *Phys Ther*. 2017;97:767–773. doi:10.1093/ptj/pzx039

26. Bhasin S, Apovian CM, Travison TG, et al. Effect of protein intake on lean body mass in functionally limited older men: a randomized clinical trial. *JAMA Intern Med*. 2018;178:530–541. doi:10.1001/jamainternmed.2018.0008

27. Snyder PJ, Bhasin S, Cunningham GR, et al.; Testosterone Trials Investigators. Effects of testosterone treatment in older men. *N Engl J Med*. 2016;374:611–624. doi:10.1056/NEJMoa1506119

28. Papanicolaou DA, Ather SN, Zhu H, et al. A phase IIA randomized, placebo-controlled clinical trial to study the efficacy and safety of the selective androgen receptor modulator (SARM), MK-0773 in female participants with sarcopenia. *J Nutr Health Aging*. 2013;17:533–543. doi:10.1007/s12603-013-0335-x

29. Woods DL, Wyma JM, Yund EW, Herron TJ, Reed B. Factors influencing the latency of simple reaction time. *Front Hum Neurosci.*. 2015;9:131. doi:10.3389/fnhum.2015.00131

30. Yarrow K, Johnson H, Haggard P, Rothwell JC. Consistent chronostasis effects across saccade categories imply a subcortical efferent trigger. *J Cogn Neurosci*. 2004;16:839–847. doi:10.1162/089892904970780