Osteoporos Int (2000) 11:914–922

Osteoporosis International

Original Article

Timed Loaded Standing: A Measure of Combined Trunk and Arm Endurance Suitable for People with Vertebral Osteoporosis

K. M. Shipp[1,2], J. L. Purser[1,2], D. T. Gold[1,3,4], C. F. Pieper[1,5], R. Sloane[1], M. Schenkman[6] and K. W. Lyles[1,7,8,9]

[1]Center for the Study of Aging and Human Development, Duke University; [2]Department of Physical and Occupational Therapy, Duke University Medical Center; [3]Department of Psychiatry and Behavior Sciences, Duke University Medical Center; [4]Department of Sociology, Duke University; [5]Department of Community and Family Medicine, Duke University Medical Center; [6]Program in Physical Therapy, University of Colorado Health Science Center, Denver, Colorado; [7]Department of Medicine, Duke University Medical Center; [8]GRECC, VA Medical Center, Durham, North Carolina; and [9]Sarah W. Stedman Center for Nutritional Studies, Duke University Medical Center, North Carolina, USA

Abstract. Chronic back tiredness or fatigue is a common complaint of people who have a history of osteoporotic vertebral fracture. Trunk muscle endurance has not been studied in people with vertebral osteoporosis, partly due to the lack of assessment tools. We developed a measure of combined trunk and arm endurance suitable for people with vertebral osteoporosis, timed loaded standing (TLS). TLS measures the time a person can stand while holding a two-pound dumbbell in each hand with the arms at 90° of shoulder flexion and the elbows extended. Intraclass correlation coefficients (ICCs) for same day inter-trial and six to ten day test-retest reliability were 0.89 (lower bound 95% confidence interval [LB 95% CI] 0.79) and 0.84 (LB 95% CI 0.68), respectively, in a sample of 21 older women with no known osteoporosis. In 127 women with vertebral fractures, the ICC for same day inter-trial reliability was 0.81 (LB 95% CI 0.75). In a sub-sample of 30 of these women with vertebral fractures, the six to ten day test-retest reliability was 0.85 (LB 95% CI 0.75). Moderately strong and statistically significant ($p \leqslant 0.05$) correlations were found between TLS and sixteen of eighteen measures of physical impairment and function. Functional reach distance, gait velocity, MOS-36 Physical Function Subscale, shoulder flexion strength, and six minute walk distance were most strongly associated with TLS time. Women with vertebral fractures who endorsed having back tiredness when standing and working with the arms in front of the body, sitting to rest because of back tiredness or pain, and planning rest periods because of back tiredness or pain had significantly lower TLS times. TLS is a simple, safe physical performance measure of combined trunk and arm endurance that demonstrates acceptable reliability (inter-trial and test- retest) and concurrent validity.

Keywords: Back injuries; Muscle weakness; Physical endurance; Osteoporosis; Spinal fractures; Spine

Introduction

One well-recognized sequela of osteoporotic vertebral fractures is back pain. Although it is estimated from population-based samples that up to two-thirds of radiographically diagnosed vertebral fractures are asymptomatic [1], people with new vertebral fractures who seek medical attention do so because of pain. A proportion of those who have acute pain after a vertebral fracture proceed to develop chronic pain. To date, the magnitude of this problem has not been well estimated, but there is evidence that those with chronic pain after vertebral fracture also have functional limitations, disability and reduced quality of life [2–7]. Chronic back tiredness or fatigue is another common complaint of people who have a history of osteoporotic vertebral

Correspondence and offprint requests to: Kathy M. Shipp, PT, MHS, Box 3003, Duke University Medical Center, Durham, NC 27710, USA. Tel: +1 (919) 660 7548. Fax: +1 (919) 684 8569. e-mail: kathy@geri.duke.edu

fractures. Activities in which the arms are used in front of the body (e.g., cooking or preparing food, washing or putting away dishes, dressing, folding clothes, ironing) often are mentioned during clinic visits as eliciting either back pain, back tiredness or both.

In our clinical experience, the complaint of back tiredness is not trivial. Rather, back fatigue is likely one factor that contributes to the functional limitations, disability and reduced quality of life which have been described in studies of people with vertebral osteoporosis. Because pain and fatigue are subjectively interpreted, the two sensations are not always well differentiated. One person may label back discomfort as fatigue while another person labels a similar sensation as pain. Although difficult to classify, both back pain and back fatigue are important sequelae of vertebral fractures that are poorly understood, in part because of the lack of assessment tools.

Limited information exists about back fatigue or its physiologic basis, trunk muscle endurance. Muscle endurance has been defined as the ability to produce work or to sustain effort over time [8]. Several tests of trunk muscle endurance (e.g., the Kraus–Weber tests for trunk flexors [8], the Sorensen test for trunk extensors [9–13], multiple repetition testing on isokinetic dynamometers [14], electromyographic analysis [11,15–17], and isometric tests using strain gauge equipment [12,13]) have been used. All these methods have limitations, however, when used with people who have had osteoporotic vertebral fractures. The limitations primarily relate to patient safety and tolerance and to the practicality of administration.

Tests that involve trunk flexion (the Kraus–Weber test [8], isokinetic dynomometers [14]) are contraindicated in osteoporotic patients because flexion increases the loads through the vertebral bodies and, therefore, puts the patient at risk for a new vertebral fracture [18–20]. Because the Sorensen trunk extensor fatigue test is performed in a prone position, it is intolerable for many people with spinal deformity after vertebral fracture. Electromyographic (EMG) analysis of the paraspinal musculature is safe for people with osteoporosis, but the method is technologically intensive and requires a precise, time-consuming protocol by a trained operator. Finally, trunk endurance assessed using a strain gauge device [12,13] requires the patient, first, to produce a reliable maximum voluntary contraction (MVC). For the endurance test, the patient produces a contraction equal to 60% of the MVC for as long as possible. For people with osteoporosis who are frail or have chronic back pain, the MVC may cause discomfort and may not be reliable. Also, this type of testing requires equipment that is not generally available in a clinic.

An ideal measure of trunk muscle endurance for use in people with vertebral osteoporosis would be safe, well tolerated by the frail and oldest-old, and able to capture the entire range of performance capabilities of people who have had osteoporotic vertebral fractures. In addition, an ideal measure would be simple, portable and quick to administer, and would possess good measurement properties. The test also should simulate functional performance of the trunk in routine daily activities, most of which require the trunk to remain erect and stable while the upper extremities are used.

With these criteria in mind, we developed a physical performance task – timed loaded standing (TLS) – a simple measure of combined trunk and arm endurance. The measure and our assessments of reliability (inter-trial and test–retest) and validity (concurrent) are described below.

Materials and Methods

Subjects

Two samples were assembled (Fig. 1). All subjects provided written informed consent, and the study was approved by the Duke University Medical Center Institutional Review Board.

Sample 1: Subjects Without Known Vertebral Fractures. A random sample of 100 women was taken from 2000 potential volunteers listed in the registry of older adults at the Center for the Study of Aging and Human Development at Duke University. Potential subjects

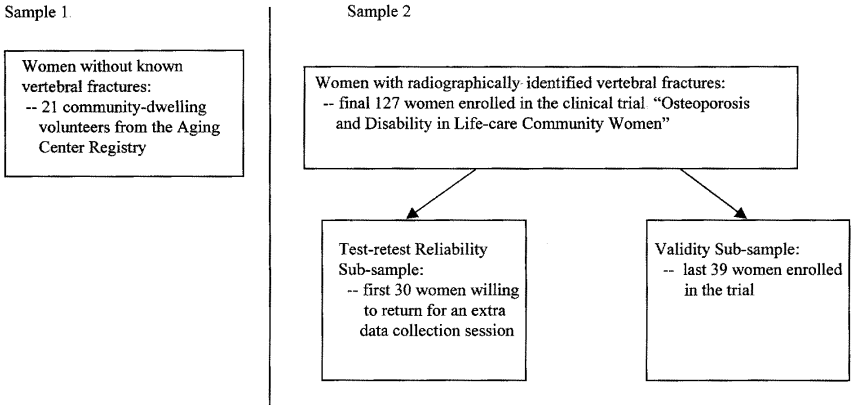

Fig. 1. Study samples.

were telephoned consecutively as they appeared on the random list, with 61% of potential subjects contacted. Twenty-seven of the 61 women contacted were scheduled to be tested at the Duke Aging Center, and 25 women ultimately came for testing. All subjects were independently ambulatory and community dwelling with no active cardiovascular, pulmonary or musculoskeletal impairments that might preclude testing. Usable data were obtained on 21 of the 25 subjects. Two women were excluded by high pre-test blood pressure readings, one subject was excluded by limited shoulder range of motion due to a previous cerebrovascular accident, and one subject could not return for the second session of testing.

No subjects in the Aging Center Registry sample reported having the diagnosis of osteoporosis or history of a vertebral fracture. However, the number of vertebral fractures for this sample was unknown because radiographs were not available. We therefore characterized this sample as having no known osteoporosis, but undiagnosed osteoporosis may have been present.

Reliability data only were collected from the Aging Center Registry sample; the validity measures were not performed by these women.

Sample 2: Subjects with a History of Vertebral Fracture. The second sample came from participants in a continuing randomized clinical trial of ''Osteoporosis and Disability in Life-Care Community Women''. The final 127 subjects recruited for the trial comprised the sample for this study. They lived in six of the nine North Carolina continuing-care retirement communities (CCRCs) enrolled in the trial. Data collection occurred at the CCRCs at the same time as the scheduled testing sessions for the trial. Women residents of the CCRCs were recruited via letters inviting them to a recruitment meeting. Women with previously diagnosed osteoporosis or non-traumatic fractures, height loss of greater than 1 inch (2.5 cm) and/or back pain were especially encouraged to attend. To be eligible for the trial, subjects had to have at least one radiographically verified vertebral fracture. To determine whether prevalent fractures existed, a skeletal radiologist reviewed a series of three radiographs – a lateral thoracic spine breathing technique, an anteroposterior lumbar spine and a lateral lumbar spine – to see whether a vertebral body had a 20% or greater reduction in anterior, central or posterior height. The methods used by the radiologist have been described previously [2]. If a potential subject had at least one vertebral fracture, a physician took a history and performed a physical examination. For admission into the study, subjects met the following inclusion criteria: one or more prevalent vertebral fractures with none occurring in the past 6 months, no hip fracture in the past 12 months, independent ambulation, corrected visual acuity of 20/40 or better, eight or more correct answers on the Short Portable Mental Status Questionnaire [21], no falls in the last 6 months, no current symptoms of active cardiac or pulmonary disease or rheumatoid arthritis, and no current disability from a chronic neurologic disease (i.e., Parkinson's disease, multiple sclerosis, history of cerebrovascular accident with residual deficits).

Table 1. Summary statistics for validity variables in subjects with known vertebral fractures (n = 127)

	Mean ± SD
Physical impairment	
Trunk extension torque (N-m)	28.0 ± 13.1
Thoracic kyphosis (deg)	49.2 ± 14.5
Lumbar lordosis (deg)	−30.0 ± 14.5
Physical activity (score)	64.4 ± 37.3
Pain (% yes)	45%
Functional performance	
Gait velocity (m/s)	1.0 ± 0.2
6 min walk (m)	361.7 ± 104.5
Functional reach (cm)	28.7 ± 5.6
Functional status	
Functional Status Index	
Assistance (score)	1.7 ± 0.5
Difficulty (score)	1.4 ± 0.5
Pain (score)	1.4 ± 0.5
Physical function subscale of MOS-36 (score)	60.1 ± 25.1

All 127 subjects attended one testing session of TLS and performed all the validity measures listed in Table 1. The test–retest reliability subsample was assembled by asking subjects to return in 1 week for an additional testing session. The first 30 subjects, of the 127-person sample with vertebral fractures, who agreed to return for the additional session, formed the test–retest reliability subsample (Fig. 1). Supplemental validity data (shoulder flexion strength, grip strength, and the questions regarding typical back tiredness and pain) were obtained on the final 39 subjects enrolled in the trial. These subjects formed the validity subsample (Fig. 1).

Timed Loaded Standing Protocol

Timed loaded standing (TLS) measures the time a person can stand while holding a 2 lb (1 kg) dumbbell in each hand with the arms at 90° of shoulder flexion, the elbows extended, and the wrists in neutral pronation/supination (Fig. 2). Equipment used included a digital stopwatch, two 2 lb dumbbells, a stethoscope and a sphygmomanometer. The subject's blood pressure and pulse were taken in a seated position prior to the test. If diastolic blood pressure was over 110 mmHg or systolic blood pressure was over 200 mmHg, the subject was excluded. A research assistant first demonstrated the task, then asked the subject to stand erect with the feet hip-width apart and handed the subject the weights. Subjects were instructed to bend the elbows to bring the weights to the shoulders, then to extend the arms and stay in this position for as long as possible. One research assistant stood to the side of the subject and began timing with the stopwatch when the subject extended the arms. A second research assistant stood in front of the subject. The role

Fig. 2. Timed loaded standing.

of the second assistant included cueing the subject to return the arms to the original starting position if the elbows bent or the arms drifted up, down or apart. If the subject could not resume the 90° of shoulder flexion with the elbows fully extended, the test was stopped. Otherwise, the timing stopped only when the subject handed the weights to the second assistant indicating the end of the test. All subjects were monitored for signs of distress and cued to continue breathing, if necessary. At 1 min intervals, the first assistant asked the subject how she was doing. Otherwise, neither the subjects nor the assistants talked during the test. Immediately after the subject released the weights, she sat in a chair and her pulse and blood pressure were obtained.

Reliability Protocol

TLS was performed twice at each session for each subject. After the first trial, the subject continued with collection of other data (e.g., questionnaire completion and height and weight measurement) for approximately 10 min; then the second trial was performed. The same research assistants administered the two trials at a session. After each trial the subject was asked whether fatigue or pain was the reason for stopping the test. In addition, each subject marked a body diagram indicating the location of the fatigue or pain.

To assess test–retest reliability, the subjects returned for another testing session 6–10 days after the first session. For the Aging Registry sample, the same research assistant administered 71% of the retest sessions, while different research assistants administered 29% of the retest sessions. For the subjects with vertebral fractures, a different research assistant administered all retest sessions.

Validity Measures

Measures of physical impairment, functional performance and functional status were correlated with TLS time to determine concurrent validity. Only the sample of women with vertebral fractures was used for assessment of validity. The physical impairment measures were age, height, weight, number of vertebral fractures, sitting isometric trunk extension torque, thoracic kyphosis, lumbar lordosis, physical activity level assessed using the Physical Activity Scale for the Elderly [22] and self-reported presence of pain. For the last 39 women enrolled in the trial (39/127 = 31%) supplemental impairment validity data on shoulder flexion strength and grip strength were collected. Functional performance measures were 6 min walk distance [23] and functional reach [24]. Functional status was assessed by the Functional Status Index [25] and the physical function subscale of the 36-item Medical Outcomes Survey (MOS-36) [26]. All validity measures were obtained within 1 week of the assessment of TLS and were administered by trained personnel following written protocols in a standard order.

Sitting isometric trunk extension peak torque was obtained using the B-200 Isostation (Isotechnologies, Hillsborough, NC), a trunk dynamometer [27]. Using an adapted protocol that has been described previously [2], subjects performed one submaximal and one maximal

practice trial followed by three maximal effort trials of 5 s duration with a 45 s rest between each trial. The mean maximum torque of the three trials was used for the analysis.

The Debrunner kyphometer [28] (Protek, Bern, Switzerland) was used to measure thoracic kyphosis and lumbar lordosis. Midpoints between T2–3, T11–12 and S1–2 were identified via palpation and marked. Subjects stood erect with feet hip-width apart and arms resting by their sides. The number of degrees with the blocks spanning T2–3 and T11–12 for thoracic kyphosis and T11–12 and S1–2 for lumbar lordosis were read directly from the scale.

Shoulder strength was assessed using a hand-held dynamometer (Chatillon, Greensboro, NC). The dominant arm was tested in 90° shoulder flexion and neutral rotation with the subject positioned supine. After a practice trial, two trials of 5 s duration were performed. The maximum force for each trial was converted to foot-pounds, and later newton-meters (N-m), using the length of the arm segment (acromion to 1 inch/2.54 cm above the antecubital crease). The mean of the maximum torques was used for analysis. Grip strength was measured using a Jamar hand dynomometer (Asimow Engineering, Los Angeles, CA). One practice and two test trials were performed with the subject sitting and the arm in 90° of shoulder flexion. The mean of the two trials was used.

The 6 min walk test [23] was administered by instructing subjects to cover as much ground as they could in the allotted time. A trained observer who provided encouragement at 1 min intervals accompanied the subject over a measured indoor course. The distances reached in inches, and later converted to centimeters, for three trials of forward reach in unsupported standing were averaged to obtain the functional reach [24] value.

Finally, the last 39 subjects to enter the trial were asked several single-item questions regarding typical back tiredness and back pain (see Table 6).

Statistical Analysis

Both same-day inter-trial reliability and 6–10 day test–retest reliability were assessed using intraclass correlation coefficients (ICCs). A paired-sample Student's *t*-test was used to detect a difference between the first and second trials of TLS. Spearman's rho was used to correlate TLS with the impairment, functional performance and functional status variables for the concurrent validity analyses. Student's *t*-tests were calculated to determine whether the mean average TLS time was different between those answering yes, compared with those answering no, to the questions regarding typical back tiredness or pain. A chi-square test was used to detect a difference between groups in the proportion of subjects who reported stopping the test due to either back fatigue or back pain. Means and standard deviations were calculated for the cardiovascular responses to the TLS task.

Results

Subject characteristics are shown in Table 2. The subjects with vertebral fractures were, not unexpectedly, older, shorter, weighed less, and had lower mean TLS times than the Aging Center Registry subjects.

Table 2. Subject characteristics

	Subjects without known vertebral fractures fractures (n = 21)	Subjects with known vertebral fractures (n = 127)	Wilcoxon Rank Sum Test, p value
Age (years)	75.4 ± 4.2	80.9 ± 5.6	0.0001
Height (cm)	159.8 ± 6.1	156.2 ± 6.1	0.04
Weight (kg)	67.1 ± 10.7	61.4 ± 10.8	0.02
Vertebral fractures (no.)	Unknown	2.3 ± 1.6	
Range (no.)	Unknown	1–9	
Timed loaded standing (s)			
Trial 1			
Mean	116.2 ± 59.6	74.7 ± 63.9	0.001
Median	98.5	55.8	
Interquartile range	74.1–144.3	25.9–102	
Range	28.6–233.6	1.4–299	
Trial 2			
Mean	105.0 ± 51.3	58.5 ± 51.8	0.0001
Median	94.4	40.7	
Interquartile range	62.7–125.4	21.6–82	
Range	23.0–229.8	3.0–293.0	
Mean of trials 1 and 2	110.6 ± 54.4	67.6 ± 56.7	0.0004
Median	110.8	50.3	
Interquartile range	70.9–129.2	23.1–90	
Range	25.7–231.7	2.7–295.0	

All data except vertebral fracture range and timed loaded standing range, median and interquartile range are the mean ± SD. n = number per group.

Table 1 contains summary statistics for the validity variables.

Reliability

Table 3 contains the inter-trial reliability results. Although the ICCs were within the range expected for a physical performance test, we noted a consistent decrement for TLS time at trial 2. The mean difference between trial 1 and trial 2 was statistically significant in both the Aging Center Registry group and the women with vertebral fractures.

Six to ten day test–retest reliability was assessed for the full sample of women without known vertebral fractures (n = 21) and for a 30 subject subsample of women with vertebral fractures (30/127 = 24% of the sample). This subsample did not differ in baseline characteristics from the remaining osteoporosis sample (data not shown). The test–retest reliability results are given in Table 4. The correlation coefficients for the sample of women with vertebral fractures reflect 6–10 day test–retest reliability as well as inter-rater reliability.

Table 3. Same day, intra-rater, inter-trial reliability of timed loaded standing

	ICC[a]	Difference trial 2 − trial 1 (s)	p value[b]
Subjects without known vertebral fractures (n = 21)	0.89 (0.79)	11.2 ± 23.6	0.04
Subjects with vertebral fractures (n = 127)	0.81 (0.75)	14.6 ± 33.6	0.0001

Difference data are expressed as the mean ± SD. n = number per group.
[a] Intraclass correlation coefficient (lower bound 95% confidence interval).
[b] Paired Student's t-test for $H_{0\ \text{Difference}} = 0$.

Table 4. Test–retest (6–10 day) reliability of timed loaded standing

	ICC	
	Mean of both trials	Trial 1 only
Subjects without known vertebral fractures (n = 21)	0.84 (0.69)[a]	0.82 (0.66)[a]
Subjects with vertebral fractures (n = 30)	0.85 (0.75)[b]	0.78 (0.63)[b]

All data are intraclass correlation coefficients (lower bound 95% confidence interval).
[a] Combination of the same and different testers at the second session.
[b] Different tester at the second session (inter-rater reliability).

Validity

The results of the concurrent validity study comparing TLS with physical impairment, functional performance and functional status are shown in Table 5. These correlations were moderately strong and statistically significant at the $p \leqslant 0.05$ level except for thoracic kyphosis and weight. In addition, in the final 39 subjects with vertebral fractures entering the osteoporosis and

Table 5. Concurrent validity: correlations between timed loaded standing and physical impairment, functional performance and functional status in subjects with vertebral fractures (n = 127)

	Spearman's rho correlation coefficient using TLS mean of trials 1 and 2	p value	Spearman's rho correlation coefficient using TLS trial 1 only	p value
Physical impairment				
Age	−0.38	0.0001	−0.38	0.0001
Height	0.19	0.03	0.20	0.03
Weight	−0.01	0.90	0.01	0.93
No. of vertebral fractures	−0.25	0.005	−0.25	0.005
Trunk extension torque	0.34	0.0001	0.32	0.0003
Thoracic kyphosis	−0.14	0.11	−0.14	0.13
Lumbar lordosis	−0.32	0.0002	−0.31	0.0005
Physical activity	0.29	0.001	0.33	0.0002
Pain	0.24	0.007	0.24	0.007
Functional performance				
Gait velocity	0.52	0.0001	0.49	0.0001
6-min walk distance	0.47	0.0001	0.46	0.0001
Functional reach distance	0.52	0.0001	0.50	0.0001
Functional status				
Functional Status Index				
Assistance	−0.38	0.0001	−0.35	0.0001
Difficulty	−0.36	0.0001	−0.36	0.0001
Pain	−0.32	0.0002	−0.32	0.0002
MOS-36 Physical Function subscale	0.49	0.0001	0.48	0.0001

Table 6. Differences in mean of trials 1 and 2 timed loaded standing time between subjects responding yes, versus no, to questions regarding typical back tiredness and pain (n = 39)

Question	% yes	Mean ± SD TLS time (s) if responded yes	% no	Mean ± SD TLS time (s) if responded no	p value for difference between TLS time, t-tests
Do you have **back tiredness** when standing and working with your arms in front of your body (during activities such as cooking or preparing food, washing or putting away dishes, ironing)?	77	48.9 ± 29.5	23	80.6 ± 47.5	0.02
Do you have **back pain** when standing and working with your arms in front of your body (during activities such as cooking or preparing food, washing or putting away dishes, ironing)?	67	48.4 ± 31.4	33	71.7 ± 41.8	0.06
During a typical day in the last month, did you **sit** down to rest because of back tiredness or back pain?	74	47.6 ± 30.0	26	81.2 ± 43.1	0.01
During a typical day in the last month, did you **lie** down to rest because of back tiredness or back pain?	56	48.7 ± 31.5	44	65.9 ± 40.7	0.14
Do you plan **rest periods** during the day to prevent back tiredness or back pain?	51	42.2 ± 31.8	49	71.0 ± 35.8	0.01
Are you more **stooped** or unable to stand erectly at the end of the day, compared with the beginning of the day?	56	47.2 ± 29.8	44	67.8 ± 41.5	0.08

disability clinical trial (39/127 = 31% of the sample), we assessed shoulder flexion and grip strength. The baseline characteristics of these subjects did not differ from those of the remaining osteoporosis sample (data not shown). Mean ± SD shoulder flexion torque was 27.7 ± 6.3 N-m. Mean ± SD grip strength was 17.8 ± 4.3 kg. The Spearman's rho correlation coefficients with mean TLS time were 0.48 (p=0.003) for shoulder flexion torque and 0.37 (p=0.02) for grip strength.

Table 6 shows the comparison between subjects responding yes versus no to the questions regarding typical back tiredness or pain. For every question, the mean TLS time was lower in those responding yes, but there was a statistically significant ($p \leqslant 0.05$) difference only for back tiredness when standing and working with the arms in front of the body, sitting to rest because of back tiredness or pain, and planning rest periods during the day to prevent back tiredness or pain.

Response to the Timed Loaded Standing Task

Based on responses immediately after each trial of TLS, 84% of subjects with vertebral fractures and 81% of the Aging Center Registry subjects reported stopping the test because of fatigue. Sixteen percent of the subjects with vertebral fractures and 19% of the Aging Center Registry subjects reported stopping the test because of pain. Forty-two percent of the subjects with vertebral fractures, compared with 9.5% of the Aging Center Registry subjects, reported the location of fatigue or pain to be in the back instead of the arms. The percentages associated with location of discomfort in the back were significantly different between the samples (p = 0.0002).

To address issues of cardiovascular safety with the TLS task, we examined the cardiovascular responses to the task with the two samples combined. At the end of TLS, the mean ± SD heart rate increase was 5.3 ± 8.9 beats/min while systolic blood pressure rose 10.6 ± 2.8 mmHg and diastolic blood pressure rose 4.5 ± 7.8 mmHg. The maximum increase for heart rate was 33 beats/min. Maximum increases for blood pressure were systolic 46 mmHg and diastolic 30 mmHg.

Discussion

Timed loaded standing, the physical performance measure of combined trunk and arm endurance described here, exhibited acceptable reliability and concurrent validity. The intraclass correlation coefficients were satisfactory for both inter-trial and test–retest reliability. Our results are similar to the reliability results reported for the Sorensen trunk endurance test [11,12,29], a test similar to TLS in that subjects sustain a submaximal

effort for as long as possible. Our reliability findings were substantially higher than those reported for an isokinetic endurance test of the trunk extensors [14]. In addition, the reliability estimates for TLS in our samples were similar to those of other physical performance measures in which the total body is used to perform a task (e.g., functional reach [24]).

The reproducibility of TLS performance is adequate for use in research and clinical settings. Although the inter-trial reliability was acceptable, there was a decrement in performance on the second trial compared with the first trial. We expect that the 10 min rest between trials was not sufficient for complete physiologic recovery. In addition, boredom or motivational factors may account for the decrement. Because the test–retest reliability using trial 1 only is nearly as good ($r = 0.78$) as the test–retest reliability using the means of the two trials ($r = 0.85$), and the correlations between TLS time and physical impairment and function were nearly identical whether trial 1 only, or the mean of trials 1 and 2, were used in the analyses (Table 5), it would be possible to have subjects perform only one trial. This would reduce the burden for the subject and save time in the clinic or in research protocols.

Assessing the validity of TLS is problematic because no gold standard for assessment of combined trunk and arm endurance exists. In the absence of a gold standard, we assessed the concurrent validity of TLS by examining the association between TLS and measures of physical impairment, functional performance and functional status. Evidence of the concurrent validity of TLS was provided by the moderately strong and statistically significant correlations between TLS and 14 of the 16 measures in the full sample and both supplemental impairment measures (shoulder flexion and grip strength) in the subsample. The strongest associations were between TLS and shoulder flexion torque and the measures of function: gait velocity, 6 min walk distance, functional reach distance and MOS-36 Physical Function Subscale. We were not surprised that TLS did not correlate highly with any single validity variable. Because no measure of combined trunk and arm endurance exists, these variables are components of the TLS task but no single variable comprehensively describes the TLS task.

The very low correlation between thoracic kyphosis and TLS score was surprising. We expected that greater kyphosis would be associated with lower TLS scores. None of our work to date offers an explanation for this finding. Future work will explore this lack of association.

The mean TLS time was lower in those responding yes to the questions regarding typical back tiredness or pain, although the differences were not statistically significant in every case. Given the small number of subjects to whom the questions were administered ($n = 39$), we believe that overall the responses to these questions support the concurrent validity of TLS.

In addition to demonstrating good reliability and validity, TLS has other desirable properties. The task is safe and well tolerated by the frail and oldest-old. To date in our laboratory, approximately 250 older women with vertebral fractures have performed over 1000 trials of TLS with only one adverse event: one subject sustained a metatarsal fracture when the research assistant dropped one of the dumbbells on the subject's foot. Since the test is self-limited, even the most frail subjects were willing to attempt the test. The cardiovascular demands are minimal. However, a large enough rise in blood pressure is possible that we recommend taking a baseline blood pressure reading prior to the test and not performing the test in persons with unacceptably high baseline blood pressures. When performed according to our protocol, the biomechanical demands of the task should not present a risk for people with established vertebral osteoporosis.

The TLS task can be administered by trained non-professional personnel in almost any setting using readily available equipment: a pair of 2 lb (1 kg) dumbbells and a stopwatch. The measure, therefore, is portable and inexpensive to administer. Although our protocol used two testers, the test can be administered by one tester performing both roles: timing the task and standing in front of the subject to monitor arm movement and to take the dumbbells at the end of the test (see Fig. 2). We have successfully used only one tester for measuring TLS in our clinic for the past 3 years.

There are several limitations to this study. Although we expect that TLS may be a good measure for all people with vertebral osteoporosis, the data presented here are for older women only. The measure will need to be tested in younger female subjects and in men. Also, the vertebral fracture subjects all lived in CCRCs. Additional work is needed to assess the measure in more general populations. Another limitation of this study is that, because the data were cross-sectional, we were unable to evaluate either the sensitivity to change of TLS or the predictive validity of the measure. Future investigation will have to address these issues.

In summary, TLS is a physical performance measure of combined trunk and arm endurance that demonstrates acceptable reliability and concurrent validity. It is a simple, safe measure of endurance that is well tolerated by older women with vertebral osteoporosis. Given its ease of administration and acceptable measurement properties, we believe that TLS has promise as an outcome measure in clinical practice and in clinical trials.

Acknowledgements. We gratefully thank the members of Measurement Core of the Duke Claude D. Pepper Older Americans Independence Center for their assistance with data collection. We acknowledge the funding support of the National Institute of Aging through the Duke Pepper Center (AG-11268), the National Institute of Child Health and Human Development (HD 30442), the Foundation for Physical Therapy, and the Department of Veteran's Affairs Medical Research Service.

Portions of this work were presented at the September 1996 meeting of the American Society for Bone and Mineral Research, Seattle, Washington, USA.

References

1. Cooper C, Melton LJ. Vertebral fractures: how large is the silent epidemic? BMJ 1992;304:793–4.
2. Lyles KW, Gold DT, Shipp KM, et al. Association of osteoporotic vertebral compression fractures with impaired functional status. Am J Med 1993;94:595–601.
3. Gold DT. The clinical impact of vertebral fractures: quality of life in women with osteoporosis. Bone 1996;18:S185–9.
4. Leidig G, Minne HW, Sauer P, et al. A study of complaints and their relation to vertebral destruction in patients with osteoporosis. Bone Miner 1990;8:217–10.
5. Cook DJ, Guyatt GH, Adachi JD, et al. Quality of life issues in women with vertebral fractures due to osteoporosis. Arthritis Rheum 1993;36:750–6.
6. Nevitt MC, Ettinger B, Black DM, et al. The association of radiographically detected vertebral fractures with back pain and function: a prospective study. Ann Intern Med 1998;128:793–800.
7. Ross PD, Ettinger B, Davis JW, et al. Evaluation of adverse health outcomes associated with vertebral fractures. Osteoporos Int 1991;1:134–40.
8. Ito T, Shirado O, Suzuki H, Takahashi M, Kaneda K, Strax TE. Lumbar trunk muscle endurance testing: an inexpensive alternative to a machine for evaluation. Arch Phys Med Rehabil 1996;77:75–9.
9. Biering-Sorensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. Spine 1984;9:106–17.
10. Nordin M, Kahanovitz N, Verderame R, et al. Normal trunk muscle strength and endurance in women and the effect of exercises and electrical stimulation. 1. Normal endurance and trunk muscle strength in 101 women. Spine 1987;12:105–11.
11. Moffroid M, Reid S, Henry SM, Haugh LD, Ricamato A. Some endurance measures in persons with chronic low back pain. J Orthop Sports Phys Ther 1994;20:81–7.
12. Holstrom E, Moritz U, Andersson M. Trunk muscle strength and back muscle endurance in construction workers with and without low back disorders. Scand J Rehab Med 1992;24:3–10.
13. Nicolaisen T, Jorgensen K. Trunk strength, back muscle endurance and low-back trouble. Scand J Rehab Med 1985;17:121–7.
14. Mayer T, Gatchel R, Betancur J, et al. Trunk muscle endurance measurement: isometric contrasted to isokinetic testing in normal subjects. Spine 1995;20:920–7.
15. Biedermann HJ, Shanks GL, Inglis J. Median frequency estimates of paraspinal muscles: reliability analysis. Electromyogr Clin Neurophysiol 1990;30:83–8.
16. Ng JKF, Richardson CA. Reliability of electromyographic power spectral analysis of back muscle endurance in healthy subjects. Arch Phys Med Rehabil 1996;77:259–64.
17. Krondraske GV, Deivanayagam S, Carmichael T, et al. Myoelectric spectral analysis and strategies for quantifying trunk muscular fatigue. Arch Phys Med Rehabil 1987;68:103–10.
18. Sinaki M, Mikkelsen BA. Postmenopausal spinal osteoporosis: flexion versus extension exercises. Arch Phys Med Rehabil 1984;65:593–6.
19. Melton LJ, Chao EYS, Lane J. Biomechanical aspects of fractures. In: Riggs BL, Melton LJ, eds. Osteoporosis: etiology, diagnosis, and management. New York: Raven Press, 1988:111–31.
20. Schultz A, Andersson GBJ, Ortengren R, Bjork R, Nordin M. Analysis and quantitative myoelectric measurements of loads on the lumbar spine when holding weights in standing postures. Spine 1982;7:390–7.
21. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficits in elderly patients. J Am Geriatr Soc 1975;23:433–41.
22. Washburn RA, Smith KW, Jette AM, et al. The physical activity scale for the elderly (PASE): development and evaluation. J Clin Epidemiol 1993;46:153–62.
23. Guyatt G, Sullivan MJ, Thompson PJ. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. J Can Med Assoc 1985;132:919–23.
24. Duncan PW, Weiner DK, Chandler JM, et al. Functional reach: a new clinical measure of balance. J Gerontol 1990;45:M192–7.
25. Jette AM. Functional Status Index: reliability of a chronic disease evaluation instrument. Arch Phys Med Rehabil 1980;61:395–401.
26. Stewart AL, Hays RD, Ware JE. The MOS short form general health survey. Med Care 1988;26:724–35.
27. Levene JA, Seeds RH, Goldberg HM, Frazier M, Fuhrman GA. Trends in isodynamic and isometric trunk testing on the Isostation B-200. J Spinal Disord 1989;2:20–35.
28. Ohlen G, Spangefort E, Tingvall C. Measurement of spinal sagittal configuration and mobility with Debrunner's kyphometer. Spine 1989;14:580–2.
29. Jorgensen K, Nicolaisen T. Two methods for determining trunk extensor endurance. Eur J Appl Physiol 1986;44:639–44.

Received for publication 27 October 1999
Accepted in revised form 4 April 2000