Journal of Gerontology: MEDICAL SCIENCES
2002, Vol. 57A, No. 5, M289–M293

Change in Physical Performance Over Time in Older Women: The Women's Health and Aging Study

Graziano Onder,[1,2] Brenda W.J.H. Penninx,[1] Pablo Lapuerta,[3] Linda P. Fried,[4] Glenn V. Ostir,[5] Jack M. Guralnik,[5] and Marco Pahor[1]

[1]Sticht Center on Aging, Section on Gerontology and Geriatrics, Department of Internal Medicine, Wake Forest University, School of Medicine, Winston-Salem, North Carolina.
[2]Department of Gerontology and Geriatrics, Catholic University of the Sacred Heart, Rome, Italy.
[3]Pharmaceutical Research Institute, Bristol-Myers Squibb, Princeton, New Jersey.
[4]Departments of Medicine and Epidemiology, The Johns Hopkins Medical Institutions, Baltimore, Maryland.
[5]Epidemiology, Demography, and Biometry Program, National Institute on Aging, Bethesda, Maryland.

Background. Although lower and upper extremity performance measures are widely used and represent validated physical function measures in older adults, there is limited information regarding the magnitude of changes in these measures over time. This study (i) assesses prospective changes in physical performance measures, (ii) defines a summary score that demonstrates a significant amount of change over time, and (iii) examines rates of decline according to age and baseline performance levels.

Methods. Data from the Women's Health and Aging Study (WHAS) were analyzed to assess change in the one third most disabled older women living in the community. Lower extremity function was assessed using walking speed, balance, and chair stands tests. The putting-on-blouse test, the lock and key test, the Purdue Pegboard test, and grip strength were used to gauge upper extremity function. Continuous and categorical summary performance scores were calculated using continuous and categorical data of lower and upper performance measures.

Results. After 3 years, lower extremity performance measures declined by 16%–27%, while upper extremity performance measures declined less (7%–24%). For lower extremity function, the continuous summary performance score showed a slightly greater 3-year decline from baseline (decline vs baseline mean: 23%; decline vs *SD* of the baseline mean: 59%) than the categorical score (22% and 41%, respectively). Older age and intermediate level of baseline performance were associated with the greatest decline, especially for lower extremity function.

Conclusions. In moderately to severely disabled women aged 65 or older, lower extremity measures show more change over 3 years than upper extremity measures. Among the lower extremity summary scores, the continuous score changes more over time than the categorical score with respect to the baseline *SD*. The lower extremity continuous summary performance score may be a useful outcome measure for clinical studies of physical performance in older women.

THE development of standardized physical performance tests has provided a valuable tool for the assessment of the ability to perform tasks required to accomplish common daily activities (1–4). Summary scores based on these tests have the potential to assess performance abilities along the full spectrum of functioning and represent ideal outcomes for studies of physical function (5–9).

However, there is little information regarding the magnitude of change in these measures over time. Such data would help predict rates of change in performance measures (in observational studies) and calculate the effect size (in intervention studies). Our aims are to assess (i) changes in lower extremity (LE) and upper extremity (UE) physical performance measures, (ii) a summary score that demonstrates a significant amount of change over time, and (iii) the rate of decline according to age and baseline performance levels.

Methods

We utilized data from the Women's Health and Aging Study (WHAS), a 3-year longitudinal study enrolling 1002 subjects presenting difficulty in at least two of four functional domains (mobility and exercise tolerance, upper extremity function, basic self-care, and higher functioning tasks of independent living) and scoring >17 on the Mini-Mental State Examination (MMSE) (10). Details of the methods and characteristics of the population are reported elsewhere (11,12).

Individual Measures

LE function was assessed using walking speed (faster of two walks), the chair stands test, and the balance test. UE function was assessed using the putting-on-blouse test, the lock and key test, the Purdue Pegboard test, and the grip strength of the dominant hand (best of three trials).

To calculate the decline over time in walking speed and grip strength, a value corresponding to the 1st percentile of baseline performance of participants completing the task was assigned to participants who were unable to perform the task or who had a performance below the 1st percentile (walking speed: 9 cm/sec; grip strength: 5 kg). Similarly, for other tasks, with the exception of the balance test, a

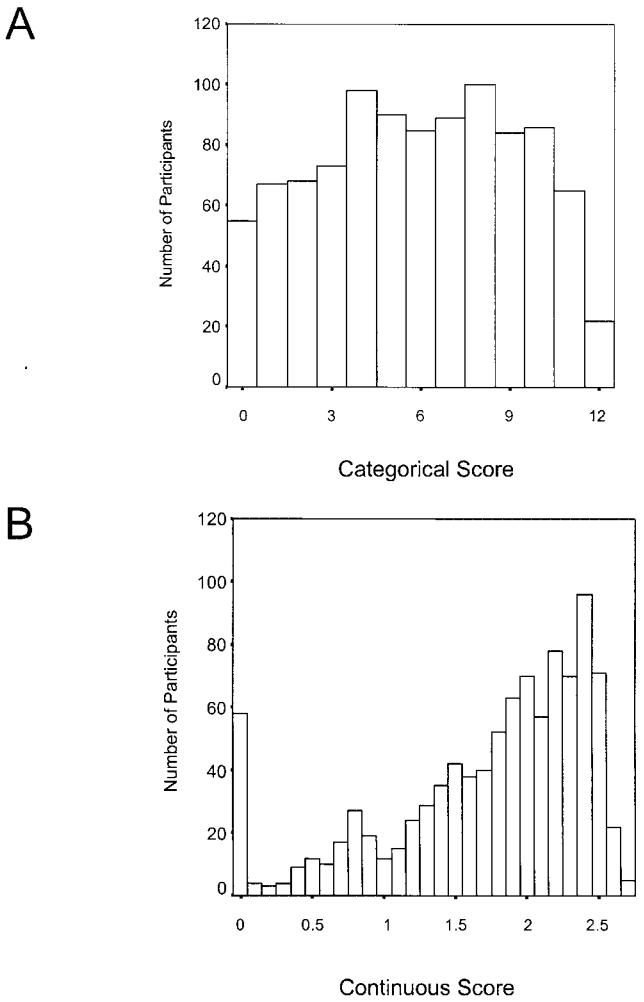

Figure 1. Distributions of the baseline lower extremity categorical (**A**) and continuous (**B**) summary performance scores.

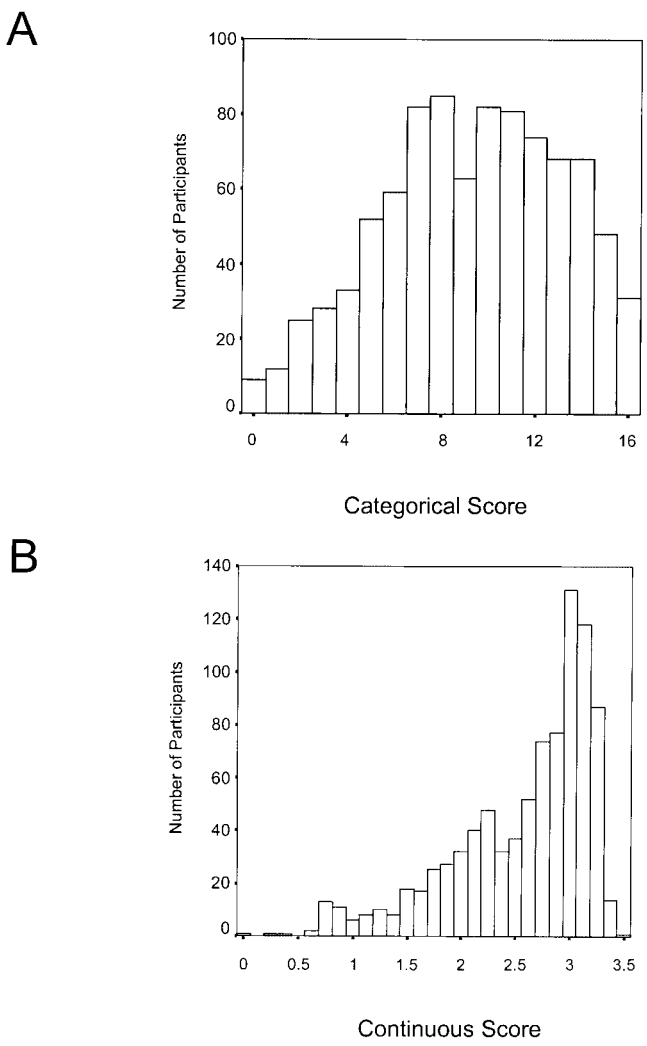

Figure 2. Distributions of the baseline upper extremity categorical (**A**) and continuous (**B**) summary performance scores.

value corresponding to the 99th percentile of baseline performance of participants completing the task was assigned to participants who were unable to perform the task or who had a performance above the 99th percentile (chair stands: 32.1 s; putting-on-blouse test: 233 s; lock and key test: 52.9 s; Purdue Pegboard test: 58.3 s).

Continuous Summary Performance Scores

After assigning arbitrary values as described above to worst performers and subjects unable to complete each task, individual measures were rescaled applying the following formulas (higher scores signify better performance):

(i) Walking speed: 1 − (9/speed in cm/s).
(ii) Chair stands test: 1 − (time in s/32.1).
(iii) Standing balance test: (time in s/30).
(iv) Putting-on-blouse test: 1 − (time in s/233).
(v) Lock and key test: 1 − (time in s/52.9).
(vi) Purdue Pegboard test: 1 − (time in s/58.3).
(vii) Grip strength test: 1 − (5/grip strength in kg).

Continuous summary performance scores for LE (baseline range 0–2.71) and UE (baseline range 0–3.49) were calculated by adding the rescaled scores for lower and upper tests.

Categorical Summary Performance Scores

To calculate a categorical score for the three LE measures, we used cut points derived from the Established Populations for Epidemiologic Studies of the Elderly (13) to construct separate 0 (unable to do test) to 4 (best performance) scales and one 0 to 12 summary score. Similarly, for the UE measures, 0 was assigned to those unable to do the test, and others received a score between 1 (worst performance) and 4 (best performance), based on quartiles of performance. The following cut-offs were used:

Putting-on-blouse test (s)	Lock and key test (s)
1. >108.7	1. >12.9
2. 78.8–108.7	2. 7.3–12.9
3. 50.6–78.7	3. 4.8–7.2
4. <50.6	4. <4.8

Table 1. One- and Three-Year Decline in Lower and Upper Extremity Function*

	Baseline Scores		1-Year Change				3-Year Change			
Extremity Function	*n*	Mean (*SD*)	*n*	Mean (*SD*)	% Change vs Baseline Mean	% Change vs Baseline *SD***	*n*	Mean (*SD*)	% Change vs Baseline Mean	% Change vs Baseline *SD***
Lower										
Balance (s)	1002	18.1 (10.3)	821	1.4 (9.6)	7.4	13.3	675	5.3 (10.6)	26.6	51.1
Chair stands (s)	998	20.1 (8.5)	815	2.2 (7.7)	11.2	25.4	666	4.0 (7.9)	21.1	47.5
Walking speed (cm/s)	987	61.2 (30.8)	811	3.2 (26.6)	5.2	10.4	693	10.3 (31.6)	16.3	33.6
Lower extremities categorical summary performance score	982	5.9 (3.3)	773	0.4 (2.4)	6.6	12.4	611	1.4 (2.8)	21.9	41.0
Lower extremities continuous summary performance	982	1.71 (0.71)	773	0.15 (0.56)	8.8	21.1	611	0.42 (0.69)	23.3	59.2
Upper										
Put on and button blouse (s)	975	115.6 (72.4)	771	4.3 (65.0)	3.9	5.9	626	24.3 (74.8)	23.9	33.5
Lock and key test (s)	994	13.9 (14.2)	782	−1.1 (14.3)	−8.4	−7.9	615	0.8 (15.2)	6.8	5.5
Purdue Pegboard (s)	997	32.9 (10.7)	820	2.4 (8.9)	7.4	22.6	675	4.5 (11.0)	14.5	42.0
Grip strength (kg)	930	19.7 (5.9)	702	0.2 (4.3)	1.0	4.0	518	1.4 (4.3)	6.8	24.5
Upper extremities categorical summary performance score	891	9.4 (3.8)	655	0.02 (2.5)	0.2	0.5	486	0.9 (2.7)	8.7	22.6
Upper extremities continuous summary performance	891	2.57 (0.65)	655	0.02 (0.45)	0.8	2.6	486	0.13 (0.50)	4.8	19.8

*Negative values signify improvement.

**Percent change vs standard deviation (*SD*) of the mean was calculated with the following formula: 100 × mean change/*SD* of mean at baseline. *SD* of all baseline participants was used for these analyses.

Hand grip strength (kg)	Purdue Pegboard test (s)
1. <17	1. >34.4
2. 17–20	2. 28.7–34.4
3. 21–24	3. 25.0–28.6
4. 24	4. < 25.0

A 0 to 16 summary score was calculated by adding up the four test scores. The baseline distributions of LE and UE scores are reported in Figures 1 and 2.

Data Analyses

We examined the average decline/year in performance measures after stratification by age and baseline performance subgroups, using mixed model analysis of covariance (SAS Version 6.12, SAS Institute, Cary, NC). We used random intercept and random slope in a growth curve model. Analyses were adjusted for baseline value of the outcome variable.

Table 2. Average Annual Decline in Lower Extremity Function Among 927 Participants With Baseline Data for All Three Lower Extremity Tests and Their Summary Scores*

		Baseline Lower Extremity Performance				
Annual Decline In	All *N* = 927	Good Performers *n* = 168	Intermediate Performers *n* = 444	Poor Performers *n* = 315	*p* Intermediate vs Good	*p* Intermediate vs Poor
Young-old (65–79 years) (*n* = 540)						
Balance test (s)	1.6	1.7	1.8	1.1	.575	.049
Chair stands (s)	1.5	1.6	2.0	0.3	.132	<.001
Usual walking speed (cm/s)	3.4	2.7	2.9	5.3	.849	.055
Lower extremities categorical summary performance score	0.4	0.4	0.5	0.3	.600	.411
Lower extremities continuous summary performance	0.12	0.11	0.13	0.11	.351	.322
Old-old (≥ 80 years) (*n* = 387)						
Balance test (s)	2.1**	2.9	2.8	1.4	.892	<.001
Chair stands (s)	2.2†	2.2	3.4	1.0	.154	<.001
Usual walking speed (cm/s)	6.2‡	6.3	6.5	6.2	.665	.617
Lower extremities categorical summary performance score	0.6‡	0.9	0.9	0.3	.500	<.001
Lower extremities continuous summary performance	0.20‡	0.19	0.25	0.16	.294	.001

Note: Good performers = categorical lower extremity score 10–12; Intermediate performers = categorical lower extremity score 4–9 and able to perform all lower extremity tests; Poor performers = categorical lower extremity score 0–3 or unable to perform one or more lower extremity tests.

*All the analyses are adjusted for baseline performance score.

***p* vs decline in young-old participants = .02.

†*p* vs decline in young-old participants = .002.

‡*p* vs decline in young-old participants < .001.

Table 3. Average Annual Decline in Upper Extremity Function Among 842 Participants With Baseline Data for All Four Upper Extremity Tests and Their Summary Scores*

Annual Decline In	All $N = 842$	Baseline Upper Extremity Performance: Good performers $n = 209$	Intermediate performers $n = 407$	Poor performers $n = 226$	p Intermediate vs good	p Intermediate vs poor
Young-old (65–79 years) ($n = 495$)						
Put on and button blouse (s)	6.9	9.1	12.0	−11.2	.242	<.001
Lock and key test (s)	0.1	0.1	−0.1	0.5	.691	.530
Purdue Pegboard (s)	1.6	1.3	1.8	1.3	.173	.350
Grip strength (kg)	0.5	0.4	0.5	0.5	.816	.625
Upper extremities categorical summary performance score	0.1	0.3	0.1	−0.2	.024	.135
Upper extremities continuous summary performance	0.03	0.04	0.05	−0.07	.131	.040
Old-old (≥ 80 years) ($n = 347$)						
Put on and button blouse (s)	13.0**	12.0	18.2	−4.6	.364	<.001
Lock and key test (s)	1.5†	2.1	2.0	0.9	.960	.641
Purdue Pegboard (s)	2.5‡	2.9	2.8	2.4	.820	.403
Grip strength (kg)	0.6	0.6	0.7	0.6	.791	.474
Upper extremities categorical summary performance score	0.5‡	1.3	0.6	0.0	.029	<.001
Upper extremities continuous summary performance	0.09‡	0.11	0.13	0.01	.102	<.001

Note: Good performers = categorical upper extremity score 12–16; Intermediate performers = categorical upper extremity score 5–11 and able to perform all upper extremity tests; Poor performers = categorical upper extremity score 0–4 or unable to perform one or more upper extremity tests.

*All the analyses are adjusted for baseline performance score. Negative values signify improvement.

**p vs decline in young-old participants = .04.

†p vs decline in young-old participants = .008.

‡p vs decline in young-old participants < .001.

Fifty-eight subjects who missed follow-up assessments were excluded from these analyses. These women were significantly older and presented a worse baseline performance compared with other participants.

Results

The mean age of the 1002 participants was 78.9 ± 8.1 years, 28.3% were black, and at baseline, 31.5% reported a lot of difficulty or were unable to perform one or more activities of daily living. The decline in the LE measures (expressed as percent change from baseline mean) ranged from 5.2% to 11.2% after 1 year and from 16.3% to 26.6% after 3 years (Table 1). The lock and key test was the only UE task presenting a 1-year improvement from baseline. The 1- and 3-year declines from baseline mean of the other UE measures ranged from 1.0% to 7.4% and from 6.8% to 23.9%. For both LE and UE, the magnitude of the decline expressed as percent change versus the baseline mean in categorical and continuous scores was similar. However, for LE performance, the average 1- and 3-year declines of the categorical summary score, expressed as percent change versus the baseline standard deviation of the mean, were substantially lower (12.4% and 41.0%) than those of the continuous summary score (21.1% and 59.2%).

Participants older than 80 years experienced greater decline in all performance measures and summary scores than women younger than 80 years (Tables 2 and 3). These results were virtually unchanged after adjustment for MMSE score. Regarding LE performance, participants from both age groups with an intermediate level of baseline performance were more likely to decline than poor performers. The only exception was walking speed, which, in the group older than 80, presented a higher decline among poor performers than among both good and intermediate performers.

Discussion

Compared with healthier populations, we described larger changes in LE measures (3,14,15), probably because the WHAS participants are all disabled and, therefore, have a higher risk of declining in function (16). The decline in UE tests was not linear, in particular for the lock and key test and the put-on-blouse test, probably because these two tests have a lower test-retest reliability than other measures (17). For this reason, and in consideration of their ability to predict incident disability (6,13,18,19), LE measures seem preferable outcomes for studies that examine prospective changes in physical function. More specifically, the continuous summary score of LE performance, which showed a larger decline from baseline *SD* of the mean than other tests, may represent a useful outcome measure for clinical studies of physical function.

Participants with intermediate baseline levels of performance were more likely to decline in LE measures and scores than poor performers. One possible explanation for this finding is that intermediate performers may have preclinical disabilities that will eventually trigger more precipitous declines in function. Alternatively, a floor effect may account for this observation, given that the poor performance group includes participants unable to perform the task, who could not further worsen.

We provide estimates of decline in physical performance measures over time. These findings are important for calculating sample sizes for studies that prospectively evaluate change in physical function in older adults. Screening participants based on physical performance and

age can identify those at greatest risk for physical performance decline.

Acknowledgments

The Women's Health and Aging Study was supported by the National Institute on Aging (NIA; Contract No. N01AG12112). The work of Dr. Onder is supported by the Wake Forest Claude D. Pepper Older Americans Independence Center (NIA Grant 5P60 AG 10484-07).

Address correspondence to Graziano Onder, MD, Section on Gerontology and Geriatrics—Sticht Center on Aging, Wake Forest University—Baptist Medical Center, Medical Center Boulevard, Winston-Salem, NC 27157. E-mail: graziano_onder@rm.unicatt.it

References

1. 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.
2. Daltroy LH, Phillips CB, Eaton HM, et al. Objectively measuring physical ability in elderly persons: the Physical Capacity Evaluation. *Am J Public Health.* 1995;85:558–560.
3. Hoeymans N, Feskens EJ, van den Bos GA, Kromhout D. Measuring functional status: cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990–1993). *J Clin Epidemiol.* 1996;49:1103–1110.
4. Guralnik JM, Seeman TE, Tinetti ME, Nevitt MC, Berkman LF. Validation and use of performance measures of functioning in a nondisabled older population: MacArthur Studies of Successful Aging. *Aging (Milano).* 1994;6:410–419.
5. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol Med Sci.* 1994;49:M85–M94.
6. 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.
7. Simonsick EM, Newman AB, Nevitt MC, et al. Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC Study. *J Gerontol Med Sci.* 2001;56A:M644–M649.
8. Cress ME, Buchner DM, Questad KA, Esselman PC, deLateur BJ, Schwartz RS. Continuous scale physical functional performance in healthy older adults: a validation study. *Arch Phys Med Rehabil.* 1996; 77:1243–1250.
9. Cress ME, Schechtman KB, Mulrow CD, Fiatarone MA, Gerety MB, Buchner DM. Relationship between physical performance and self-perceived physical function. *J Am Geriatr Soc.* 1995;43:93–101.
10. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.* 1975;12:189–198.
11. Kasper JD, Shapiro S, Guralnik JM, Bandeen-Roche KJ, Fried LP. Designing a community study of moderately to severely disabled older women: the Women's Health and Aging Study. *Ann Epidemiol.* 1999;9:498–507.
12. Guralnik JM, Fried LP, Simonsick EL, Kasper JD, Lafferty ME. *The Women's Health and Aging Study. Health and Aging Characteristics of Older Women with Disability.* Bethesda, MD: National Institute of Aging; 1995. NIH Pub No. 1995:95-4009. Available at: http://www.nih.gov/nia/edb/whasbook. Accessed December 2000.
13. 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 Med Sci.* 2000;55A:M221–M231.
14. Seeman TE, Charpentier PA, Berkman LF, et al. Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur Studies of Successful Aging. *J Gerontol Med Sci.* 1994;49:M97–M108.
15. Furuna T, Nagasaki H, Nishizawa S. Longitudinal change in the physical performance of older adults in the community. *J Jpn Phys Ther Assoc.* 1999;1:1–5.
16. Fried LP, Herdman SJ, Kuhn KE, Rubin G, Turano K. Preclinical disability: hypotheses about the bottom of the iceberg. *J Aging Health.* 1991;3:285–300.
17. Guralnik JM, Ferrucci L, Penninx BW, et al. New and worsening conditions and change in physical and cognitive performance during weekly evaluations over 6 months: the Women's Health and Aging Study. *J Gerontol Med Sci.* 1999;54A:M410–M422.
18. Penninx BW, Ferrucci L, Leveille SG, Rantanen T, Pahor M, Guralnik JM. Lower extremity performance in nondisabled older persons as a predictor of subsequent hospitalization. *J Gerontol Med Sci.* 2000; 55A:M691–M697.
19. Ostir GV, Markides KS, Black SA, Goodwin JS. Lower body functioning as a predictor of subsequent disability among older Mexican Americans. *J Gerontol Med Sci.* 1998;53A:M491–M495.

Received October 15, 2001
Accepted January 14, 2002