Journal of Gerontology: MEDICAL SCIENCES
1994, Vol. 49, No. 2, M85–M94

A Short Physical Performance Battery Assessing Lower Extremity Function: Association With Self-Reported Disability and Prediction of Mortality and Nursing Home Admission

Jack M. Guralnik,[1] Eleanor M. Simonsick,[1] Luigi Ferrucci,[2] Robert J. Glynn,[3] Lisa F. Berkman,[4] Dan G. Blazer,[5] Paul A. Scherr,[6] and Robert B. Wallace[7]

[1]Epidemiology, Demography, and Biometry Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland.
[2]Geriatric Department, Hospital I Fraticini, INRCA, Florence, Italy.
[3]Channing Laboratory, Department of Medicine, Harvard Medical School.
[4]Department of Epidemiology, Yale University School of Medicine.
[5]Department of Psychiatry, Duke University School of Medicine.
[6]Aging Studies Branch, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control, Atlanta, Georgia.
[7]Department of Preventive Medicine and Environmental Health, University of Iowa.

Background. A short battery of physical performance tests was used to assess lower extremity function in more than 5,000 persons age 71 years and older in three communities.

Methods. Balance, gait, strength, and endurance were evaluated by examining ability to stand with the feet together in the side-by-side, semi-tandem, and tandem positions, time to walk 8 feet, and time to rise from a chair and return to the seated position 5 times.

Results. A wide distribution of performance was observed for each test. Each test and a summary performance scale, created by summing categorical rankings of performance on each test, were strongly associated with self-report of disability. Both self-report items and performance tests were independent predictors of short-term mortality and nursing home admission in multivariate analyses. However, evidence is presented that the performance tests provide information not available from self-report items. Of particular importance is the finding that in those at the high end of the functional spectrum, who reported almost no disability, the performance test scores distinguished a gradient of risk for mortality and nursing home admission. Additionally, within subgroups with identical self-report profiles, there were systematic differences in physical performance related to age and sex.

Conclusion. This study provides evidence that performance measures can validly characterize older persons across a broad spectrum of lower extremity function. Performance and self-report measures may complement each other in providing useful information about functional status.

THE assessment of physical functioning and disability is a critical component in the evaluation of older persons in both clinical and research settings. Although measures of self- and proxy-reported functioning have proved extremely valuable, standardized tests of physical performance have been employed with increasing frequency in recent years (1–9). These physical performance measures may offer advantages over self-report measures in terms of validity, reproducibility, sensitivity to change, applicability to cross-national and cross-cultural studies, and the ability to characterize high levels of function, although much methodologic work remains to be done in this area (1). Understanding the relationship between self-report and performance-based measures is an important aspect of this issue.

The purpose of this report is to investigate the association of self-report and performance-based measures of lower extremity functioning. We present findings from a short battery of performance tests of lower extremity functioning found to be efficient, practical, and safe in three large cohorts of older persons. These tests were administered by single interviewers in the homes of more than 5,000 older persons in about 10 minutes. The association between self-report and performance measures will be shown, followed by analyses addressing the advantages of the performance measures in more finely classifying levels of functioning and in classifying persons not only at the disabled end of the functional spectrum but along the entire spectrum of functioning. Finally, the relative contributions of the self-report and performance measures in predicting death and nursing home admission will be demonstrated.

METHODS

Study populations. — Data for this report are from older adults in the original three communities of the Established Populations for Epidemiologic Studies of the Elderly (EPESE), a collaborative longitudinal study of aging initiated and funded by the Epidemiology, Demography, and Biometry Program of the National Institute on Aging. Baseline household interviews, conducted between 1981 and 1983, obtained information on demographic characteristics, medical history, the use of prescription and nonprescription drugs, health behaviors, and functional status. These interviews were sought from all people aged 65 and older living in the community in East Boston, Massachusetts, and Iowa and Washington counties in rural Iowa, and a stratified random sample in New Haven, Connecticut. In the three communities, between 80 and 85% of those eligible participated in the baseline interview. The details of the methods used in the baseline surveys have been described previously (10).

At the sixth annual follow-up, conducted in 1988 and 1989, information on health, functional status, and medical care utilization was ascertained in a home interview. Performance measures of functioning were added to the protocol at that time. During the 6 years of follow-up, between 26.2 and 35.3% of participants died at the three sites. Of those not known to have died, follow-up six assessments were conducted by personal or proxy interview on between 91.5 and 94.0% of persons who participated in the baseline interview (6,563/7,093 at all sites, 2,345/2,563 in East Boston, 2,547/2,711 in Iowa, and 1,671/1,819 in New Haven). Among these persons, 491 resided in a nursing home; 792 lived at home but required a proxy informant and were not administered the performance measures uniformly across sites; and 106 had a telephone interview. Thus, there were a total of 5,174 persons at all sites who were living at home and able to respond personally to the sixth follow-up interview (East Boston, 1,800; Iowa, 2,101; New Haven, 1,273). Of this group, less than 1.5% in East Boston, 0.2% in Iowa, and 3.7% in New Haven refused to participate in one or more of the performance measures. The proportion of persons unable to complete a task or who did not attempt it because they or the interviewer felt the task was potentially unsafe varied according to the performance measure. No injuries resulted from the administration of the performance tests.

Mortality was ascertained at all three sites through the end of 1989 through review of local obituary notices and linkage with the National Death Index. Nursing home admissions were ascertained in Iowa through 1990 by periodic visits to all area nursing homes.

Physical performance measures. — Performance measures for this study were adapted from previously used measures (3,11–14), with the consideration that the assessments would be administered by a single lay interviewer in a home setting with limited unobstructed space. Interviewers were specially trained in the administration of the performance measures used in the study. To ensure uniformity of administration, a videotape was produced; it provided detailed instructions for administering and scoring the tests, as well as instructions on maintaining the safety of the subjects. Total time for a single interviewer to perform these tests in the home ranged from 10 to 15 minutes.

Tests of standing balance included tandem, semi-tandem, and side-by-side stands. For each stand, the interviewer first demonstrated the task, then supported one arm while participants positioned their feet, asked if they were ready, then released the support and began timing. The timing was stopped when participants moved their feet or grasped the interviewer for support, or when 10 seconds had elapsed. Each participant began with the semi-tandem stand, in which the heel of one foot was placed to the side of the first toe of the other foot, with the participant choosing which foot to place forward. Those unable to hold the semi-tandem position for 10 seconds were evaluated with the feet in the side-by-side position. Those able to maintain the semi-tandem position for 10 seconds were further evaluated with the feet in full tandem position, with the heel of one foot directly in front of the toes of the other foot.

To test walking speed, an 8-foot walking course, with no obstructions for an additional 2 feet at either end, was denoted by placing a rigid 8-foot carpenter's rule to the side of the course. Participants were instructed to "walk to the other end of the course at your usual speed, just as if you were walking down the street to go to the store." Participants could use assistive devices if needed, and each participant was timed for two walks. The faster of the two is used for analyses here.

To test the ability to rise from a chair (termed the chair stand), a straight-backed chair was placed next to a wall; participants were asked to fold their arms across their chest and to stand up from the chair one time. If successful, participants were asked to stand up and sit down five times as quickly as possible, and were timed from the initial sitting position to the final standing position at the end of the fifth stand.

Categories of performance were created for each set of performance measures to permit analyses that included those unable to perform a task. For the 8-foot walk and repeated chair stands, those who could not complete the task were assigned a score of 0. Those completing the task were assigned scores of 1 to 4, corresponding to the quartiles of time needed to complete the task, with the fastest times scored as 4. The three tests of standing balance were considered as hierarchical in difficulty in assigning a single score of 0 to 4 for standing balance (Table 1). A summary performance scale was created by summing the category scores for the walking, chair stand, and balance tests.

Self-reported physical functioning. — Activities of daily living (ADLs) used in these analyses were limited to those that require lower extremity function as a major component. Participants were classified as having a disability in ADLs if they needed help from a person or were unable to perform one or more of the following activities: walking across a small room, bathing, transferring from bed to chair, or using the toilet (15,16). To assess higher level mobility, participants were asked if they were able to walk up and down stairs to the second floor and walk a half mile without help (17).

Statistical analysis. — The reliability of the summary

Table 1. Categories of Performance in Tests of Standing Balance

Category	Side-by-Side Stand (seconds)	Semi-tandem Stand (seconds)	Full Tandem Stand (seconds)	*n*	Percent
0	0–9 Tried but unable Not attempted	< 10 Tried but unable Not attempted	–	486	9.5%
1	10	0–9 Tried but unable Not attempted	–	738	14.5
2	–	10	0–2 Tried but unable Not attempted	662	13.0
3	–	10	3–9	705	13.8
4	–	10	10	2513	49.2
Total				5104	100.0

performance scale was assessed, in terms of internal consistency, by computing Cronbach's alpha (18). The independent associations of age, sex, and self-reported functioning with score on the summary performance scale were evaluated using a multiple linear regression model. Rates of mortality and nursing home admission were calculated for each possible score on the summary performance scale. Mortality surveillance continued through the end of 1989; nursing home surveillance continued through the end of 1990. Because follow-up 6 assessments were performed over an 18-month period in 1988–1989, length of follow-up varied considerably. Thus, rates were calculated per person year of follow-up and were age- and sex-adjusted using the direct method, with the total population serving as the standard population. The independent associations of self-report and performance measures of function with mortality and nursing home admission were evaluated using Cox proportional hazards regression models, with time of follow-up considered as the time from functional assessment to death or the end of 1989, whichever came first (for mortality analyses), and the time from functional assessment to first nursing home admission, death, or the end of 1990, whichever came first (for the Iowa nursing home admission analyses). Overall estimates of relative risk were calculated from proportional hazards models stratified by community, using the BLOCK option of the SAS PHGLM procedure to obtain summary estimates across the communities (19).

RESULTS

Figures 1 and 2 show the distribution of times for the 8-foot walk and the repeat chair stands and the proportion unable to complete these tasks by age group and sex. Mean and median times for women were significantly longer than for men in both age groups for both of these tasks, and a greater proportion of women was unable to complete the tasks. Of note, however, is the large amount of overlap between men and women and between the two age groups. The cutpoints used to classify categories of performance are depicted in Figures 1 and 2. Table 1 shows the distribution of scores for the five categories assigned to performance on the tests of standing balance. Just under 10% were unable to hold the side-by-side position for 10 seconds and were assigned category 0. Nearly half of participants were able to maintain a full tandem stand for 10 seconds and were assigned category 4. Cutpoints were selected on the full tandem stand so that categories 2 and 3 had about equal numbers of persons in each.

For reference purposes, distributions of performance on each test are provided for age and sex subgroups in Appendix Tables A1 to A3. For the 8-foot walk and the repeat chair stands, percent unable, and mean time and percentile distributions of time for those able to complete the task are shown. Cutpoints for the classification into categories of performance are also included.

Scores on the three performance tasks were significantly correlated. Spearman correlation coefficients for the categorical scores were: walking and chair stands, 0.48; walking and standing balance, 0.39; chair stands and standing balance, 0.39 (all significant at $p < .001$). The summary performance scale, created by summing the category scores on the three individual performance tests, ranged from 0 to 12, with a mean of 7.12, a median of 8, and an interquartile range of 5 (difference between 25th and 75th percentile). The internal consistency of this scale, assessed as Cronbach's alpha, was 0.76.

Table 2 shows the percentage of persons reporting disability in ADLs, walking one-half mile, and climbing stairs according to categories on the performance tests. Increasingly better performance on the timed walk, chair stands, and tests of standing balance was associated with a very clear stepwise decrease in the proportion reporting disability. Extremely high rates of disability were reported by those in category 0 and extremely low rates by those in category 4. The summary performance score also showed a very strong association with measures of self-reported disability (Table 3). The associations shown in Tables 2 and 3 were very

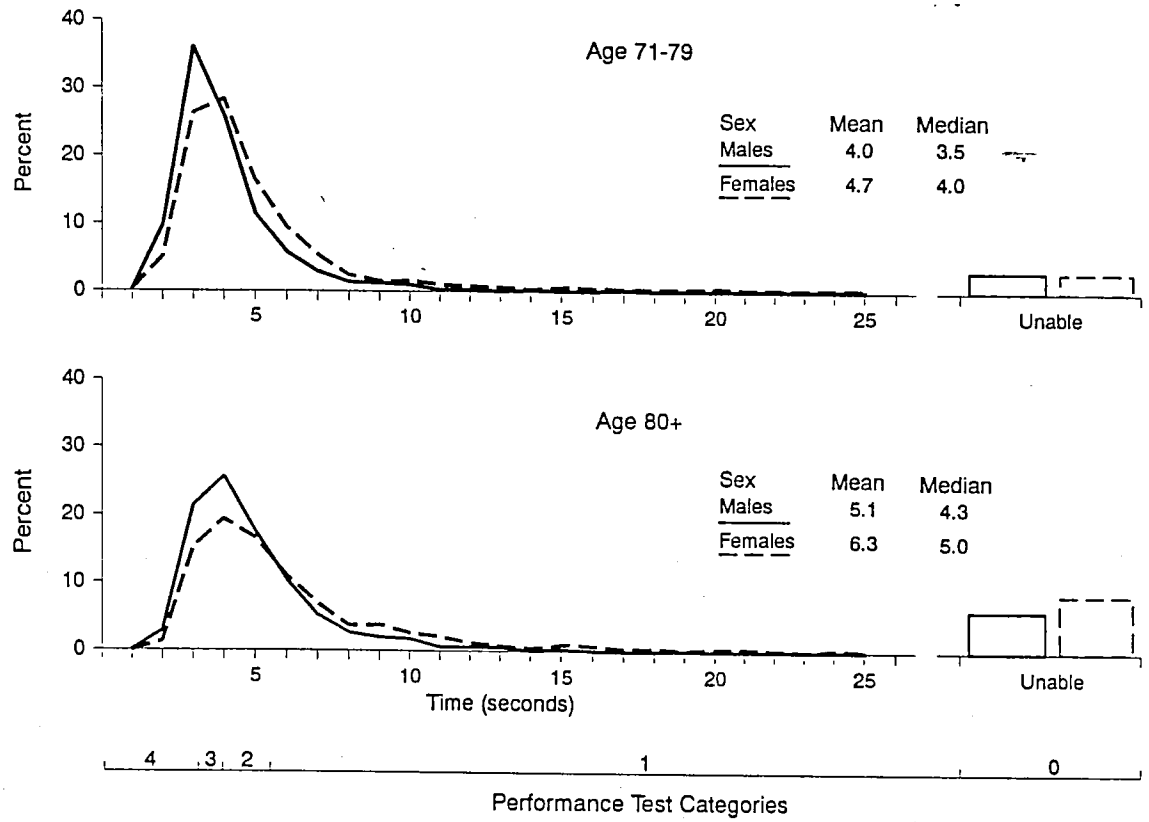

Figure 1. Distribution of times to walk 8 feet according to age group and sex.

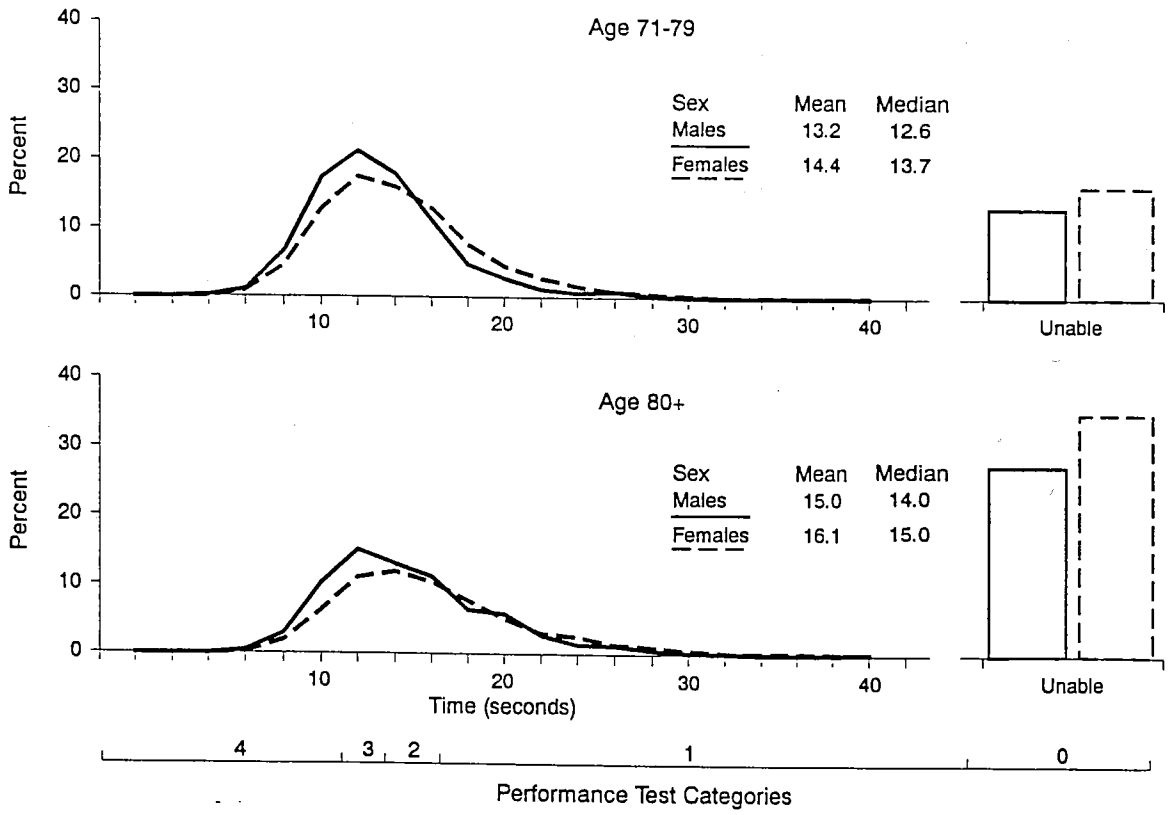

Figure 2. Distribution of times to complete five chair stands according to age group and sex.

Table 2. Percent of Persons Reporting Need for Help in ADLs, Walking 1/2 Mile, and Climbing Stairs According to Performance Test Categories; All Sites Combined

Performance Test Score	ADLs*	Walking 1/2 Mile	Climbing Stairs
Walk			
0	51.4	86.6	64.7
1	15.2	61.3	31.3
2	3.2	32.5	8.2
3	1.2	17.2	3.5
4	0.7	7.0	1.1
Chair stands			
0	25.9	73.6	42.5
1	5.2	39.4	11.2
2	2.2	21.4	6.8
3	0.8	12.8	3.0
4	0.6	10.7	1.7
Standing balance			
0	42.0	86.0	59.8
1	11.5	57.2	25.7
2	4.2	35.5	12.1
3	2.4	26.9	6.6
4	1.6	15.5	3.5

*Need help in one or more of the following: Walking, transferring, bathing, using the toilet.

Table 3. Percent of Persons Reporting Need for Help in ADLs, Walking 1/2 Mile, and Climbing Stairs According to Performance Summary Measure; All Sites Combined

Performance Test Summary Score	ADLs*	Walking 1/2 Mile	Climbing Stairs
0	56.4	91.3	70.8
1	39.5	91.4	66.9
2	21.1	82.1	40.8
3	15.5	66.8	30.2
4	7.4	57.3	21.6
5	4.9	48.4	15.5
6	3.2	36.4	9.3
7	3.3	27.1	6.7
8	1.2	17.2	3.8
9	1.4	15.3	1.9
10	0.5	9.4	2.4
11	0.2	6.8	0.2
12	0.0	3.7	0.9

*Need help in one or more of the following: Walking, transferring, bathing, using the toilet.

consistent across the three individual sites (results not shown).

To further assess the association of self-report and performance measures, multiple linear regression analyses were performed that included age categories, sex, and self-report items as independent variables and the summary performance score as the dependent variable (Table 4). In Model I, self-report measures alone explained 42% of the variance in the summary performance score. In Model II, adding age and sex increased the proportion of the variance explained to 46%. Of particular importance in interpreting the results of Model II is that, even after controlling for self-report of disability, both age and sex were associated with performance. The indicator variables for age show that, after adjustment for self-reported disability, each higher age group scored lower on the summary performance scale. For example, after adjusting for self-report, those age 75 to 79 years scored 0.47 points lower on the summary performance scale than the reference group, those age 71 to 74 years. Those age 90 and older scored 2.11 points lower than the reference group, after adjustment for self-report.

This is illustrated graphically in Figure 3, which shows actual mean summary performance scores according to age and sex for (a) those with no self-reported disability on the items considered here and (b) for those reporting the need for help only in walking one-half mile. For both these subgroups of participants there is a clear decline in summary performance score with age. In those reporting no disability, women have mean summary performance scores about one-half point lower than men at all ages (Figure 3a). The difference between men and women is inconsistent for those reporting the need for help only in walking one-half mile (Figure 3b). Other subgroups defined by self-report were too small to yield stable estimates for the 5 age groups of men and women and are not shown.

Although the proportion of the variance in the summary performance score explained by the self-report measures used here is quite high, there is still a substantial amount of variability in performance that remains unexplained. This is illustrated graphically in Figure 4, which shows the distribution of summary performance scores for three groups defined by self-reported disability: (a) ADL and mobility disability; (b) no ADL disability but mobility disability; (c) no ADL or mobility disability. In this categorization, mobility disability is defined as needing help in walking one-half mile, climbing stairs, or both. This classification scheme is a hierarchical or Guttman scale in that there were very few participants reporting ADL disability who did not also report mobility disability. These participants (< 1% of the sample) were not included in this analysis. The distributions for the three groups defined by self-report are very different, yet there is a great deal of overlap. A small proportion of participants in the group with the greatest degree of self-reported disability had performance scores that were higher than some members of the group reporting no disability. Summary performance scores for the middle-disability group are spread across the entire spectrum.

Finally, the relative ability of self-report and performance measures to predict adverse outcomes was evaluated for mortality and nursing home admission. Of the 5,174 persons from the three sites considered in these analyses, 300 (5.8%) died between the follow-up 6 assessment and the end of 1989. Average length of follow-up to death or the end of 1989 was 1.39 years. Figure 5a shows a stepwise decline in the age- and sex-adjusted mortality rate with increasing score on the summary performance scale. The associations of age, self-reported disability, and summary performance score with death, as demonstrated using proportional hazards models, are shown for men and women in Table 5. In Model I, self-reported disabilities in ADLs and walking one-half mile are significant predictors of death in men, and disability in walk-

Table 4. Multiple Linear Regression Model Showing Association of Age, Sex, and Self-Reported Disability with Summary Performance Scale

	Model I		Model II	
Independent Variable	β Coefficient	*SE* of β	β Coefficient	*SE* of β
Intercept	8.1*	0.04	9.73*	0.13
Age 75–79†			–0.47*	0.09
Age 80–84†			–1.11*	0.10
Age 85–89†			–1.89*	0.13
Age 90 + †			–2.11*	0.20
Sex (female vs male)			–0.41*	0.07
ADL (one or more)	–2.26*	0.13	–2.08*	0.15
Climbing stairs, needs help	–2.28*	0.13	–2.17*	0.12
Walking 1/2 mile, needs help	–2.84*	0.09	–2.46*	0.09
	R^2 = 0.42		R^2 = 0.46	

†Compared to reference group, age 71–74 years.
*$p < .001$.

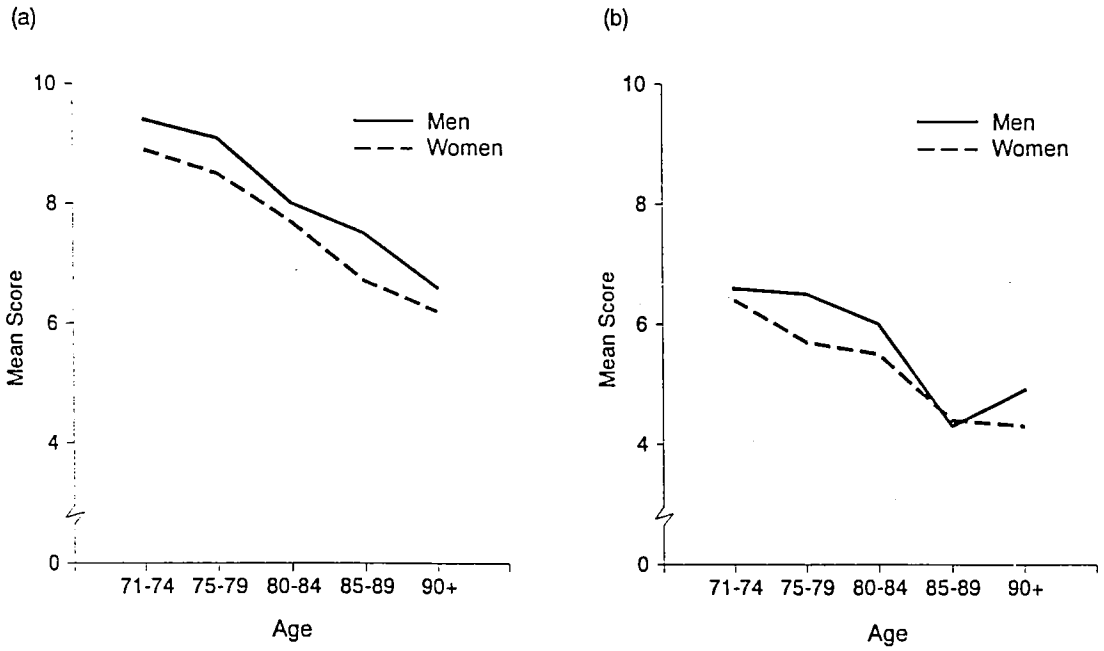

Figure 3. Mean scores on summary performance scale according to age group and sex for (a) those who report needing no help with activities of daily living, climbing stairs, and walking one-half mile, and (b) those who report they can not walk one-half mile without help but need no help with activities of daily living and climbing stairs.

ing one-half mile is significant in women. In Model II, the summary performance score is significantly associated with death, with those at the 25th percentile being more than two times more likely to die than those in the 75th percentile (5-point difference in score). In Model III, in which self-report and performance measures are entered together, ADLs, walking one-half mile, and the summary performance score remain significantly associated with mortality in men, but only the summary performance score reaches significance as a predictor of mortality in women, with walking one-half mile almost statistically significant ($p = .056$). In a subgroup analysis limited to those at the high end of the functional spectrum (summary performance score of 8 to 12), the summary performance score was significantly associated with mortality (results not shown).

A similar pattern was seen in analyses examining predictors of nursing home admission in Iowa, the only site to have comprehensive assessment of nursing home admission after the follow-up 6 assessment. Between the time of the follow-up 6 assessment in the home and the end of 1990, 286/2097 (13.6%) of Iowa participants entered a nursing home. Average length of follow-up to nursing home admission, death, or the end of 1990 was 2.46 years. Figure 5b shows a general stepwise decline in the age- and sex-adjusted nursing home admission rate with increasing score on the summary performance scale. Table 6 demonstrates in Model III that significant predictors of nursing home admission include both self-report (the question on climbing stairs in men and the question on walking one-half mile in women) and the summary performance scale. In an analysis limited to those with

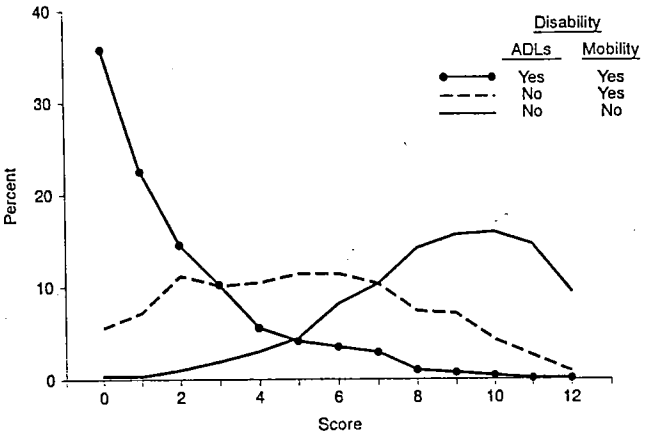

Figure 4. Distribution of summary performance score for three groups defined by self-reported disability. Those with no mobility disability report they can climb a flight of stairs and walk one-half mile without help.

a summary performance score of 8 to 12, this score was a significant predictor of nursing home admission (results not shown).

Discussion

The performance measures of functioning administered to this population aged 71 years and older assessed gait, balance, and lower extremity strength and endurance. These tests were found practical and safe for well-trained interviewers to administer in a home setting. The gait and balance tests could potentially have led to falls, and the repeat chair stand test required a vigorous effort, but there was a high level of compliance and no injuries or adverse outcomes were encountered. Results may be interpreted as continuous variables or, in order to include individuals in the analyses who were unable to perform the tests, as categorical variables.

Those age 80 and older had poorer performance than those age 71–79, but the amount of overlap was great, with a large proportion of the older age group functioning at or above the mean levels of the younger group. The prevalance of self-reported disability has been found to be higher in older women than men (10,20). These results were consistent, showing slightly slower times for women and a higher proportion of women in the older age group unable to complete the tasks. However, there is considerable overlap here, as well.

A strong association was found between performance measures of lower extremity function and self-reported disability in this study. The performance measures employed in this research were not designed to examine objectively the same tasks as assessed by the self-report measures. This has been done in several studies designed specifically to compare a participant's ability to complete standardized tasks representing particular ADLs with their self-reported disability in these tasks (21–23). These studies found a substantial amount of discordance between performance tests and self-report that may represent inaccurate reporting and measurement error, although there are some limitations in making these kinds of comparisons. The simple tests of performance used in these studies could not fully assess all aspects

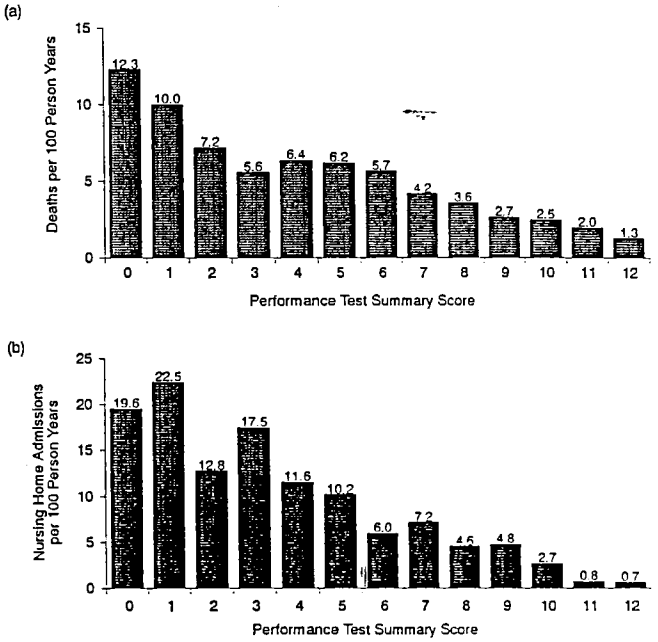

Figure 5. (a) Age- and sex-adjusted death rate according to summary performance score for East Boston, Iowa, and New Haven sites. (b) Age- and sex-adjusted nursing home admission rate according to summary performance score for Iowa site.

of an ADL that might influence a participant's self-report. Furthermore, two of the studies (21,22) were conducted on persons at the end of an acute care hospitalization, when their ability to recognize their abilities and limitations might have been reduced, especially for activities such as dressing and bathing.

Although our study was not designed to assess discordance between performance and self-report measures of the same task, discordance was found for assessments of walking. Of those unable to complete the measured walk (walk test category 0, Table 2), 13.6% of persons stated they could walk a half mile without help. The apparent discordance in these 33 individuals may have multiple causes, including the participant misunderstanding the question or answering the question with reference to his or her usual state of functioning while suffering from a temporary incapacity, and the interviewer miscoding the self-report response or the reason for nonperformance of the walk test. An analysis restricted to those unable to perform the walk that compared those who stated they could and could not walk a half mile found no differences between these two groups in cognitive functioning, depression, and educational status (analyses not shown).

Rather than assessing the same task in several ways, the measures used in our study can probably best be conceptualized as representing two points on the pathway from disease to disability. In a framework proposed by an Institute of Medicine committee (24) and derived from the work of Nagi (25), the pathway progresses from disease to impairment to functional limitation to disability. The performance measures are objective assessments of functional limitations (defined as a restriction or lack of ability to perform an action or activity), while the self-report measures reflect disability

Table 5. Relative Risks Relating Age, Self-Reported Disability, and Summary Performance Score to Mortality

	Model I		Model II		Model III	
	Men	Women	Men	Women	Men	Women
	Relative Risk* (95 percent confidence interval)					
Age (10 years)	1.7 (1.3–2.1)	1.7 (1.3–2.3)	1.5 (1.1–2.0)	1.5 (1.1–2.0)	1.5 (1.1–2.0)	1.5 (1.1–2.0)
Disabled in:						
ADLs (one or more)	1.9 (1.2–3.1)	1.4 (0.9–2.2)			1.7 (1.0–2.8)	1.1 (0.6–1.7)
Walking 1/2 mile	2.2 (1.5–3.3)	2.2 (1.5–3.3)			1.7 (1.1–2.6)	1.5 (1.0–2.3)
Climbing stairs	1.4 (0.9–2.3)	1.3 (0.9–2.0)			1.1 (0.7–2.0)	1.0 (0.7–1.6)
Summary performance score (5 points poorer)			2.3 (1.8–2.9)	2.6 (2.0–3.5)	1.6 (1.1–2.1)	2.2 (1.6–3.2)

*Based on proportional hazards model.

Table 6. Relative Risks Relating Age, Self-Reported Disability, and Summary Performance Score to Nursing Home Admission. Iowa EPESE.

	Model I		Model II		Model III	
	Men	Women	Men	Women	Men	Women
	Relative Risk* (95 percent confidence interval)					
Age (10 years)	2.1 (1.5–2.9)	2.3 (1.8–2.9)	1.9 (1.4–2.6)	1.9 (1.5–2.5)	1.8 (1.3–2.4)	1.9 (1.4–2.4)
Disabled in:						
ADLs (one or more)	1.7 (0.7–3.3)	1.5 (1.0–2.4)			1.1 (0.5–2.1)	1.2 (0.8–2.0)
Walking 1/2 mile	2.3 (1.4–3.8)	2.1 (1.5–2.8)			1.2 (0.7–2.2)	1.4 (1.0–2.0)
Climbing stairs	2.8 (1.5–5.1)	1.7 (1.1–2.4)			1.9 (1.0–3.4)	1.2 (0.8–1.7)
Summary performance score (5 points poorer)			3.4 (2.5–4.6)	2.8 (2.2–3.6)	2.7 (1.8–4.0)	2.2 (1.6–3.0)

*Based on proportional hazards model.

(defined as an inability or limitation in performing socially defined activities and roles). Since not all functional limitations are expected to lead to disability, a perfect correlation between the kind of performance measures employed here and disability would not be expected. A fruitful line of research to improve our understanding of the pathway, however, is to identify other factors that influence the impact of functional limitations on disability. Jette and colleagues have employed performance measures which assess joint range of motion and have demonstrated the association of these measures with disability (11) and the impact of the change in these measures with change in disability (26).

In regard to the pathway discussed above, it should be noted that the regression analyses used to evaluate the relationship between self-report and performance measures were not constructed to imply a causal pathway but were simply used to evaluate their cross-sectional association. Considering how few self-report items were used and their limited response categories, it is striking how much of the variance in the summary performance measure was explained. It is possible that even more of the variance in the summary performance scale would be explained if these items also assessed perceived level of difficulty or if additional self-report items were used that evaluated more difficult activities, such as those assessing vigorous and moderate activities in the full Medical Outcomes Study instrument (27) or Reuben and colleagues' Advanced Activities of Daily Living (28). These findings raise the practical question of just how much information the performance measures add to the self-report measures.

This report provides evidence that the performance measures do indeed contribute information beyond that obtained from self-report. The most important finding here, that has not been previously demonstrated, is that a comprehensive scale of lower extremity performance can provide information across the entire spectrum of functioning. This classification of level of functioning appears to be valid, as evidenced by the clearly demonstrated gradients across the summary performance scale of prevalence of self-reported disabilities and incidence of nursing home admission and death. That those scoring 12 on the summary performance scale have a lower mortality than those scoring 11, who have a lower mortality than those scoring 10 (Figure 5) is quite remarkable. No self-reported measures of functioning have been demonstrated that can define a gradient of risk for those who not only are not disabled but rank at the highest end of the functional spectrum, although it is conceivable that self-report measures assessing very vigorous activities might work in this way.

Other findings also support the value of performance measures. For certain subgroups with identical self-report profiles, there were systematic differences in performance related to age and sex (Figure 3). Within each of the three hierarchical levels of functioning defined by self-report there was a wide distribution of functioning on the performance tests (Figure 4). Even more refined self-report measures

would probably not appreciably affect rankings derived from self-report, so it is likely that performance measures are providing information not obtainable from self-report. This is also evidenced in the analyses of predictors of death and nursing home admission (Tables 5 and 6), in which the summary performance scale is a strong and consistent predictor of these outcomes even after adjustment for self-reported disability. These findings are consistent with findings from one of the earliest applications of performance measures, in which a performance battery was more highly predictive of death and hospitalization than self-report of disability (29). Other studies have shown the association of performance measures with death (30) and nursing home admission (31), but did not directly compare the influence of performance and self-report measures in a single model. Two studies of the predictors of falls found that performance measures were important predictors of this outcome, even after adjustment for self-reported disability (2,3).

This study has demonstrated that a short battery of performance measures of lower extremity function characterizes community-dwelling older persons across a broad spectrum of functional status and predicts mortality and institutionalization across this spectrum. The research also supports the conclusions of other studies (4,30) that self-report and performance measures complement one another, and that both add to our understanding of older persons' functional status. Standardized measurement of performance thus appears to be a valuable addition to gerontologic assessment, but its widespread use may be premature. Performance measures may prove useful in comparing the functional status of different populations and tracking function over time, but these applications require additional methodologic work. Self-report assessments require less training and are quicker to obtain and may adequately serve the purposes of much research and clinical evaluation. The value of performance measures in assessing requirements for care and placement needs of an individual patient can not be inferred from these population studies, and further investigation in this area would also be valuable.

ACKNOWLEDGMENTS

The research on which this article is based was performed pursuant to contract numbers N01-AG-0-2105, N01-AG-0-2106, and N01-AG-0-2107 with the National Institute on Aging, NIH. The authors wish to acknowledge the important contribution of Yong Lee and Geane Schubert to the analysis of these data.

The performance measure protocol, data entry form, and videotape used to train interviewers are available from the first author.

Address correspondence to Dr. Jack M. Guralnik, National Institute on Aging, 7201 Wisconsin Ave., Room 3C-309, Bethesda, MD 20892.

REFERENCES

1. Guralnik JM, Branch LG, Cummings SR, Curb JD. Physical performance measures in aging research. J Gerontol Med Sci 1989;44: M141–6.
2. Tinetti ME, Speechley M, Ginter S. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988; 319:1701–7.
3. Nevitt MC, Cummings SR, Kidd S, Black D. Risk factors for recurrent nonsyncopal falls. A prospective study. JAMA 1989,261:2663–8.
4. Kelly-Hayes M, Jette AM, Wolf PA, D'Agostino RB, Odell PM. Functional limitations and disability among elders in the Framingham Study. Am J Public Health 1992;82:841–5.
5. Williams ME, Gaylord SA, McGaghie WC. Timed manual performance in a community elderly population. J Am Geriatr Soc 1990; 38:1120–6.
6. Reuben DB, Siu AL. An objective measure of physical function of elderly patients. The physical performance test. J Am Geriatr Soc 1990;38:1105–12.
7. Podsiadlo D, Richardson S. The timed "up and go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991; 39:142–8.
8. Weiner DK, Duncan PW, Chandler J, Studenski SA. Functional reach: a marker of physical frailty. J Am Geriatr Soc 1992;40:203–7.
9. Siu AL, Hays RD, Ouslander JG, et al. Measuring functioning and health in the very old. J Gerontol Med Sci 1993;48:M10–14.
10. Cornoni-Huntley J, Brock DB, Ostfeld A, Taylor JO, Wallace RB, eds. Established Populations for Epidemiologic Studies of the Elderly, Resource data book. NIH Pub. No. 86-2443, 1986. Bethesda, MD: National Institutes of Health.
11. Jette AM, Branch LG. Impairment and disability in the aged. J Chron Dis 1985;38:59–65.
12. Csuka M, McCarty DJ. Simple method for measurement of lower extremity muscle strength. Am J Med 1985;78:77–81.
13. Tinetti ME. Performance oriented assessment of mobility problems in elderly patients. J Am Geriatr Soc 1986;34:119–26.
14. Cummings SR, Nevitt MC, Black DM, et al. The study of osteoporotic fractures: A multicenter prospective study of predictors of hip and wrist fractures. In: Christiansen C, Johansen JS, Riis BJ, eds. Osteoporosis 1987. International Symposium on Osteoporosis. Copenhagen: Osteopress ApS, 1987:58–9.
15. Katz SC, Ford AB, Moskowitz RW, Jack BA, Jaffe MW. Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. JAMA 1963;185:914–19.
16. Branch LG, Katz S, Kniepmann K, Papsidero JA. A prospective study of functional status among community elders. Am J Public Health 1984;74:266–8.
17. Rosow I, Breslau N. A Guttman health scale for the aged. J Gerontol 1966;21:556–9.
18. Spector PE. Summated rating scale construction. Newbury Park, CA.: Sage Publications, 1992.
19. Harrell FE. The PHGLM procedure. In: SAS Institute, Inc., SUGI supplemental library user's guide, Version 5 ed. Cary, NC: SAS Institute, Inc., 1986:437–66.
20. Fulton JP, Katz S, Jack SS, Hendershot GE. Physical functioning of the aged: United States, 1984. Vital Health Stat. Series 10, No. 167. DHHS publication no. (PHS) 89-1595. Hyattsville, MD: National Center for Health Statistics, March 1989.
21. Elam JT, Graney MJ, Beaver T, El Derwi D, Applegate WB, Miller ST. Comparison of subjective ratings of function with observed functional ability of frail older persons. Am J Public Health 1991;81:1127–30.
22. Sager MA, Dunham MC, Schwantes A, Mecum L, Halverson K, Marlowe D. Measurement of activities of daily living in hospitalized elderly: a comparison of self-report and performance based methods. J Am Geriatr Soc 1992;40:457–62.
23. Dorevitch MI, Cossar RM, Bailey FJ, et al. The accuracy of self and informant ratings of physical functional capacity in the elderly. J Clin Epidemiol 1992;45:791–8.
24. Pope AM, Tarlov AR, eds. Disability in America: Toward a national agenda. Washington, DC: National Academy Press, 1991.
25. Nagi SZ. An epidemiology of disability among adults in the United States. Milbank Mem Fund Q 1976;6:493–508.
26. Jette AM, Branch LG, Berlin J. Musculoskeletal impairments and physical disablement among the aged. J Gerontol Med Sci 1990;45:M203–8.
27. Stewart AL, Ware JE. Measuring functioning and well-being. The Medical Outcomes Study Approach. Durham, NC: Duke University Press, 1992.
28. Reuben DB, Laliberte L, Hiris J, Mor V. A hierarchical exercise scale to measure function at the Advanced Activities of Daily Living level. J Am Geriatr Soc 1990;38:855–61.
29. Kuriansky JB, Gurland BJ, Fleiss JL. The assessment of self-care

capacity in geriatric psychiatric patients by objective and subject. J Clin Psych 1976;32:95–102.
30. Reuben DB, Siu AL, Kimpau S. The predictive validity of self-report and performance-based measures of function and health. J Gerontol Med Sci 1992;47:M106–10.
31. Williams ME, Hornberger JC. A quantitative method of identifying older persons at risk for increasing long term care services. J Chron Dis 1984;37:705–11.

Received March 8, 1993
Accepted June 17, 1993

Appendix

Table A1. Performance on Timed 8-Foot Walk: Percentage Unable To Complete and Mean Times and Percentile Times for Those Completing the Task

	All	Men			Women		
	71+	71+	71–79	80+	71+	71–79	80+
n	5097	1785	1239	546	3312	2030	1282
Unable to complete (percent)	4.9%	4.2%	3.3%	6.2%	5.3%	3.2%	8.7%
Times (sec)							
Mean	5.0	4.4	4.0	5.1	5.3	4.7	6.3
1st percentile	19.0	14.1	12.7	15.6	20.1	16.3	24.3
5th percentile	10.3	8.6	7.6	9.8	11.4	9.1	15.3
10th percentile	7.9	6.7	6.1	7.7	8.7	7.1	10.9
25th percentile	5.7*	4.9	4.5	5.7	5.9	5.3	7.0
50th percentile	4.1*	3.8	3.5	4.3	4.3	4.0	5.0
75th percentile	3.2*	3.0	2.9	3.4	3.4	3.2	3.8
90th percentile	2.7	2.5	2.4	3.0	2.9	2.8	3.1
95th percentile	2.4	2.3	2.2	2.6	2.6	2.5	2.8
99th percentile	2.0	1.9	1.9	2.1	2.1	2.0	2.3

*Used to define categories of performance. Category 1: ≥ 5.7 sec.; Category 2: 4.1–5.6 sec.; Category 3: 3.2–4.0 sec.; Category 4: ≤ 3.1 sec.

Table A2. Performance on Timed Repeat Chair Stands: Percentage Unable To Complete and Mean Times and Percentile Times for Those Completing the Task

	All	Men			Women		
	71+	71+	71–79	80+	71+	71–79	80+
n	5106	1786	1239	547	3320	2033	1287
Unable to complete (percent)	21.6%	17.7%	13.3%	27.6%	23.7%	16.4%	35.2%
Times (sec)							
Mean	14.5	13.7	13.2	15.0	14.9	14.4	16.1
1st percentile	30.0	27.9	26.6	31.9	31.9	29.1	35.6
5th percentile	23.2	21.5	20.2	24.5	24.1	22.6	26.6
10th percentile	20.4	19.0	18.0	21.0	21.1	20.1	23.3
25th percentile	16.7*	15.4	15.0	17.3	17.3	16.6	18.7
50th percentile	13.7*	13.0	12.6	14.0	14.1	13.7	15.0
75th percentile	11.2*	10.8	10.6	11.5	11.5	11.1	12.3
90th percentile	9.5	9.2	9.0	10.0	9.8	9.5	10.3
95th percentile	8.6	8.3	8.0	9.0	8.8	8.6	9.4
99th percentile	6.7	6.6	6.5	7.2	7.0	6.8	7.2

*Used to define categories of performance. Category 1: ≥ 16.7 sec.; Category 2: 13.7–16.6 sec.; Category 3: 11.2–13.6 sec.; Category 4: ≤ 11.1 sec.

Table A3. Distribution (%) of Performance Categories on Tests of Standing Balance

	All	Men			Women		
Category*	71+	71+	71–79	80+	71+	71–79	80+
0	9.5	7.9	5.9	12.5	10.4	5.7	17.8
1	14.5	10.5	7.2	18.0	16.6	12.0	23.9
2	13.0	10.6	8.6	15.0	14.3	12.6	16.8
3	13.8	12.6	12.1	12.3	14.7	15.4	13.6
4	49.2	58.9	66.2	42.3	44.1	54.3	28.0
	100	100	100	100	100	100	100
n	5104	1784	1238	546	3320	2030	1290

*See Table 1 for definition of categories.