Physical Performance Measures in the Clinical Setting

Stephanie Studenski, MD, MPH,† Subashan Perera, PhD,*‡ Dennis Wallace, PhD,‡‖ Julie M. Chandler, PhD, PT,¶ Pamela W. Duncan, PhD, PT,*§ Earl Rooney, MD, MHA,†#†† Michael Fox, ScD,*§ and Jack M. Guralnik, MD, PhD***

OBJECTIVES: To assess the ability of gait speed alone and a three-item lower extremity performance battery to predict 12-month rates of hospitalization, decline in health, and decline in function in primary care settings serving older adults.

DESIGN: Prospective cohort study.

SETTING: Primary care programs of a Medicare health maintenance organization (HMO) and Veterans Affairs (VA) system.

PARTICIPANTS: Four hundred eighty-seven persons aged 65 and older.

MEASUREMENTS: Lower extremity performance Established Population for Epidemiologic Studies of the Elderly (EPESE) battery including gait speed, chair stands, and tandem balance tests; demographics; health care use; health status; functional status; probability of repeated admission scale (Pra); and primary physician's hospitalization risk estimate.

RESULTS: Veterans had poorer health and higher use than HMO members. Gait speed alone and the EPESE battery predicted hospitalization; 41% (21/51) of slow walkers (gait speed <0.6 m/s) were hospitalized at least once, compared with 26% (70/266) of intermediate walkers (0.6–1.0 m/s) and 11% (15/136) of fast walkers (>1.0 m/s) ($P < .0001$). The relationship was stronger in the HMO than in the VA. Both performance measures remained independent predictors after accounting for Pra. The EPESE battery was superior to gait speed when both Pra and primary physician's risk estimate were included. Both performance measures predicted decline in function and health status in both health systems. Performance measures, alone or in combination with self-report measures, were more able to predict outcomes than self-report alone.

CONCLUSION: Gait speed and a physical performance battery are brief, quantitative estimates of future risk for hospitalization and decline in health and function in clinical populations of older adults. Physical performance measures might serve as easily accessible "vital signs" to screen older adults in clinical settings. **J Am Geriatr Soc 51:314–322, 2003.**

Key words: geriatric assessment; risk assessment; locomotion; health maintenance organizations; hospitalization; activities of daily living

Older people are at increased risk for worsening health, declining function, and high healthcare use, but this age group is diverse; interventions to prevent functional decline or avoid unnecessary use are costly and should be targeted toward those most at risk, making screening for high-risk older people a priority.[1] Medical diagnoses have limited capacity to define high-risk groups.[2] Functional status screening identifies older adults who need assistance, but functional assessment has not been incorporated into most clinical settings and may be insensitive to early disability.[3–11] Risk for hospitalization in older adults can be estimated using a measure that includes indicators for chronic diseases, health status, social support, and prior use.[12–14] Physical performance measures have potential as clinical risk indicators.[15–19] In a large epidemiological study, the Established Populations for Epidemiological Study of the Elderly (EPESE), the lower extremity performance battery, combining gait speed, chair rise capacity, and balance skills, discriminated risk of death and nursing home placement in the total older population.[20] In the subset of older adults with no self-reported disability reported from EPESE studies, the performance battery predicted functional decline and hospitalization over 4 years.[21,22] Evidence from epidemiological studies

From the *Center on Aging, †Department of Internal Medicine, ‡Department of Preventive Medicine, §Department of Health Policy and Management, University of Kansas Medical Center, Kansas City, Kansas; ‖Rho World, Inc., Chapel Hill, North Carolina; ¶Merck Research Laboratories, Blue Bell, Pennsylvania; #Department of Veterans Affairs, Kansas City, Missouri; and **Epidemiology, Demography and Biometry Program, National Institute on Aging, Bethesda, Maryland. ††Deceased.

Supported by Merck Research Laboratories, the Kansas Claude D. Pepper Older Americans Independence Center (AG 14635), and contracts from the National Institute on Aging.

Address correspondence to Stephanie Studenski, MD, MPH, Department of Medicine (geriatrics), VA Pittsburgh GRECC, 3471 Fifth Avenue Suite 500, Pittsburgh, PA 15213. E-mail: studenskis@msx.dept-med.pitt.edu

© 2003 by the American Geriatrics Society
0002-8614/03/$15.00

may not apply to clinical settings, where much more is known about the health of the individual. The purpose of this study was to determine whether simple physical performance measures can be used in real-world healthcare programs for older adults to predict future healthcare use, health status, and functional status after accounting for other information available to the healthcare system. This study also attempted to determine the most efficient approach, by assessing whether gait speed alone provided most of the information derived from the EPESE lower extremity performance battery.

METHODS

Overview

In this prospective cohort study, subjects were recruited from two primary care clinic sites and followed for 36 months. Results are presented here for the first 12 months of follow-up. The Veteran's Affairs (VA) and university institutional review boards approved this study. Informed consent was obtained from all participants.

Subjects

Participants for this study were recruited from a VA network site and a Medicare health management organization (HMO) serving a common geographic area. Eligible persons were aged 65 and older, lived in the community within a 20-mile radius of the ambulatory clinic site, and had been in the same healthcare system for at least 1 year. Mental status eligibility criteria were based on the need to maintain a use diary and self-report health. The Mini-Mental State Examination (MMSE)[23] was used to define groups with scores of 24 or higher (eligible), scores of 16–23 (eligible if they had a caregiver to maintain a diary), and scores less than 16 (excluded). Mobility screening, based on walking ability, excluded persons who were unable to walk at least 4 meters and those considered to be extremely fit or extremely fragile (gait speed >1.3 m/s or <0.2 m/s).

Of 572 individuals who agreed to be screened for this study, 487 met all criteria and entered the study. Of the 85 excluded, 21 failed mental status criteria, 16 failed mobility criteria, eight refused home visits, five had not been in care for 1 year, four lived outside the study radius, three received most of their care outside the system, and 28 were excluded for other reasons.

Measures

Participants had baseline assessments of demographic characteristics, cognition by MMSE,[23] depression by Geriatric Depression Scale,[24] social support,[25] self-report of comorbid conditions, an estimate of hospitalization risk, (the probability of repeated admission (Pra)).[12] The Pra includes age, sex, global health, social support, hospitalization or more than six physician visits in the last year, and presence of diabetes mellitus and coronary heart disease. The decision to collect Pra data was made after subject accrual began, so data on the first 45 patients were not obtained. Baseline and outcome rates did not differ between participants with and without Pra data.

Physical performance, health status, functional status, and symptoms were assessed at baseline and at 3, 6, 9, and 12 months. Health status was assessed using the European Quality of Life questionnaire (Euroqol)[26] and the global health item from the Medical Outcomes Study 36-item short-form health survey (SF-36), which asks respondents if their current health is excellent, very good, good, fair, or poor.[27,28] Functional status was assessed using items from the National Health Interview Survey (NHIS)[29] and the physical function index of the SF-36.[27,28] The NHIS explores difficulty and dependence in 16 activities related to activities of daily living (ADLs) and instrumental ADLs. The SF-36 physical function index examines limitations in 10 activities related to mobility and physical movements.

Performance testing was conducted twice during the baseline period, first during an outpatient clinic visit for eligibility and consent and second during a comprehensive home visit within 1 week. Testing was based on the EPESE lower extremity performance battery.[20] Gait speed for the separate gait speed variable was done over 4 meters, with a 1-meter start-up before starting timing.[30] All gait speed measures used the instruction to "walk at your usual speed."[20] Baseline physical performance is presented here as the average of the clinic and home observations. Performance testing continued at all scheduled clinic visits but is not further reported here.

Interrater and test-retest reliability over 2 weeks was assessed on the entire interview battery. Intraclass correlation coefficients (ICCs) were above 0.9 for interrater reliability of all items. Test-retest reliability ICCs were above 0.9 for all items except the Euroqol (ICC = 0.875) and the EPESE summary score (ICC = 0.723). Reliability of clinic-home baseline by ICC was 0.84 for gait speed and 0.82 for EPESE summary score.

Physician's risk estimate was used as another marker of healthcare system knowledge about a patient's future risks. Each participant's primary care physician was asked to estimate the likelihood of being hospitalized during the coming year on a scale ranging from 0% to 100% in increments of 10%. The decision to collect data on the physician risk estimate was made after subject accrual started. Data were obtained on 365 subjects: 73% of HMO participants and 89% of VA participants. Baseline characteristics and outcome rates did not differ between participants with and without physician risk assessments except for sex and race. Those with physician estimates were more likely to be veterans, male, and black than those without estimates.

Use was monitored in two ways. Health system data included hospitalizations and deaths. The participants or caregivers maintained healthcare diaries, which included hospitalizations and nursing home stays. The main purpose of the diaries was to capture out-of-system use, especially in the VA system. The two data sources were merged and discrepancies resolved by review of original medical documents and additional patient and provider interviews.

The three major outcome domains were use, health status, and functional status. A consensus panel of two physicians, two physical therapists, and a biostatistician preselected major outcomes, including criteria for significant change, using cross-sectional data and prior experience. The two measures of use were one or more hospitalizations and one or more adverse events, including hospitalization, nursing home placement, and death. The outcome "one or more adverse events" was designed to

avoid misclassification of persons with "competing events" such as nursing home placement or death, which might occur instead of hospitalization in frail persons. The two measures of decline in health status were Euroqol score decreased 0.1 points and global health declined two levels or reached the lowest level. The two measures of decline in functional status were NHIS functional status, indicating new difficulty in a basic ADL (eat, dress, bathe, toilet, groom, and transfer), and SF-36 physical function index decline of 10 points. Outcome events were considered to have occurred if they were detected at any point during the 12 months of follow up.

ANALYSIS

All analyses were performed using SAS software, version 8.0 (SAS Institute, Inc., Cary, NC). The statistical analyses were designed to assess the relationship between two baseline measures of physical performance (gait velocity and the EPESE summary score) and the six major outcome measures described above. Preliminary descriptive analyses of the baseline data were used to characterize the sample. Contingency tables and logistic regression models were used to assess relationships between physical performance and the longitudinal outcome measures.

Initial cross-sectional logistic regression analyses of baseline performance data against indicators of functional status were used to determine clinically valid cutpoints of 0.6 and 1.0 m/s on gait velocity.[31] EPESE ordinal categories were based on the original recommendations.[20] Ordinal scales based on these values were used in the subsequent univariate descriptive analyses for 1-year outcomes. The performance measures were assessed as continuous variables in all logistic regression analyses.

Analyses of the relationships between the two ordinal physical performance measures and each of the dichotomous outcome measures were conducted using contingency tables with formal inference based on chi-square statistics.[32] These analyses were initially stratified by provider system, with final analyses collapsed across systems if no system effect was observed. Logistic regression models were then used to assess the adjusted relationships between the two physical performance measures in their original continuous form and each of the six dichotomous variables.[32,33] An interaction term between physical performance and baseline status was included in each of the logistic regression models. It was significant only for the model addressing SF-36 physical function. Therefore, the analysis of SF-36 physical function is stratified by two levels of baseline status.

Finally, to compare the contribution of performance and self-report measures as predictors of all outcomes, models were fit that included one factor at a time and then combinations of factors. Differences in areas under the receiver operating characteristic (ROC) curves for these models were used to assess added value.[33] For all model-based analyses, gait velocity and the EPESE score were treated as continuous variables.

RESULTS

There were 487 participants: 140 from the VA system and 347 from the Medicare HMO (Table 1). The VA participants were all male and tended to be in worse health than those from the Medicare HMO. There were 34 participants (7%) without complete 12-month follow-up or major outcomes such as hospitalization or death; 20 changed provider systems, 12 withdrew, and two moved out of the study area.

Major predefined outcomes including use, decline in health status, and decline in function are described in Table 2. The VA population had higher rates of use and of

Table 1. Baseline Characteristics of the Sample: Total Population and By Site

Characteristic	Total Sample N = 487	Medicare Health Maintenance Organization n = 347	Veterans Affairs n = 140
Age, mean ± SD	74.1 ± 5.7	74.1 ± 5.8	74.2 ± 5.7
Female, n (%)*	213 (43.7)	213 (61.7)	0
Black, n (%)*	90 (18.5)	33 (9.5)	57 (40.7)
Education (high school or more)*	319 (65.5)	249 (71.8)	70 (50.0)
Gait speed, m/sec, mean ± SD†	0.88 ± 0.22	0.89 ± 0.21	0.83 ± 0.24
Established Population for Epidemiologic Studies of the Elderly score, mean ± SD†	8.4 ± 2.3	8.6 ± 2.1	7.9 ± 2.7
Mini-mental state examination cognition score, mean ± SD	27.5 ± 2.3	27.7 ± 2.3	27.1 ± 2.4
Depression score, mean ± SD	2.3 ± 2.7	2.4 ± 2.9	2.0 ± 2.4
Global health: fair or poor, n (%)*	108 (22.0)	58 (16.6)	50 (35.2)
Probability of repeated admission score, mean ± SD (n = 408)*‡	0.32 ± 0.11	0.29 ± 0.11	0.37 ± 0.11
SF-36, physical function, mean ± SD	64.3 ± 29.1	65.9 ± 28.8	60.3 ± 30.0
Congestive heart failure, n (%)	49 (10.0)	32 (9.1)	17 (12.1)
Diabetes mellitus, n (%)	84 (17.3)	53 (15.4)	31 (22.1)
Arthritis, n (%)	284 (58.3)	193 (55.6)	91 (65.0)

Difference between Veteran's Affairs and Medicare HMO groups, based on a two-sample *t* test for continuous measures and chi-square test for categorical measures.
* $P < .001$; † $P < .01$; ‡ $P < .05$.
SD = standard deviation; SF-36 = Medical Outcomes Study 36-item short form health survey.

Table 2. One-Year Cumulative Outcome Rates: Total Population and By Site

Outcome	Total Sample n = 453	Medicare Health Maintenance Organization n = 323	Veterans Affairs n = 130
	n (%)		
Use			
≥1 hospitalizations*	106 (23.2)	62 (19.2)	44 (33.9)
Deaths	18 (4.0)	10 (3.1)	8 (6.2)
≥1 adverse events (hospitalization, nursing home placement, death)*	115 (25.4)	68 (21.1)	47 (36.2)
Health status			
Significant decline in health status (Euroqol score decline ≥0.1)†	140 (31.4)	103 (32.3)	37 (29.1)
Significant decline in global health (drop two levels or become poor)*‡	53 (12.2)	28 (9.0)	25 (20.5)
Functional status			
New difficulty with personal care*†	122 (27.4)	72 (22.3)	50 (39.4)
Significant decline in physical function (36-item short-form health survey, physical function index change ≥10 points)†	122 (26.9)	86 (26.6)	36 (27.7)

* Site effect $P < .001$.
† n = 446, excludes seven persons who had adverse events in the first three months of the study.
‡ n = 435, excludes 11 persons who were in poor health at baseline and seven who had adverse events in the first 3 months of the study.

worsening health and function than did the Medicare HMO population.

Use

One or more hospitalizations occurred in 23% of participants, more in the VA (34%) than in the Medicare HMO (19%) (Table 2). Gait speed and the EPESE score predicted hospitalization in the Medicare HMO (Table 3). The EPESE score predicted hospitalization in the VA sample, whereas gait speed alone did not, although the direction of effect was sufficiently similar to allow data to be pooled. Because the EPESE score includes measures of chair rise and balance as well as gait speed, the better predictive ability in the VA is likely due to the chair or balance tasks. Gait speed and EPESE summary score remained predictors after adjusting for the Pra hospitalization risk score and age (Table 4). When the primary physician's prediction of hospitalization risk is included (Table 4), the EPESE

Table 3. Rates of Hospitalization Over 1 Year By Predictors and By Site

Factor	Medicare Health Maintenance Organization n = 323	Veterans Affairs n = 130
	n (%)	
Gait speed, m/sec		
<0.6	12 (41.4)	9 (40.9)
0.6–1.0	45 (23.6)	25 (33.3)
>1.0	5 (4.9)	10 (30.3)
P-value	$P < .0001$	Not significant
Established Population for Epidemiologic Studies of the Elderly score		
<7	19 (61.2)	20 (47.6)
7–9.5	32 (18.8)	15 (31.9)
10–12	11 (10.6)	9 (22.0)
P-value	$P < .0001$	$P = .0131$

Table 4. Gait Speed and Established Population for Epidemiologic Studies of the Elderly Performance Battery as Predictors of Hospitalization, Sites Combined, Adjusted for Pra Alone and then Pra and Physician Estimate

Variable	Risk Adjusted for Pra n = 408		Risk Adjusted for Pra and Physician's Estimate n = 328	
	Odds Ratio	*P*-Value	Odds ratio	*P*-Value
Gait speed (unit = 0.2 m/sec)	0.62	.002	0.6	.16
Hospitalization risk by Pra (unit = 0.2)	1.9	.03	1.8	.046
Physician risk estimate (unit = 10%)			1.2	.033
EPESE (unit = 2 points)	0.68	.0004	0.79	.05
Hospitalization risk by Pra (unit = 0.2)	1.9	.002	1.6	.07
Physician risk estimate (unit = 10%)			1.14	.05

Note: All analyses are age adjusted.
Pra = probability of repeated admission scale.

Table 5. Change in Health Status Over 1 Year, Sites Combined

Predictor	Decline in Global Health n = 435	Decline in Euroqol Score n = 446
	n (%)	
Univariate		
Gait speed (m/sec)		
<0.6	16 (35.6)	24 (50.0)
0.6–1.0	29 (11.3)	82 (31.1)
>1.0	8 (6.0)	34 (25.4)
P-value	<.0001	.0039

score remains a predictor of hospitalization, whereas gait speed alone becomes marginal. Analyses of the other use outcome, one or more adverse events including hospitalization, nursing home placement, and death, revealed results similar to hospitalization alone in both healthcare systems (results not shown).

Health Status

The two health status outcomes were decline in global health and decline in Euroqol score (Table 2). Persons with the lowest global health level at baseline are not included in the global health analysis because further decline from the lowest initial level cannot be detected. Global health declined in 12% of the participants. The Euroqol score declined at least 0.1 points in 30% of the participants. Gait speed and EPESE score effects were similar in the VA and Medicare HMO populations; consequently, the results are pooled across systems, and only those for gait speed are shown (Tables 5 and 6). Gait speed and EPESE battery were strong univariate predictors of global health decline, less so for Euroqol score decline. After adjustment for baseline status and age (Table 4), physical performance measures were strong predictors of global health decline and Euroqol decline.

Functional Status

The two functional status outcomes were new difficulty in personal care and decline in physical function. New difficulty in personal care activity occurred in 28% of participants, whereas physical function decline developed in 27%. Gait speed and EPESE score were strong predictors of new personal care difficulty in both the Medicare HMO and VA samples; consequently, results are pooled (Table 7 and 8, only those for gait speed shown). Slow walkers had a 69% incidence of new personal care difficulty; only 12% of the fast walkers developed such difficulty. Physical performance remained a strong predictor of new ADL difficulty after adjustment for baseline function and age. The ability of physical performance measures to predict decline in physical function, as measured by the SF-36, was heavily influenced by the initial SF-36 score. Those with higher baseline function (SF-36 physical function score ≥70) and poor performance had a slightly increased risk for decline. In those persons who reported lower baseline physical function (SF-36 score <70), performance measures did not predict who would experience further decline.

Table 7. Change in Function Over 1 Year, Sites Combined

Predictor	New Difficulty in Personal Care n = 446	Decline in Physical Function n = 446	
		SF-36 <70 n = 200	SF-36 ≥70 n = 246
	n (%)		
Gait speed (m/sec)			
<0.6	33 (68.8)	12 (27.9)	3 (60.0)
0.6–1.0	73 (27.6)	32 (25.8)	44 (32.4)
>1.0	16 (11.9)	10 (34.5)	20 (19.0)
P-value	<.0001	Not significant	.0101

SF-36 = 36-item short-form health survey; OR = odds ratio.

Table 6. Change in Health Status Over 1 Year, Adjusted for Age and Baseline Health

Predictor	Decline in Global Health n = 435		Decline in Euroqol Score n = 446	
	Odds Ratio	*P*-value	Odds Ratio	*P*-value
Gait speed (unit = 0.2 m/sec)	0.49	.0001	0.63	.0001
Baseline health (baseline raw score)	0.66	.02	1.39	.0001

Comparative Predictive Ability of Self-Report, Performance, and Combinations

The comparative ability of self-report measures and performance measures, alone and in combination, to predict 1-year outcomes is described in Table 9. Predictive ability in logistic models with dichotomous outcomes can be assessed using the area under the ROC curve.[33] These areas estimate the ability of the proposed variables to predict the outcome beyond chance and run from about 0.5 to 1.0. A 2.5% (0.025) difference in areas is considered meaningful. In Table 9, the first column on the left gives the outcome to be predicted, the second column gives the area predicted by the self-report measure alone, the next two columns give the areas for each performance measure alone (gait speed or EPESE score), and the last two columns give the areas for gait speed or EPESE score in combination with the self-report measure. Analyses to predict hospitalization were performed using the Pra score (a self-report of eight factors found to predict hospitalization), primary physician hospitalization risk estimate, and the two together. Results are provided for the HMO sample, the VA sample, and the pooled sample.

Table 8. Change in Function Over 1 Year, Adjusted for Age and Baseline Function

	New Difficulty in Personal Care n = 446		Decline in Physical Function n = 446			
			SF-36 <70 n = 200		SF-36 ⩾70 n = 246	
Predictor	OR	*P*-value	OR	*P*-value	OR	*P*-value
Gait speed (unit = 0.2 m/sec)	0.62	.0003	0.26	.23	.72	.06
Baseline function (unit = raw score)	3.01	.0001	1.7	.0001	1.04	.81

OR = odds ratio; SF-36 = 36-item short form health survey.

When predicting hospitalization in the HMO using the Pra score, EPESE score was at least as good as, and gait speed was marginally better than, Pra (area for Pra = 0.649, gait speed = 0.677, and EPESE = 0.656). The combination of EPESE score or gait speed with Pra score was somewhat better (0.687 and 0.686, respectively). In the VA sample, the EPESE score was far superior to gait speed or Pra (EPESE score area = 0.630, Pra = 0.578, and gait speed = 0.532). Adding Pra score to EPESE score did not improve predictive ability (EPESE score alone = 0.630, EPESE and Pra = 0.631). Primary physician estimate of risk demonstrated similar results. The combination of Pra and physician risk is superior to EPESE score and about the same as gait speed in the HMO (area for Pra and physician risk= 0.691, gait speed = 0.683, and EPESE score = 0.663). Adding performance measures to PRA and physician risk estimate does not improve predictive ability. In the VA sample, neither EPESE score nor the combination of Pra and physician risk is a good predictor (EPESE = 0.599, Pra and physician risk = 0.601). Overall, in the HMO, the predictive ability of EPESE score and Pra (0.687) were approximately equivalent to that of Pra and primary physician risk estimate (0.691). In the VA, EPESE score alone (0.630) did as well as any combination of measures.

For models predicting decline in health, using the self-report of global health or the Euroqol score, baseline self-report of health status was a better predictor than performance in the HMO, but performance was better than self-report in the VA. The combination of self-report and performance was better than either alone.

For functional decline, defined as new difficulty in personal care, EPESE score alone was at least as good as self-report of functional status in the HMO setting (EPESE score = 0.772, self-report of function = 0.761), and the combination did not substantially improve the area under the curve (0.784). In the VA, EPESE score (0.785) was at least as good as self-report (0.771), but the combination was substantially better than either alone (0.838). When functional decline was defined as decline in SF-36 score, the effects were different depending on whether the initial SF-36 score was high or low. In persons who self-reported lower physical function at baseline, performance measures did not aid in predicting decline. In persons who reported high physical function at baseline, the EPESE score alone was a predictor in the HMO (area = 0.677), whereas in the VA, self-report and EPESE score together were most able to predict decline (area = 0.666).

Gait speed did at least as well as EPESE score in predicting hospitalization in the HMO, but EPESE was better in the VA. Gait speed and EPESE score were equivalent in predicting health decline in both the VA and the HMO. EPESE was superior to gait speed in predicting functional decline.

Feasibility and Acceptability

In the outpatient clinic, gait speed testing took less than 2 minutes in three-quarters of participants, and only one person required more than 3 minutes. The full EPESE battery testing took less than 5 minutes in more than three-quarters of participants, and the longest time to complete it was 7 minutes. Healthcare professionals and staff (n = 32) from each site were asked about the acceptability of performance testing in their setting, where these tests were done on their patients at baseline and during regularly scheduled follow-up visits. More than half felt the testing was very acceptable, and the rest felt it was acceptable. Sixty participants were asked about the acceptability of performance testing during clinic visits. Only two felt it was less than acceptable.

DISCUSSION

Physical performance measures are independent predictors of use, change in health status, and decline in function in a primary care clinical environment, even after accounting for baseline status, age, a well-established hospital risk estimator, and primary physician's risk estimate. Performance measures alone, or in combination with self-report measures, were able to predict outcomes more than self-report alone. Performance measures have the potential to be incorporated into clinical practice as convenient global markers or "vital signs" for health-related risk in older adults.

Why would physical performance be a rational marker of future health risk in older adults? Physical function measures in general appear to integrate the effects of multiple facets of health and aging, including disease processes, nutritional status, fitness, and emotional state.[4–7,16,18] Physical performance measures distinguish outcomes in persons who self-report no functional problems, thus complementing self-report alone.[21,22] More-detailed assessments of dis-

Table 9. Area Under the Receiver Operating Characteristic Curve for Major Outcomes using Self-Report Measures, Gait Speed or Established Population for Epidemiologic Studies of the Elderly (EPESE) Score Alone, and Measures Combined

Outcome	Self-Report Measure	Gait Speed	EPESE Score	Gait Speed and Self-Report	EPESE Score and Self-Report
Pra score* (n = 408)					
Hospitalization					
HMO	0.649	0.677	0.656	0.686	0.687
VA	0.578	0.532	0.630	0.580	0.631
Pooled	0.644	0.632	0.656	0.656	0.677
Physician risk estimate (n = 365)					
Hospitalization					
HMO	0.665	0.683	0.663	0.690	0.669
VA	0.599	0.506	0.599	0.603	0.617
Pooled	0.642	0.630	0.647	0.657	0.662
Pra score* and physician risk estimate (n = 328)					
Hospitalization					
HMO	0.691	0.683	0.663	0.701	0.704
VA	0.601	0.506	0.599	0.602	0.611
Pooled	0.672	0.630	0.647	0.673	0.679
Global health (n = 435)					
Health-global change					
HMO	0.678	0.625	0.623	0.701	0.686
VA	0.664	0.743	0.756	0.820	0.799
Pooled	0.656	0.691	0.689	0.741	0.742
Euroqol score (n = 446)					
Health-Euroqol change					
HMO	0.624	0.578	0.597	0.662	0.674
VA	0.493	0.631	0.629	0.672	0.683
Pooled	0.592	0.587	0.602	0.653	0.676
Functional status (dependence in basic or instrumental activities of daily living) (n = 446)					
Function—new personal care difficulty					
HMO	0.761	0.690	0.772	0.773	0.784
VA	0.771	0.748	0.785	0.830	0.838
Pooled	0.778	0.721	0.752	0.798	0.807
SF-36 physical function score (n = 446)					
SF-36 <70 (n = 200)					
Function—SF-36 change					
HMO	0.751	0.485	0.511	0.758	0.758
VA	0.712	0.582	0.586	0.719	0.745
Pooled	0.734	0.521	0.534	0.747	0.744
SF-36 ≥70 (n = 246)					
Function—SF-36 change					
HMO	0.594	0.639	0.677	0.649	0.680
VA	0.617	0.569	0.574	0.638	0.666
Pooled	0.539	0.592	0.653	0.594	0.652

Note: All analyses are age adjusted.

*Pra score = hospitalization risk by probability of repeated admission (Pra); HMO = health maintenance organization; VA =Veterans Affairs; SF-36 = 36-item short-form health survey.

ease presence and severity or reports of subtle functional change can be sensitive markers of future decline and mortality,[11,34] but they take time and resources. In this study, physical performance remained an independent predictor of hospitalization even after including the physician's global estimate. Even though physicians are likely to have information unavailable in epidemiological studies or medical records, performance data might add further to their ability to identify high-risk older adults.

How might physical performance data be used clinically? If performance is to be useful as a vital sign or a screening test, then a differential diagnosis of the causes of poor performance and interventions to change outcomes must be available. Ferrucci et al. have proposed a structured approach to gait disorders that includes cardiopulmonary, neurological, and musculoskeletal disorders.[35] It may be possible to develop and test an evaluation and intervention process for abnormalities of gait and balance.[36]

Physical performance, like blood pressure, may be a general indicator that reflects several underlying physiological processes and predicts a range of important future events. Clinical acceptance of high blood pressure as an important medical condition required developing an approach to measurement, assembling an evidence base, and creating a diagnostic term (hypertension).[37,38] Physical performance techniques have evolved an approach to standardized measurement and continue to assemble an evidence base. Is it time to name a condition? Geriatricians recognize "dysmobility" as a condition.[39] "Bradypedia" is brief and descriptive.

Which performance measure makes sense in a clinical setting? The EPESE battery had some additional explanatory power over gait speed alone for several outcomes, although at the cost of additional time spent. The value of the entire battery over gait speed alone was examined in a pooled EPESE analysis predicting disability; gait speed contributed most but not all the explanatory power.[40] In another epidemiological study, gait speed alone performed as well as the whole battery in predicting mortality.[41] The whole battery takes two to three times as long to complete as does gait speed alone. The two additional tasks, chair rise and tandem stands, have strong face validity. The complete battery may discriminate risk within subgroups with high gait speeds and may help further define diagnostic possibilities.

What are some benefits and costs of using performance measures in the clinical setting? Benefits might include risk prediction and aid with diagnosis or treatment. Performance measures, with and without self-report, help predict risk of decline and adverse events. Performance measures may provide a basis for a medical differential diagnosis and for planning medical or restorative interventions. Potential costs include staff time and burdens to older adults due to the health risk of testing or fatigue. Using performance measures in the clinical setting would demand staff time but little equipment or health risk. Self-report measures might use less staff time if they are self-administered, but some self-report measures require staff time for administration or scoring. In the future, further research may identify better combinations of measures, whether self-report or performance based, that offer even better clinical value for cost. This study provides information about the value of performance measures relative to some common self-report measures against important geriatric outcomes.

The limitations of this study include the use of a single metropolitan area, the differences in use patterns in the two systems, and lack of data on hospitalization risks for a subset of subjects. Although only one metropolitan area was used, there were two provider systems, and participants had a wide range of socioeconomic, health, and functional states, thus improving generalizability.

Hospitalization rates in the VA setting were higher overall. Gait speed did not predict hospitalization in this setting, whereas EPESE score did. Because the EPESE score includes two tasks in addition to gait speed, it is likely that the chair rise, the tandem stand, or both provide the additional explanatory value in this setting and population. The decision to hospitalize in the VA setting may differ from the Medicare HMO setting. Characteristics of the veteran population itself, such as higher rates of social and mental health problems, might contribute to these findings. Practice varies regionally in the VA, as in other provider systems, so this finding may not apply to the VA system in general.[42]

The decision to add additional measures of hospital risk was made after the study began. Thus the Pra score and physician global risk estimate are missing in the first 10% and 20% of subjects who entered the study. The baseline characteristics and hospitalization rates of subjects who did and did not have these measures were examined, and no significant differences were found. Although power was reduced because the sample size was smaller, the estimates of effect have a reasonable likelihood of being valid.

If performance was tested in the clinical setting, some patients might intentionally alter their performance by moving more quickly or slowly. This study used a standard "usual gait speed" based on the EPESE protocol. Numerous approaches to assessing performance have been used to study older adults, and further refinement of protocols and instructions is an active area of study.[43] Although motivation affects many clinical measures, including pulmonary function tests and aerobic fitness tests, they are still useful in a large proportion of the population.

This study averaged baseline performance across clinic and home. Although agreement rates were high between clinic and home performance, with intraclass correlations greater than 0.8, some slow walkers tended to walk more quickly in the clinic setting. This may represent an equivalent of "white coat hypertension," where values in a clinical setting are not truly representative. Repeated measures of blood pressure across time and settings are now recommended before applying a diagnosis of hypertension.[44] It was felt that a repeated baseline measure of physical performance across settings was the best estimate of the truth. If performance continues to be considered a potentially useful clinical indicator, then further examination of site effects is warranted.

Performance measures may be basic integrators of the health of older people. Through the ages, common people have celebrated the older gentleman who "still has a spring in his step" as doing exceptionally well and expressed concern for the aging neighbor who "seems to be slowing down." Perhaps it is time to bring this folk wisdom into mainstream clinical practice with simple quantitative measures. Use of brief performance measures as "vital signs" could offer providers and healthcare systems a new tool for recognizing impending needs of older adults.

REFERENCES

1. Stuck AE, Siu AL, Wieland GD et al. Comprehensive geriatric assessment: A meta-analysis of controlled trials. Lancet 1993;342:1032–1036.
2. Waite K, Oddone E, Weinberger M et al. Lack of association between patient's measured burden of disease and risk for hospital readmission. J Clin Epidemiol 1994;11:1229–1236.
3. Ikegami N. Functional assessment and its place in health care. N Engl J Med 1995;332:598–599.
4. Applegate WB, Blass JP, Williams TF. Instruments for the functional assessment of older patients. N Engl J Med 1990;322:1207–1214.
5. Inouye SK, Paduzzi PN, Robison JT et al. Importance of functional measures in predicting mortality among older hospitalized patients. JAMA 1998;279: 1187–1193.
6. Scott WK, Macera CA, Cornman CB et al. Functional health status as a predictor of mortality in men and women over 65. J Clin Epidemiol 1997;50: 291–296.

7. Silliman RA, Barry PB. Outpatient comprehensive geriatric assessment:; An intervention whose time has come, or has it? J Am Geriatr Soc 1999;47:371–372.
8. Williams MA. Why screen for functional disability in elderly persons? Ann Intern Med 1990;112:639.
9. Lachs MA, Feinstein AR, Cooney LM et al. A simple procedure for general screening for functional disability in elderly patients. Ann Intern Med 1990; 112:699–706.
10. Davis RB, Iezzoni LI, Phillips RS et al. Predicting in-hospital mortality: The importance of functional status information. Med Care 1995;33:906–921.
11. Fried LP, Bandeen-Roche K, Williamson JD et al. Functional decline in older adults. Expanding methods of ascertainment. J Gerontol A Biol Sci Med Sci 1996;51A:M206–M214.
12. Pacala JT, Boult C, Boult L. Predictive validity of a questionnaire that identifies older persons at risk for hospital admission. J Am Geriatr Soc 1995;43: 374–377.
13. Pacala JT, Boult C, Reed RI et al. Predictive validity of the Pra instrument among older recipients of managed care. J Am Geriatr Soc 1997;45:614–617.
14. Coleman EA, Wagner EH, Grothaus LC et al. Predicting hospitalization and functional decline in older health plan enrollees: Are administrative data as accurate as self-report? J Am Geriatr Soc 1998;46:419–425.
15. Hall WJ. Update in geriatrics. Ann Intern Med 1996;125:390–397.
16. Reuben DB, Siu A, Kimpau S. The predictive validity of self-report and performance based measures of function and health. J Gerontol 1992;47:M106–M110.
17. Williams ME. Identifying the older person likely to require long term care services. J Am Geriatr Soc 1987;35:761–766.
18. Guralnik JM, Branch LG, Cummings SR et al. Physical performance measures in aging research. J Gerontol 1989;44:M141–M146.
19. Gill TM, Williams CS, Tinetti ME. Assessing risk for the onset of functional dependence among older adults: The role of physical performance. J Am Geriatr Soc 1995;43:603–609.
20. 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 placement. J Gerontol 1994;49:M85–M94.
21. Guralnik JM, Ferrucci L, Simonsick EM et al. 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.
22. Penninx BWJH, Ferrucci L, Leveille SG et al. Lower extremity performance in nondisabled older persons as a predictor of subsequent hospitalization. J Gerontol A Biol Sci Med Sci 2000;55A:M691–M697.
23. 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.
24. Yesavage JA, Brink TL, Rose TL et al. Development and validation of a geriatric depression screening scale. J Psychiatr Res 1982;17:37–49.
25. Pfeiffer E, ed. Multidimensional Functional Assessment: The OARS Methodology A Manual, 2nd Ed. Durham, NC: Duke University Center for Study of Aging and Human Development, 1978.
26. The Euroqol Group. Euroqol—a new facility for the measurement of health related quality of life. Health Policy 1990;16:199–208.
27. Ware JF, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): 1. Conceptual framework and item selection. Med Care 1992;30:473–483.
28. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Med Care 1988;26:724–735.
29. Fitti JE, Kovar MG. The Supplement on Aging to the 1984 National Health Interview Survey. Vital Health Stat 1 1987;21:1–115.
30. Rigler S, Studenski S, Wallace D. Translating gait speed measures. J Am Geriatr Soc 1997;45:S27.
31. Studenski S, Wallace D, Belt J et al. Efficient use of performance measures in the clinical setting. J Am Geriatr Soc 1997;45:S115.
32. Agresti A. Categorical Data Analysis. New York: Wiley, 1990.
33. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley, 2000.
34. Fried LP, Kronmal RA, Newman AB et al. Risk factors for 5-year mortality in older adults. JAMA 1998;279:585–592.
35. Ferrucci L, Bandinelli S, Benvenuti E et al. Subsystems contributing to the decline in ability to walk: Bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. J Am Geriatr Soc 2000;48:1618–1625.
36. Rubenstein LZ, Powers CM, Maclean CH. Quality indicators for the management and prevention of falls and mobility problems in vulnerable elders. Ann Intern Med 2001;135:686–693.
37. Crenner CW. Introduction of the blood pressure cuff into U.S. medical practice: Technology and skilled practice. Ann Intern Med 1998;128:488–493.
38. Master AM, Garfield CI, Walters MB. Normal Blood Pressure and Hypertension. Philadelphia: Lea & Febiger Publishers, 1952.
39. Mouton CP, Espino DV. Health screening in older women. Am Fam Physician 1999;59:1747–1748,1750.
40. Guralnik JM, Ferrucci L, Pieper CF et al. Lower extremity function and subsequent disability: Consistency across studies, predictive models and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 2000;55A:M221–M231.
41. Markides KS, Balck SA, Ostir GV et al. Lower body function and mortality in Mexican-American elderly people. J Gerontol A Biol Sci Med Sci 2001; 56A:M243–M247.
42. Ashton CM, Petersen NJ, Souchek J et al. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. N Engl J Med 1999;340:32–39.
43. Alexander NB. Gait disorders in older adults. J Am Geriatr Soc 1996;44: 434–451.
44. Rigaud A, Forette B. Hypertension in older adults. J Gerontol A Biol Sci Med Sci 2001;56A:M217–M225.