Clinical Assessment of Function among Women with a Recent Cerebrovascular Event: A Self-Reported versus Performance-Based Measure

Pamela L. Owens, PhD; Elizabeth H. Bradley, PhD; Sarah M. Horwitz, PhD; Catherine M. Viscoli, PhD; Walter N. Kernan, MD; Lawrence M. Brass, MD; Philip M. Sarrel, MD; and Ralph I. Horwitz, MD

Background: **Self-reported functional status is a commonly used health measure in clinical settings, yet the optimal approach for assessing function is often debated.**

Objective: **To examine the agreement between a self-reported and a performance-based measure of function and the relative ability of each measure to predict long-term health outcomes.**

Design: **Prospective cohort study.**

Setting: **20 hospitals in Connecticut and Massachusetts.**

Participants: **620 postmenopausal women (46 to 91 years of age) who had experienced a stroke or transient ischemic attack.**

Measurements: **A self-reported and a performance-based measure of function were assessed at baseline (before intervention) by using the Barthel index and the Physical Performance Test.**

Results: **Disagreement between the self-reported and performance-based measure of function was common (slight disagreement, 55.0%; substantial disagreement, 19.3%). Most women (95.4%) overreported their level of function. Women who were more clinically impaired (risk ratio [RR] for more comorbid conditions, 1.52 [95% CI, 1.17 to 1.97]; RR for recent stroke, 2.33 [CI, 1.45 to 3.73]; and RR for cognitive impairment, 1.76 [CI, 1.34 to 2.32]); who were less educated (RR = 1.30 [CI, 1.02 to 1.67]); and who were of nonwhite ethnicity (RR 1.43 [CI, 1.07 to 1.91]) were more likely to overreport their level of function. An impaired performance-based measure of function predicted subsequent stroke or death (hazard ratio, 1.50, [CI, 1.06 to 2.11]); however, an impaired self-reported measure of function was not likely to predict these outcomes.**

Conclusions: **Clinicians should be aware that results of self-reported and performance-based measures of function can differ in women who have experienced a recent cerebrovascular event. Although more difficult to collect, results of a performance-based measure may provide information about long-term health outcomes that is not available from a self-reported measure.**

Ann Intern Med. 2002;136:802-811. www.annals.org

For author affiliations, see end of text.

Self-reported measures of function are commonly used by clinicians and clinical researchers to assess an individual's level of independence in basic self-care skills (1). Self-reported indices are inexpensive, easy to administer, able to detect severe levels of disability, and able to provide information about usual performance of basic tasks in the home environment (2). On the basis of these assessments, clinicians and clinical researchers often categorize individuals' levels of functional independence and then use these categories to determine long-term prognoses, appropriateness of medical treatments and surgical therapies, and need for rehabilitative services or nursing home admission (1–4).

Although self-reported function measures are often used in clinical settings, the optimal approach for assessing functional status is still being debated (5, 6). Some clinicians and researchers prefer self-reported functional status indices, whereas others have argued that performance-based measures are more objective, capable of detecting slight disabilities, free of reporting bias, sensitive to change, easily reproducible, and clinically relevant for determining treatment effectiveness (2, 5–9). Thus, the question of which type of measure (self-reported or performance-based) is most appropriate for clinical settings remains unresolved.

Although many clinicians and clinical researchers believe that self-reported and performance-based measures of function assess the same capabilities, several researchers have reported that the two types of functional status indices measured on continuous scales produce different results (8, 10–22). Only a few studies have measured agreement between more clinically relevant self-reported and performance-based levels of function and identified the source of disagreement (23–25). These studies are limited, however, by small sample size or lack of report of overall agreement (rather than agreement between specific tasks). For example, in an early study of 35 psychiatric inpatients, Kuriansky and coworkers (23) found poor agreement (unweighted κ, 0.00) between self-reported and performance-based

 © 2002 American College of Physicians–American Society of Internal Medicine

measures of function. Another group of researchers (24) documented that some participants reported higher levels of function than they actually performed on specific tasks, but the researchers did not report κ values. Another study indicated moderate agreement (κ, 0.41 to 0.55) between specific self-reported and performance-based tasks; however, this study focused on specific tasks rather than overall function (25). No study has classified individuals into functional levels and explicitly examined the overall agreement and reasons for disagreement between self-reported and performance-based functioning.

In addition to assessing agreement between a self-reported and performance-based measure of function, we sought to understand the relative contribution of these measures to long-term outcomes. Only one study has examined self-reported and performance-based function simultaneously in a multivariate model predicting mortality (8). In this study of lower-extremity function among older women, researchers found self-reported disability to be associated with increased mortality, adjusting for age. These effects were attenuated, but remained significant, in the presence of performance-based functioning. Although the study provided a foundation for assessing the predictive validity of self-reported function relative to performance-based function, the lack of adjustment for other clinical conditions limited its usefulness.

The objectives of our study were to 1) assess the agreement between a self-reported and a performance-based measure of function, using clinically relevant levels of function; 2) examine the predictors of disagreement between the two measures; and 3) examine the association of each measure of function with subsequent stroke and death among postmenopausal women who participated in a randomized clinical trial.

Context

Assessment of patients' functional capacity is essential to identify needs for care after a cerebrovascular event.

Although often used, measures based on patient report of functional status may not adequately assess some patients' function.

Contribution

Researchers compared function measured by self-report and by a clinician watching the patients perform standardized activities. The two measures often disagreed.

Women tended to report better function than the performance measure showed.

A performance-based test predicted future stroke or death. A self-reported measure of function did not.

Implications

Clinicians should be aware that women may report better function after a cerebrovascular event than a performance test would show.

–The Editors

METHODS

Design

Data for this study were obtained from the Women's Estrogen for Stroke Trial (WEST). Described in detail elsewhere (26), this randomized, double-blind, placebo-controlled trial of 652 postmenopausal women who experienced a recent stroke or transient ischemic attack was designed to determine the effects of estradiol, as compared with placebo, on the risk for stroke or death. As part of the trial, women completed physical, cognitive, and psychosocial testing at the time of randomization (within 90 days of their stroke or transient ischemic attack). At baseline, all of the women (n = 652) completed a self-assessment of function, and almost all (n = 640) completed a performance-based assessment of function. Consistent with the work by Kempen and coworkers (17), women who had scores less than 17 (n = 16) on the Mini-Mental State Examination (MMSE) (27) and women who had an incomplete MMSE (n = 4) were excluded from the study because it was uncertain whether they could provide valid self-reports of function. Thus, the sample for the analyses included 620 women.

Measures

Dependent Variables

The outcome of interest for the first two objectives of our study was the extent of disagreement between a self-reported and a performance-based measure of function. The 10-item Barthel index, originally designed as a hospital discharge index for stroke patients (28) (**Appendix Figure 1**, available at www.annals.org), was used to determine the baseline self-reported measure of func-

tion. Interviewers asked each woman about her ability to eat, transfer, groom, use a toilet, bathe, walk, climb stairs, dress, and control bowel and bladder functions. Following standard practice, interviewers asked women for additional information on the basis of their responses and categorized the level of independence for each activity. Consistent with earlier research (29, 30), Barthel index scores were categorized into four levels. A score of 20 indicated independence, a score of 18 to 19 indicated slight dependence, a score of 16 to 17 indicated moderate dependence, and a score of 15 or less indicated severe dependence. Additional analyses of the frequency distribution of the subscale scores confirmed that these cut-points were appropriate.

The Physical Performance Test (31) (**Appendix Figure 2**, available at www.annals.org), given after the Barthel index, was used for the baseline performance-based measure of function. Scores on the seven-item Physical Performance Test were based on the time required for completion of each task (writing, eating, lifting, dressing, bending, turning, and walking). Consistent with earlier research (with 8-point adjustments made for a scale of 0 to 28 rather than 0 to 36) (32), Physical Performance Test scores were categorized into four levels. A score of 24 to 28 indicated independence, a score of 17 to 23 indicated slight dependence, a score of 9 to 16 indicated moderate dependence, and a score of 8 or less indicated severe dependence. Additional analyses of the frequency distribution of the subscale scores confirmed that these cut-points were appropriate.

Disagreement between the two measures was divided into three categories. Clinical agreement (no clinical disagreement) between the two scales was defined as the same level of function (independent, slight dependence, moderate dependence, and severe dependence) on both scales. Slight clinical disagreement between the two scales was defined as a difference in one functional level between the scales. Substantial clinical disagreement between the two scales was defined as a difference in two or more functional levels between the scales.

The outcome of interest for the third objective of our study was the time to stroke or time to death—confirmed end points of the clinical trial. Every 3 months, researchers contacted participants or, when not available, their proxies to determine whether new events or deaths had occurred.

Independent Variables

Features of interest for the second and third objectives of our study included baseline sociodemographic, clinical, and psychosocial characteristics. Sociodemographic characteristics included age (46 to 64 years, 65 to 74 years, or ≥75 years), ethnicity (nonwhite or white), and education (<12 vs. ≥12 years). Clinical characteristics included number of self-reported chronic conditions (including cardiac conditions, cerebrovascular conditions, and diabetes [2 to 3 vs. 0 to 1]); type of index event (stroke or transient ischemic attack); severity of index event, as measured by the National Institutes of Health stroke scale (severe [≥4] vs. less severe [0 to 3]) (33); cognitive status, as measured by the MMSE (<24 vs. ≥24) (27, 34); smoking status (current vs. former or never); and randomized treatment assignment (active vs. placebo). Psychosocial characteristics included depressive symptoms, as measured by the Center for Epidemiologic Studies Depression Scale (CES–D) (≥16 vs. <16) (35); health locus of control, as measured by strong agreement with the statement, "I am directly responsible for my health"; religiosity, as measured by strong agreement with the statement, "Religion is a great source of strength"; marital status (married or living with partner vs. other); number of close friends and relatives (at least one to three friends or relatives vs. none); and social support (adequate support [defined as having someone to talk to, provide basic care needs, and turn to for advice] vs. inadequate support).

The primary independent variables for the third objective of our study were self-reported functional status, as measured by the Barthel index (28), and performance-based functional status, as measured by the Physical Performance Test (31). Because clinicians and clinical researchers are predominantly concerned with the long-term effects of moderate and severe functional dependence, the functional status indices were dichotomized: moderate and severe dependence (Barthel index <18; Physical Performance Test <17) vs. independent and slight dependence (Barthel index ≥18; Physical Performance Test ≥17).

Statistical Analysis

After completing a test for marginal homogeneity, we examined agreement between the Barthel index and Physical Performance Test using the Fleiss–Cohen

Table 1. **Comparison of Functional Status Classification Based on the Barthel Index and the Physical Performance Test for 620 Women***

Physical Performance Test (Scores)	Barthel Index (Scores)				Total
	Independent (Score, 20)	Slight Dependence (Score, 18–19)	Moderate Dependence (Score, 16–17)	Severe Dependence (Score, ≤15)	
	← *n (%)* →				
Independent (score, 24–28)	97†	11‡	0§	0§	108 (17.4)
Slight dependence (score, 17–23)	257‡	43†	4‡	0§	304 (49.0)
Moderate dependence (score, 9–16)	89§	49‡	7†	6‡	151 (24.3)
Severe dependence (score, ≤8)	15§	16§	14‡	12†	57 (9.2)
Total	458 (73.9)	119 (19.2)	25 (4.0)	18 (2.9)	620

* Test of marginal homogeneity: chi-square test, 140.4; weighted $\kappa = 0.279$ (95% CI, 0.223 to 0.335).
† Women who had clinical agreement between self-reported and performance-based measures of function (159 of 620 women [25.7%]).
‡ Women who slightly overreported or underreported physical performance (341 of 620 women [55.0%]).
§ Women who substantially overreported or underreported physical performance (120 of 620 women [19.3%]).

weighted κ statistic (36). The magnitude of the weighted κ is interpreted in the same fashion as an unweighted κ ($\kappa \geq 0.75$ = excellent agreement; $\kappa \leq 0.40$ = poor agreement) (37).

The association between baseline characteristics and baseline disagreement in functional status measures was assessed in both bivariate (Pearson chi-square tests) and multivariate analyses (logistic regression models). To identify the potential correlates of disagreement between functional status measures, two logistic regression models (slight clinical disagreement vs. no clinical disagreement; substantial clinical disagreement vs. no clinical disagreement) were derived. Sociodemographic characteristics were considered for inclusion in the models first, followed by clinical (except treatment assignment) and psychosocial variables. Treatment assignment was not included in the models because disagreement in functional status was derived from baseline measures administered before treatment. Because of the high prevalence of disagreement, odds ratios (which can easily be derived from logistic regression models) overestimate the relative risks for disagreement. Thus, to better understand the magnitude of difference between exposed and unexposed groups, the proportion of disagreement for each group was tabulated. Marginal standardized risk ratios and 95% CIs were calculated (38).

The association between each type of functional status measure (self-reported or performance-based) and subsequent stroke or death was examined by using bivariate techniques (Pearson chi-square tests) (results not shown) and multivariate techniques (proportional hazards models). To examine the association of each measure of function on time to stroke or time to death, Cox regression models for censored survival data (39) were derived. The effect of each functional status measure on time to stroke or time to death was estimated, sequentially controlling for age, sociodemographic characteristics (ethnicity and education), treatment assignment, and clinical and psychosocial variables, as well as the other functional status measure. All statistical analyses were performed by using SAS software, version 8.0 (SAS Institute, Inc., Cary, North Carolina).

Role of the Funding Sources

The National Institute of Neurological Diseases and Stroke and Meade Johnson Laboratories provided the funding for the collection of the clinical trial data. None of the funding agencies had a role in the design, analysis, or interpretation of data or in the decision to submit this article for publication.

RESULTS

Most women who participated in WEST were 65 years of age or older (range, 46 to 91 years) (75.5%), white (85.5%), and had at least a high school education (67.8%). Fewer than half of the women were married (41.3%), although 43.1% were widows. Although 74.3% had had a stroke (index event), 51.8% had a severe index event, and 39.2% had two or more comorbid conditions, few women had cognitive impairment (11.1%) or currently smoked (12.7%). More than one third of women reported depressive symptoms (36.8%) and indicated that "religion was not a major source of

strength" (32.5%), but fewer than one quarter reported the absence of a health locus of control (23.4%), few friends and relatives (8.8%), or inadequate social support (17.3%). Consistent with previous studies (40), more than one quarter of the women died (12.7%) or had a stroke (15.3%) during the nearly 3 years of follow-up.

The test of marginal homogeneity indicated that the marginal percentages of the Barthel index and the Physical Performance Test presented in **Table 1** were statistically different ($P < 0.001$). As a result, agreement between the two functional status measures was poor, as indicated by a weighted κ of 0.28 (95% CI, 0.22 to 0.33). Regardless of the weights used, the κ values suggested poor agreement. Self-reported function was the same as measured by the Physical Performance Test for 25.7% (159 of 620) of women; there was slight (one level) clinical disagreement between the Barthel index and Physical Performance Test for 55.0% (341 of 620) of women; and clinical disagreement was substantial (two or more levels) between the Barthel index and Physical Performance Test for 19.3% (120 of 620) of women. Although 93.8% of women with slight clinical disagreement between the Barthel index and Physical Performance Test overreported their function, all women with substantial clinical disagreement between the measures overreported their function (high Barthel index score [≥18] and low Physical Performance Test scores [<17]).

In bivariate analyses, several sociodemographic and clinical characteristics but none of the psychosocial characteristics were associated with disagreement (**Table 2**). Women with substantial clinical disagreement (that is, women who substantially overreported their physical performance) were more likely to be nonwhite and have less education than women with slight or no clinical disagreement (**Table 2**). In addition, women with substantial clinical disagreement were more likely to have two or more comorbid conditions, stroke as the index event, a severe index event, and cognitive impairment than were women with slight or no clinical disagreement (**Table 2**).

The factors associated with slight clinical disagreement and substantial clinical disagreement in a multivariate model are shown in **Table 3**. The relationship between fewer than 12 years of education and slight

Table 2. **Characteristics of 620 Women in the Study Sample***

Characteristic	Extent of Clinical Disagreement between the Barthel Index and the Physical Performance Test			*P* Value
	No Disagreement (n = 159 [25.7%])	Slight Disagreement (n = 341 [55.0%])	Substantial Disagreement (n = 120 [19.3%])	
	← *n (%)* →			
Sociodemographic				
Age				
46–64 y	40 (25.2)	87 (25.5)	25 (20.8)	>0.2
65–74 y	53 (33.3)	98 (28.7)	40 (33.3)	>0.2
≥75 y	66 (41.5)	156 (45.8)	55 (45.8)	>0.2
Nonwhite ethnicity	14 (8.8)	45 (13.2)	31 (25.8)	<0.001
<12 years education	40 (25.2)	112 (32.9)	47 (39.5)	0.037
Clinical				
≥2 comorbid conditions	54 (34.0)	131 (38.4)	58 (48.3)	0.047
Stroke as index event	100 (62.9)	253 (74.2)	108 (90.0)	<0.001
Severe index event	70 (44.0)	163 (47.8)	88 (73.3)	<0.001
Cognitive impairment (MMSE score <24)	9 (5.7)	34 (10.0)	26 (21.7)	<0.001
Current smoker	22 (13.8)	42 (12.3)	15 (12.5)	>0.2
Active treatment assignment	75 (47.2)	167 (49.0)	65 (54.2)	>0.2
Psychosocial				
Depressive symptom (CES–D score ≥16)	60 (38.2)	120 (35.9)	44 (37.3)	>0.2
No health locus of control	30 (18.9)	82 (24.4)	32 (26.9)	>0.2
Religion not a source of strength	49 (30.8)	118 (34.6)	34 (28.6)	>0.2
Unmarried	88 (55.3)	205 (60.1)	71 (59.2)	>0.2
Few friends and relatives	14 (8.9)	28 (8.3)	12 (10.1)	>0.2
Inadequate social support	27 (17.9)	55 (16.4)	22 (19.3)	>0.2

* CES–D = Center for Epidemiologic Studies Depression Scale; MMSE = Mini-Mental State Examination.

Table 3. **Factors Associated with Slight and Substantial Clinical Disagreement (Compared with No Clinical Disagreement)***

Characteristic	Model 1: Slight Clinical Disagreement (n = 341) vs. No Clinical Disagreement (n = 159)†		Model 2: Substantial Clinical Disagreement (n = 120) vs. No Clinical Disagreement (n = 159)‡	
	Women, *n/n (%)*§‖	Adjusted RR (95% CI)	Women, *n/n (%)*§‖	Adjusted RR (95% CI)
Ethnicity				
Nonwhite			31/45 (68.9)	1.43 (1.07–1.91)
White			89/234 (38.0)	
Education				
<12 y	112/152 (73.7)	1.12 (0.99–1.27)	47/87 (54.0)	1.30 (1.02–1.67)
≥12 y	228/347 (65.7)		72/191 (37.7)	
Comorbid conditions				
≥2			73/130 (56.1)	1.52 (1.17–1.97)
<2			47/149 (31.5)	
Index event				
Stroke	253/353 (71.7)	1.20 (1.03–1.39)	108/208 (51.9)	2.33 (1.45–3.73)
Transient ischemic attack	88/147 (59.9)		12/71 (16.9)	
Cognitive impairment				
Present (MMSE score <24)			26/35 (74.3)	1.76 (1.34–2.32)
Absent (MMSE score ≥24)			94/244 (38.5)	
Depressive symptoms				
Present (CES–D score ≥16)			44/104 (42.3)	0.81 (0.62–1.06)
Absent (CES–D score <16)			74/171 (43.3)	

* CES–D = Center for Epidemiologic Studies Depression Scale; MMSE = Mini-Mental State Examination; RR = risk ratio.
† Model 1 included only education and index event. Ethnicity, comorbid conditions, cognitive impairment, and depressive symptoms did not significantly change the fit of the model and were therefore excluded.
‡ Model 2 included ethnicity, education, comorbid conditions, index event, cognitive impairment, and depressive symptoms.
§ The proportions presented in this table are derived from Table 2. Table 2 presents the proportion of women with a particular characteristic (exposure) among those women with (slight or substantial) disagreement or without disagreement (outcome). Table 3 presents the proportion of women with a particular type of disagreement (slight [Model 1] or substantial [Model 2]) among those with and without a particular characteristic. These latter proportions can be used to estimate the risk ratio (unadjusted).
‖ Denominations may not add to total sample size (n = 500 for Model 1; n = 279 for Model 2) due to missing values.

clinical disagreement was marginally significant (risk ratio [RR], 1.12 [CI, 0.99 to 1.27]). The relationship between stroke as the index event and slight clinical disagreement (compared with no clinical disagreement), however, was statistically significant (RR, 1.20 [CI, 1.03 to 1.39]). (Although not shown, results were similar for analyses excluding the 21 women who underreported their function.) In contrast, several sociodemographic and clinical characteristics were associated with substantial clinical disagreement (compared with no clinical disagreement). Women who were of nonwhite ethnicity (RR, 1.43 [CI, 1.07 to 1.91]), were less educated (RR, 1.30 [CI, 1.02 to 1.67]), had two or more comorbid conditions (RR, 1.52 [CI, 1.17 to 1.97]), had stroke as the index event (RR, 2.33 [CI, 1.45 to 3.73]), and were cognitively impaired (RR, 1.76 [CI, 1.34 to 2.32]) were more likely to have substantial disagreement than women who were white, more educated, and less clinically impaired. The indicator for depressive symptoms, although nonsignificant, was retained because it substantially changed the association between all other independent variables in the model and substantial disagreement. Significant interaction effects were not evident.

The effect of self-reported and performance-based function on stroke and death, sequentially adjusted for sociodemographic and clinical characteristics, is shown in **Table 4**. Self-reported function (moderate and severe dependence vs. independent and slight dependence), as measured by the Barthel index, was not significantly associated with time to stroke and time to death; and the effect size was small (unadjusted hazard ratio, 1.30 [CI, 0.78 to 2.18]). This hazard ratio was attenuated further and remained not significant after adjustment for age, ethnicity, education, treatment assignment, comorbid conditions, index event, cognitive impairment, and physical performance. (Psychosocial characteristics did not substantially alter the models.) In contrast, performance-based function (moderate and severe dependence vs. independent and slight dependence), as measured by the Physical Performance Test, was a statistically significant predictor of time to stroke and time to death and the effect size was substantial (unadjusted hazard ratio, 1.89 [CI, 1.41 to 2.55]). The hazard ratio was attenu-

ated but remained statistically significant and elevated (hazard ratio, 1.50 [CI, 1.06 to 2.11]) after adjustment for age, ethnicity, education, treatment assignment, comorbid conditions, index event, cognitive impairment, and self-reported function.

DISCUSSION

Our study found that there may be substantial differences between self-reported and performance-based measures of function, using categorical scales of functional levels, among women with a recent cerebrovascular event. Disagreement between the two measures was common, and a performance-based, not a self-reported, measure of function had predictive validity for long-term health outcomes.

These results have important implications for clinical practice and research. On the basis of our findings, a self-reported measure of function, as determined by the Barthel index, may overestimate the functional status of women who have had a recent cerebrovascular event and who need assistance with their basic self-care activities. Clinicians and clinical researchers must be aware of the potential biases in self-reported measures of function and use caution when basing clinical decisions solely on self-reported measures. Performance-based measures of function may provide more accurate assessments of function.

The factors associated with disagreement between the Barthel index and Physical Performance Test varied depending on the amount of disagreement. Consistent with previous studies that found disagreement to be related to clinical features (24), women who had had a recent stroke (index event) were more likely to demonstrate slight disagreement. Women who were more clinically impaired (more comorbid conditions, recent stroke, and cognitive impairment), were less educated, and were of nonwhite ethnicity were more likely to demonstrate substantial clinical disagreement (and substantial overreporting of physical performance).

The latter finding may be the result of an inability of these women to accurately assess their functional skills or an error in measurement of the self-reported or performance-based measure of function. Although women with severe cognitive impairment were excluded from the analysis, it is possible that women with slight cognitive deficits were less able to accurately identify functional deficits. In addition, women who were more clinically impaired may have adapted to their limitations and no longer viewed themselves as having functional impairments. Furthermore, there may be differences in the value that women place on being independent in their activities of daily living. For example, some women may report independence on specific self-care tasks because they are required to complete the tasks independently to maintain their living arrangements.

In addition to identifying substantial overreporting of function among women with a recent cerebrovascular event, our study suggests that a physical performance

Table 4. **Relationship of Self-Reported and Performance-Based Function to Time to Stroke and Death, Adjusting for Sociodemographic and Clinical Characteristics, in 620 Women (Proportional Hazards Models)**

Model*	Hazard Ratio for Relationship of Function to Time to Stroke and Death (95% CI)	
	Self-Reported Moderate to Severe Dependence (Barthel Index Score <18)	Performance-Based Moderate to Severe Dependence (Physical Performance Test Score <17)
Function only	1.30 (0.78–2.18)	1.89 (1.41–2.55)
Function and age	1.19 (0.71–1.99)	1.77 (1.31–2.39)
Function, age, and sociodemographic characteristics	1.29 (0.76–2.16)	1.81 (1.33–2.46)
Function, age, sociodemographic characteristics, and treatment assignment	1.29 (0.76–2.17)	1.81 (1.33–2.46)
Function, age, sociodemographic characteristics, treatment assignment, and clinical variables	1.11 (0.65–1.91)	1.48 (1.06–2.05)
Function, age, sociodemographic characteristics, treatment assignment, clinical variables, and function using other measure	0.93 (0.53–1.62)†	1.50 (1.06–2.11)‡

* Sociodemographic characteristics include ethnicity and education; clinical variables include comorbid conditions, stroke as the index event, and cognitive impairment. For all models, the variable "Function" is self-reported or performance-based function, according to subsequent column title.
† Other measure of function was moderate and severe dependence, as measured by the Physical Performance Test (score <17).
‡ Other measure of function was moderate and severe dependence, as measured by the Barthel index (score <18).

measure of function provides predictive information, which is not available from a self-reported measure of function, about long-term health outcomes. If clinicians and researchers rely solely on a self-reported measure of function and fail to detect functional impairments, the clinical management of patients with functional deficits may be negatively affected. Patients may not be linked with necessary resources or monitored appropriately. For example, hospitalized patients may be discharged to their homes before being able to function safely in them. Despite criticisms that performance-based measures of function are expensive and time-consuming (2), we found that a performance-based measure, such as the Physical Performance Test, could easily be administered by nurses or trained interviewers in approximately 5 minutes.

Several limitations should be considered when interpreting our results. Most important, they may not apply to all self-reported and performance-based measures. Our study examined women with a recent cerebrovascular event and was based on one self-reported measure of function and one performance-based measure of function. In addition, the two measures used in this study may assess different aspects of function, such as various activities of daily living or different approaches to task completion. Only three tasks (dressing, walking, and eating) were common to both measures. Although not shown, agreement between the task-specific measures was low (22% to 26%), similar to our overall findings. Alternatively, differences in these measures may result from the specific aspect of task completion being emphasized. The Barthel index focuses on whether assistance is required to complete the task, whereas the Physical Performance Test focuses on the time required to complete the task. It is plausible that women may be able to complete a task without assistance but not in a reasonable length of time. Both indices, however, have been shown to be valid and reliable measures of functional performance in older individuals who are disabled or who recently had a stroke (30–32, 41). In addition, clinicians and researchers routinely use the complete scales of these types of measures, not specific subscales, to make clinical judgments about long-term prognoses, treatment regimens, and discharge plans (2, 3, 7, 12, 16, 28, 29, 31, 32, 42).

Furthermore, there may be an error in measurement in the measurement of function or of comorbid conditions. There are floor and ceiling effects in measurement of function (whether self-reported or performance-based) that directly influence measurement of the extent of disagreement. For example, women who reported functional independence on the Barthel index could not be categorized as underreporting their physical performance. This may introduce measurement error because women may have been misclassified in terms of their extent of disagreement between self-reported and performance-based function. Because the distribution of self-reported function was skewed toward independence, few women in our study underreported their physical performance, thereby limiting our ability to examine the factors associated with underreporting. In addition, the self-reported measurement of comorbid conditions may have been biased. Given that many women overreported their functional status, it is possible that they also underreported or overreported the number of chronic conditions. Unlike self-reported function, however, the self-report of comorbid conditions was based on a woman's indication that a doctor had told her that she had a particular condition.

In conclusion, our study addresses an important clinical question—the comparative value of a self-reported and a performance-based measure of function. Noting that there were substantial differences between the two measures of function for specific groups of women, we found that a performance-based measure of function, not a self-reported measure, was predictive of stroke and death. Our findings suggest that a performance-based measure of function may be more appropriate than a self-reported measure in clinical settings that address the needs of women with a recent cerebrovascular event. Thus, clinicians and researchers should carefully consider whether a self-reported measure should be used routinely for all patients. A performance-based measure, particularly with certain groups of older women, may more accurately assess functional status, enable clinicians to more effectively manage patients' care, and provide better estimates of community health service needs.

From Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland; Yale University School of Medicine, New Haven, Connecticut; and the Veterans Administration Connecticut Healthcare System, West Haven, Connecticut.

Note: Dr. Ralph I. Horwitz is a member of John D. and Catherine T. MacArthur Foundation Network on Socioeconomic Status and Health.

Disclaimer: Although the principal author is currently employed at the Agency for Healthcare Research and Quality, this study was conceived and designed, the data were analyzed, and the manuscript was written while the principal author was a doctoral student at Yale University and postdoctoral fellow at Johns Hopkins University.

Grant Support: By a grant from the National Institute of Neurological Diseases and Stroke (RO1-N531251), a grant from Meade Johnson Laboratories, and training grants from the National Institute of Mental Health (5T32-MH15783 and 5T32-MH19545).

Requests for Single Reprints: Pamela L. Owens, PhD, Agency for Healthcare Research and Quality, 2101 East Jefferson, Suite 605, Rockville, MD 20852; e-mail, POwens@ahrq.gov.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Huber MR, Radosevich DM.** The role of functional status. In: Huber M, The Health Outcomes Institute, eds. Faulkner & Gray's Medical Outcomes and Practice Guidelines Library. Measuring Medicine: An Introduction to Health Status Assessment and a Framework for Application. vol. 2. Washington, DC: Faulkner & Gray; 1994:9-18.

2. **McDowell I, Newell C.** Physical disability and handicap. In: Measuring Health: A Guide to Rating Scales and Questionnaires. 2nd ed. New York: Oxford Univ Pr; 1996:47-50.

3. **LaPlante MP, Carlson D.** Disability in the United States: Prevalence and Causes, 1992. Washington, DC: U.S. Department of Education, National Institute on Disability and Rehabilitation Research; 1996.

4. **Gill TM, Robison JT, Tinetti ME.** Difficulty and dependence: two components of the disability continuum among community-living older persons. Ann Intern Med. 1998;128:96-101. [PMID: 9441588]

5. **Guralnik JM, Branch LG, Cummings SR, Curb JD.** Physical performance measures in aging research. J Gerontol. 1989;44:M141-6. [PMID: 2768767]

6. **Keith RA.** Functional status and health status. Arch Phys Med Rehabil. 1994; 75:478-83. [PMID: 8172511]

7. **Rozzini R, Frisoni GB, Bianchetti A, Zanetti O, Trabucchi M.** Physical Performance Test and Activities of Daily Living scales in the assessment of health status in elderly people. J Am Geriatr Soc. 1993;41:1109-13. [PMID: 8409158]

8. **Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, 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. 1994;49:M85-94. [PMID: 8126356]

9. **Guralnik JM, Reuben DB, Buchner DM, Ferrucci L.** Performance based measures of physical function in comprehensive geriatric assessment. In: Rubenstein LZ, Wieland D, Bernabei R, eds. Geriatric Assessment Technology: The State of the Art. Milan, Italy: Editrice Kurtis; 1995:59-74.

10. **Little AG, Hemsley DR, Volans PJ, Bergmann K.** The relationship between alternative assessments of self-care ability in the elderly. Br J Clin Psychol. 1986; 25(Pt 1):51-9. [PMID: 3955274]

11. **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. [PMID: 1951822]

12. **Reuben DB, Siu AL, Kimpau S.** The predictive validity of self-report and performance-based measures of function and health. J Gerontol. 1992;47:M106-10. [PMID: 1624692]

13. **Sager MA, Dunham NC, Schwantes A, Mecum L, Halverson K, Harlowe 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. [PMID: 1634697]

14. **Myers AM, Holliday PJ, Harvey KA, Hutchinson KS.** Functional performance measures: are they superior to self-assessments? J Gerontol. 1993;48: M196-206. [PMID: 8366262]

15. **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. [PMID: 7836655]

16. **Reuben DB, Valle LA, Hays RD, Siu AL.** Measuring physical function in community-dwelling older persons: a comparison of self-administered, interviewer-administered, and performance-based measures. J Am Geriatr Soc. 1995; 43:17-23. [PMID: 7806733]

17. **Kempen GI, Steverink N, Ormel J, Deeg DJ.** The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies. J Gerontol B Psychol Sci Soc Sci. 1996; 51:P254-60. [PMID: 8809001]

18. **Kempen GI, van Heuvelen MJ, van den Brink RH, Kooijman AC, Klein M, Houx PJ, et al.** Factors affecting contrasting results between self-reported and performance-based levels of physical limitation. Age Ageing. 1996;25:458-64. [PMID: 9003883]

19. **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;49:1103-10. [PMID: 8826989]

20. **Merrill SS, Seeman TE, Kasl SV, Berkman LF.** Gender differences in the comparison of self-reported disability and performance measures. J Gerontol A Biol Sci Med Sci. 1997;52:M19-26. [PMID: 9008665]

21. **Kivinen P, Sulkava R, Halonen P, Nissinen A.** Self-reported and performance-based functional status and associated factors among elderly men: the Finnish cohorts of the Seven Countries Study. J Clin Epidemiol. 1998;51:1243-52. [PMID: 10086816]

22. **Sherman SE, Reuben D.** Measures of functional status in community-dwelling elders. J Gen Intern Med. 1998;13:817-23. [PMID: 9844079]

23. **Kuriansky JB, Gurland BJ, Fleiss JL, Cowan D.** The assessment of self-care capacity in geriatric psychiatric patients by objective and subjective methods. J Clin Psychol. 1976;32:95-102. [PMID: 1249244]

24. **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. [PMID: 1533995]

25. **Ferrer M, Lamarca R, Orfila F, Alonso J.** Comparison of performance-based and self-rated functional capacity in Spanish elderly. Am J Epidemiol. 1999;149: 228-35. [PMID: 9927217]

26. **Kernan WN, Brass LM, Viscoli CM, Sarrel PM, Makuch R, Horwitz RI.** Estrogen after ischemic stroke: clinical basis and design of the Women's Estrogen for Stroke Trial. Journal of Stroke and Cerebrovascular Diseases. 1998;7:85-95.

27. **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;12:189-98. [PMID: 1202204]

28. **Mahoney FI, Barthel DW.** Functional evaluation: the Barthel Index. Md Med J. 1965;14:61-5.

29. **Wade DT, Hewer RL.** Functional abilities after stroke: measurement, natural history and prognosis. J Neurol Neurosurg Psychiatry. 1987;50:177-82. [PMID: 3572432]

30. **Shah S, Vanclay F, Cooper B.** Improving the sensitivity of the Barthel Index for stroke rehabilitation. J Clin Epidemiol. 1989;42:703-9. [PMID: 2760661]

31. **Reuben DB, Siu AL.** An objective measure of physical function of elderly

outpatients. The Physical Performance Test. J Am Geriatr Soc. 1990;38:1105-12. [PMID: 2229864]

32. **Brown M, Sinacore DR, Binder EF, Kohrt WM.** Physical and performance measures for the identification of mild to moderate frailty. J Gerontol A Biol Sci Med Sci. 2000;55:M350-5. [PMID: 10843356]

33. **Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, et al.** Measurements of acute cerebral infarction: a clinical examination scale. Stroke. 1989;20:864-70. [PMID: 2749846]

34. **Folstein M, Anthony JC, Parhad I, Duffy B, Gruenberg EM.** The meaning of cognitive impairment in the elderly. J Am Geriatr Soc. 1985;33:228-35. [PMID: 3989183]

35. **Radloff LS.** The CES–D Scale: A self-report depression scale for research in the general population. Applied Psychological Measurement. 1977;1:385-401.

36. **Fleiss JL, Cohen J.** The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement. 1973;33:613-9.

37. **Fleiss JL.** Statistical Methods for Rates and Proportions. 2nd ed. New York: J Wiley; 1981:212-36.

38. **Flanders WD, Rhodes PH.** Large sample confidence intervals for regression standardized risks, risk ratios, and risk differences. J Chronic Dis. 1987;40:697-704. [PMID: 3597672]

39. **Hosmer DW, Lemeshow S.** Applied Survival Analysis: Regression Modeling in Time to Event Data. New York: J Wiley; 1999:87-112.

40. **Kernan WN, Horwitz RI, Brass LM, Viscoli CM, Taylor KJ.** A prognostic system for transient ischemia or minor stroke. Ann Intern Med. 1991;114:552-7. [PMID: 2001088]

41. **Dombovy ML, Sandok BA, Basford JR.** Rehabilitation for stroke: a review. Stroke. 1986;17:363-9. [PMID: 2940735]

42. **Rodgers H, Curless R, James OF.** Standardized functional assessment scales for elderly patients. Age Ageing. 1993;22:161-3. [PMID: 8503310]