Journal of Gerontology: MEDICAL SCIENCES
1997, Vol 52A, No. 1, M19–M26

Gender Differences in the Comparison of Self-Reported Disability and Performance Measures

Susan S. Merrill,[1] Teresa E. Seeman,[2] Stanislav V. Kasl,[3] and Lisa F. Berkman[4]

[1]Institute of Gerontology, University of Michigan.
[2]Ethel Percy Andrus Gerontology Center, University of Southern California.
[3]Department of Epidemiology and Public Health, Yale University.
[4]School of Public Health, Harvard University.

Background. Gender differences in functioning among older adults have been well documented. Differential reporting of functional problems by men and women may contribute to this observed difference. The purpose of this study was to examine the gender differences in functional ability by comparing self-reported function to observed performance of physical tasks.

Methods. In 1988, 1,458 men and women ages 71 and older from the New Haven site of the Established Populations for the Epidemiologic Study of the Elderly (EPESE) self-reported activities of daily living (ADL) disability and functional limitations. Subjects' ability to perform 7 tasks was observed. Gender differentials in "accurate" and "inaccurate" reporting were determined by examining comparable self-reports and performance measures. Logistic regression determined how much of the gender differential in self-reported function was explained by performance ability.

Results. More women than men reported disability and functional limitation, and women had poorer performance scores for every task. Compared to similar performance items, self-reports of function were accurate for the majority of men and women. However, among those who inaccurately reported function, more men than women underreported disability and more women than men overreported disability. Overall performance explains all of the gender difference in ADL disability and most of the difference in functional limitation.

Conclusions. These results suggest that both men and women generally report their disability accurately, and women's higher prevalence of reported functional problems is probably a reflection of true disability for most disability measures.

GENDER differences in function among older adults are well known, but not well understood. Women consistently report more functional problems than men (1–8). Several reasons for this gender difference have been hypothesized. One hypothesis posits that the gender difference in function is due to women's greater prevalence of nonfatal but disabling conditions and diseases, such as arthritis and migraine headaches (2,3,9,10). Verbrugge and Balaban (11) have shown that more nonfatal conditions contribute to women's disability, whereas men's disability is associated with life-threatening disease. Although chronic conditions account for some of the gender differential in function, a substantial difference remains unexplained (12).

Another possible explanation for observed gender differences in function is differential reporting of functional problems by men and women; in other words, reported functional problems may inaccurately reflect real functional deficits. Most clinical and epidemiological studies of function among older adults utilize patient or subject self-reports of functional status, such as activities of daily living (ADLs), instrumental activities of daily living (IADLs), and functional limitations. A number of investigators have hypothesized that women are more likely to interpret physical discomforts as symptoms and have a greater tendency to recall and report those symptoms (2,13–15). Gender differences in reported morbidity and disability may stem from differential perceptions of symptoms between men and women (1–3,10). Women are typically more involved with health and health care all their lives than men; females are generally socialized to pay attention to and acknowledge pain and discomfort, while males are socialized to ignore physical discomfort (10). Tests of these hypotheses are sparse, and none have focused on disability.

Guralnik et al. (16) suggest that the assessment of functional disability can be improved if researchers and clinicians use objective physical performance measures to supplement the traditional self-reports of disability. They reviewed the few studies to date that made such comparisons and found good correlation between self-reports and performance. They note that performance tests can overcome some of the limitations of self-reports, such as respondent reporting bias due to perceptions of socially acceptable or appropriate responses, and performance tests are less influenced by culture, language, and education level.

A few recent studies have compared self-reports to performance of tasks, but none have addressed gender differences. For example, Elam et al. (17) compared patients' self-reported ADL disability with family member and physician assessment of ADL and with performance of ADL tasks. Patients' self-reports were similar to their ability to perform tasks; however, compared to task performance, patients slightly overrated their ability. Defining performance as the

standard, the authors found that patients were most accurate in their assessment of disability, followed by family member and physician. Another study comparing self-reported ADL disability with performed ADL tasks found that among discrepancies between the two measurements of ADL, 89% of the subjects ranked their disability greater than observed performance indicated (18).

Though the studies that have examined the correlation between self-reported functional status and performance measures have found generally good agreement between measures, there are some discrepancies. Examining those discrepancies more closely provides a means to test the theory that gender differences in the way functional problems are reported contribute to observed gender differences in function. Specifically, when compared with performance measures, women may overestimate and report more, while men may underestimate and report fewer functional problems.

The purpose of this study was to examine gender differences in the reporting of functional problems. We examined self-reported ADL disability and functional limitation in relation to actual performance measures. Three specific questions are asked: Do women report more disability and have poorer performance scores than men? Comparing reported disability to specific performance tasks, do women have higher false positives (overreporting of disability) than men, and do men have higher false negatives (underreporting of disability) than women? Do problems performing physical tasks contribute to the reported gender differential in disability?

METHODS

Data. — Data are drawn from the 1988 annual interview of the New Haven site of the National Institute on Aging's Established Populations for the Epidemiologic Study of the Elderly (EPESE). The New Haven EPESE is a community-based sample of noninstitutionalized men and women aged 65 and older who lived in the New Haven, Connecticut, area in 1982. The cohort is based on a stratified probability sample of three distinct housing strata: general community housing, age-restricted private housing for the elderly, and age- and income-restricted public housing for the elderly. The sampling design has been described in detail elsewhere (19). Weighted data from this sample, adjusted for oversampling in the two age-restricted strata, provide population estimates for the noninstitutionalized elderly population of New Haven. The analyses presented here are based on the 1988 interview data, as it is the only wave of EPESE to include performance-based measures of function.

In 1988, 1,795 cohort members were alive and recontacted. Of these, 1,491 had complete (95%) or partially complete (5%) interviews (172 proxy responses, 127 interview refusals, and 5 persons lost to follow-up), for a response rate of 83%. Thirty-three of the partially complete interviews were missing both self-reported function and performance data and thus were excluded from analyses. Sample size for this study is 1,458 (32% men, 68% women).

Variable measurement. — Self-reported functional status is measured by activities of daily living (ADLs) and functional limitations. Ability in seven ADLs was ascertained using a modified version of the Katz ADL (20). Respondents were asked if they were able to bathe, groom, eat, dress, walk across a room, use a toilet, and transfer from a bed to a chair without help. Each item is scored dichotomously: does not need help, needs help, or is unable. A two-level ADL summary disability measure is constructed: needs no help with any ADL versus needs help with or is unable to do one or more ADLs. The ADL summary measure is dichotomized to represent ability or any disability in the realm of personal care.

Functional limitations were ascertained in two ways: three items were adopted from an instrument developed by Rosow and Breslau (21) and five items came from an instrument developed by Nagi (22). For this study, the Rosow and Breslau items are defined as gross mobility limitations, and the Nagi items are defined as range-of-motion limitations. Gross mobility limitation items include ability to do heavy housework, walk up and down stairs, and walk half a mile without help. Two-level variables indicate score for each item: able, unable. A summary gross mobility limitation item is categorized as no limitation for any item versus inability for one or more of these items. Range-of-motion (ROM) limitation is measured by respondent degree of difficulty (none, a little, some, a lot, unable) to push or pull a large object, carry 10 pounds, raise arms above shoulders, write or handle small objects, and stoop, crouch, or kneel. Each item is scored dichotomously: has little or no difficulty, has some or a lot of difficulty, or is unable. Summarized ROM limitation is categorized as little or no difficulty with all items versus some or a lot of difficulty or inability with one or more items. The gross mobility and ROM limitation summary items represent ability or any disability in the general functional areas of mobility and range of motion.

Respondents were asked to perform a number of physical tasks related to balance, gait, and lower and upper extremity movement. Balance comprised three measures. Respondents were first asked to hold a semi-tandem stand (i.e., ball of one foot to heel of other foot) for 10 seconds. If able to successfully perform this task, they were asked to hold a tandem stand (i.e., ball of one foot to toe of other foot) for 10 seconds. If respondents were unable to hold the semi-tandem stand for 10 seconds, they were asked to stand with their feet side by side for 10 seconds. Each balance measure was coded dichotomously, indicating ability to hold each stand for 10 seconds. Those who tried but were unable to hold a stand at all or who were not asked to attempt to hold a stand for safety reasons were coded as unable. A balance summary measure adds the score of the 3 stands and ranges in value from 0–2: 0 = inability to hold any stand for 10 seconds; 1 = ability to hold either the semi-tandem *or* the side-by-side stand for 10 seconds; and 2 = ability to hold both the semi-tandem *and* tandem stands for 10 seconds.

Respondents were timed as they walked 8 feet and back, and allowed any aid (e.g., cane or walker) normally used, if needed. Walking score is number of seconds to complete the walk. For those who only walked one way (n = 2), walk is scored as number of seconds doubled plus an additional 10% of that time. A 3-level walking summary was constructed, ranging in score from 0 to 2: 0 = inability to walk or walk

not attempted; 1 = slower walking times (> median time); and 2 = quicker times (≤ median time).

Lower extremity strength was assessed by ability to rise from a chair 5 consecutive times without use of arms; due to unavailability of a suitable chair, 102 respondents were not asked to perform this task. Initially, respondents were asked to stand once from a sitting position without using arms to push off. A dichotomous measure indicates ability to rise once from a chair. If the single chair rise was successfully completed, respondents were timed while rising from the chair in the same manner 5 consecutive times. A 2-level measure distinguishes between ability to complete 5 rises and completing 4 or fewer rises. Time to complete consecutive rises represents overall score. A 3-level chair-rise summary measure ranges from 0 to 2: 0 = inability to perform a single chair rise, performed 4 or fewer consecutive rises, tried but was unable, or did not attempt; 1 = slower chair times (> median time); and 2 = quicker times (≤ median time).

Upper extremity dexterity was assessed by respondents' ability to rotate their shoulders. Respondents placed both hands behind their head, at ear level, with fingers touching, and tried to raise their arms up parallel to the floor, pointing elbows out to the side. For each shoulder, performance is scored as completely able, partially able, or unable to perform the task. Those who did not attempt are scored unable. Correlation between scores for right and left shoulders is nearly perfect (R = .97). Thus, only one shoulder score (the higher score) is utilized in analyses as a 3-level measure: 0 = unable, 1 = partially complete rotation, 2 = complete rotation.

An overall summary measure of all performance tasks was constructed by adding the four individual summary items. The physical performance measure ranges in value from 0–8, with lower scores indicating poorer performance. While this summary measure is based on the tasks performed, it may also serve as a crude measure of general performance ability (e.g., those with very low scores [trouble with all tasks] are likely to have trouble with other, untested tasks).

Refusal rates for task performance (those who refuse may be more disabled) and resulting potential bias are of concern. However, range of refusal rates among tasks is quite low: 0.4–1.5% (men) and 1.3–3.9% (women). In addition, distribution of refusals over self-reported disability is reasonably even. Among both men and women who refused to perform tasks, slightly less than half reported ADL disability, and slightly more than half reported gross mobility or ROM limitation. Thus, resulting bias from these refusals is minimal.

Analytic strategy. — Prevalence of reported functional deficits and performance ability for men and women are described; chi-square and t-tests determine significance of gender differences.

Comparisons are made between 3 self-reported function measures and comparable tasks: (a) reported ability to walk across a room (an ADL item) with ability to walk 8 feet and back; (b) reported difficulty stooping, crouching, or kneeling (ROM limitation item) with ability to perform consecutive rises from chair; and (c) reported difficulty reaching or extending arms above head (ROM limitation item) with ability to rotate shoulders. For these analyses, performance of task is considered the "true positive"; to determine accuracy of reported functional problems, sensitivity and specificity (and, conversely, false negative and false positive proportions) are calculated for each comparison. Log-linear analyses provide significance testing for gender differences in comparisons of self-reported function and performance measures. Since we are interested in the difference that exists between men and women for false negative and false positive reports of disability, this test of association is appropriate because it allows for the testing of all individual and interaction terms in the models. Significance of gender differences in the over- and underreporting of disability is determined by the significance of the interaction between gender, disability, and performance.

Logistic regression is used to evaluate the association between performance and reported function in a different manner. A series of regression models are analyzed for each of the three summarized self-reported function outcomes (ADL, gross mobility limitation, and ROM limitation). The initial model determines the unadjusted gender differential in function. The second model assesses the effect of overall performance (physical performance score) on the gender differential. In these analyses, observed performance serves as a proxy for actual, overall disability. Any remaining gender difference is interpreted as an inaccurate perception of true functional problems and may be explained by other factors.

Analyses are conducted using the Survey Data Analyses (SUDAAN) software, which can account for the New Haven EPESE sampling design (multiple housing strata, oversampling of men). Results reflect weighted data, though actual numbers (ns) of subjects are presented.

RESULTS

Distributions of individual and summary ADL disability and functional limitations are presented in Table 1. Women report more disability and functional limitations than men; gender differences are significant for all function items. The gender differentials are smallest for ADL disability and largest for summarized ROM limitation. Gender differences for summary measures of ADL disability, gross mobility limitation, and ROM limitation are, respectively, 12%, 18%, and 23%.

Table 2 presents the distribution of performance measures among men and women. Women perform more poorly than men for every measure. Significant gender differences are seen for the semi-tandem stand, full tandem stand, single rise from a chair, 8-foot walk, and shoulder rotation. Significantly more women than men had poor scores for all summary measures.

Comparison between reported ability to walk across a room and actual ability to walk 8 feet and back is presented in Table 3. Specificity is quite high (.91 for men and .92 for women), indicating that reported ability to walk across a room is accurate. Sensitivity is fairly high for both men and women (.71 and .82, respectively), indicating reported inability to walk across a room is a good indicator of actual inability to walk a short distance. False negatives were fairly

Table 1. Gender Differences in the Distribution of Disability and Functional Limitations

Disability and Functional Limitation	Men (N = 522) n	Men %	Women (N = 936) n	Women %
ADL items; (% need any help)				
Walking across a small room	81	10.6	219	20.2***
Bathing	76	9.8	221	21.3***
Dressing	58	7.7	136	13.8***
Transferring from bed to chair	50	6.6	156	15.3**
Using the toilet	43	6.2	125	12.0**
Grooming	29	3.5	100	8.5*
Eating	30	3.3	91	7.7**
% Need help with one or more ADL	111	15.5	288	27.9***
Gross mobility limitations items; (% unable)				
Do heavy housework	197	33.9	453	43.9***
Walk half a mile	162	26.6	481	47.8***
Walk up and down stairs	116	15.5	338	30.8***
% Unable in one or more gross mobility limitations	236	41.3	582	58.7***
Range-of-motion limitation items; (% difficulty)				
Lifting < 10 pounds	153	25.0	498	48.3***
Pushing or pulling heavy objects	155	25.0	465	44.6***
Stooping, crouching, or kneeling	143	22.4	454	43.4***
Reaching or extending arms above head	56	8.2	181	16.0***
Writing or using small objects	37	6.0	150	13.9**
% With difficulty in one or more range-of-motion limitations	237	40.2	640	63.3***

Note: Ns are unweighted, % are weighted values.
*$p < .05$, **$p < .01$, ***$p < .001$.

low; 29% of men and 18% of women reported ability but were unable to walk 8 feet. This 11% gender difference is marginally significant ($p < .07$), but is in the predicted direction for the hypothesized greater underreporting of disability among men.

The top portion of Table 4 presents the comparison between reported difficulty to stoop, crouch, or kneel and ability to rise 5 consecutive times from a chair. Sensitivity and specificity are moderately high for both men and women, indicating fairly accurate reporting. However, among those who reported inaccurately, women have a higher proportion of false positives (16% difference) and men have a greater proportion of false negatives (5% difference). These gender differences are highly significant ($p < .001$) and are in the predicted direction for the hypothesized female overreporting and male underreporting of disability. Because reported difficulty represents degree of difficulty rather than inability to stoop, crouch, or kneel, a second comparison was made between reported difficulty in stooping and slow or quick times to rise from a chair; timing represents a graded measurement which is more comparable to graded degree of difficulty (Table 4, bottom half). Specificity is high, but sensitivity is very low. The gender differences in false positive and negative proportions (11% and 17%, respectively) are quite significant ($p < .0001$). More women than men report difficulty but perform the task quickly; more men than women report no difficulty but perform the task slowly.

The top half of Table 5 presents the comparison between difficulty reaching or extending arms above the head and ability to rotate shoulders. Sensitivity is fairly high for men and women (.67 and .74, respectively), and specificity quite high (.92 and .86, respectively), indicating accurate reporting. The gender differences for false positive and false negative proportions are significant at $p < .03$. Six percent more women than men underreported and 4% more men than women overreported ability to extend their arms above their shoulders. Comparison of those reporting difficulty extending arms with partial or complete ability to rotate shoulders is presented in the bottom half of Table 5. Again, since reported difficulty is a graded measure, partial and complete ability to rotate a shoulder allows comparison with a graded measure. For this comparison, sensitivity is very low and specificity is very high. False positive proportions for men and women were similar (.07 and .10, respectively). A 14% gender differential in false negatives (.85 vs .71) for those who report no difficulty but are only partially able to rotate a shoulder is significant ($p < .02$).

For each of the comparisons, overreporting of disability is more common than underreporting, and both men and women over- and underreport their ability; however, more women than men underreport ability and more men than women overreport ability. These differences in reporting may explain all or some of the reported differences in self-reports of functional disability. A direct test of the effect of inaccurate reporting on the gender difference in reported function is not available. We can, however, assess the effect on the gender difference in reported disability by adjusting for performance in logistic regression models. In these regression models, the performance summary score is used as a proxy for "true" dysfunction.

Table 6 presents the influence of physical performance on the gender differential for summary ADL disability and functional limitations. For ADL disability, the unadjusted odds ratio for gender is 1.80. Adjusting for performance summary score not only reduces the odds ratio to nonsignificance, but reverses the direction of the effect. This suggests that women's greater reported ADL disability is real. For gross mobility limitation, the unadjusted gender effect is 1.95 (95% confidence limit = 1.66–2.27). Adding performance to the model reduces the gender effect to 1.16, statistically nonsignificant at $\alpha < .05$. This suggests that women's greater reports of gross mobility limitation reflect an actual female excess in this type of functional limitation. Further adjusting for age reduces the gender effect only slightly. For ROM limitation, the unadjusted gender odds ratio is 2.57. Adjusting for performance reduces the gender effect to 2.21 (95% confidence limit = 1.71–2.54), indicating that inability to perform tasks explains some of the reported difference. When the model is further adjusted for age, the odds ratio for gender decreases to 1.72.

Discussion

The purpose of this study was to examine gender differences in the reporting of functional problems. We evaluated the accuracy of reported functional status for women and

Table 2. Gender Differences in the Distribution of Task Performance

Disability and Functional Limitation	Men (N = 522)		Women (N = 936)		Refusals: Men		Refusals: Women	
	n	%	n	%	n	%	n	%
Individual tasks†:								
Semi-tandem stand (% unable)	141	20.2	389	37.9***	10	1.5	36	3.9
Tandem stand (% unable)‡	120	29.4	216	38.2*	0	0.0	1	0.0
Feet side-by-side stand (% unable)§	74	46.6	211	47.0	9	4.8	23	6.9
Single rise from chair (% unable)	84	13.0	224	23.9**	2	0.4	23	2.3
5 consecutive rises from chair (% unable)‖	19	3.4	45	5.1	3	1.0	16	2.4
Walk 8 feet (% unable)	55	7.7	149	13.7**	5	1.0	32	3.1
Shoulder rotation (% unable)	25	3.1	86	7.7*	3	0.4	17	1.3
Summary measures¶:								
Standing balance:								
0 = unable	76	10.3	215	20.8**				
1 = able to hold 1 stand	188	33.9	413	42.6				
2 = able to hold 2 stands	250	55.7	285	36.6				
Rises from chair:								
0 = unable	95	14.8	260	27.4**				
1 = slow	165	32.8	330	41.0				
2 = quick	218	52.3	235	31.6				
Walking 8 feet:								
0 = unable	55	7.7	149	13.7**				
1 = slow	152	24.6	410	40.1				
2 = quick	310	66.7	343	43.1				
Shoulder rotation:								
0 = unable	25	3.1	86	7.8*				
1 = partially able	86	12.6	188	21.0				
2 = completely able	408	84.2	645	71.2				
Overall performance summary††								
0	22	2.6	77	6.8				
1	14	1.7	51	5.8				
2	29	4.3	66	6.5				
3	24	3.1	75	5.9				
4	42	6.3	114	11.4				
5	54	8.8	127	13.5				
6	82	15.8	157	15.6				
7	135	29.9	164	21.0				
8	118	27.4	96	13.5				
Mean score	5.7		4.7***					

Note: Ns are unweighted, % are weighted values.

†Chi-square tests of gender differences.

‡Proportions of 371 men and 510 women who successfully held semi-tandem stand for 10 seconds.

§Proportions of 141 men and 389 women who held semi-tandem stand < 10 seconds.

‖ Proportions of 404 men and 619 women who successfully completed one rise from chair.

¶, Mantel-Haenzel Chi-square tests of gender differences.

††Overall performance summary is sum of 4 individual summary measures (lower scores indicate poorer performance of tasks); T-test of gender difference.

*$p < .05$; **$p < .01$; ***$p < .001$.

men by comparing self-reported function with performance measures. For all measures of ADL disability and functional limitations, significantly more women than men reported functional problems. For all seven performance measures, women performed more poorly than did men. This in itself suggests that women are truly more disabled than men, if performance measures are assumed to reflect true ability. In other words, women's poorer performance of tasks may be attributed to their greater problems in functioning.

To assess differences in the reporting of disability, we compared performance of three tasks to comparable self-reported items. In these data, while most self-reports of functional problems are quite accurate when compared to performance (defined as the standard), there are genuine gender differences among those who inaccurately reported their functional status. Among the discrepancies, both men and women over- and underreported functional problems. However, more women overreported and more men underreported these problems.

There are several possible explanations for these differences. Women may have a greater symptom sensitivity and readiness to perceive physical sensations as symptoms of

Table 3. Comparison of Reported Ability To Walk Across a Small Room and Performed Ability To Walk 8 Feet and Back

Reported Ability To Walk Across Small Room	Performance: 8-Foot Walk					
	Unable *n*	Able *n*	Sensitivity	Specificity	False Positive	False Negative
Men						
Unable	39	40	.71	.91	.09	.29
Able	16	422				
Total	55	462				
Women						
Unable	122	85	.82	.92	.08	.18
Able	27	668				
Total	149	753				

Model significance: $p < .07$.

Table 4. Comparison of Reported Difficulty To Stoop, Crouch, or Kneel and Performed Ability To Rise From Chair 5 Consecutive Times

Reported Difficulty To Stoop, Crouch, or Kneel	Performance: Consecutive Rises From Chair					
	Unable *n*	Able *n*	Sensitivity	Specificity	False Positive	False Negative
Men						
Difficulty	10	49	.53	.87	.13	.47
No Difficulty	9	333				
Total	19	382				
Women						
Difficulty	26	165	.58	.71	.29	.42
No Difficulty	19	399				
Total	45	564				

Model significance: $p < .001$

Reported Difficulty To Stoop, Crouch, or Kneel	Performance: Consecutive Rises From Chair					
	Slow *n*	Quick *n*	Sensitivity	Specificity	False Positive	False Negative
Men						
Difficulty	35	14	.21	.94	.06	.79
No Difficulty	129	204				
Total	164	218				
Women						
Difficulty	125	40	.38	.83	.17	.62
No Difficulty	204	195				
Total	329	235				

Model significance: $p < .0001$.

illness or disability (3,23). Typically, females are socialized to pay attention to and acknowledge pain and discomfort, while males are socialized to ignore physical discomfort (2). While the literature cited primarily focuses on symptoms and illness rather than disability, it is likely that the same mechanisms apply to disability since pain, discomfort, and weakness are all associated with disability. Thus, women may be more willing to report and talk about their symptoms or dysfunction, whereas men may be more hesitant to admit physical limitations.

Perceptual differences may exist for response categories, particularly ordinal responses, such as the degree of difficulty scaling common to the range-of-motion limitation items. One person's definition of "a little" difficulty may be another's definition of "some difficulty," and this perceptual discrepancy may contribute to observed gender differences in reported function.

Over- and underreports of disability are based on performance being defined as the standard for comparison purposes; however, performance measures may not truly repre-

Table 5. Comparison of Reported Difficulty To Reach or Extend Arms and Performed Ability To Rotate Shoulders

Reported Difficulty To Reach or Extend Arms	Performance: Shoulder Rotation					
	Unable	Able				
	n	*n*	Sensitivity	Specificity	False Positive	False Negative
Men						
Difficulty	16	40	.67	.92	.08	.33
No Difficulty	8	453				
Total	24	493				
Women						
Difficulty	59	114	.74	.86	.14	.29
No Difficulty	24	716				
Total	83	830				

Model significance: $p < .03$

Reported Difficulty To Reach or Extend Arms	Partial	Complete				
	n	*n*	Sensitivity	Specificity	False Positive	False Negative
Men						
Difficulty	13	28	.15	.93	.07	.85
No Difficulty	73	380				
Total	86	408				
Women						
Difficulty	53	61	.29	.90	.10	.71
No Difficulty	132	584				
Total	185	645				

Model significance: $p < .02$.

Table 6. Influence of Scaled Physical Performance on the Gender Differences in Summarized ADL Disability*, Gross Mobility Limitation†, and Range of Motion Limitation‡

Model: Gender, Covariates	ADL Disability		Gross Mobility Limitation		Range of Motion Limitation	
	OR (w/m)	95% CI	OR (w/m)	95% CI	OR (w/m)	95% CI
Gender	1.80	1.29–2.50	1.95	1.66–2.27	2.57	2.14–3.09
Gender, performance summary	0.85	0.49–1.48	1.16	0.89–1.59	2.21	1.71–2.54
Gender, performance summary, age	0.84	0.52–1.38	1.11	0.79–1.56	1.72	1.28–2.30

*Need help with one or more ADL items.
†Unable to do one or more gross mobility items.
‡Difficulty with one or more functional limitation items.

sent the "gold standard." Performance items may not reflect usual daily activities, or study subjects may try harder to perform tasks than they would in normal daily life. Cognitively impaired subjects may not understand the task they are asked to perform. The chance of misunderstanding is greatly reduced, however, since the interviewer demonstrates each task before asking the study subject to attempt it. Guralnik et al. (16) have suggested that use of performance measures offers a supplemental approach to obtaining valid data on physical functioning in cognitively impaired persons. Several studies indicate the high correlation between self-reports and performance measures (e.g., 16,24); thus, while performance measures may not truly be the "gold standard," they provide a good means for comparison with reported ability.

We were unable to directly test the effect of inaccurate reporting on the gender effect for all disability measures as appropriate comparisons were available for only three functional measures. However, the effect of overall performance (a proxy for true disability) on the gender differences in function was assessed. Overall performance accounted for all the gender difference in ADL disability and gross mobility limitation, and some of the difference in ROM limitation, again suggesting that functional problems women report reflect real functional deficits. The remaining gender difference in ROM limitation is likely due to differences in the perception of ability among men and women, detailed previously. In these data, while most self-reports of functional problems are quite accurate when compared to performance, there are genuine gender differences among those who inaccurately reported their functional status.

A limitation of these analyses must be recognized. Not all the performance measures assessed are the same activities

queried in functional status items. Few performance measures, and fewer items that are comparable to ADL disability and functional limitation items, do not allow complete examination of gender differences in the reporting of functional status. Ideally, multiple and varied performed tasks should replicate self-reported function items so that an increased number of comparisons can be examined and so that self-reports and performance measures will be more comparable. Despite the limitation, this is the first study, to our knowledge, that has made such comparisons to provide another way of examining the gender differences in function. While there are significant gender differences in the way function is reported, these results suggest that men and women are generally accurate in their reported disability and that women's higher prevalence of reported functional problems is probably a reflection of true disability.

ACKNOWLEDGMENTS

This research was supported by National Institutes of Health contract 1-AG-0-2105, NIH training grants AG00114-10, AG00153-04, and 1-F32-HS00114-01, and the MacArthur Foundation Research Network on Successful Aging through a grant from the John D. and Catherine T. MacArthur Foundation. An earlier version of this article was presented at the Chronic Diseases and Changing Care Patterns in an Ageing Society meeting, Amsterdam, The Netherlands, June 1993.

Address correspondence to Dr. Susan S. Merrill, California Medical Review, Inc., 60 Spear Street, Suite 400, San Francisco, CA 94105.

REFERENCES

1. Verbrugge LM. Recent, present and future health of American adults. Annu Rev Public Health 1989;10:333–61.
2. Verbrugge LM. Gender and health: an update on hypotheses and evidence. J Health Soc Behav 1985;26:156–82.
3. Verbrugge LM, Wingard DL. Sex differentials in health and mortality. Women Health 1987;12:103–45.
4. Barusch AS, Spaid WM. Gender differences in caregiving: why do wives report greater burden? Gerontologist 1989;29:667–76.
5. Hall RG, Channing DM. Age, pattern of consultation, and functional disability in elderly patients in one general practice. Br Med J 1989;301:424–8.
6. Kaplan RK, Anderson JP, Wingard DL. Gender differences in health-related quality of life. Health Psychol 1991;10:86–93.
7. Jette AM, Branch LG. The Framingham Disability Study: II. Physical disability among the aging. Am J Public Health 1981;71:1211–6.
8. Hanis T, Kovar MG, Suzman R, Kleinman JC, Fledman JJ. Longitudinal study of physical ability in the oldest old. Am J Public Health 1989;79:698–702.
9. Kandrack MA, Grant KR, Segall A. Gender differences in health related behaviour: some unanswered questions. Soc Sci Med 1991; 32:579–90.
10. Wingard DL. The sex differential in morbidity, mortality, and lifestyle. Annu Rev Public Health 1984;5:433–58.
11. Verbrugge LM, Balaban DJ. Pattern of change in disability and well-being. Med Care 1989;27:S128–47.
12. Waltz M, Badura B. Subjective health, intimacy, and perceived self-efficacy after heart attack: predicting life quality five years afterwards. Social Indicators Res 1988;20:303–32.
13. Gove WR. Gender differences in mental and physical illness: the effect of fixed roles and nurturant roles. Soc Sci Med 1984;19:77–84.
14. Hibbard JH, Pope CR. Gender roles, illness orientation and use of medical services. Soc Sci Med 1983;17:129–37.
15. Verbrugge LM. Sex differences in complaints and diagnoses. J Behav Med 1980;3:327–55.
16. Guralnik JM, Branch LG, Cummings SR, Curb JD. Physical performance measures in aging research. J Gerontol Med Sci 1989; 44:M141–6.
17. 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.
18. 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.
19. Berkman LF, Berkman CS, Kasl SV, et al. Depressive symptoms in relation to physical health and functioning in the elderly. Am J Epidemiol 1986;124:372–88.
20. Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. Gerontologist 1970;1:20–30.
21. Rosow I, Breslau N. A Guttman Health Scale for the aged. J Gerontol 1969;21:557–60.
22. Nagi SZ. An epidemiology of disability among adults in the United States. Milbank Mem Fund Q 1976;54:439–68.
23. Gijsbers van Wijk CM, van Vliet KP, Kold AM, Everaerd WT. Symptom sensitivity and sex differences in physical morbidity: a review of health surveys in the United States and The Netherlands. Women Health 1991;17:91–124.
24. Ford AB, Folmar SJ, Salmon RB, Medalie JH, Rov AW, Galazka SS. Health and function in the old and very old. J Am Geriatr Soc 1988;36:187–97.

Received July 21, 1995
Accepted May 5, 1996