Comparison of Performance-based and Self-rated Functional Capacity in Spanish Elderly

Montserrat Ferrer, Rosa Lamarca, Francesc Orfila, and Jordi Alonso

Recent data have shown differences between Spain and the United States in the prevalence of reported disability among community elderly. Differences in reporting functional capacity by culture may contribute to these observed differences. The purpose of this study was to estimate the agreement between self-report of disability and performance-based measures for some basic mobility tasks in the community-dwelling elderly of a Mediterranean country. Interviews containing questions about difficulty for walking and rising from a chair, and performance-based measures (walking speed and chair stand tests) were carried out in 626 individuals aged 72 years and older in Barcelona, Spain. Kappa statistics were calculated, and logistic regression models were constructed to identify possible factors associated with under- and overreporting functional capacity. Moderate kappas (0.41–0.55) were found between self-report and performance-based measures. Patients who rated their health as "poor or very poor" were less likely to underreport disability (adjusted odds ratio (OR) = 0.2, 0.4) but more likely to overreport it (adjusted OR = 23.4, 9.9). No significant agreement differences by sex or informant source were found. These findings suggest that Spanish elderly self-report functional capacity accurately and that, contrary to previous results among US elderly, the direction of the observed disagreement is not systematic. Am J Epidemiol 1999;149:228-35.

In the past two decades, assessment of disability has been increasingly used to characterize the health status and health service needs of older populations. Disability in older persons has been generally assessed through self-reported indicators of functioning on different activities of daily living. However, it is only recently that physical performance measures have been developed. Performance-based measures consist of objective observations of functional capacity and are claimed to be applicable cross-culturally because they seem less likely to be as influenced by culture, language, and educational level as are self-reported measures (1–3). A few comparisons between self-reports and performance test results conducted in the United States (4–8) and Canada (9) have been published, but with no consistent conclusions. However, good correspondence (overall agreement >80 percent) has been found in US studies conducted among community-dwelling elderly (7, 8).

In Spain, recent data (10) have shown that prevalence of reported disability in community elderly is somewhat lower than that reported in the US studies (11–13), but comparability may be limited due to differences in methods and definitions. A potential source of this difference is that the Spanish elderly underreport disability. This could be due to a lower availability of long-term care and social services and/or a different cultural approach to aging (for instance, a much higher proportion of Spanish elderly live with their family (10) and institutionalization rates among the elderly are much lower in Spain than in the United States).

While performance-based measures are not a gold standard of disability nor functional capacity, they seem reliable enough and are probably less culturally dependent than self-reports in order to make comparisons across countries (8, 14). For instance, if prevalence of functional limitations as assessed by performance-based measures are similar, differences of prevalences based on self-reports are more likely to be due to cultural or environmental factors.

The purpose of this paper was: 1) to evaluate the agreement between self-report of disability and performance-based measures for some basic mobility tasks.
in the community-dwelling elderly of a Mediterranean country; and 2) to examine the factors which may influence self-report independently of physical function.

MATERIALS AND METHODS

Population sample

A cohort of elderly adults included in the “Health Interview Survey of Barcelona” (15) was studied. Details about the study have been described elsewhere (10, 16, 17). Briefly, from a total sample of 1,632 eligible individuals aged 65 years or older who resided in the city of Barcelona (population = 1.5 million inhabitants), 1,315 (80.6 percent) participated in the baseline interview in 1986 and constituted the study cohort (15). Survivors were considered eligible for a reinterview and physical examination conducted between June 1993 and June 1994, after a median of 7.5 years of follow-up. The analyses presented here are based on the 1993–1994 reinterview data, because this was the only wave on which performance-based and self-report measures were included.

At the moment of the reassessment, from the 1,315 individuals of the initial cohort, 424 had died and 19 were institutionalized. From the 872 individuals alive and living in the community, 735 (84.3 percent) were interviewed (68 refused the interview and 69 could not be contacted). Compared with participants, persons who had no second interview were similar in terms of age, sex, and level of education, and also reported similar perceived health and functional capacity to perform nine basic activities of daily living at the baseline interview. Participants were interviewed at their homes, and performance-based measures were assessed after the interview. Performance-based data were missing for 109 of the interviewees (86 were refusals, 18 did not live in Barcelona, and five presented cognitive incapacity) and thus were excluded from the analyses. The final sample size for this analysis was 626.

Interview

Variables recorded in the interview included: perceived health status, functional capacity, health-related practices, and health-services utilization, as well as sociodemographic and other information. Self-perceived health was assessed by the question, “In general, how would you rate your health: very good, good, fair, poor, or very poor?” (15). Functional capacity was assessed by asking the participants to rate their level of difficulty to carry out nine basic activities of daily living (“What level of difficulty do you have to perform the activity without help or devices? none, a little, moderate or unable to perform the function”). Only the two activities corresponding with performance measures (see below) were chosen for the comparison ("walking" and "standing up and sitting down from a chair"). Two dichotomous variables were created: individuals were first classified as having "self-reported difficulty" if they had reported any level of difficulty, or "without difficulty" otherwise; subsequently, they were considered as having "self-reported need for help" if they had reported being unable to perform the activity without help or devices, or "without need for help" otherwise. Thus, both variables assessed disability, with "self-reported need for help" reflecting a more severe level.

Physical performance measures

Performance measures were adapted from the Established Populations for the Epidemiologic Study of the Elderly (EPESE) performance test, a short battery of tests to assess lower extremity function, which has been used widely in the United States (1, 18) and more recently in the Netherlands (19) to assess functions needed to perform routine daily activities. The measures were designed to be applied by lay examiners in a home setting with limited space (1). Interviewers were specially trained in the administration of the performance measures used in the study. To ensure uniformity of administration, in training interviewers, we used the videotape produced for the EPESE with detailed instructions for administering and scoring the tests, as well as instructions on maintaining the safety of the subjects.

Walking speed test followed procedures used by the EPESE, with the exception of the distance: rather than a course of 8 ft (2.44 m), we used 4 m, as in another study (20). The walking speed test therefore consisted of having the participant walk over a course of 4 m and recording the time needed to complete the entire path. The test was repeated twice and the shortest time was used in the analysis. Participants were instructed to "walk at their usual speed, just as if they were walking down the street to go to the store," and they could use assistive devices if they needed them.

The chair stand test was carried out by having the participant stand up and sit down from a chair five times as quickly as possible and recording the total time required. Chairs were those at the interviewees' homes.

For both performance measures, a three-level ordinal variable was constructed, ranging in score from 0 to 2: 0 = individuals who could not complete the task or the task was not attempted; 1 = slower times (>1st quartile time); and 2 = quicker times (≤1st quartile time).
Analysis

Differences in sociodemographic characteristics and functional status by sex were tested using the chi-square test for proportions.

Comparisons were made between self-reported function measures and similar physical performance tests: 1) reported difficulty to walk was compared with the observed ability to walk 4 m; 2) reported difficulty to stand up and sit down from a chair was compared with the observed ability to rise five times from a chair. Crude agreement and Kappa coefficients were calculated to estimate the agreement between interview and physical examination data (21, 22). To determine accuracy of reported functional problems, sensitivity and specificity (23) were calculated for each comparison. For these analyses, physical performance tests were considered the “true positive,” although this is clearly an arguable assumption. Confidence intervals for these proportions were calculated using the exact binomial formula. Analyses were performed using the statistical package Epi Info (USD Inc., Stone Mountain, Georgia). All p values are two-tailed.

In addition to agreement, analyses examined the percent of bias in subject self-reports compared with observed performance (14, 24). Percent bias was calculated as the ratio of the difference between the proportion of reported disability and the proportion of observed limitation to perform the task, expressed as percent of the proportion of subjects with observed limitation. A positive percent bias indicates that subjects reported disability more often than functional limitation to perform the task was observed. Bias was tested for statistically significant departures from zero using McNemar’s test (25).

Logistic regression models were constructed to identify possible factors associated with under- and overestimation of functional limitation from reported disability. Subject’s sex, age, perceived health, level of education, type of informant (self- or proxy-report), number of chronic conditions, and health-related practices (smoking, alcohol consumption, and physical activity) were included as independent variables. Analyses were performed using the statistical package SPSS-PC (SPSS, Inc., Chicago, Illinois).

RESULTS

The mean age of participants was 79 years (standard deviation (SD) = 5.16). Their sociodemographic characteristics, perceived health status, and both self-report and performance-based measurements of functional capacity are shown in table 1. The proportion of individuals who reported disability was relatively low even when considering the less restrictive definition (31 percent reported “difficulty” for walking and 19 percent “difficulty” for standing up). Women reported more disability than men and also obtained significantly worse scores on the physical performance measures. Individuals with and without physical performance data were of similar age and level of education, and reported similar perceived health and disability, although information for individuals with missing performance test was more frequently provided in interview responses by a proxy.

The comparison between reported “need of help” to walk and the ability to walk 4 m is presented in the top portion of table 2. Specificity (98 percent) was high and kappa was moderate (0.55), indicating that reported disability to walk is acceptably accurate. Nevertheless, false negatives were quite frequent: 42 percent of participants who were unable to complete the walk 4 m in the performance test did not report need for help. For the comparison between reported “difficulty” and the proportion of subjects who walked slowly (table 2, middle third), specificity was again quite high, kappa was moderate, and false negatives were fairly frequent (40 percent of participants who walked slowly did not report difficulty).

The bottom third of table 2 presents the comparison between reported difficulty to stand up or sit down from a chair and observed times to perform five consecutive rises from a chair. For this comparison, specificity (92 percent) was again high and kappa was moderate (0.55), indicating that reported disability to stand up is accurate. Nevertheless, false negatives were moderately frequent (37 percent of participants who were unable to rise five times from a chair did not report difficulty).

For each of the comparisons, although false positive rates were higher than false negative rates, a lack of systematic direction of the disagreement was observed (bias was less than 25 percent and was not statistically significant), indicating that around 50 percent of subjects overreported their limitation (52 percent for “need of help to walk,” 57 percent for “difficulty to walk,” and 49 percent for “difficulty to rise from a chair” (table 2)) while the other half underreported it.

Multiple logistic regression analysis (table 3) showed that perceived health was the variable most strongly and consistently associated with disagreement between performance and self-reports. Those patients who rated their health as “fair,” “poor,” or “very poor” were less likely to under report but more likely to overreport disability than patients who perceived health as “very good” or “good.” In addition, proxy respondents, subjects in the older age groups, and educated subjects tended to underreport significantly less “difficulty” in walking (table 3, first column).
In table 4, results from physical performance tests for the sample studied in this paper ("Spanish sample") are compared with those in a previous US study (1) ("US sample") which used exactly the same procedures with the exception of the distance assessed for the walking speed test (2.44 m instead of 4 m). Both samples showed a similar prevalence of individuals who were unable to perform these tests and required a similar time to complete the chair stand test. Nevertheless, US elderly performed somewhat more poorly than Spanish elderly in the walking speed test.

**DISCUSSION**

The purpose of this study was to examine the agreement between self-report of disability and performance-based measures in the community-dwelling elderly of a Mediterranean community. Although performance measures were considered as the standard criterion for this comparison, some conceptual differences between the two measures exist: while subject self-reports are closer to the concept of disability because they reflect subjective performance within a sociocultural context, the assessment of an external observer is closer to objective functional limitation. An individual with a limitation in a specific task might have adapted to his/her environment and may not perceive this limitation as a disability. Also, a limitation may not translate into disability if the specific activity is not very relevant for the individual in his/her daily life. Therefore, perfect agreement was not expected.
TABLE 2. Agreement between reported disability and observed performance among 626 elderly adults, Barcelona, Spain, 1993–1994

<table>
<thead>
<tr>
<th>Reported performance</th>
<th>Observed performance</th>
<th>Sensitivity* (95% CI)</th>
<th>Specificity* (95% CI)</th>
<th>% Agreement (95% CI)</th>
<th>Kappa (SE)</th>
<th>% bias</th>
<th>p value†</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Need of help</em> to walk</td>
<td></td>
<td>4 meters walk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unable</td>
<td>15</td>
<td>12</td>
<td>0.58</td>
<td>0.98</td>
<td>96</td>
<td>0.55</td>
<td>4.0</td>
</tr>
<tr>
<td>Able</td>
<td>11</td>
<td>571</td>
<td>(0.37–0.77)</td>
<td>(0.96–0.99)</td>
<td>(94–98)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>583</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Difficulty</em> to walk</td>
<td></td>
<td>4 meters walk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>85</td>
<td>75</td>
<td>0.60</td>
<td>0.83</td>
<td>78</td>
<td>0.41</td>
<td>13.4</td>
</tr>
<tr>
<td>No difficulty</td>
<td>56</td>
<td>367</td>
<td>(0.52–0.68)</td>
<td>(0.79–0.86)</td>
<td>(74–81)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>141</td>
<td>442</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Difficulty</em> to stand up from chair</td>
<td></td>
<td>5 rises from chair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>71</td>
<td>41</td>
<td>0.63</td>
<td>0.92</td>
<td>86</td>
<td>0.55</td>
<td>−0.9</td>
</tr>
<tr>
<td>No difficulty</td>
<td>42</td>
<td>455</td>
<td>(0.53–0.72)</td>
<td>(0.89–0.94)</td>
<td>(83–89)</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td>496</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Sensitivity and specificity were calculated considering the physical performance test as the gold standard.
† CI, confidence interval; SE, standard error.
‡ p value for a test of zero bias (McNemar’s test).
§ Slow = subjects who spent >7.5 seconds to walk 4 m; quick = subjects who spent ≤7.5 seconds to walk 4 m.

although comparisons between the two measures were considered informative in order to improve the assessment of disability in the elderly (8, 14).

We found moderate agreement between performance-based and self-report measures (kappa ranging from 0.41 to 0.55). Percent agreement (ranging from 78 to 96 percent) is well in accordance with that reported in comparable studies published in the United States (7, 8). Nevertheless, while the direction of the differences observed in those studies was systematic (most of the time when a disagreement was identified, the reported disability was greater than the functional limitations observed), in our sample, the direction of the differences was not systematic. Around 50 percent of the time when a disagreement between self-report and performance was identified, the reported disability was higher than the functional limitation observed. Around 50 percent of the time when a disagreement between self-report and performance was identified, the reported disability was higher than the functional limitation observed and the opposite was found for the other half. Thus, the level of overreporting of disability observed in our sample was lower than that observed in the US studies. These results, together with the fact that there seems to be no relevant differences in performance-based functional capacity between Spanish and US elderly, suggest that Spanish elderly are truly as functionally limited as their US peers. Therefore, the lower prevalence of self-report disability among Spanish elderly may be attributed to their lower tendency to overreport observed limitations.

Detailed discussion of the causes of a cultural effect when reporting disability in basic mobility activities is beyond the scope of this article. Nevertheless, it may be speculated that different expectations, and differences in the availability and access to health services may contribute to differences in disability reporting for a similar level of objective functional limitation. In any case, our results do support the conclusion that both self-report and performance-based measures of functional limitation provide quite similar aggregate information of non-institutionalized elderly in our culture. This is useful in planning health services. On the other hand, our results also indicate that information based on self-reported disability should be used cautiously for individual assessments.

As in previous studies (7, 26), disagreement between self-report and performance-based measures was influenced by the level of perceived health status. Subjects who rated their health as poor ("fair," "poor," or "very poor") were more likely to overreport functional limitation while subjects who perceived their health as "very good" or "good" tended to underreport limitations. A possible explanation is that individuals with illnesses that affect their general health status may perceive physical symptoms (pain or discomfort) as functional limitation. Nevertheless, the influence of perceived health status remained after adjusting by the number of chronic conditions. Our results suggest that subjects...
whether proxy reports can be used interchangeably with self-reports is an important issue in community elderly surveys. In our study, proxy respondents were less likely to underreport functional limitation than self-respondents, but they did not present significant differences for overreporting. This finding is in contrast to previous studies (4, 6, 14, 33) which have shown that proxy-reports were less accurate. Although cultural differences could explain the higher accuracy of proxy respondents found in our study, a methodological issue should be taken into account. Our study differs from the works mentioned above because instead of comparing the information of the same individual from two sources (proxy and self-report), we obtained information from a proxy only when the eligible participant was unable to self-respond (7.5 percent of the sample). In previous studies (4, 6, 14, 33), they did not present significant differences for overreporting. This finding is in contrast to previous studies (4, 6, 14, 33) which have shown that proxy-reports were less accurate. Although cultural differences could explain the higher accuracy of proxy respondents found in our study, a methodological issue should be taken into account. Our study differs from the works mentioned above because instead of comparing the information of the same individual from two sources (proxy and self-report), we obtained information from a proxy only when the eligible participant was unable to self-respond (7.5 percent of the sample). In fact, Corder et al. (34), who also obtained their data through proxy respondents, found in our study, a methodological issue should be taken into account. Our study differs from the works mentioned above because instead of comparing the information of the same individual from two sources (proxy and self-report), we obtained information from a proxy only when the eligible participant was unable to self-respond (7.5 percent of the sample). In fact, Corder et al. (34), who also obtained their data from a unique information source, found small differences by proxy respondent in the accuracy of reporting service use. Proxy respondents are probably individuals who are very involved with the subject’s care.

Am J Epidemiol Vol. 149, No. 3, 1999
because the majority of proxy responses were due to the difficulty of obtaining a response from the eligible participant because of health problems, and therefore, they are likely to better know the participant’s functional capacity. This finding is important since it suggests that in severely impaired individuals the proxy responses are valid.

Some limitations of our study should be discussed. First, only agreement for two specific mobility tasks has been examined. Because functional capacity is typically assessed through a global score obtained from several tasks and agreement could differ by specific tasks and by the manner in which measures are scored (14), our results are limited and inferences have to be made cautiously. Second, comparisons with observed functional ability to perform a task among US elderly are based on point estimations. Nevertheless, for ability to rise from a chair, estimations were similar, with the US means being included in the Spanish confidence intervals; and for walking speed, although US means were lower and lay outside the Spanish confidence intervals, it should be pointed out that observed speeds were not directly comparable because walked distances were different. Finally, with regard to the representativeness of our results, despite the fact that the percent of subjects with missing performance-based data was large (15 percent), these subjects reported similar disability levels to those who completed physical performance tests and it is unlikely that their level of agreement was different.

Despite these limitations, our findings suggest that Spanish elderly self-report functional capacity accurately and, contrary to previous results among US elderly (7, 8), the direction of the observed disagreement is not systematic. Our study has also shown that women were not more likely to overreport disability, suggesting that the higher prevalence of reported disability among women reflects a higher rate of functional limitation instead of differences in reporting by sex.

ACKNOWLEDGMENTS

This study was supported by a grant from Fondo de Investigaciones Sanitarias (FIS Expte. 91/0629). Montserrat Ferrer was supported by a grant from Generalitat de Catalunya (CIRIT/PQS-1996). Rosa Lamarca was supported by a grant from Instituto Carlos III (97/4364).

The authors thank Dave Macfarlane for help with data management and in preparing the manuscript for publication and Dr. Antonia Domingo for comments on a previous version of the manuscript.

REFERENCES


Am J Epidemiol Vol. 149, No. 3, 1999