Performance-Based and Self-Reported Physical Functioning in Low-Functioning Older Persons: Congruence of Change and the Impact of Depressive Symptoms

Gertrudis I. J. M. Kempen, Mark Sullivan, Eric van Sonderen, and Johan Ormel

1Department of Medical Sociology, Maastricht University, The Netherlands.  
2Psychiatry and Behavioral Sciences, University of Washington, Seattle, USA.  
3Northern Center for Healthcare Research and 4Department of Psychiatry, University of Groningen, The Netherlands.

RESULTS from cross-sectional studies of older persons clearly indicate that levels of self-reported physical functioning are only moderately related to levels of performance-based physical functioning (e.g., Cress et al., 1995; Deeg, 1994; Jette & Branch, 1985; Kelly-Hayes, Jette, Wolf, D’Agostino, & Odell, 1992; Kempen, Steverink, Ormel, & Deeg, 1996; Mendes de Leon, Seeman, Baker, Richardson, & Tinetti, 1996; Myers, Holliday, Harvey, & Hutchinson, 1993; Reuben, Valle, Hays, & Siu, 1995; Sager et al., 1992). This indicates a lack of congruence between objective and subjective measures of functioning. Why the association between self-reported and performance-based levels of physical functioning is not stronger is unclear. One explanation has focused on the fact that depressive symptoms are more strongly related to self-reported but not on performance-based physical functioning. There is some evidence that part of the discrepancy between self-reported and performance-based physical functioning is due to the effects of depressive symptoms on self-reported but not on performance-based physical functioning. Several cross-sectional studies showed, for example, that older persons with depressive symptoms reported lower levels of physical functioning than those without such symptoms while controlling for levels of performance-based physical functioning (Cress et al., 1995; Kempen, Van Heuvelen, et al., 1996; Kempen, Steverink, et al., 1996). Other researchers reported substantial negative associations between depressive symptomatology and self-reported physical functioning (e.g., Griffiths et al., 1987; Kempen & Suurmeijer, 1991; Laukkanen, Kappinen, Era, & Heikkinen, 1993; Rozzini, Frisoni, Bianchetti, Zanetti, & Trabucchi, 1993; Turner & Noh, 1988; Wells et al., 1989). Cress and colleagues (1995) found modest negative correlations between depressive symptoms and performance-based physical functioning, whereas the negative correlations between depressive symptoms and self-reported functioning were higher. Penninx and colleagues (1998) recently showed that older persons who reported depressive symptoms were at greater risk of subsequent physical decline, but that the relationship between depressive symptoms and performance-based physical decline was not very strong. Rozzini and colleagues (1993) and Kempen, Steverink, and colleagues. (1996) found no significant relationship between performance-based physical functioning and depressive symptomatology in older persons. The results of these studies suggest that depressive symptoms are more strongly related to self-reported physical functioning than to performance-based physical functioning due to differences in the context of functioning (demand properties) or differences in report methods.

To date, most research on the associations between self-reported and performance-based physical functioning and depressive symptoms has been cross-sectional. Longitudinally, there might be a weaker association between self-reported and performance-based physical functioning (congruence of change) because of the possible impact of confounders in cross-sectional analyses and increased measurement error in change variables. However, in older persons with stable levels of depressive symptoms, the congruence in measured change may be relatively strong because there is no change in depressive symptoms that may influence it.

The objectives of this article are threefold: (a) to examine the congruence in measured change between self-reported physical functioning and performance-based physical functioning (arrow a in Figure 1); (b) to examine whether subsequent change in self-reported physical functioning can be predicted from initial levels of and change in depressive symptoms (arrows b and d in Figure 1); and (c) to study whether the congruence in measured
change between self-reported and performance-based physical functioning is related to initial levels of and change in depressive symptoms (arrows c and e in Figure 1).

In fact, we are examining whether the cross-sectional findings mentioned previously can be confirmed with data from a two-wave panel study with a 2-year interval. We seek to answer the following specific questions: (a) How strongly are levels of self-reported and performance-based physical functioning longitudinally associated? Is this longitudinal association weaker or stronger than the cross-sectional one? (We predict a weaker association.); (b) Can changes in self-reported physical functioning be predicted from initial levels of depressive symptoms while controlling for changes in performance-based levels of physical functioning?; (c) Can changes in self-reported physical functioning be predicted from changes in depressive symptoms while controlling for changes in performance-based physical functioning?; and (d) Does the degree of congruence in measured change between performance-based and self-reported physical functioning depend on either initial levels and/or changes in depressive symptoms? We selected a low-functioning older sample, as changes in physical functioning within a 2-year interval are expected to be greater in moderately disabled older persons than in healthy elders.

METHODS

Sample

Data were obtained from a subsample \( n = 753 \) of baseline participants \( N = 5,279 \) in the Groningen Longitudinal Aging Study (GLAS) in 1993. GLAS is a population-based prospective follow-up study of the determinants of health-related quality of life of elderly people, with special emphasis on physical and social disability and well-being (Kempen, Jelicic, & Ormel, 1997; Kempen, Ormel, Brilman, & Relyveld, 1997; Ormel et al., 1998; Ormel et al., 1997). The primary objective of GLAS is to identify the psychosocial factors that influence the trajectory of quality of life, either independently or in combination with disease-related factors. The source population of GLAS is defined as people aged \( \geq 57 \) years living either on their own or in adapted housing for elderly people in the north of The Netherlands in 1993. Research participants with severe cognitive impairments at baseline (MMSE < 17; Folstein, Folstein, & McHugh, 1975) were excluded \( n = 78 \).

The subsample in this paper comprises 753 elderly people (14.3%) with the lowest scores on the six-item physical functioning scale of the MOS Short-Form General Health Survey (SF-20; Stewart, Hays, & Ware, 1988). The selected persons reported four (35.9%), five (45.7%), or six (18.5%) physical limitations on this scale in 1993 indicating substantial levels of disability (Kempen, Steverink, et al., 1996). Examples of items are: “Has your health limited you in moderate activities, like moving a table, carrying groceries or bowling?”; “Walking uphill or climbing a few flights of stairs?”; “Bending, lifting or stooping?”; and “Walking one block?” Five hundred seventy-five persons (76.4%) participated in follow-up interviews in 1994 and 1995, 1 and 2 years after the baseline. Attrition \( n = 178 \) was due to mortality \( n = 58 \), bad physical and/or mental health \( n = 66 \), and refusal \( n = 54 \). One other participant has
been excluded because of a missing value on one of the variables in 1995 (i.e., depressive symptomatology).

The subsample in the present study consisted of 428 women (mean age = 72.8; range = 57–91; SD = 7.3 in 1993) and 146 men (mean age = 71.3; range = 57–93; SD = 8.6). Of the research participants, 110 (19.2%) were younger than 65 years of age, and 218 persons (38.0%) were 75 years of age or older.

Measures
Self-reported functioning was assessed during all three waves with the 11-item ADL subscale of the Groningen Activity Restriction Scale (GARS; Kempen & Suurmeijer, 1990; Kempen, Miedema, Orlunigum, & Molenaar, 1996). Examples of items are: “Can you, fully independently, dress yourself?”, “Can you, fully independently, stand up from sitting in a chair?”, “Can you, fully independently, get around in the house?”, and “Can you, fully independently, get on and off the toilet?” The theoretical score range varies from 11 (no limitations) to 44 (maximum of limitations). The internal reliability estimate (Cronbach’s alpha) in 1993 was .80. Results from previous studies show that the GARS meets the stochastic cumulative scalability criteria of the Mokken model. (Kempen, Myers, & Powell, 1995; Kempen & Suurmeijer, 1990; Kempen, Miedema, et al., 1996; Suurmeijer et al., 1994).

Performance-based physical functioning in all three waves was assessed with three simple tests. The tests were obtained from the Longitudinal Aging Study Amsterdam (Deeg, Knipscheer, & Van Tilburg, 1993) and described in Kempen, Steverink, and colleagues. (1996). The tests consist of (a) putting on and taking off a jacket, (b) walking 6 meters (including a 180-degree turn after 3 meters), and (c) standing up and sitting down 5 times from a kitchen chair, without using one’s arms. The score in each test is the number of seconds required for the performance of the task. Although putting on a jacket and taking it off again were performed separately, the scores of both tests were added up. Higher scores reflect poorer performance. Subjects for whom no timed score was available (i.e., they tried but were unable to do the test, were told to stop for safety reasons, or they refused) were assigned a score equal to the worst score from the three tests (see Results). Correlation analyses revealed that these 178 nonparticipants were significantly older (mean age of 75.2 years vs 72.5 years at baseline; t test, p < .001), had poorer levels of performance-based physical functioning at baseline (GARS: 19.5 vs 16.9; t test, p < .001), had reported higher levels of depressive symptoms at baseline (7.0 vs 5.9; t test, p < .001) than those who participated in the follow-up. The differences in depressive symptoms, self-reported and performance-based physical functioning for those lost to follow-up remain significant when adjusted for age (Analyses of Variance, p < .01). However, an additional analysis (not presented in this article) showed that the correlations between self-reported physical functioning, performance-based physical functioning and depressive symptoms in 1993 (n = 754, see Results section) were very similar to those computed with the total sample of 1993 (n = 753) and fairly similar to those of the nonparticipants in 1995 (n = 178).

Statistical Analysis
Summary scores for performance-based physical functioning were created using the sumscores of the three tests for each wave. This procedure was justified after principal component analyses with the three tests (see Results). Correlation and regression analyses were used to analyze bivariate and multivariate associations. First-order interaction terms, indicating whether the association between changes in performance-based and changes in self-reported physical functioning is dependent on (the changes in) depressive symptoms, were included in the regression equations. In addition, baseline levels of self-reported physical functioning, age, gender, and the number of chronic medical conditions at baseline were included as covariates. The last variable was measured with a list of 18 conditions that is used by the Dutch Central Bureau for Statistics in their health interview surveys (for detailed information: see Kempen, Jelicic, et al., 1997; and Kempen, Ormel, et al., 1997).

Change scores were created by subtracting absolute baseline scores from absolute follow-up scores. All analyses were conducted with SPSS/PC 5.0.2 (Norusis, 1992).

Results
The mean level of self-reported physical functioning decreased from 16.9 in 1993 to 17.4 in 1994 (t value: −3.28, p = .001), and from 17.4 in 1994 to 18.0 in 1995 (t value: −3.43, p = .001). Paired t tests showed no significant differences in the decrease of physical functioning between the 1993–1994 and 1994–1995 intervals for the total sample (t value: 0.19, p = .852). We conclude that the decline in physical functioning may be considered linear across the three measurement waves (e.g., Kempen, Van Sonderen, & Ormel, 1999). Similar results were
obtained for performance-based physical functioning and depressive symptoms. As a result, we decided to include only data from the 1993 and 1995 waves in the present article.

Table 1 gives an overview of the outcomes of the principal component analyses for the three performance-based physical functioning tests in 1993 and 1995. Only the first component had an Eigenvalue > 1 in both waves, and there was a large gap between the first and the second component (Eigenvalue of .77 and .69, respectively; not in the table). The results clearly showed one underlying dimension in both waves. As a result, summary scores for both waves were created using the sum-scores of the three tests in each wave.

Table 2 shows the descriptive statistics for self-reported physical functioning, for performance-based physical functioning, and for depressive symptoms in 1993 and 1995. Levels of functioning for all three variables declined significantly between 1993 and 1995. The range of change between 1993 and 1995 (not in the table) varied for self-reported physical functioning (GARS) from -21 to +24 (mean = 1.1; SD = 4.4), for performance-based physical functioning from -426 to +412 seconds (mean = 14.0; SD = 80.8), and for depressive symptoms (HADS) from -12 to +14 (mean = 0.7; SD = 3.7).

The Pearson correlation coefficients for self-reported with performance-based functioning in the same wave were .57 in 1993 and .71 in 1995 (difference statistically significant, \( p < .001 \); testing procedure according to Lindeman, Merenda, & Gold, 1980). The correlation coefficients for concurrent functioning with depressive symptoms were .17 (self-reported functioning in 1993, \( p = .000 \)), .09 (performance-based functioning in 1993, \( p = .037 \)), .30 (self-reported functioning in 1995, \( p = .000 \)), and .20 (performance-based functioning in 1995, \( p = .000 \)), respectively. The differences between the 1993 and 1995 correlations were statistically significant (\( p < .05 \)).

The correlation analyses showed moderate but significant correlations between changes in self-reported physical functioning and changes in performance-based physical functioning: .28 (\( p = .000 \)). The correlation coefficients for changes in functioning with changes in depressive symptoms were .26 (self-reported, \( p = .000 \)) and .13 (performance-based, \( p = .001 \)); the difference between these two correlation coefficients is statistically significant (\( p < .01 \)).

Multiple regression analysis was conducted to examine whether change in self-reported functioning can be predicted from (a) baseline levels of depressive symptoms and (b) changes in depressive symptoms while controlling for changes in performance-based physical functioning. Two regression equations were estimated according to the two models of Figure 1. Tables 3 and 4 show the results of both regression analyses. Table 3 shows that baseline levels of depressive symptoms are not significantly related to subsequent change in self-reported physical functioning. The significance of the interaction term in Table 4 indicates that the strength of the association between changes in performance-based and changes in self-reported physical functioning is dependent on changes in depressive symptoms. The variance inflation factor (VIF) was used to measure collinearity in both regression models. The highest VIF was 4.0 and much lower than 10.0, which is considered the maximum acceptable (Kleinbaum, Kupper, & Muller, 1988).

We used the original, unstandardized time scores of the performance-based tests before aggregating them. This may have led to inflation to the overall score by a particular measure because their time ranges were different. However, additional regression analyses with standardized test scores before aggregating them showed hardly any other outcomes as compared with the results in Tables 3 and 4.

Finally, we analyzed the significant interaction effect of changes in performance-based physical functioning and changes in depressive symptoms on changes in self-reported physical functioning in detail. The total sample was divided into three groups with a substantial number of participants in each group: participants who improved on depressive symptoms between 1993 and 1995 (sumscore on the HADS subscale at least two points lower in 1995 compared to 1993, \( n = 143 \)), participants who remained stable (maximum acceptable (Kleinbaum, Kupper, & Muller, 1988)).

### Table 1. Summary of Principal Component Analyses for Three Performance-Based Physical Functioning Measures in 1993 and 1995 (\( N = 574 \))

<table>
<thead>
<tr>
<th>Measure</th>
<th>Loadings on First Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1993</td>
</tr>
<tr>
<td>Putting on/taking off jacket</td>
<td>.77</td>
</tr>
<tr>
<td>Walking 6 meters</td>
<td>.87</td>
</tr>
<tr>
<td>Standing up from chair</td>
<td>.69</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>1.81</td>
</tr>
<tr>
<td>Amount of variance explained</td>
<td>60.3%</td>
</tr>
</tbody>
</table>

### Table 2. Mean Scores and Standard Deviations for Self-Reported and Performance-Based Physical Functioning and Depressive Symptoms in 1993 and 1995 (\( N = 574 \))

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Self-reported physical functioning</td>
<td>16.9</td>
<td>5.0</td>
<td>18.0</td>
</tr>
<tr>
<td>Performance-based functioning</td>
<td>75.5</td>
<td>71.4</td>
<td>89.5</td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>5.9</td>
<td>3.7</td>
<td>6.5</td>
</tr>
</tbody>
</table>

*All 1993-1995 differences in mean scores are statistically significant \((t \text{ test}, p < .001)\).

*Higher scores indicate poorer function.

*Number of required seconds.

*Poorer indicates poorer functioning in 1995 compared to 1993 by at least 2 points or 5 seconds on the scale; stable indicates a maximum difference between 1995 and 1993 by 1 point or 5 seconds on the scale; better indicates better functioning in 1995 compared to 1993 by at least 2 points or 5 seconds on the scale.
point on the HADS subscale between 1993 and 1995, 574), and participants whose depressive symptoms worsened (sum-score on the HADS subscale at least two points higher in 1995 compared to 1993, 574). Pearson correlations between changes in performance-based and self-reported physical functioning were computed for each of the three groups. The three correlations were: (improved depressive symptoms) .01 (p = .873), (stable depressive symptoms) .20 (p = .004), and (worse depressive symptoms) .42 (p < .000). Table 5 shows the 1993 and 1995 mean scores and standard deviations for self-reported functioning, performance-based functioning and depressive symptoms for each of the three depression groups. Participants who remained stable or worsened with respect to depressive symptoms reported significantly lower levels of physical functioning in 1995 compared to 1993. Participants whose depressive symptoms improved reported similar levels of physical functioning in 1993 and 1995. Table 2 already showed that the self-reported levels of physical functioning in the total sample declined from 1993 to 1995. Performance-based physical functioning of the three groups follows a less consistent pattern. Substantial decline in performance-based physical functioning is particularly seen in participants whose depressive symptoms worsened between 1993 and 1995. We found a smaller decline in performance-based functioning in participants who improved with respect to depressive symptoms (p = .014).

**DISCUSSION**

What can be concluded from the results of this study? First, there is only a moderate association between changes in self-reported physical functioning and changes in performance-based physical functioning (r = .28), much lower than the cross-sectional associations (r = .57 in 1993 and r = .71 in 1995; difference is significant). This was expected, because of the possible effects of confounders in particular cross-sectional analyses and increased measurement error in change variables. Second, there is a significantly stronger association between changes in depressive symptoms and changes in self-reported functioning (r = .26) than between changes in depressive symptoms and changes in performance-based physical functioning (r = .13; difference is significant), which is consistent with previous cross-sectional studies. Third, baseline levels of depressive symptoms are not predictive for change in self-reported physical functioning. Fourth, the strength of the association between changes in self-reported and performance-based physical functioning, respectively, depends on the changes in depressive symptoms. Changes in self-reported and performance-based physical functioning are most strongly related in elderly persons whose depression worsened. This may be due to the fact that levels of both self-reported and performance-based physical functioning declined most for participants whose depressive symptoms worsened. Although it should be stressed that causal inferences cannot be drawn from our outcomes, the results support the idea that worsening depression has a clearer and more significant effect on physical functioning in the elderly population than stable or improved depressive symptoms. Preventing a decline in depressive symptoms may thus prevent a decline in physical functioning to some extent.

Although our results suggest that older persons with higher levels of depressive symptomatology report also higher levels of physical dysfunctioning, one can argue whether the results can be interpreted as depressive realism (e.g., Haaga & Beck, 1995): depressed persons probably may be more accurate in describing their level of functioning, whereas nondepressive persons may overestimate their level of functioning.

Our study has several limitations. First, nonresponse analyses showed significantly poorer levels of functioning for nonresponders than responders at baseline. This supports the hypothesis that older persons with high frailty levels in particular were not able or refused to participate in the follow-up. Although it is not entirely clear how this affected the outcome of our study, the results of an additional correlation analysis of 1993 levels of self-reported and performance-based physical functioning and depressive symptoms in follow-up participants (n = 574) were
greatly similar to those computed for the total sample in 1993 (n = 753) and fairly similar to those who did not participate in 1995 (n = 178). From this point of view, there is no reason to believe that the outcome was strongly biased by attrition. Second, note that for the assessment of performance-based levels of physical functioning only three basic but quite commonly used tests were applied in the present study. Although we consider these tests as representative indicators of physical functioning in daily life, they do not exactly reflect the items of the self-report questionnaire. However, using cross-sectional data Kempen, Steverink, and colleagues. (1996) recently showed that the associations between similar aspects of self-reported and performance-based functioning are only moderate. It is important to find out whether and how depressive symptoms affect more complex levels of functioning such as instrumental activities of daily living and mobility. Additional research is necessary to study this matter. Third, our conceptualization of the time-to-complete measure for performance-based physical functioning may be problematic, because (as the cognitive aging and neurological literatures have shown) there is a general behavioral slowing that is somewhat independent of increasing physical impairment. Moreover, in tasks like the brief walk task, slowing reflects not only impairment, but also cautiousness, personal style, etc. Therefore, performance-based and self-reported evaluations of functioning may be different constructs. Unfortunately, in our present study we did not ask our participants how long it took to conduct specific tasks. It would be of interest to include such a measure in future studies.

The results of our study suggest that preventing a decline in depressive symptoms may prevent declining physical functioning in low-functioning older persons to some extent. To increase our understanding of this association, more studies are needed, in which interventions to reduce depressive symptoms are evaluated with respect to their ability to improve physical functioning in older persons.

**ACKNOWLEDGMENTS**

This research is part of the Groningen Longitudinal Aging Study (GLAS). GLAS was conducted by the Northern Centre for Healthcare Research (NCH) and various Departments of the University of Groningen. The primary departments involved are Health Sciences (Prof. Dr W. J. A. van den Heuvel), Family Medicine (Prof. Dr. B. Meyboom-de Jong), Psychiatry (Prof. Dr. J. Ormel), Sociology (Prof. Dr. S. M. Lindenberg), and Human Movement Sciences (Prof. Dr. P. Rispens). GLAS and its subsudies are financially supported by the Dutch government (through NESTOR), the University of Groningen, the Faculty of Medical Sciences in Groningen, the Dutch Cancer Foundation, and the Netherlands Organization for Scientific Research.

Address correspondence to Dr. G. I. J. M. Kempen, Department of Medical Sociology, Faculty of Health Sciences, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: G.Kempen@zw.unimaas.nl

**REFERENCES**


Received July 20, 1998
Accepted June 14, 1999