Eur J Ageing (2006) 3: 15–21
DOI 10.1007/s10433-006-0021-5

ORIGINAL INVESTIGATION

Stefan Fors · Mats Thorslund · Marti G. Parker

Do actions speak louder than words? Self-assessed and performance-based measures of physical and visual function among old people

Published online: 22 February 2006

Abstract Self-assessments and performance-based tests are methods commonly used to assess physical function in health surveys of older people. It has been suggested that the choice of method could affect the results, particularly in certain groups. This study compares results attained using self-assessed and performance-based measures of upper and lower body function and vision. The influence of sex, age, education and cognitive status is explored. This is done by studying the prevalence rates of self-reported and observed limitations in function, the prevalence rates of discrepancies between measures and the prevalence rates as well as the odds ratios of discrepancies depending on sex, age, education and cognitive status. Data are from a nationally representative sample of the Swedish population aged 77 or above ($n=492$). The results show that discrepancies occur among a minority of the sample and with no distinctive bias toward either under- or overestimations of functional ability at the cross-sectional level. Cognitive impairment seemed to increase the risk of discrepancies. Women showed an increased tendency toward discrepancies between measures of upper body function. Age and education showed associations with some discrepancies but were not significant in the multiple regression models. In conclusion, there is a risk of systematic biases in the association between self-assessed and performance-based measures of function. At the cross-sectional level, however, these differences are small.

Keywords Methodology · Objective measures · Self-reports

S. Fors (✉) · M. Thorslund · M. G. Parker
Aging Research Center, Karolinska Institute and Stockholm University, Gävlegatan 16, 113 30, Stockholm, Sweden
E-mail: stefan.fors@ki.se
Tel.: +46-8-6906878
Fax: +46-8-6906889

Introduction

There has been an increasing number of studies focused on physical function and disability among older people. These studies sometimes reveal contradictory tendencies, for example on health trends in the older population. Some of these inconsistencies could be explained by the use of different indicators of health (Freedman et al. 2004; Thorslund and Parker 2005). A wide array of instruments has been developed in order to measure physical function in large populations. There are two major traditions, one based on self-assessments and the other on tests of performance. Typically the instruments based on self-assessments consist of a series of questions about activities of daily life (ADL) (e.g., shopping, bathing) or mobility (e.g., walking a given distance, climbing stairs). Performance-based instruments are standardized tests designed either to mimic ADLs (e.g., spooning beans into a can to simulate eating) or to measure more specific dimensions of function (e.g., hand strength, reach, balance).

The benefits of self-assessments are obvious: they offer a quick, inexpensive and easily administered path to information on a wide range of conditions, both general and specific. The information can be gathered by face-to-face or telephone interviews as well as by questionnaires. If a person is too ill or cognitively impaired to answer directly, the same items can be posed to a proxy. However, a review article by (Guralnik et al. 1989) showed that self-assessments were of limited value in identifying clinically significant change. Additionally, the self-assessed measures sometimes lack reliability, causing fluctuations in results that do not reflect any actual change in the level of physical function. Furthermore, it is suggested that self-assessed measures of physical function are sensitive to the influence of cognitive impairment, culture, language and education.

Instruments based on standardized performance tests seem to be gaining ground. Several studies have developed and used tests that appear to be safe, quick and easy to administer to large samples (Guralnik et al. 1994; Reuben et al. 1992). It has been suggested that, compared to self-assessments, performance-based measures have clear face validity for the tasks being performed, better reproducibility and sensitivity to change and are less sensitive to cognitive impairment and sociocultural factors (Guralnik et al. 1989; Guralnik et al. 1994; Reuben et al. 1992).

Self-assessed measures of function may encompass several subjective factors, such as individual expectations and aspirations, comparisons to peers and level of functioning in earlier life. Standardized performance tests, on the other hand, measure the functional ability, or capacity, to perform the task in question, but such results may not reflect ability as realized in the everyday life of the participants (Glass 1998). There is a difference between performing a task once during a monitored test and actually functioning in the context of everyday life (Daltroy et al. 1999; Myers et al. 1993). Performance-based measures could produce both under- and overestimations of functional limitations. Some individuals may develop coping strategies or adapt their living situation to enable them to perform certain tasks in their everyday life despite limited capacity, while others might push themselves during a test to perform in a way that they would not be able to on a regular basis.

In the theoretical framework of the International Classification of Functioning, Disability and Health (ICF) there is a distinction made between what an individual does in her ordinary environment (*performance*) and what the individual is capable of in a standardized environment (*capacity*). Different environments can restrict or facilitate performance (WHO 2001). Self-assessed measures most likely draw on the participants' experiences from their everyday life whereas performance-based measures are more specific and standardized. Consequently, some differences between self-assessed and performance-based measures of function could be due to differences in environment.

Several studies have compared self-assessments with other forms of measures, such as proxy reports, observations made by clinicians and performance tests (Brach et al. 2002; Elam et al. 1991; Hong et al. 2004; Kempen et al. 1996a). Conventionally these studies compare a 'subjective' self-assessment measure to 'objective' measure (e.g., observation by a clinician, performance tests), the results are then used as a measure of validity (Sayers et al. 2004; Wijlhuizen and Ooijendijk 1999). Such an approach is, however, not entirely consensual. The notion that performance on a set of standardized tasks could be considered a 'gold standard' has been disputed. In a comparative study Myers et al. (1993) did not find performance-based measures to be psychometrically superior, more acceptable to participants, easier to administer, or easier to interpret than self-assessed measures. Other studies have suggested that self-assessments and performance-based measures tap distinctly separate aspects of function and should thus either be interpreted differently or combined in a complementary fashion (Kivinen et al. 1998; Klein et al. 2000; Reuben et al. 2004). One hypothesis says that the different methods of measuring physical function capture physical deterioration at different stages of the disablement process, as described by Nagi (1991) and Verbrugge and Jette (1994). This hypothesis implies that performance-based measures might detect physical impairment before it is perceived by the participant as a disability (Guralnik et al. 1994; Hoeymans et al. 1996).

In another study, van den Brink et al. (2003) compared the association between self-reported disability and performance-based limitations among older men in Finland, Italy and the Netherlands. The results showed country differences in the association between self-reports and performance tests, suggesting a cultural component in the self-assessed measures of function. Whereas differences in socio-economic status, chronic diseases and performance-based function explained some of the variance, some variation between the countries remained unexplained. Differences in role expectations, availability of assistive devices, linguistic and cultural interpretations of the questions and assessments were suggested as possible explanations.

Yet another study (Merrill et al. 1997) suggested that gender might affect the association between self-assessed and performance-based measures of physical function. Whereas men as well as women both under- and overreported disability, women were more likely to overreport and men were more likely to underreport functional problems. Daltroy et al. (1999) found a similar pattern: women reported more disabilities than men even when controlling for observed function.

There is also the possibility of a social gradient in the subjective assessment of physical function. Iburg et al. (2001) compared self-reported and physician-assessed levels of mobility in the NHANES III study. The study showed that, given the same level of mobility (as assessed by a physician), participants with higher income were less likely to report limitations in mobility. Similarly Melzer et al. (2004) found that participants with higher income had a higher threshold for reporting disability. Iburg et al. (2001) proposes that this effect could be understood as "wishful thinking", i.e., wealthier participants believe they should be healthy and may thus be less likely to perceive difficulties. A social gradient could also reflect the fact that people who are better off are better able to compensate their functional limitations with assistive technology or housing adaptations.

An additional concern when using self-assessments from old people is the influence of cognitive impairment. Since cognitive impairment is associated with aging and since even the simplest self-assessed measure is dependent on reasonable judiciousness on the part of the participant there seems to be good ground for concern. The results from several studies indicate that cognitive impairment is correlated with a higher degree of discrepancies between observed and self-assessed function (Cress et al. 1995; Kelly-Hayes et al. 1992; Parker et al. 1996).

Whereas a fair amount of research has been concerned with comparisons between different methods of measuring physical function very little research has been done regarding different measures of vision. A single item for self-assessment of visual function showed low sensitivity in identifying visual impairment (Hiller and Krueger, 1983). Results from the Beaver Dam Eye Study showed moderate correlations between self-assessed and performance-based measures of visual function (Klein et al. 1999). Similar results were reported by Kempen et al. (1996b). The association between self-assessed and performance-based measures can be expected to differ depending on the nature of the function being measured.

Regardless of whether we accept the notion of performance tests as a 'gold standard', making comparisons between methods is one way to understand the differences between various measures of function. While similarities between different methods suggest validity, studying the discrepancies between them can reveal the structural differences between measures and the results they produce. Most studies mentioned above have compared summarized indexes of self-reports and performance scores. In this study single items, developed to measure specific functions using self-reports and performance tests, are compared.

The aim of this study is to explore potential systematic biases affecting the measurement of physical and visual function among old people. Do sex, age, education and cognitive status affect the association between self-assessed and performance-based measures of physical and visual function?

Methods

Data

The study is based on the Swedish longitudinal study of living conditions of the oldest old (SWEOLD) survey from 2002. SWEOLD is a continuation of the Swedish level of living surveys (LNU), a longitudinal study founded in 1968. The SWEOLD sample consists of all persons from the LNU sample who have reached the age of 75 or above, regardless of whether or not they had been interviewed previously. The sample is representative on a national level and comprises institutionalized as well as community dwelling persons (Lundberg and Thorslund 1996). Proxy interviews were done when direct interviews were not possible due to poor health or cognition. Telephone interviews were conducted when the respondent was reluctant to invite interviewers into their home.

At the time of the 2002 SWEOLD survey, 736 individuals were eligible for participation (previously included in the LNU sample, aged 75 or above and still alive). Interviews were conducted with 621 persons, yielding a response rate of 84.4%.

Due to the nature of this particular study only those who were given direct, face to face interviews were included, a total of 492 participants. Eighty-two proxy interviews and forty-seven telephone interviews were thus excluded.

Measures

Six items were used to measure three areas of physical function: lower body function, upper body function and vision. Each area of function was measured by one self-assessed measure and one performance-based measure. The self-report questions measuring upper and lower body function were posed before the performance tests, with several unrelated items intervening. The self-report question on visual function, on the other hand, were posed after the performance test, these measures were also intervened by several unrelated items. In the performance tests the ability to perform the task in question was observed and assessed by the interviewer.

- Lower body function was measured by the question: *Can you rise from a kitchen chair (without armrests) without difficulties?* The possible answers were: *Yes* or *No*. In the corresponding performance test the participant was asked to rise from a chair with arms crossed over the chest. The interviewer's alternatives were: *Managed without difficulties, Managed with difficulty* or *Did not manage*.
- Self-assessed upper body function was measured with the question: *Can you pour from a coffee pot or bottle without difficulties?* The possible answers were: *Yes* or *No*. The corresponding performance-test had the participant turn her hand holding a packet of salt (weighing 1 kg) from vertical to horizontal and back again. The interviewer's alternatives were as above.
- Self-assessed vision was measured with the question: *Can you read a newspaper without difficulties?* The given answers were: *Yes, without glasses*, *Yes, with glasses*, *No, have certain difficulties* or *No, not at all*. In the performance-test the participant was asked to read the instructions from a bottle of medicine. The font used on the bottle was similar to newspaper text. The interviewer's alternatives were: *Managed without difficulties* (with or without glasses), *Managed with difficulty* (i.e., strained to read), *Could read, but not correctly* or *Did not manage*.

The variables were dichotomized, separating those who managed or reported ability to manage the task from those who reported or exhibited difficulty with or inability to perform the task. The measures of visual function were dichotomized by ability to perform the task without difficulties (i.e., the first two alternatives of the self-assessed measure and the first alternative of the performance-based measure regardless of whether the participant uses glasses or not). The self-assessed and performance-based measures were then paired. If, for example, a participant claimed to be able to stand up from a chair without difficulties but failed to perform the test without difficulty, the result would be noted as an underestimated lower body functional limitation.

Education was measured by self-report and dichotomized to differentiate those with a grade school education from those with an education beyond grade school.

Cognitive status was measured by items from the Folstein mini mental state examination (MMSE) (Folstein et al. 1975; Parker et al. 1996). Due to time constraints, items were selected for a total of 18 of the 30 original points. From the total possible score of 18, a cut-off point was determined using data from the HARMONY study (Gatz et al. 2005). In HARMONY, identically scored MMSE items were examined against clinical dementia diagnoses and a cut-off was found that best distinguished demented from non-demented. Those who score below the cutoff (<12) can be regarded as cognitively impaired, corresponding to having at least mild dementia.

Statistical methods

Statistical analyses were conducted in order to explore the associations between the sex, age, education, cognitive status and the risk of exhibiting discrepancies between self-assessed and performance-based measures of physical and visual function. Bivariate associations (as shown in Tables 1 and 3) were analysed using Cohens Kappa and Chi-square. Multivariate models (as shown in Table 4) were analyzed using logistic regressions.

Results

The first step of the analysis was to explore the prevalence rates of the observed as well as the self-assessed limitations for each of the three areas of physical function and the correlation between the two methods of measurement.

As shown in Table 1, there were minor differences in the prevalence rates of functional limitations between the different measurements. All paired measurements were significantly correlated, the strongest correlation was found between the measurements of vision ($\kappa=0.56$) and the weakest between the measures of upper body function ($\kappa=0.22$).

Discrepancies were found in all three pairs of measurements, but with no clear tendency toward under- or overestimations (Table 2). The least degree of discrepancies was found between the measures of vision, where 10.5% of the self-assessments were discordant in relation to the performance test. The highest degree of discordance was found between the measures of lower body function, where almost 15% of the reports were at odds with the performance.

Table 3 shows the distribution of discrepancies broken down by sex, age, education and cognitive status. Women were more likely than men to overestimate limitations in lower body function and to both under- and overestimate limitations in upper body function. With the exception of overestimated limitations of lower body function, the oldest age group exhibited more discrepancies than the younger groups for all the pairs of measures, however this tendency was only statistically significant for the measures of vision. Participants with an education beyond grade school were less likely to underestimate limitations in upper body function and seemed more likely to overestimate visual limitations. Cognitively impaired participants were more likely to underestimate visual limitations and both under- and overestimate upper body function. For lower body function there was also a non-significant tendency among the cognitively impaired to both under- and overestimate limitations.

Table 4 shows multivariate models exploring the odds ratios of exhibiting discrepancies depending on sex, age, education and cognitive status. When comparing the three pairs of measurements, it seemed women were more likely to exhibit discrepancies, both under- and overestimated limitations, for upper and lower body function, but this tendency was statistically significant only for the measures of upper body function. Men, on the other hand, seemed somewhat more likely to overestimate visual limitations, but not significantly. The oldest age group showed an increased, non-significant risk of discrepancies in all categories with the exception of overestimated limitations in lower body function and underestimated limitations in upper body function.

Education showed only non-significant relationships. Whereas the associations with lower and upper body

Table 1 Prevalence rates (percentages in parenthesis) and correlations of exhibiting/reporting limitations in three areas of physical functioning

	n^{a}	Performance-based limitations	Self-assessed limitations	Kappa	Standard error
Lower body function	430	76 (17.7%)	80 (18.6%)	0.50***	0.054
Upper body function	479	47 (9.8%)	42 (8.8%)	0.22***	0.067
Vision	488	61 (12.5%)	72 (14.8%)	0.56***	0.055

***$P \leq 0.001$; **$P \leq 0.01$; *$P \leq 0.05$; † ≤ 0.10

[a]Due to the nature of the performance test assessing lower body function, participants unable to stand without support were excluded (this group included a significant portion of the cognitively impaired). Combined with a somewhat higher rate of refusals, this explains the lower rate of participation for lower body function. A small number of bedridden participants were also excluded from the performance tests measuring body function

Table 2 Frequencies (percentages in parenthesis) of underestimated and overestimated limitations in three areas of physical functioning

	n	Underestimated disability	Overestimated disability	Discrepancies (total)
Lower body function	430	30 (7.0%)	34 (7.9%)	64 (14.9%)
Upper body function	479	34 (7.1%)	29 (6.1%)	63 (13.2%)
Vision	488	20 (4.1%)	31 (6.4%)	51 (10.5%)

function were mixed, those with an education beyond grade school seemed more likely to exhibit discrepancies between the measures of visual function. Cognitive status showed a strong association with under- and overestimated limitations in upper body function as well as with underestimated limitations in visual function. There was also a non-significant tendency among the cognitively impaired to overestimate limitations in lower body function and vision.

Discussion

This study explored discordance between self-assessed and performance-based measures of physical function and vision. The results showed that sex and cognition seem to affect the relationship between self-assessed and performance-based measures of physical function. Women were more likely than men to both under- and overestimate limitations in upper body function. Cognitive impairment was associated with an increased risk of discrepancies in all categories. However, the effect of cognitive impairment on the assessment of lower body function was considerably smaller and non-significant,

Table 3 Frequencies (percentages in parenthesis) of exhibiting discrepancies between self-assessed and performance-based ability to perform in three different areas of physical function

Variable	Category	Lower body function			Upper body function			Vision		
		n	Underestimated limitation (%)	Overestimated limitation (%)	n	Underestimated limitation (%)	Overestimated limitation (%)	n	Underestimated limitation (%)	Overestimated limitation (%)
Sex	Male	185	11 (5.9)	10 (5.4)†	208	8 (3.8)†	7 (3.4) *	210	8 (3.8)	16 (7.6)
	Female	245	19 (7.8)	24 (9.8)	271	26 (9.6)	22 (8.1)	278	12 (4.3)	15 (5.4)
Age group	77–79	124	6 (4.8)	10 (8.1)	132	9 (6.8)	6 (4.5)	131	1 (0.8) †	8 (6.1) *
	80–84	180	11 (6.1)	14 (7.8)	200	11 (5.5)	9 (4.5)	203	11 (5.4)	7 (3.4)
	85+	126	13 (10.3)	10 (7.9)	147	14 (9.5)	14 (9.5)	154	8 (5.2)	16 (10.4)
Education	Grade school	281	18 (6.4)	26 (9.3)	310	28 (9.0)*	18 (5.8)	321	12 (3.7)	18 (5.6)
	Beyond grade school	143	12 (8.4)	8 (5.6)	163	6 (3.7)	11 (6.7)	161	7 (4.3)	13 (8.1)
Cognition	Normal	380	26 (6.8)	29 (7.6)	405	23 (5.7)**	18 (4.4)***	406	6 (1.5)***	24 (5.9)
	Impaired	49	4 (8.2)	5 (10.2)	67	10 (14.9)	10 (14.9)	76	14 (18.4)	6 (7.9)

***$P \leq 0.001$; **$P \leq 0.01$; *$P \leq 0.05$; † ≤ 0.10. In case of variables with three categories (i.e., age group) the P values refer to the whole effect of the independent variable

Table 4 Odds ratios of exhibiting discrepancies for each pair of measurements, depending on sex, age education and cognitive status

Variable	Category	Area of physical function					
		Lower body function		Upper body function		Vision	
		Underestimated limitation OR	Overestimated limitation OR	Underestimated limitation OR	Overestimated limitation OR	Underestimated limitation OR	Overestimated limitation OR
Sex	Male	–	–	–	–	–	–
	Female	1.28	1.83	2.28†	2.78*	1.06	0.61
Age group	77–79	–	–	–	–	–	–
	80–84	1.34	0.89	0.61	0.86	5.20	0.58
	85+	2.28	0.86	1.03	1.43	2.94	1.76
Education	Grade school	–	–	–	–	–	–
	Beyond grade school	1.40	0.60	0.43†	1.23	1.59	1.53
Cognition	Normal	–	–	–	–	–	–
	Impaired	1.07	1.37	2.68*	3.43**	14.41***	1.21
Nagelkerke R^2		0.03	0.03	0.09	0.09	0.24	0.05
		423	423	466	466	476	476

***$P \leq 0.001$; **$P \leq 0.01$; *$P \leq 0.05$; † ≤ 0.10

than for the other areas of physical function. Since a significant portion of the cognitively impaired were excluded from this particular performance test this could be an effect of selection.

Another possible reason for the discrepancy is the participants' and interviewers' interpretations of 'difficulty' (Reuben et al. 2004). In the self-assessed as well as the performance-based measures the question is whether the participant is able to manage the tasks *without difficulties*. The subjectivity of 'difficulty' could explain some of the variance between groups. The sometimes hypothetical nature of the self-assessment items should also be taken into consideration. Even simple self-report items can often be hypothetical, for example, participants may usually use support when rising from a chair.

The variance in the different measures in this study is limited by a ceiling effect, i.e., the majority of the sample, especially the younger groups, neither report nor exhibit disabilities. This, in turn, makes it difficult to attain statistical significance in the analyses.

While it is not surprising that cognitive impairment affects the ability to assess functional ability, the effect of the other factors calls for further investigation. Besides the significant associations between sex and discrepancies, there are also the low levels of explained variance in the models to consider. These suggest that much of the variance in discrepancies can not be explained by the variables in the models, but require further investigation.

One possible explanation springs from the hypothesis that self-assessed measures of physical function are less sensitive to change than performance-based measures (Guralnik et al. 1989; Reuben et al. 1992). Whereas, this hypothesis generally assumes changes over longer periods of time (e.g., between surveys), it is possible that functional ability is subject to fluctuations over shorter time periods as well. Sleep disturbances, exhaustion and temporary stiffness are but a few conditions that could be expected to affect performance. An example from the actual study illustrates this quite well: during the interview a participant was asked whether he was able to stand up from a chair without using his arms. Uncertain of his own ability to do so, he tried and succeeded. Later in the interview during the performance tests, he was unable to rise from a chair without using his arms, due to fatigue.

Such fluctuations could explain the tendency among women to exhibit discrepancies between the measures of upper body function. This particular task, i.e., turning the wrist while holding a weight, utilizes a complex structure of muscles and joints in the hand and wrist. This is an area especially problematic for older women, which may explain why women were significantly more likely to exhibit discrepancies between the measures of upper body function than men. The existence of a 'gray zone' between ability and disability characterized by temporary fluctuations in physical function, may explain why it can be difficult even for groups of cognitively intact elders to accurately assess their level of physical functioning at any given time.

This gray zone has implications for both self-assessed and performance-based measures. Whereas, self-assessed measures probably are less sensitive to this kind of short-term change, performance-based measures could be less reproducible since they are more influenced by such fluctuations. To make matters worse, there is no reasonable way around this problem. Large-scale surveys rarely possess the means necessary for long-term observations, which would be the obvious solution to this problem. The good news is that, even in this sample of very old participants, this seems to be a limited problem, affecting a minority of the population.

How these discrepancies should be taken into consideration by the researchers remains an open question, should they be considered as measurement errors or do they reflect something else? Is it possible that discrepancies could serve as indicators of "sub-clinical" physical and visual limitations? Do discrepancies predict future limitations, both observed and self assessed?

This study shows that a minority of the sample exhibit discrepancies between self-assessed and performance-based measures of function but that sex and cognition are significantly correlated with some discrepancies. There is no distinct bias toward over- or underestimations of functional limitation except in vision for the cognitively impaired. In other words, choice of method does not affect the results significantly on the aggregated level, e.g., when presenting population prevalences. However, in studies concerned with certain sub-groups or that follow individuals over time the choice of method could affect the results. Thus, when analyzing data from populations including cognitively impaired groups, researchers should be aware of the complex nature of physical function in these groups as well as the possible effect of sex.

Acknowledgements We are grateful for the financial support from the Swedish Council for Working Life and Social Research, grants 2003-0539 and 2002-0919 and from the Swedish Research Council, grant 2000-0893/28. The Karolinska Institute Ethics Committee approved the project.

References

Brach JS, VanSwearingen JM, Newman AB, Kriska AM (2002) Identifying early decline of physical function in community-dwelling older women: performance-based and self-report measures. Phys Ther 82(4):320–328

Cress ME, Schechtman KB, Mulrow CD, Fiatarone MA, Gerety MB, Buchner DM (1995) Relationship between physical performance and self-perceived physical function. J Am Geriatr Soc 43(2):93–101

Daltroy LH, Larson MG, Eaton HM, Phillips CB, Liang MH (1999) Discrepancies between self-reported and observed physical function in the elderly: the influence of response shift and other factors. Soc Sci Med 48(11):1549–1561

Elam JT, Graney MJ, Beaver T, el Derwi D, Applegate WB, Miller ST (1991) Comparison of subjective ratings of function with observed functional ability of frail older persons. Am J Public Health 81(9):1127–1130

Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12(3):189–198

FreedmanVA, Crimmins E, Schoeni RF, Spillman BC, Aykan H, Kramarow E, et al (2004) Resolving inconsistencies in trends in old-age disability: report from a technical working group. Demography 41(3):417–441

Gatz M, Fratiglioni L, Johansson B, Berg S, Mortimer JA, Reynolds CA, et al (2005) Complete ascertainment of dementia in the Swedish twin registry: the HARMONY study. Neurobiol Aging 26(4):439–447

Glass TA (1998) Conjugating the "tenses" of function: discordance among hypothetical, experimental and enacted function in older adults. Gerontologist 38(1):101–112

Guralnik JM, Branch LG, Cummings SR, Curb JD (1989) Physical performance measures in aging research. J Gerontol A Biol Sci Med Sci 44(5):M141–M146

Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al (1994) A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol A Biol Sci Med Sci 49(2):M85–M94

Hiller R, Krueger DE (1983) Validity of a survey question as a measure of visual acuity impairment. Am J Public Health 73(1):93–96

Hoeymans N, Feskens EJ, van den Bos GA, Kromhout D (1996) Measuring functional status: cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990–1993). J Clin Epidemiol 49(10):1103–1110

Hong TB, Zarit SH, Malmberg B (2004) The role of health congruence in functional status and depression. J Gerontol B Psychol Sci Soc Sci 59(4):P151–157

Iburg KM, Salomon JA, Tandon A, Murray CJL (2001) Cross-population comparability of self-reported and physician-assessed mobility levels: Evidence from the third National Health and Nutrition Examination Survey (No. 14). World Health Organization, Geneva

Kelly-Hayes M, Jette AM, Wolf PA, D'Agostino RB, Odell PM (1992) Functional limitations and disability among elders in the Framingham Study. Am J Public Health 82(6):841–845

Kempen GI, Steverink N, Ormel J, Deeg DJ (1996a) 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 51(5):P254–P260

Kempen GI, van Heuvelen MJ, van den Brink RH, Kooijman AC, Klein M, Houx PJ, et al (1996b). Factors affecting contrasting results between self-reported and performance-based levels of physical limitation. Age Ageing 25(6):458–464

Kivinen P, Sulkava R, Halonen P, Nissinen A (1998) Self-reported and performance-based functional status and associated factors among elderly men: the Finnish cohorts of the seven countries study. J Clin Epidemiol 51(12):1243–1252

Klein BE, Klein R, Jensen SC (2000) A short questionnaire on visual function of older adults to supplement ophthalmic examination. Am J Ophthalmol 130(3):350–352

Klein BE, Klein R, Lee KE, Cruickshanks KJ (1999) Associations of performance-based and self-reported measures of visual function. The beaver dam eye study. Ophthalmic Epidemiol 6(1):49–60

Lundberg O, Thorslund M (1996) Fieldwork and measurement considerations in surveys of the oldest old. Experience from the Swedish level of living surveys. Soc Indic Res 37:165–167

Melzer D, Lan TY, Tom BD, Deeg DJ, Guralnik JM (2004) Variation in thresholds for reporting mobility disability between national population subgroups and studies. J Gerontol A Biol Sci Med Sci 59(12):1295–1303

Merrill SS, Seeman TE, Kasl SV, Berkman LF (1997) Gender differences in the comparison of self-reported disability and performance measures. J Gerontol A Biol Sci Med Sci 52(1):M19–M26

Myers AM, Holliday PJ, Harvey KA, Hutchinson KS (1993) Functional performance measures: are they superior to self-assessments? J Gerontol A Biol Sci Med Sci 48(5):M196–M206

Nagi SZ (1991) Disability concepts revisited: implications for prevention. In: A. M. T. Pope, A.R. (ed), Disability in America: toward a national agenda for prevention. National Academy Press, Washington

Parker MG, Gatz M, Thorslund M (1996) Brief cognitive screening in a field survey of the oldest old. Aging (Milano) 8(5):354–359

Reuben DB, Seeman TE, Keeler E, Hayes RP, Bowman L, Sewall A, et al (2004) Refining the categorization of physical functional status: the added value of combining self-reported and performance-based measures. J Gerontol A Biol Sci Med Sci 59(10):1056–1061

Reuben DB, Siu AL, Kimpau S (1992) The predictive validity of self-report and performance-based measures of function and health. J Gerontol A Biol Sci Med Sci 47(4):M106–M110

Sayers SP, Jette AM, Haley SM, Heeren TC, Guralnik JM, Fielding RA (2004) Validation of the Late-Life Function and Disability Instrument. J Am Geriatr Soc 52(9):1554–1559

Thorslund M, Parker MG (2005) Förbättras eller försämras egentligen de äldres hälsa? En genomgång av aktuell svensk och internationell forskning. [Getting better or getting worse? Conflicting evidence and a possible resolution.]. Läkartidningen, (in press)

van den Brink CL, Tijhuis M, Kalmijn S, Klazinga NS, Nissinen A, Giampaoli S, et al (2003) Self-reported disability and its association with performance-based limitation in elderly men: a comparison of three European countries. J Am Geriatr Soc 51(6):782–788

Verbrugge LM, Jette AM (1994) The disablement process. Soc Sci Med 38(1):1–14

WHO (2001) International classification of functioning, disability and health. World health Organization, Geneva

Wijlhuizen GJ, Ooijendijk W (1999) Measuring disability, the agreement between self evaluation and observation of performance. Disabil Rehabil 21(2):61–67