Eur J Ageing (2006) 3: 3–14
DOI 10.1007/s10433-006-0018-0

ORIGINAL INVESTIGATION

P. Hämäläinen · J. Suni · M. Pasanen
J. Malmberg · S. Miilunpalo

Changes in physical performance among high-functioning older adults: a 6-year follow-up study

Published online: 9 February 2006

Abstract Purposes of the study were (1) to investigate changes in physical performance during 6 years follow-up among high-functioning older adults and (2) to describe the selection of study sample with reference to measured performance. Subjects ($n = 1{,}133$) born during 1917–1941 participated in the battery of health-related fitness (HRF) tests (6.1-m walk, stair climbing, backwards walk, trunk side-bending, dynamic back extension, 1-km walk and body mass index) in 1996. Six hundred and six subjects were retested in 2002. In general, poorer fitness in the baseline assessment predicted non-participation in retesting as well as test exclusions and interruptions in retesting. The 6-year changes in the HRF showed a linear trend ($P < 0.01$) according to age group: performance of older groups deteriorated on average more than the performance of younger groups. In most of the tests, gender was statistically significantly ($P < 0.05$) associated with the changes in performance. The mean performance of the women deteriorated in all tests during the follow-up, while the mean performance of the men deteriorated only in the trunk side-bending, 6.1-m walk and 1-km walk tests. It can be concluded that among the subjects who participated in the follow-up testing, older age and being a woman increased deterioration in several components of HRF. Considering the selection of the subjects, the deteriorations identified are very likely underestimations of real fitness changes among this sample.

Keywords Health-related fitness · High-functioning · Older adults · Follow-up · Selection

P. Hämäläinen (✉) · J. Suni · M. Pasanen · J. Malmberg
S. Miilunpalo
UKK Institute for Health Promotion Research, P.O. Box 30, 33501 Tampere, Finland
E-mail: pauliina.hamalainen@uta.fi
Tel.: +358-3-2829219
Fax: +358-3-2829200

P. Hämäläinen · J. Malmberg · S. Miilunpalo
Tampere School of Public Health, University of Tampere, Tampere, Finland

S. Miilunpalo
Kiipula Foundation,
Vocational Training and Rehabilitation Centre,
Turenki, Finland

Introduction

During the last few decades, premature mortality has decreased and life-expectancy of the population has increased. The number of older people has increased and this trend is going to accelerate in the future: As an example in Finland, the population projection for 2002–2040, suggests that the proportion of people aged 65 years and older will increase from 15.8 to 26.6% (Statistical Yearbook of Finland 2004). There is evidence that functional ability of the people decreases (Ferrucci et al. 1996) and the recovery from disability impairs with increasing age (Leveille et al. 2000). In order to successfully intervene this expanding public health burden, we need further knowledge about the development of functional disabilities among different population groups.

Poor muscle strength and balance as well as difficulties in walking have been identified as risk factors for impaired mobility function (Rantanen et al. 1994; Guralnik et al. 1995, 2000), falls (Wolfson et al. 1995), nursing home admissions (Guralnik et al. 1994), hospitalizations (Penninx et al. 2000) and mortality (Guralnik et al. 1994). Studying longitudinal changes in health-related fitness (HRF) among older adults is essential since the change in measured physical performance may be associated with concurrent dependency in activities of daily living as suggested by Gill et al. (1997). Identifying changes in HRF may therefore be useful in predicting future mobility difficulties and disability.

Previous longitudinal studies have reported selection of the study population to healthier and more mobile subjects (Aniansson et al. 1983; Rantanen et al. 1997; Rantanen and Heikkinen 1998; Hughes et al. 2001;

Paterson et al. 2004). The selection bias will lead to underestimations of the performance and fitness changes with aging (Rantanen et al. 1997), and will cause the population estimates of physical performance, if used as norm reference values, to be too high.

The purpose of the present study was twofold. First, the purpose was to describe the selection of the study sample during the follow-up with reference to fitness. Secondly, the target was to analyse the changes in several components of HRF during 6 years of follow-up among high-functioning older adults. The main purpose was to identify gender and age categories that are most vulnerable to diminishing fitness.

Materials and methods

Study sample

The study forms a part of the Kainuu Study on Living Habits and Health. A systematic and regionally representative sample of all residents between the ages of 19 and 63 was drawn from the 1979 census data of a medium-sized industrial town and two rural municipalities in northeast Finland (Oja et al. 1994). The sample included 6,787 men and women, of whom 5,259 (77.5%) answered the baseline questionnaire in 1980. According to national census data from the Central Statistical Office of Finland, a total of 490 men and women (9.3% of the study cohort) died between baseline survey on 1 March 1980 and follow-up survey on 30 September 1996 (Malmberg et al. 2002b). In 1996 and 2002, assessment of HRF for cohort members aged 55 years and older was included into the study. The sample of the present study consisted of all the men and women who were born during 1917–1941 and who participated in the assessment of HRF in 1996 ($n=1{,}133$). The representativeness of the sample has been described by Malmberg et al. (2002a). In the present study selection of HRF test participants was analysed after 6 years. Six hundred and six subjects of the original 1,133 could be retested after the follow-up. The progression of the study is presented in Fig. 1.

Procedures

Postal questionnaires were sent to cohort members and the respondents' ability to participate in the assessment of HRF was pre-screened on the basis of their answers. The exclusion criteria for HRF assessment in both 1996

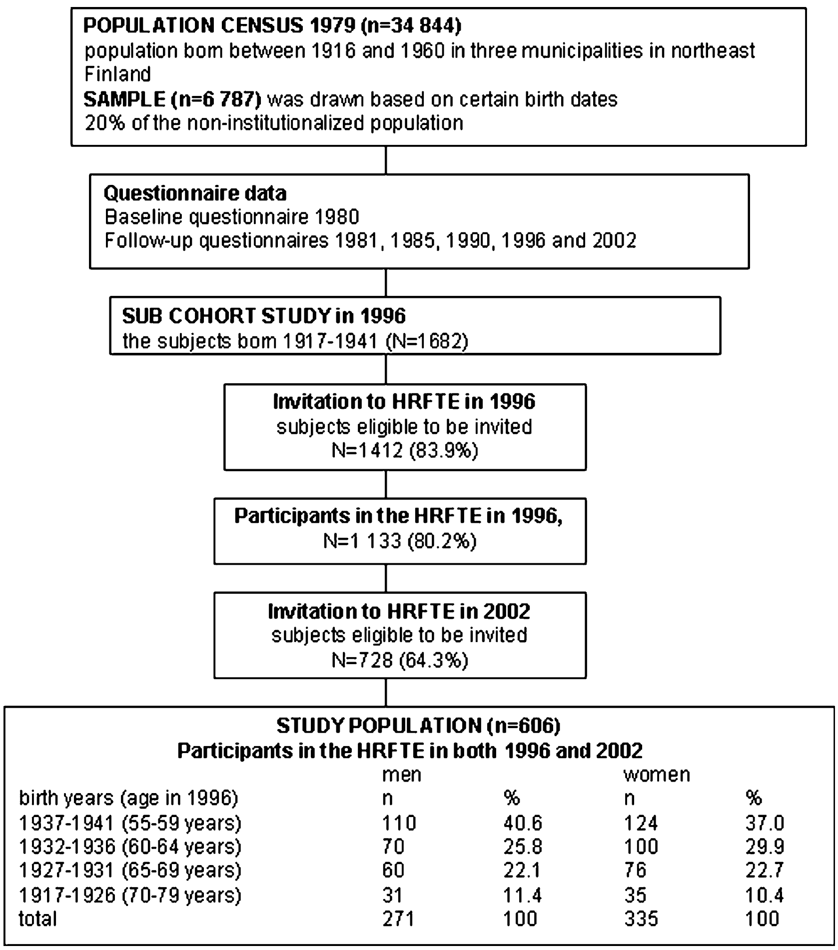

Fig. 1 Data collection for the Kainun Study on living habits and health, and selection of the study population for the present study

and 2002 were "living in an institution" or "severe difficulties or inability to walk independently outdoors and on stairs". Invitation for retesting of HRF was sent to all eligible subjects who answered the follow-up questionnaire. Based on these criteria 16.1% ($n=270$) respondents did not get an invitation to the testing session in 1996. In 2002 35.7% ($n=405$) of the subjects, tested in 1996, could not be invited for retesting (307 subjects did not respond to the follow-up questionnaire in 2002 and 98 subjects did not meet the inclusion criteria). In both 1996 and 2002, the invitation contained information about the purpose of the study and the option to interrupt the testing session at any time. Each participant signed the statement of informed consent before participating in testing. The study was approved by the Ethical Committee of the UKK Institute for Health Promotion Research in 1995 and by the Ethical Committee of Pirkanmaa Hospital District in 2002.

A team of six (in 1996) and four (in 2002) trained health and fitness professionals, all of whom had a degree in sport or health sciences, screened and tested the participants individually at a local gymnasium in each of the three target communities. The testing sites and the pretesting health screening were the same in 1996 and 2002. A detailed description of the health screening process has been previously reported (Suni et al. 1998a; Malmberg et al. 2002a). Briefly, using selected information from the questionnaires and health screening, the testers applied a systematic safety procedure to exclude non-eligible participants from various tests (Suni et al. 1998a; Malmberg et al. 2002a). For example, the participant was regarded as non-eligible for several tests if she/he had any of the following: severe cardiorespiratory or musculoskeletal symptoms or diseases; risk factors for exercise-induced cardiovascular complications, such as significant obesity with inactivity; poor visual acuity hindering safe mobility and testing; or poor hearing accuracy causing difficulties in understanding test instructions.

Assessment of health-related fitness

The proposed battery of HRF tests for elderly (HRFTE) is based on the concept of HRF (Bouchard and Shephard 1994), and it was developed on the basis of a former HRF test battery for the middle-aged adults (Suni et al. 1996, 1998a, b, 1999; Suni 2000). The proposed HRFTE included six tests: 6.1-m (20-ft) walk (Fiatarone et al. 1990; Bassey et al. 1992), stair climbing (Salen et al. 1994), backwards walk (Nelson et al. 1994), trunk side-bending (Suni et al. 1996), dynamic back extension (Mälkiä et al. 1983) and 1-km walk (Oja et al. 1991). Description of the testing methods is presented in Appendix. The safety and feasibility (Suni et al. 1998a; Malmberg et al. 2002a), health (Suni et al. 1998b; Malmberg et al. 2002a) and physical activity (Suni et al. 1999) -related content validity and reliability (Oja et al. 1991; Suni et al. 1996; Rinne et al. 2001) of the different tests have been reported elsewhere.

The main purpose of the HRFTE is to identify individuals with signs of risks for mobility difficulties among the high-functioning older adults. There is evidence that decline in mobility function precedes changes in activities of daily living (Dunlop et al. 1997). The proposed HRFTE, including all the main components of HRF (cardiovascular fitness, muscular fitness and motor fitness; Bouchard and Shephard 1994) was also designed to serve as a tool in promoting physical activity. The fitness assessments help to target exercise individually to specific components of fitness, which are important to mobility function.

In both 1996 and 2002, the testers were trained to conduct the tests in a standard manner. The test order and instructions were consistent for all subjects performing a given test. In timed tests (6.1-m walk, stair climbing, backwards walk, dynamic back extension and 1-km walk) the participants were consistently encouraged to perform the test as fast as possible without risking their health to obtain their best test performance.

Covariates

Age, education, physical activity, perceived health status and body mass index (BMI) were regarded as potential confounding factors when describing selection of the study sample during the follow-up. Education was measured by a self-administered questionnaire in 1980 by answering this following question: "How long is your vocational education?" The original response alternatives were (1) no vocational training or education, (2) vocational training (preparatory courses), (3) secondary education (middle or high school/vocational institute) for 2 years or less, (4) secondary education (middle or high school/vocational institute) for more than 2 years and (5) higher education (university/college). Physical activity and perceived health status were measured by a self-administered questionnaire in 1996. Respondents were asked to report their level of physical activity by answering to this following question: "Which of the following categories best describes your physical activity during the past 12 months? Consider all types of physical activity, including walking and cycling to and from work, if the trip takes at least 15–20 min one way". The original response alternatives were (1) vigorous activity at least twice a week, (2) vigorous activity once a week and in addition other light activities, (3) light intensity activity weekly and (4) no regular weekly activity. Vigorous activity was described in the instructions as intense enough to cause perspiration or breathlessness. Perceived health status was measured by a question that asked: "How do you rate your current state of health?" The original response alternatives were (1) good, (2) fairly good, (3) average, (4) fairly poor and (5) poor. BMI was calculated from the measured weight and height (weight in kilograms divided by height in metres squared) in 1996.

Data analysis

The subjects were divided into four categories according to their age in 1996: 55–59 years (subjects born 1937–1941), 60–64 years (subjects born 1932–1936), 65–69 years (subjects born 1927–1931) and 70–79 years (subjects born 1917–1926).

The selection of the study sample

Differences of group means (1996) between the retest participants and non-participants were analysed by the analysis of variance. The selection of the study sample tested in 1996 was analysed by multinomial logistic regression analysis separately among the men and women. Baseline HRF test results were regarded as continuous independent variables and subjects' ability to be retested in 2002 was regarded as a dependent variable (1 = tested, 2 = test-specific exclusion, 3 = interruption, 4 = not eligible to be invited and 5 = non-response). Group of tested subjects (1) was used as a reference category. Test-specific exclusions were mainly based on cardiovascular and musculoskeletal symptoms and diseases. Interruptions were mainly due to subjects' inability to perform the test according to test instructions. Odds ratios (OR) were computed according to one unit change in test results. Units are presented as sub-indexes of OR (OR_x). The analyses were adjusted for covariates. Due to small number of subjects in some categories of covariates, the original response alternatives were combined into two categories for these analyses (original response alternatives are presented in parenthesis): the education variable was categorized to the subjects with (2–5) and without (1) vocational training or education, the physical activity variable was categorized to the subjects who reported (1–2) and who did not report (3–4) vigorous activity and the health status variable was categorized to the subjects with good (1–2) and with average or poor (3–5) perceived health.

Changes in health-related fitness

The age and gender differences of 6-year changes in the HRF were analysed by the two-way analysis of covariance with log-transformed test variables (due to the skewed distributions of some variables). The test results in 2002 were used as dependent variables, age group and gender as factors and the test result in 1996 as a covariate. Both interaction and main effects of age and gender were tested. Interactions with P values less than 10% ($P < 0.10$) and main effects less than 5% ($P < 0.05$) were considered statistically significant. The anti-log transformation of mean differences in log-transformed variables gives an estimate of the ratio of group means. This also allows the presentation of results as relative changes (percentages) of the test variables. Additionally, changes in HRF were analysed according to gender- and age-specific thirds of the test results in 1996 by the analysis of covariance. The absolute changes in the test results (result in 2002 minus result in 1996) were used as dependent variables, test-specific thirds as factors, and age and gender as covariates.

Results

Selection of the study sample with reference to assessments of health-related fitness

The retested subjects were on an average 4 years younger, had better perceived health status, were less likely to be current smokers and more likely to be physically active than the subjects who could not be retested (Table 1). The crude analysis showed that the subjects who could be retested in 2002 had better baseline HRF than the other subjects (Table 2).

According to confounder-adjusted multinomial logistic regression analysis poorer baseline performance in the 1-km walk ($OR_{min} = 1.32$, 95% CI 1.02–1.71 among the men, $OR_{min} = 1.32$, 95% CI 1.02–1.70 among the women) increased the risk for questionnaire non-response between both genders in 2002. Additionally, poorer performance in the dynamic back extension ($OR_{repetitions} = 1.07$, 95% CI 1.00–1.14) and backwards walk ($OR_{sec} = 1.07$, 95% CI 1.03–1.12) tests increased the risk among the men. Poorer baseline performance in the 1-km walk ($OR_{min} = 2.11$, 95% CI 1.43–3.12 among the men, $OR_{min} = 1.78$, 95% CI 1.37–2.32 among the women), dynamic back extension ($OR_{repetitions} = 1.19$, 95% CI 1.07–1.32 among the men and $OR_{repetitions} = 1.14$, 95% CI 1.03–1.26 among the women) and stair climbing ($OR_{sec} = 1.16$, 95% CI 1.09–1.25 among the men and $OR_{sec} = 1.08$, 95% CI 1.04–1.13 among the women) tests increased the risk for not being eligible to be invited in retesting. In addition, poorer baseline performance in the 6.1-m walk ($OR_{sec} = 2.00$, 95% CI 1.22–3.27) increased the risk among the men. (see also Table 2).

In 2002 the average percentage of test exclusions was the lowest in the 6.1-m walk and the highest in the dynamic back extension (Fig. 2). The overall exclusion rate increased with age, being on average 4.4% in the subjects aged 55–59 years and 16.2% in the subjects aged 70–79 years. The percentage of subjects who were unable to perform the tests according to test instructions (interruption) was the lowest in the 6.1-m walk and the highest in the backwards walk test (Fig. 3). The rate of being unable to perform the test correctly increased with age, being 3.1% in subjects aged 55–59 years and 14.8% in the subjects aged 70–79 years.

Poorer baseline performance in the 1-km walk ($OR_{min} = 2.10$, 95% CI 1.49–2.97 among the men, $OR_{min} = 1.54$, 95% CI 1.14–2.08 among the women) increased the risk for retest exclusion in both genders. Additionally, poorer baseline performance in the dynamic back extension test among the men

Table 1 The general characteristics of the study sample

		Retested in 2002, *n* = 606 (%)	Tested only in 1996, *n* = 527 (%)	*P*
Gender	Female	55	56	
	Male	45	44	0.716
Age group	55–59 years	39	22	
	60–64 years	28	21	
	65–69 years	22	25	
	70–79 years	11	32	< 0.001
Marital status	Single	8	10	
	Married	85	79	
	Widowed	3	6	
	Divorced	4	5	0.050
Education	No vocational training or education	29	38	
	Vocational training (preparatory courses)	39	43	
	Secondary education (middle or high school/vocational institute), for			
	2 years or less	17	9	
	More than 2 years	10	7	
	Higher education (university/college)	5	4	< 0.001
Physical activity	Vigorous activity at least twice a week	31	18	
	Vigorous activity once a week and in addition other light activities	29	20	
	Light intensity activity weekly	39	59	
	No regular weekly activity	1	3	< 0.001
Smoking	Never smoke	65	61	
	Current smoker	9	16	
	Past smoker	26	23	0.001
Perceived health status	Good	12	7	
	Fairly good	36	29	
	Average	49	55	
	Fairly poor	3	9	
	Poor	0	0	< 0.001
BMI	Mean (SD)	27.2 (3.6)	28.4 (4.2)	< 0.001

BMI body mass index (kg/m^2)

($OR_{repetitions}$ = 1.11, 95% CI 1.03–1.21) and in the stair climbing (OR_{sec} = 1.06, 95% CI 1.02–1.11) and trunk side-bending (OR_{cm} = 1.59, 95% CI 1.04–2.42) tests among the women increased the risk for exclusion.

Baseline level of HRF was also associated with inability to follow the test instructions in retesting. Among the women poorer performance in the 1-km walk (OR_{min} = 1.61, 95% CI 1.09–2.40) and dynamic back extension ($OR_{repetitions}$ = 1.17, 95% CI 1.02–1.35) tests increased the risk for test interruption. Among the men poorer performance in the backwards walk test (OR_{sec} = 1.09, 95% CI 1.04–1.14) increased the risk for interruption.

Changes in health-related fitness

Among the subjects who were tested in both years, test results were on an average better among the younger age groups when compared with the older groups, and among the men when compared with the women in both 1996 and 2002. The mean changes in the test results and their 95% confidence intervals from the analysis of covariance by test, age group and gender are presented in Table 3 and Fig. 4. Positive percentages indicate improvement and negative percentages indicate deterioration of the results.

In general, the gender-adjusted analysis of covariance showed that there was a linear trend in the changes of HRF test results according to the age group: the performance of the older groups deteriorated on an average to a greater extent than the performance of the younger groups. Confidence intervals were the largest in the oldest age groups, which indicates a smaller number of subjects in this group and also a larger variability in the test results of the oldest subjects. According to the age group-adjusted analysis, the difference in the changes of the test results between the genders varied between the tests.

The only statistically significant interaction between age group and gender was found in the 1-km walk test (Fig. 4). In general, the test results deteriorated in the subjects aged 60–79 years, but not in the subjects aged 55–59 years. The statistically significant differences in the changes between the men and women were found in the youngest age group, where the men improved their performance but there was no change among the women. In the oldest age group (subjects aged 70–79 years) women deteriorated their performance to a greater extent than the men. In the subjects aged 60–69 years there was no differences between the genders.

With reference to age group, the mean of the test results in the trunk side-bending test deteriorated in the

Table 2 The unadjusted mean test results and the number of tested subjects in 1996 according to whether subject participated in retesting in 2002 or not and whether the retest participants had test results or not

	Retest participants		Retest non-participants								
	Tested mean (SD)	*n*	Excluded mean (SD)	*n*	Interrupted mean (SD)	*n*	Not invited mean (SD)	*n*	Non-response mean (SD)	*n*	*P**
6.1 m (s)	3.6 (0.8)[a,c,d]	601	4.7 (2.3)[a]	4	–	–	4.3 (1.1)[c,i]	98	4.0 (1.0)[d,i]	425	< 0.001
Stairclimbing (s)	29.3 (6.1)[a,c,d]	534	38.9 (11.2)[a,g]	62	34.3 (6.4)	3	41.0 (17.9)[c,i]	92	35.5 (10.5)[d,g,i]	409	< 0.001
Backwards walk (s)	27.8 (9.8)[b,d]	396	32.9 (7.0)	5	35.2 (13.0)[b]	82	30.9 (13.1)	43	34.0 (15.0)[d]	254	< 0.001
Trunkside-bending(cm)	14.9 (3.6)[a,c,d]	592	11.1 (2.6)[a]	7	14.6 (2.3)	2	13.2 (3.5)[c]	94	13.5 (3.8)[d]	418	< 0.001
Dynamic back extension (repetitions)	16.2 (5.8)[a,b,c,d]	452	12.5 (4.2)[a]	89	11.6 (4.1)[b]	23	11.2 (4.6)[c]	63	12.5 (5.0)[d]	353	< 0.001
1-km walk (min)	9.9 (1.4)[a,b,c,d]	492	11.4 (1.8)[a,e,f]	65	12.8 (1.5)[b,e,h]	21	12.1 (2.3)[c,f,i]	73	11.4 (2.0)[d,h,i]	379	< 0.001

*Analysis of variance, *P* value for the overall group difference

Post hoc tests of between-group comparisons (Tukey's test $P < 0.05$): [a]tested versus excluded, [b]tested versus interrupted, [c]tested versus not invited, [d]tested versus non-response, [e]excluded versus interrupted, [f]excluded versus not invited, [g]excluded versus non-response, [h]interrupted versus non-response, [i]not invited versus non-response

subjects aged 60–79 years, while there was no change in the mean of the test results in the subjects aged 55–59 years. The mean of the 6.1-m walk test results deteriorated in all age groups. The deteriorations in these tests (6.1-m and trunk side-bending) were on an average greater in the older age groups than in the younger groups.

The mean of the results in backwards walk improved in the subjects aged 55–59 years, stayed unchanged in the subjects aged 60–64 years and deteriorated in the subjects aged 65–79 years. The mean of the test results in dynamic back extension improved in the subjects aged 55–59 years, remained unchanged in the subjects aged 60–64 years and deteriorated in the subjects aged 65–69 years, but the change in the subjects aged 70–79 years was not statistically significant. In the stair climbing test the mean of the test results improved in the subjects aged 55–64 years, while the mean of the test scores in the subjects aged 65–79 years deteriorated.

With reference to gender, the mean results in the backwards walk test, dynamic back extension and stair climbing test deteriorated among the women, while there were no statistically significant changes in the means of the results among the men. In the 6.1-m walk, the means of the results deteriorated in both genders, but among the women the deterioration was on an average greater than among the men. In trunk side-bending, there were no statistically significant difference between the men and the women.

The change of test results from 1996 to 2002 was associated with the test performance in 1996 in all tests (Table 4). Results in the 6.1-m walk, trunk side-bending and 1-km walk tests deteriorated during the follow-up among the subjects who belonged to the best performing (highest) and intermediately performing (middle) thirds in 1996. Results in the trunk side-bending test improved among the subjects who belonged to the poorest performing (lowest) third in 1996, but there was no change in the results of 6.1-m walk and 1-km walk tests in the lowest third. Results in the backwards walk and dynamic back extension tests deteriorated in the highest third, remained unchanged in the middle third and improved in the lowest third. In the stair climbing test, the results remained unchanged in the highest and middle thirds and deteriorated in the lowest third. Despite the changes of the test results, the performance of the subjects in the initially highest third remained on an average better in 2002 than the performance of the subjects in the middle and lowest thirds.

Discussion

We investigated 6-year changes (between 1996 and 2002) in several components of measured HRF among high-functioning older Finnish men and women. The aim of the study was twofold: (1) to describe the selection of the study sample with reference to the assessment of HRF and (2) to identify gender and age categories with

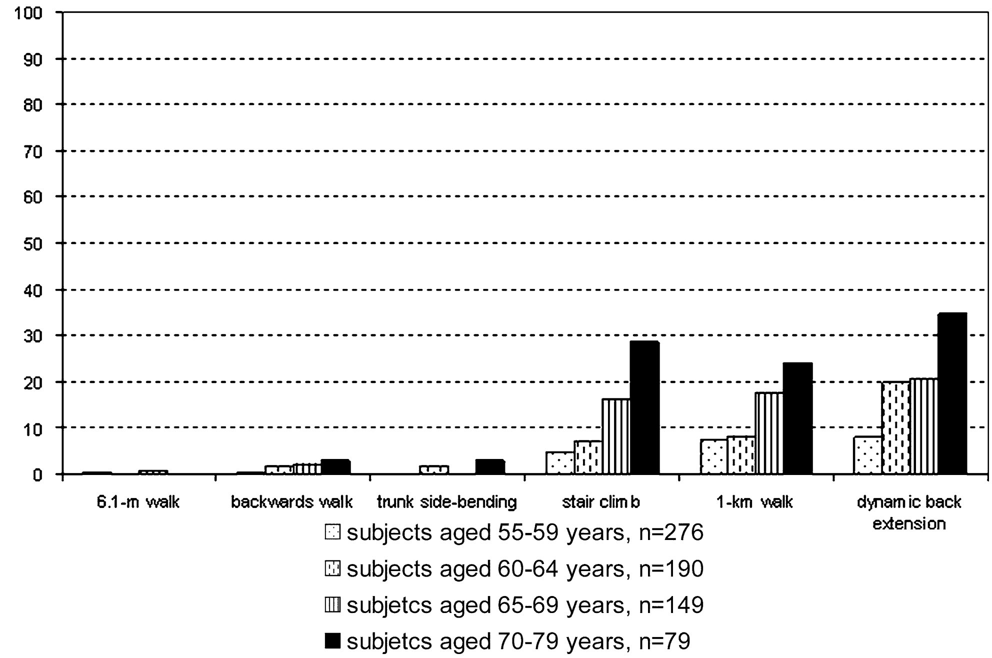

Fig. 2 Percentange of subjects with test-specific exclusions in 2002

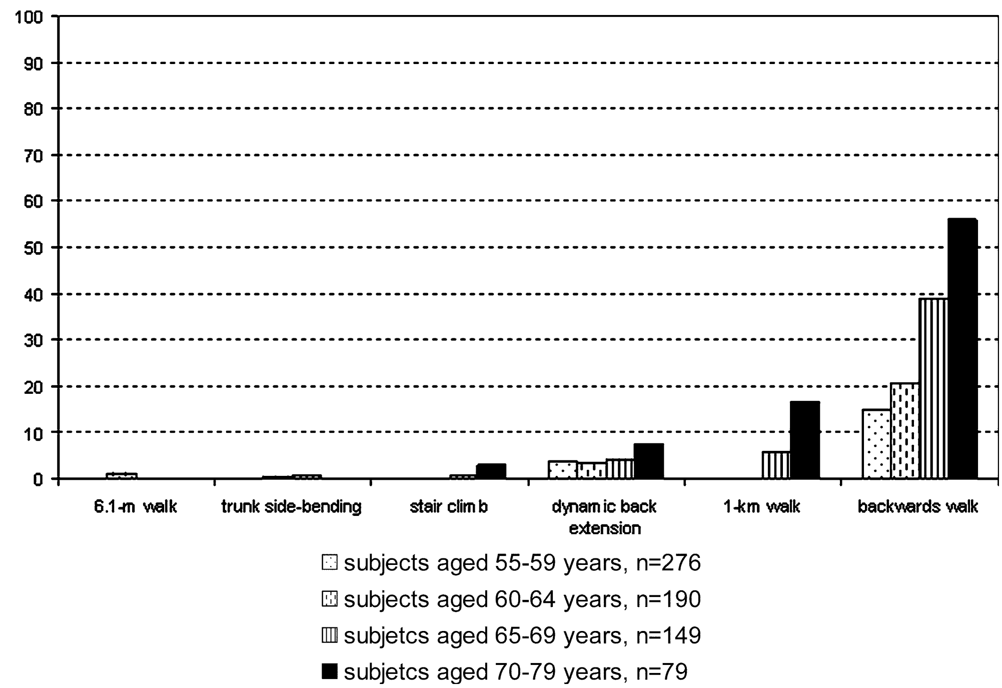

Fig. 3 Percentages of subjects who were unable to perform the tests according to test instructions (=interruptions) in 2002

diminishing HRF among the subjects who could be retested in 2002.

Selection of the study sample with reference to assessments of health-related fitness

The sample of the present study consisted of the high-functioning older men and women without severe self-reported mobility difficulties in 1996. The selection of the sample between 1980 and 1996 has been described by Malmberg et al. (2002a, b). The HRF test results in 1996 served as a baseline to assess the selection for participation in the retesting in 2002. Poorer performance in 1996 was associated with lower participation in retesting. In addition, subjects with poorer baseline performance were more likely to interrupt or be excluded from specific tests in 2002.

The subjects who undertook the HRF testing in both years had on an average better self-rated health status and were physically more active than the subjects who were tested only in 1996. However, the present study still

Table 3 Changes (%) in the mean test results according to age group and gender from the analysis of covariance

	Change % (95% CI) 6.1-m walk (s)	Stair climbing (s)	Backwards walk (s)	Trunk side-bending (cm)	Dynamic back extension (rpt./30 s)	1-km walk (min)
Age group						
55–59 years	−3.4 (−5.3 to −1.4)	9.0 (6.9–11.0)	8.5 (4.7–12.2)	−0.8 (−3.2 to 1.6)	3.9 (0.4–7.4)	0.7 (−0.5 to 1.9)
60–64 years	−7.6 (−10.0 to −5.3)	2.7 (0.1–5.2)	1.5 (−3.5 to 6.2)	−4.4 (−7.0 to −1.6)	−2.8 (−6.8 to 1.3)	−1.8 (−3.3 to −0.4)
65–69 years	−14.0 (−16.8 to −11.2)	−6.4 (−9.8 to −3.2)	−10.3 (−17.8 to −3.2)	−9.5 (−12.3 to −6.7)	−7.9 (−12.2 to −3.4)	−6.4 (−8.3 to −4.6)
70–79 years	−17.4 (−21.8 to −13.2)	−10.1 (−15.9 to −4.6)	−19.3 (−36.3 to −4.4)	−15.9 (−19.7 to −11.9)	−6.8 (−14.2 to 1.4)	−7.4 (−10.6 to −4.4)
P value*	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001‡
P value for linearity**	< 0.001	< 0.001	< 0.001	< 0.001	0.008	< 0.001
n						
55–59 years	233	220	184	230	198	206
60–64 years	168	157	125	164	127	149
65–69 years	135	112	70	135	95	99
70–79 years	65	45	17	63	32	38
Gender						
Men	−7.2 (−9.1 to −5.2)	2.1 (−0.1 to 4.3)	4.5 (−0.3 to 9.1)	−8.7 (−10.8 to −6.5)	−0.5 (−4.1 to 3.2)	2.4 (−3.7 to −1.1)
Women	−13.8 (−15.8 to −11.9)	−4.1 (−6.5 to −1.8)	−14.0 (−20.2 to −8.1)	−7.0 (−8.9 to −5.0)	−6.4 (−9.7 to −3.0)	−5.0 (−6.3 to −3.7)
P value†	< 0.001	< 0.001	< 0.001	0.223	0.010	0.002‡
n						
Men	269	255	194	264	207	221
Women	332	279	202	328	245	271

Positive percentages indicate improvement and negative percentages deterioration of the results
* *P* value for the differences of changes between cohorts
** *P* value for linear trend in changes according to birth cohort
† *P* value for differences of changes between men and women
‡ Statistically significant interaction between age group and gender ($P = 0.041$), see Fig. 4

included subjects with chronic diseases and physically inactive lifestyles. Corresponding selection bias has been reported in other studies (Aniansson et al. 1983; Rantanen et al. 1997; Rantanen and Heikkinen 1998; Hughes et al. 2001; Paterson et al. 2004).

The selection of the study sample during 6 years follow-up period causes underestimation of the true deterioration in fitness. In the present study, the deterioration in HRF was the greatest in those test items where the exclusion rates were lowest, i.e. in 6.1-m walk, backwards walk and trunk side-bending (see Fig. 2). These are the physically least strenuous test items of the battery. The bias is bigger in the physically more strenuous tests that had the highest exclusion rates, i.e. the dynamic back extension, 1-km walk and stair climbing (see Fig. 2). Since the exclusion rates in general increase with age, the feasibility of these strenuous tests may be limited among the older age groups.

Changes in health-related fitness

There was a linear trend in the changes in HRF according to age group: the older groups deteriorated to a greater extent than the younger groups. The overall deterioration in test performance among the older age groups in the present study is in line with other studies (Aniansson et al. 1983; Rantanen et al. 1997), although the study designs and testing procedures were not exactly the same. In some tests (backwards walk, stair climbing and dynamic back extension) the younger subjects even improved their performance. Improvements in these age groups might be due to increased level of physical activity after recent retirement. This is supported by the findings of a large national survey in Finland, which reported that physical activity was most common in the youngest age groups of those subjects who had reached retirement age (Health 2000). The other explanation could be the regression to mean hypothesis as suggested by Hebert et al. (1999), who studied functional decline and improvement among older Canadians. Accordingly, in most of the tests analysed in the present study the greater deterioration of the results in the initially highest third may be an indication of a regression to mean effect: the best performers had greater reserve to deteriorate. However, it is important to notice that despite the deterioration, the subjects in the initially highest third performed on an average better than the subjects in the middle and lowest initial thirds after the follow-up.

Fig. 4 Changes (%) in the health-related fitness test results between the years 1996 and 2002. Age groups indicate the subjects' ages in 1996. Letter *M* refers to males and letter *F* to females

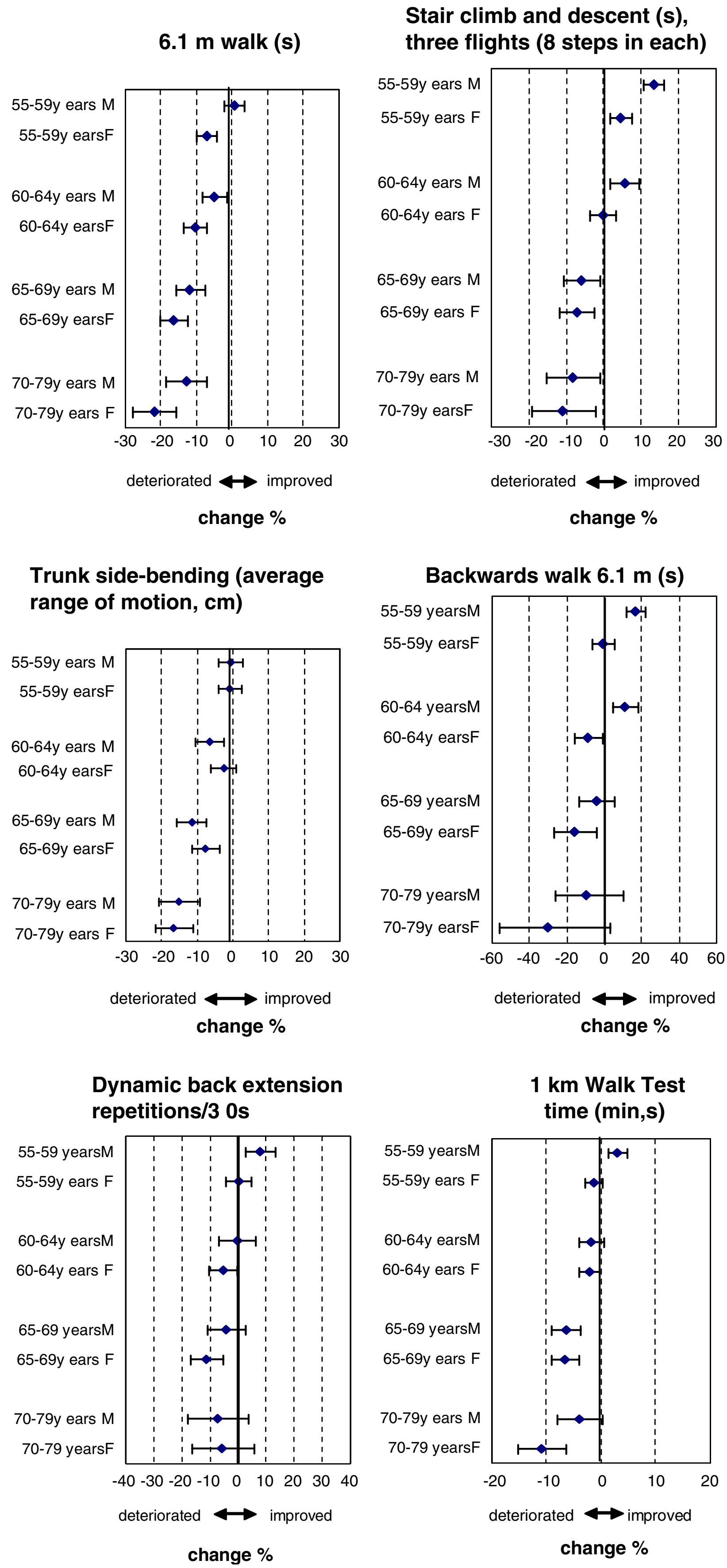

Table 4 Age- and gender-adjusted changes in mean test results (result in 2002 minus result in 1996) according to test-specific thirds from the analysis of covariance

	6.1-m walk (s)	Stair climbing (s)	Backwards walk (s)	Trunk side bending (cm)	Dynamic back extension (rpt./30 s)	1-km walk (min)
Test-specific thirds[a]						
Highest third	0.55(0.47–0.64)	0.54(−0.17 to 1.26)	4.08(2.65–5.52)	−1.70(−2.02 to −1.38)	−3.02(−3.51 to −2.53)	0.48(0.36–0.61)
Middle third	0.38(0.30–0.46)	−0.47(−1.23 to 0.29)	0.02(−1.57 to 1.60)	−0.79(−1.12 to −0.47)	0.42(−0.18 to 1.02)	0.15(0.01–0.29)
Lowest third	−0.05(−0.14 to 0.04)	−2.10(−3.08 to −1.11)	−6.59(−8.26 to −4.91)	0.38(0.02–0.74)	3.56(2.79–4.33)	−0.16(−0.35 to 0.02)
P value[b]	< 0.001	< 0.001	< 0.001	< 0.001	< 0001	
n						
Highest third	206	223	155	213	218	218
Middle third	219	195	127	210	146	176
Lowest third	176	116	114	169	88	98

In the 6.1-m walk, stair climbing, backwards walk and 1-km walk tests positive change indicates deterioration of the result

[a]Highest third includes the best performing third of the subjects in 1996, middle third includes the intermediate performers and the lowest third includes the subjects who belonged to the poorest performing third

[b]*P* value for the differences of changes between the thirds

With reference to gender, the women had lower level of HRF in all tests than the men in 1996. During the follow-up period, the average test performances of the women deteriorated to a greater extent than the average performances of the men, especially in the backwards walk, dynamic back extension and stair climbing tests. This may provide one mechanism to explain previous findings that disability is more common among women than among men (Leveille et al. 2000).

Performance in 6.1-m walk, stair climbing, backwards walk and trunk side-bending showed marked deterioration in the older age groups during the follow-up period. In the present study the stair climbing test is an indicator of lower extremity strength, the backwards walk test indicates dynamic balance and 6.1-m walk is an indicator of walking performance. Several previous studies have shown that the tests of balance, lower extremity strength and walking over short distances are valuable in predicting mobility limitations and disabilities in activities of daily living (Guralnik et al. 1995, 2000). Walking speed has been suggested to be the best single measure of these for predicting the onset of mobility limitations and disabilities (Guralnik et al. 1995, 2000). The findings of the present study indicate marked deteriorations in all of these mobility-related performances.

Previous studies among older populations have paid little attention to flexibility. In the present study, the mean flexibility in trunk side-bending deteriorated markedly in all age groups. Phillips and Haskell (1995) and Holland et al. (2002) also reported age-related declines in flexibility at the trunk and at several joints (shoulder, hip, knee, angle) with functional importance. According to former cross-sectional findings among the present study sample (Malmberg et al. 2002a), poor flexibility in trunk side-bending is associated with poor self-rated mobility function. Loss of height (Suominen 1997) due to the degenerative changes and the loss of bone in the spine can cause changes in posture as well as deterioration of spinal flexibility (Holland et al. 2002). Both of these can affect postural control. Thus, a decrease in trunk side-bending may also reflect decreased postural stability and unwillingness to lean over.

Conclusions

During the 6-year-follow-up, the study sample was selected towards better performing, healthier, physically more active and younger subjects. Among the subjects who could be retested the deteriorations in HRF were on an average greatest in the 6.1-m walk, stair climbing, backwards walk and trunk side-bending tests. It is not possible to indicate which of these factors of HRF starts to deteriorate first. However, it can be suggested that older women, subjects who have poor baseline HRF and subjects who are approaching the age of 70 years, are the most vulnerable group. These groups should be the priority groups in targeting the preventive actions to maintain the mobility function of the older adults. It is worth noticing that the selection in follow-up studies among older populations is marked and has effects on the performance-level differences between age groups.

Acknowledgements Pekka Oja PhD and Ilkka Vuori MD are acknowledged for their valuable work in the earlier stages of the Kainuu Study on Living Habits and Health. This work was supported by The Finnish Ministry of Education, Juho Vainio Foundation, Doctoral Programs of Public Health.

Appendix

Description of the health-related fitness tests and test-specific exclusion criteria.

Test: 6.1-m (20-ft) walk for assessing the ability to walk (Fiatarone et al. 1990; Bassey et al. 1992).
Method: subject walks the course two times 1) at his/her ‘‘usual’’ pace and 2) as fast as possible, starting from a standstill.
Outcome: performance time (sec) of the second trial (as fast as possible) measured by a stopwatch.
Exclusion criteria: severe dizziness and severe symptoms of the spine, hip and knee that may be aggravated by the test.
Test: stair climb and descent for assessing the ability to climbing stairs (Salen et al. 1994).
Method: subject walks up and down a standard flight of stairs (8 steps, a′ 17 cm)[a] first once, and if successful, then three times, using the handrail for support only if needed.
Outcome: performance time (s) measured by a stopwatch from the initial standing position to the end of the third descent.
Exclusion criteria: severe dizziness, severe diseases or symptoms of the spine, hip and knee that may be aggravated by the test.
Test: backwards walk to measure postural control in movement (Nelson et al. 1994)
Method: subject walks backwards along a marked 6.1-m (20-ft) line with tandem steps as quickly as possible. After a 2-m practice trial the subject performs three test trials.
Outcome: walking times (s) of the three test trials as measured by a stopwatch from the starting position to the end of the line. The best time is the final result.
Exclusion criteria: severe dizziness, severe symptoms of the spine or lower extremities that may be aggravated by the test.
Test: trunk side-bending to the right and left for measuring the average range of motion in lateral flexion of the thoracic and lumbar spine and pelvis (Suni et al. 1996).
Method: subject stands on marked lines (15 cm apart) with the back against the wall and arms and fingers straight at the sides of the body (baseline). Subject slides the middle finger along the lateral thigh to the right and then to the left as far as possible, keeping shoulders and buttocks in contact with the wall and heels in contact with the floor. The tester measures the distance between baseline and maximum slide of the middle finger tip.
Outcome: the average distance (cm) between the maximal right and left side-bending range of motion measured by a cloth tape measure.
Exclusion criteria: severe dizziness and severe spinal symptoms that may be aggravated by the test movement.
Test: dynamic back extension for assessing trunk extensor muscle endurance (Mälkiä 1983).
Method: subject lies in a semi-inclined body position (50°) in a portable standing hyper extensor[b] with hips and lower legs supported and fingers crossed behind the neck. Subject raises the upper body off the table to a straight back level and returns to the starting position as quickly as he/she can.
Outcome: maximum number of repetitions in 30 s.
Exclusion criteria: moderate to severe diseases or symptoms of the cardiovascular system and severe spinal, hip and knee symptoms that may be aggravated by the test movement.
Test: 1-km walk for assessing submaximal aerobic capacity (Oja et al. 1991).
Method: subject walks as fast as possible on a flat surface using his/her normal walking style.
Outcome: walking time (min) measured by a stopwatch.
Exclusion criteria: severe diseases or symptoms of the cardiovascular system, severe dizziness and severe symptoms of the spine, hip and knee that may be aggravated by the test.

Suppliers: [a]RT-88-10470, Construction standard 1992, Finland, [b]Standing Hyper Extensor, HUR Ltd., Patamäentie 4, 67100 Kokkola, Finland

References

Aniansson A, Sperling L, Rundgren Å, Lehnberg E (1983) Muscle function in 75-year-old men and women, a longitudinal study. Scand J Rehab Med (Suppl 9):92–102
Bassey E, Fiatarone M, O'Neill E, Kelly M, Evans W, Lipsitz L (1992) Leg extensor power and functional performance in very old men and women. Clin Sci 82:321–327
Bouchard C, Shephard R (1994) Physical activity, fitness, and health: the model and key concepts. In: Bouchard C, Shephard R, Stephens T (eds) Physical activity, fitness and health. Champaign, IL: Human Kinetics pp 7–88
Dunlop D, Hughes S, Manheim L (1997) Disability in activities of daily living: patterns of change and a hierarchy of disability. Am J Public Health 87:378–383
Ferrucci L, Guralnik J, Simonsick E, Salive M, Corti C, Langlois J (1996) Progressive versus catastrophic disability: a longitudinal view of disablement process. J Gerontol A Biol Sci Med Sci 51A(3):M123-M130
Fiatarone M, Marks E, Ryan N, Meredith C, Lipsitz L, Evans W (1990) High-intensity strength training in nonagenarians. Effects on skeletal muscle. JAMA 263(12):3029–3034

Gill T, Williams C, Mendes de Leon C, Tinetti M (1997) The role of change in physical performance in determining risk for dependence in activities of daily living among non-disabled community-living elderly persons. J Clin Epidemiol 50(7):765–772

Guralnik J, Simonsick E, Ferrucci L, Glynn R, Berkman L, Blazer D, Scherr P, Wallace R (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

Guralnik J, Ferrucci L, Simonsick E, Salive M, Wallace R (1995) Lower-extremity function in persons over the age of 70 years as predictor of subsequent disability. N Engl J Med 332(2):556–561

Guralnik J, Ferrucci L, Pieper C, Leveille S, Markides K, Ostir G, Studenski S, Berkman L, Wallace R (2000) Lower extremity function and subsequent disability: consistency across studies, predictive models and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 55A(4):M221–M231

Health 2000: Health and functional capacity in Finland. Baseline results of the Health 2000 health examination survey. Publications of National Public Health Institute, KTL B12/2004. Finland

Hebert R, Brayne C, Spiegelhalter D (1999) Factors associated with functional decline and improvement in a very elderly community-dwelling population. Am J Epidemiol 150(5):501–510

Holland, Tanaka K, Shigematsu R, Nakagaichi M (2002) Flexibility and physical functions of older adults: a review. J Ageing Phys Act 10(2):169–206

Hughes V, Frontera W, Wood M, Evans W, Dallal G, Roubenoff R, Fiatarone Singh M (2001) Longitudinal muscle strength changes in older adults: influence of muscle mass, physical activity and health. J Gerontol A Biol Sci Med Sci 56A(5):B209–B217

Leveille S, Penninx B, Melzer D, Izmirlian G, Guralnik J (2000) Sex differences in the prevalence of mobility disability in old age: the dynamics of incidence, recovery and mortality. J Gerontol Soc Sci 55B(1):S41–S50

Malmberg J, Miilunpalo S, Vuori I, Pasanen M, Oja P, Haapanen-Niemi N (2002a) A health-related fitness and functional performance test battery for middle-aged and older adults: feasibility and health-related content validity. Arch Phys Med Rehabil 83:666–677

Malmberg J, Miilunpalo S, Vuori I, Pasanen M, Oja P, Haapanen-Niemi N (2002b): Improved functional status in 16 years of follow-up of middle aged and elderly men and women in north eastern Finland. J Epidemiol Community Health 56: 905–912

Mälkiä E (1983) Muscular performance as a determinant of physical ability in Finnish adult population. Helsinki (Finland): Publications of the Social Insurance Institution. Report No. AL:23

Nelson M, Fiatarone M, Morganti C, Trice I, Greenberg R, Evans W (1994) Effects of high-intensity strength training on multiple risk factors for osteoporotic fractures. JAMA 272(24):1909–1914

Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I (1991) A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. Int J Sports Med 12(4):356–62

Oja P, Miilunpalo S, Vuori I, Pasanen M, Urponen H (1994) Trends in health-related physical activity in Finland: 10-year follow-up and adult cohort in eastern Finland. Scand J Med Sci Sports 4:75–81

Paterson D, Govindasamy D, Vidmar M, Cunningham D, Koval J (2004) Longitudinal study of determinants of dependence in an elderly population. JAGS 52:1632–1638

Penninx B, Ferrucci L, Leveille S, Rantanen T, Pahor M, Guralnik J (2000) Lower extremity performance in non-disabled older persons as a predictor of subsequent hospitalization. J Gerontol A Biol Sci Med Sci 55(11):M691–M697

Phillips W, Haskell W (1995) "Muscular fitness"—easing the burden of disability for elderly adults. J Ageing Phys Act 3:261–289

Rantanen T, Era P, Heikkinen E (1994) Maximal isometric strength and mobility among 75-year-old men and women. Age Ageing 23:132–137

Rantanen T, Era P, Heikkinen E (1997) Physical activity and changes in maximal isometric strength in men and women from the age of 75 to 80 years. JAGS 45:1439–1445

Rantanen T, Heikkinen E (1998) The role of habitual physical activity in preserving muscle strength from age 80 to 85 years. J Ageing Phys Act 6:121–132

Rinne M, Pasanen M, Miilunpalo S, Oja P (2001) Test–retest reproducibility and inter-rater reliability of a motor skill test battery for adults. Int J Sports Med 22:192–200

Salen B, Spangfort E, Nygren Å, Nordemar R (1994) Disability rating index: an instrument for the assessment of disability in clinical settings. J Clin Epidemiol 47(12):1423–1434

Statistical Yearbook of Finland 2004. Tilastokeskus 2004, vol 99 (new series). Hämeenlinna. Karisto Oy

Suni J, Oja P, Laukkanen R, Miilunpalo S, Pasanen M, Vuori I, Bös K (1996) Health-related fitness test battery for adults: aspects of reliability. Arch Phys Med Rahabil 77:399–405

Suni J, Miilunpalo S, Asikainen T-M, Laukkanen R, Oja P, Pasanen M, Bös K, Vuori I (1998a) Safety and feasibility of a health-related fitness test battery for adults. Phys Ther 78:134–148

Suni J, Oja P, Miilunpalo S, Pasanen M, Vuori I, Bös K (1998b) Health-related fitness test battery for adults: associations with perceived health, mobility, and back function and symptoms. Arch Phys Med Rehabil 9:559–569

Suni J, Oja P, Miilunpalo S, Pasanen M, Vuori I, Bös K (1999) Health-related fitness test battery for middle-aged adults: associations with physical activity patterns. Int J Sports Med 20:183–191

Suni J (2000) Health-related fitness test battery for middle-aged adults—with emphasis on musculoskeletal and motor tests [dissertation]. University of Jyväskylä

Suominen H (1997) Changes in physical characteristics and body composition during 5-year follow-up in 75- and 80-year-old men and women. Scand J Soc Med Suppl (53):19–24

Wolfson L, Judge J, Whipple R, King M (1995) Strength is a major factor in balance, gait and the occurrence of falls. J Gerontol A Biol Sci Med Sci 50A(Special issues):64–67