Jaana Suni

Health-related Fitness Test Battery for Middle-aged Adults

with Emphasis on Musculoskeletal and Motor Tests

STUDIES IN SPORT, PHYSICAL EDUCATION AND HEALTH 66

Jaana Suni

Health-related Fitness Test Battery for Middle-aged Adults

With Emphasis on Musculoskeletal and Motor Tests

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2000

Editors
Harri Suominen
Department of Health Sciences, University of Jyväskylä
Kaarina Nieminen
Publishing Unit, University Library of Jyväskylä

Cover picture:
Niko Airaksinen

URN:ISBN:978-951-39-7926-3
ISBN 978-951-39-7926-3 (PDF)
ISSN 0356-1070

ISBN 951-39-0612-4
ISSN 0356-1070

Jyväskylä University Printing House, Jyväskylä
and ER-Paino Ky, Lievestuore 1999

To my mother Pirkko
and
father Keijo, in memoriam

ABSTRACT

Suni, Jaana
Health-related fitness test battery for middle-aged adults with emphasis on musculoskeletal and motor tests
Jyväskylä: University of Jyväskylä, 2000, 96 p.
(Studies in Sport, Physical Education and Health,
ISSN 0356-1070; 66)
ISBN 951-39-0612-4
Diss.

The purpose of this study was to develop a reliable, safe, feasible and valid field-based musculoskeletal and motor health-related fitness (HRF) test battery for middle-aged adults. The subjects were a representative sample of men (n=246) and women (n=254) between the ages of 37 and 57 years, 83% of whom also participated in 3-year follow-up study. Leisure-time physical activity (LTPA) and self-rated health were assessed with questionnaires. HRF was measured with 9 standard field tests consisting of musculoskeletal (one-leg squat, vertical jump, static back extension, modified push-up, hamstring muscle extensibility, trunk side-bending), motor (one-leg balance), cardiorespiratory (Walk Test) and morphological (body mass index) measures of fitness. The inter-rater and test-retest reliability of the tests was evaluated with a small sample (n=42) of volunteers. Safety and feasibility were assessed in terms of acute complications, heart rate after each test, post-test muscular soreness, subject exclusion rate and testing time requirements. Content and predictive validity was evaluated by studying the cross-sectional and follow-up associations between HRF and self-rated health. In addition, the cross-sectional relations between LTPA and HRF were assessed. Four of the 7 musculoskeletal and motor tests possessed acceptable reliability. With the aid of the standard health screening built into the testing procedure the HRF assessment was safely and effectively conducted with minor physician participation. Seven of the 9 proposed tests showed prudent associations with current or future self-rated perceived health, mobility in stair climbing, back functioning and back pain. Three of the tests also showed physical-activity-related validity for both sexes, and 2 had a corresponding result for the women only. The results suggest that the developed test battery is a promising field-based method for the reliable, safe, feasible and valid assessment of HRF among adult populations. With respect to validity, the development is an ongoing process.

Key Words: health-related fitness, field assessment, musculoskeletal, motor, musculoskeletal health, physical function, health promotion, adult populations.

Author's address	Jaana Suni Urho Kaleva Kekkonen Institute for Health Promotion Research, Tampere, Finland
Supervisors	Docent Pekka Oja Urho Kaleva Kekkonen Institute for Health Promotion Research, Tampere, Finland
	Professor Markku Alen Department of Health Sciences University of Jyväskylä, Jyväskylä, Finland
	Professor Tapio Videman University of Alberta, Edmonton, Canada
Reviewers	Docent Hannu Alaranta University of Helsinki, Helsinki, Finland
	Professor Clas-Håkan Nygård University of Tampere, Tampere, Finland
Opponent	Professor Willem van Mechelen Free University, Amsterdam, the Netherlands

ACKNOWLEDGEMENTS

This study was carried out at the UKK Institute for Health Promotion Research in Tampere, Finland in collaboration with the Sports Research Insitute of University of Frankfurt, Frankfurt, Germany.

First I want to thank my supervisor docent Pekka Oja, Ph.D., Scientific Director of the UKK Institute, who in 1991 recruited me to work with the "International Study on Physical Activity, Fitness and Health". I am grateful to him for encouraging and conducting me to do this thesis as part of that study, otherwise it would not exist. Along the way I did experience some desperate moments when I doupted my ability to finish this work. As my supervisor, Pekka often said that it is natural to have difficulties with new study concepts. That gave me the courage to continue.

I wish to express my warmest gratitude to the whole FINGER-study group at the UKK Institute and at the Franfurt University: to co-authors Tuula-Maria Asikainen, M.D., professor Klaus Bös, docent Raija Laukkanen, Ph.D., docent Seppo Miilunpalo, M.D., and professor Ilkka Vuori, M.D., Director of the UKK Institute; to Ms. Katriina Ojala, M.Sc. Jarmo Malberg, and M.Sc. Sanna Roine for their invaluable help in conducting the fitness measurements; to the skillful ladies at our laboratory for taking the blood samples and other health measurements; to Dr. Alexander Woll for sharing the experience of being the "responsible junior researcher" and also for being my best German friend; to M.Sc. Susanne Tittlbach for bringing new energy to our study group. Especially I appreciated the expertise and guidance of our statistician Matti Pasanen, M.Sc. Your uncompromising criticism on statistical methods gave me confidence in writing the manuscripts.

I am extremely grateful to my official referees docent Hannu Alaranta M.D. and professor Clas-Håkan Nygård, Ph.D. for their devoted work in giving me constructive criticism and expert advise. The process was a true learning experience which helped me to improve the thesis.

I also want to thank my supervisors from the University of Jyväskylä. Professor Tapio Videman gave me expert advice on selecting the outcome measures for back functioning and pain at the early stage of the study. Professor Markku Alen supported me in the final stage of my doctoral studies and the preparation of my thesis.

I owe special thanks to Ms. Riitta Simukka for her skilful technical assistance in giving the finishing layout for the figures of this thesis and to Mr. Ismo Lapinleimu for patiently helping with all kinds of "computer-related" problems.

Warm thank also belongs to Georgianna Oja, E.L.S. for editing the English language in this thesis.

My thanks to all the subjects who participated in the studies. They made this thesis possible.

I am greatly indebted to the 300 testers who have participated in the "Health-related fitness assessment" courses at the UKK Institute. Their enthusiasm and professional skills have been enormous encouragement for my scientific work. Special thanks also to Ms. Päivi Moisio, M.Sc. for the strong support in developing these courses.

To the whole staff at the UKK Institute I want say that it is a privilege to work with you. It is wonderful to have friends at your workplace. "Liberian lady", who could do without you. Eija, you belong to the group of "invaluables".

Bill Tuxworth, you are a real English gentleman. I thank you for your kind support to my scientific work and to our international fitness workshops.

"The not so young squash girls" Helena, Sanna, and Tiina. Thank you for sharing your positive attitudes and unique friendship during all these years. Hurrah for our high density bones!

Outi, your are a dear friend. We have shared a lot during our lives. Thank you for taking interest in my work, too.

My dearest brother Juha, and your wife Anna. Thank you so much for sharing many precious moments of your life with me. Lotta and Ella, I enjoy your company and I am so proud to be your aunt.

I want to dedicate this thesis with my deepest gratitude to my mother and late father, who have raised me to respect knowledge and education. Most importantly you gave me a loving home and happy childhood. That foundation has carried me through the troubles of life in recent years. To my deep sorrow my father, due to lung cancer, did not have enough time to see me reach this goal. I miss you much.

Finally, Jari dear, we have travelled in a roller-coaster the last two years. I love you and I hope with all my heart that we can stay together.

Tampere, December 1999

ABBREVIATIONS

ACSM	American College of Sports Medicine
AHA	American Heart Association
BMI	body mass index
CI	confidence interval
DOMS	delayed onset of muscle soreness
HRF	health-related fitness
ICC	intraclass correlation coefficient
LBT	low-back trouble
LTPA	leisure-time physical activity
MVC	maximal voluntary contraction
OA	osteoarthritis
OR	odds ratios
ROM	range of motion
SEM	standard error of measurement
UKK Institute	Urho Kaleva Kekkonen Institute for Health Promotion Research
VO_{2max}	maximal aerobic power; maximal oxygen uptake
$\%HR_{max}$	percentage of age-predicted maximal heart rate

LIST OF ORIGINAL ARTICLES

The thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I 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. Archives of Physical Medicine and Rehabilitation 77, 399-405.

II Suni, J., Miilunpalo, S., Asikainen, T-M., Laukkanen, R., Oja, P., Pasanen, M., Bös, K. & Vuori, I. 1998. Safety and feasibility of a health-related fitness test battery for adults. Physical Therapy 78, 134-148.

III Suni, J., Oja, P., Miilunpalo, S., Pasanen, M., Vuori, I.& Bös, K. 1998. Health-related fitness test battery for adults: associations with perceived health, mobility, and back function and symptoms. Archives of Physical Medicine and Rehabilitation 79, 559-569.

IV Suni J., Oja, P., Miilunpalo, S., Pasanen, M., Vuori, I. & Bös, K. 1999. Predictive value of fitness for 3-year changes in musculoskeletal functioning. Manuscript.

V 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

In addition, some unpublished data are presented.

CONTENTS

1 INTRODUCTION

Scientific knowledge on the significance of physical activity and physical fitness with respect to health and physical functional status has increased and become more specified during the past decade [Bouchard et al. 1990, Bouchard & Shephard 1994, U.S. Department of Health and Human Services 1996, ACSM (American College of Sports Medicine) 1998]. Accordingly, increasing the physical activity level of the general population has become an important issue in today's health promotion (U.S. Department of Health and Human Services 1996). Simultaneously the concept of health-related fitness (HRF) has been introduced (Bouchard & Shephard 1994). HRF refers to the components of fitness that are affected by habitual physical activity and relate to health status.

Effective methods to increase physical activity are needed. It has been proposed that assessing and monitoring relevant aspects of fitness may have an important role in this type of health promotion (Jette et al. 1992, King & Senn 1996). When HRF assessment is used as a method to increase physical activity, it is aimed at serving (a) as a means to assess the need for physical activity with reference to health and physical function, (b) as a safe basis for individual physical activity counselling and exercise prescription, (c) as a means with which to follow individual changes in HRF, (d) as a tool to educate and motivate individual persons and groups with respect to regular physical activity (Jette et al. 1992, ACSM 1995, Oja & Tuxworth 1995, King & Senn 1996). The *emphasis* of HRF testing is more on enhancing physical functional capacity for everyday life than on avoiding specific diseases (Breslow 1999).

With regard to middle-aged population physical activity and fitness levels need to be increased to promote their health and musculoskeletal functioning, to maintain their work ability and to prevent premature functional disability in later life (Pate et al. 1995). Low-back pain is the major source of work disability and activity limitations among them (Heliövaara et al. 1989, Mäkelä et al. 1993, Hagen & Thune 1998, Aromaa et al. 1999).

Among the rapidly increasing elderly populations osteoarthritis (OA) of the knees and hips are major sources of mobility limitation (Davis et al. 1991, Mäkelä et al. 1993, Aromaa et al. 1999), which can substantially limit the ability to remain independent (Davis et al. 1991, Launer et al. 1994, Schroll 1994, Guralnik et al. 1995). Falls leading to bone fractures are the most serious health consequences of mobility-related limitations (Province et al. 1995). All of the mentioned problems cause substantial human suffering and also economic and social costs to society.

Increasingly, the physical functional status of adults has been characterized through the use of objective physical performance (or fitness or capacity) tests (Guralnik et al. 1994 and 1995, Rejeski et al. 1995, Huang et al. 1998, Simmonds et al. 1998). The objective measurements have shown that decline in musculoskeletal fitness can be marked after the age of 50 years (Nygård et al. 1991). However, physical fitness has not yet been extensively examined with respect to physical disability, especially in middle-aged populations (Huang et al. 1998). Physical fitness testing offers a means for exploring the relationship between lifestyle and physical functional status among middle-aged and elderly men and women (Huang et al. 1998, Morey et al. 1998b).

Thus population-based HRF assessment methods are needed for both promotional purposes and epidemiological and intervention studies. Before testing can be applied to large populations, the methods need to be safe, economic and easy to administer under conditions available in ordinary communities (Skinner & Oja 1994, King & Senn 1996). The assessment of HRF is indicated primarily for unfit and physically inactive adults. Therefore, the maximal nature of many existing flexibility, strength and aerobic fitness tests may limit the safe administration of the tests (Skinner & Oja 1994) and may also have adverse effects on exercise motivation (Dishman & Sallis 1994). The methods have to be repeatable if reliable information is to be obtained about the fitness level of individual persons or populations. Most importantly, the fitness tests must show meaningful relationships with health (Reuben et al. 1992, Phillips & Haskell 1994). Such relationships are necessary if the contribution of particular components of fitness to important health outcomes are to be determined and the test scores are to be interpreted in terms of the adequacy of fitness with respect to health.

Several validated functional performance and HRF test batteries for the elderly have already been developed (Bravo et al. 1994, Guralnik et al. 1994 and 1995, Reuben et al. 1994). Recently, an expert group of the Council of Europe proposed the test battery *Eurofit for Adults,* which is aimed at assessing the HRF of middle-aged and older adults (Oja & Tuxworth 1995). However, the repeatability, safety and validity of the Eurofit test battery has not been systematically studied.

The Urho Kaleva Kekkonen Institute for Health Promotion Research (UKK Institute) has proposed a health-related fitness test battery for apparently healthy middle-aged adults with special emphasis on musculoskeletal and motor tests. This study examines the reliability, safety, feasibility, and health-related validity of the test battery with reference to self-rated perceived health, mobility function, back functioning and back pain.

2 REVIEW OF THE LITERATURE

2.1 Concept of health-related fitness

2.1.1 Concept and definitions of health-related fitness, physical activity and health

The term "physical fitness" has been defined in many ways. The terms "physical performance" and "physical capacity" are considered to be synonyms with physical fitness in the present review. Most definitions of physical fitness refer strictly to the capacity of movement: "a set of attributes that people have or achieve" (Caspersen et al. 1985, p.128). In general terms, fitness can be conceived as the matching of the physical capacity of the individual person to his or her physical and social environment (Bouchard & Shephard 1994). Physical fitness has almost always been viewed as a multifactorial construct that includes several components (Marsh 1993, ACSM 1995). It can be understood in terms of components that should be taken into consideration for its assessment according to the context in which the concept of fitness is operationalized (Bouchard & Shephard 1994).

The concept of HRF was first introduced by what is known as the Toronto model on physical activity, fitness, and health (Bouchard et al. 1990, Bouchard & Shephard 1994). The health effects of physical activity can be examined and understood through this model, which is presented in Figure 1. According to the model habitual physical activity can influence physical fitness, which can, in turn, modify the level of physical activity. The model also specifies that fitness is related to health reciprocally manner. It not only influences health, but health status also influences both habitual physical activity and fitness level. HRF includes components of fitness that are related to health and can be affected by regular physical activity. More specifically (Bouchard & Shephard 1994, p. 81):

> health-related fitness refers to those components of fitness that are affected favourably or unfavourably by habitual physical activity and relate to health-status. It has been defined as a state characterised by (a) an ability to perform daily activities with vigour and (b) demonstration of traits and capacities that are associated with a low risk of premature development of hypokinetic diseases and conditions.

Other types of life-style behavior, physical and social environmental conditions, personal attributes, and genetic characteristics also affect the major components of the basic model and determine their relationships (Bouchard & Shephard 1994). Despite of the genetic component (Bouchard & Preusse 1994, Thomis et al. 1998) and other factors, physical fitness is, to some extent, a physiological marker of the behavior of physical activity. Activity of sufficient frequency, intensity, and duration will lead to increased fitness. However, the ability to improve fitness through activity will depend on each person's genetic endowment (Bouchard & Pérusse 1994, Launderdale et al. 1997, Simonen et al. 1998, Thomis et al. 1998). Leisure-time physical activity (LTPA) was recently been shown to be associated with reduced mortality, even after genetic and other familial factors had been taken into account (Kujala et al. 1998).

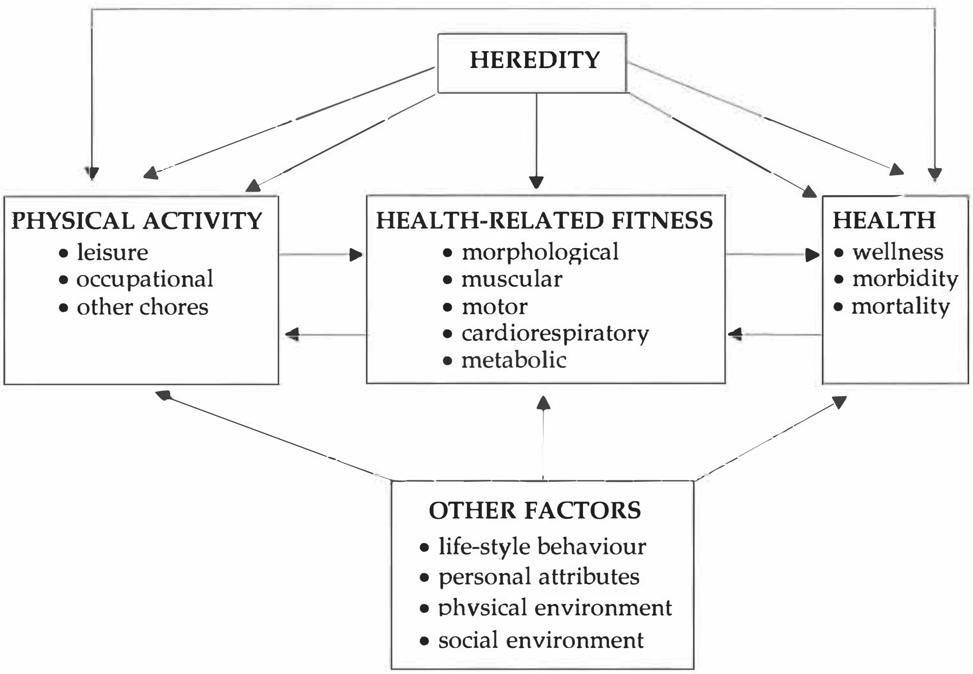


FIGURE 1 Toronto model on physical activity, fitness and health (Bouchard & Shephard 1994, p. 78)

In the Toronto model "*physical activity* comprises any body movement produced by the skeletal muscles that result in a substantial increase over the resting energy expenditure" (Bouchard & Shephard 1994, p. 77). Thus habitual physical activity includes exercise and other everyday physical activities during leisure time, occupational work and associated active transportation (walking, cycling). *LTPA* is an activity undertaken in a person's discretionary time, and it is selected on the basis of personal needs and interests (Bouchard & Shephard 1994). *Exercise,* a subset of LTPA, is planned, structured, repetitive, and purposive in the sense that improvement or maintenance of one or more components of physical fitness is an objective (Caspersen et al. 1985).

Any measurement of health status depends on the *definition of health,* a particularly complex concept which remains a major challenge. The well known, broad definition of the World Health Organization that health is a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity, points out that health encompasses more than not being physically ill (Caspersen et al. 1994). In the Toronto model (Bouchard & Shephard 1994, p. 84) health is defined as:

> a human condition with physical, social, and psychological dimensions each characterized on a continuum with positive and negative poles. Positive health pertains to the capacity to enjoy life and to withstand challenges; it is not merely the absence of disease. Negative health pertains to morbidity and, in the extreme, with premature mortality.

The *traditional surveillance measures of public health* (incidence rates, prevalence rates, mortality rates, risk factor prevalence rates of disease and death, and disability rates) reflect negative aspects of health (Caspersen et al. 1994, Skinner & Oja 1994). A broader spectrum of measures including more positive states of health has been adopted within the HRF concept. An important part of positive health is the concept of *quality of life,* which includes measures of *function* (physical, mental, functional activities) and *well-being* (bodily, emotional, self-concept, global perceptions of health and well-being) (Caspersen et al. 1994). Thus the concept of quality of life has many aspects, being related to health and the ability to perform significant activities but also to the general feeling of well-being. It is probable that good physical function and maintained ability in various daily activities influence the general well-being of a person (Skinner & Oja 1994, Grimby 1995). Disability, function, and well-being provide important areas for research aimed at studying the health effects of physical activity (Caspersen et al. 1994, Morey et al. 1998b).

Regarding physical function, the early *models of disability* (stages from pathology to impairment to functional limitation leading to disability) focused on disease-specific factors as the primary course of disability. With this respect, measures of flexibility are typically assessments of impairment, objective performance tests are assessments of functional limitations, and self-reported functional assessments reflect disability (Guranick et al. 1994, Kivinen et al. 1998, Simmonds et al. 1998). Jette (1994) has described the relationship of quality of life to the disablement concepts. More recent models have shown that many factors other than disease or pathology may lead to disability (Lawrence & Jette 1996), and measures of physical function (or functional limitations) have been used to determine the causes of disability (Reuben et al. 1992, Fried et al. 1994).

The most recent findings (Morey et al. 1998b) suggest that low-fitness (cardiorespiratory fitness, muscular strength, flexibility) is a risk factor for functional decline independent of disease processes (or pathology). Furthermore, a cross-sectional population study (Rantanen et al. 1999a) indicated that fitness (muscular strength) has a mediating role between physical activity and disability: disability correlated with physical inactivity, which correlated with lower muscular strength, which in turn was associated with a greater degree of disability. In addition, objective assessments of physical function and fitness have been successfully used as predictors of subsequent disability among non-disabled elderly (Guralnik et al. 1995) and middle-aged persons (Huang et al. 1998), and also as means for targeting intervention toward those at high risk for falls (Tinetti et al. 1994).

2.1.2 Components and factors of health-related fitness

According to the Toronto model (see Figure 1), the components of HRF are morphological, musculoskeletal, motor, cardiorespiratory and metabolic fitness (Bouchard & Shephard 1994, Skinner & Oja 1994). The specific factors related to these fitness components are presented in Table 1. The emphasis is on fitness components shown by research to be related to health in contrast to performance-related components, which contribute to optimal job or sports performance (Skinner & Oja 1994). However, fitness can be related to both performance and health. For example, adequate strength and balance are needed for independent living or for the retention of function. They also help to avoid a higher risk of falling, which may have adverse effects on health (Skinner & Oja 1994).

TABLE 1 Components and factors of health-related fitness

Components	Factors
Morphogical fitness	Body composition
	Bone strength
Musculoskeletal fitness	Muscular strength and endurance
	Flexibility
Motor fitness	Postural control
Cardiorespiratory fitness	Maximal aerobic power
	Submaximal cardiorespiratory capacity
Metabolic fitness	Carbohydrate metabolism
	Lipid metabolism

Source: Skinner & Oja (1994), p.160

Critical concerns in the study of physical fitness are the definition of the construct and the selection of appropriate indicators (Marsh 1993). Almost all textbooks of fitness emphasize components of fitness based in part on the classic factor analytic studies by Fleishman conducted in 1964, and reviewed by Marsh in 1993. However, according to Marsh (1993) only a few studies using the factor analytic approach have been published in the last 3 decades.

Within the field of HRF, Hagan et al. (1991) conducted a factor analytic study between physical fitness and heart disease risk among fire fighters (n=779, ages 18 to 64 years). The findings provided primary evidence that measures of physical fitness other than aerobic endurance capacity may also guard against coronary heart disease. Shephard and Bouchard (1994) measured a wide selection of potential markers of sedentary living from healthy sedentary adults (n=350) and found 4 HRF factors: obesity, cardiovascular fitness, total cholesterol and high density lipoprotein cholesterol. Rejeski et al. (1995) examined the factor structure of physical activity restrictions in patients (n=203) with knee osteoarthrosis. Four tests were selected into the performance battery (walking for 6 minutes, stair climbing, lifting and carrying, getting into and out of a car) based on the data from the factor analysis.

The concept of HRF is new, and its construct (components, factors) is an "agreement" made by researcher on the basis of current scientific evidence on the associations of fitness with different dimensions of health. In other words, it is based on studies attempting to establish a theoretically consistent, logical pattern of relations between measures of physical fitness and health (Marsh 1993). It is a construct that needs to be continually developed according to new research knowledge.

At present, the role of endurance types of physical activity and the respective fitness components, aerobic and metabolic fitness, in preventing common chronic diseases like coronary heart disease, hypertension or non-insulin dependent diabetes has been scientifically well established (Pate et al. 1995, U.S. Department of Health and Human Services 1996).

The knowledge needed to promote physical activity, exercise and fitness to enhance musculoskeletal health and prevent functional disability is inconclusive and inconsistent. A certain amount of loading is vital to the musculoskeletal system, whereas both overstrain and immobilization can lead to premature degeneration and illness (Troup & Videman 1989, Panush 1994, Kannus et al. 1992). The understanding of the meaning of physical activity in the maintenance of bone strength and the prevention of osteoporosis has rapidly increased in recent years. Regular physical exercise can reduce the risk of osteoporosis and delay the physiological decrease of bone mineral density. It has been shown that exercise training programs prevented or reversed almost 1% of bone loss per year in both lumbar spine and femoral neck (Wolff et al. 1999).

However, the degree to which individual fitness components and factors contribute to back and mobility functioning and disability, the main interest of the present study, has been studied only in limited detail (Morey et al. 1998b). New fitness factors like reaction time (Taimela et al. 1993, Venna et al. 1994, Luoto et al. 1995a) and neuromuscular control of the trunk muscles (Hides et al. 1996, Hodges & Richardson 1996) have only recently shown to be important with respect to low-back pain.

Despite of the aforementioned facts, the experts, in their recommendations concerning physical activity and public health (Pate et al. 1995), have stated that:

> clinical experience and limited studies suggest that people who maintain or improve their strength and flexibility may be better able to perform daily activities, may be less likely to develop back pain, and may be better able to avoid disability, especially as they advance into older age. Regular physical activity also may contribute to better balance, coordination, and agility, which in turn may help prevent falls in the elderly.

Sections 2.5.2 and 2.5.3 review the literature concerning the cross-sectional and prospective associations of physical activity and fitness with low-back and mobility functioning. In regard to physical activity the emphasis is on LTPA.

2.2 Methodological requirements for developing a field-based health-related fitness test battery for adults

Before large acceptance, all items of any fitness test battery should be subjected to evaluations of their inherent characteristics of reliability, safety, feasibility, validity, and sensitivity to change (Reuben et al. 1992). The last characteristic is beyond the limits of the present study. Distinction is typically made between large sample epidemiology – like studies of physical fitness that rely on easily administered field tests and small sample laboratory studies, which emphasize technically sophisticated measures that require expensive equipment (March 1993). The sophisticated measures are not necessarily more reliable or more valid indicators of physical fitness. Rather, both the field tests and laboratory measures of fitness are merely indicators, whose reliability and validity should be systematically evaluated on the basis of their effective use for selected purpose (March 1993, Atkinson & Nevill 1998).

Important aspects and current problems related to the assessment of the reliability of fitness test methods are presented in section 2.3. Contrary to typical laboratory studies, factors related to safety and feasibility are a major concern in field-based HRF testing of adult populations. These issues are discussed in section 2.4. In regard to validity, several methods are commonly employed. The methods relevant within the concept of HRF are presented in section 2.5.1.

Characteristics of several performance test batteries designed for the elderly (typically including tests of walking, chair climbing, stair climbing, balance) have been evaluated in the systematic manner for example: (a) the reliability of all test items and criterion-related validity of the cardiorespiratory endurance test of the Functional Fitness Assessment Battery of the American Alliance of Health, Physical Education, and Dance (Bravo et al. 1994, Shaulis et al. 1994), (b) the reliability, safety, feasibility, content validity, predictive validity and sensitivity to change of the Tinetti Physical Performance Test (Reuben et al. 1992, Koch et al. 1994, Tinetti et al. 1994), and (c) the reliability, safety, feasibility, content validity and predictive validity of the Physical Performance Battery Assessing Lower Extremity Function (Guralnik et al. 1994 and 1995, Ferrucci et al. 1997).

Two disease-specific performance test batteries were recently evaluated. Rejeski et al. (1995) described the development and validation of a test battery for evaluating activity restrictions in patients with knee OA. The tasks include a 6 minute walk, a stair climb, a lifting and carrying task and getting into and out of a car. Simmonds et al. (1998) assessed the reliability, validity, and potential clinical use of a test battery designed for patients with low back pain. The tasks included repeated trunk flexion, repeated sit-to-stand, timed up-and-go, loaded reach, unloaded reach, 50-foot walk, 5-minute walk, and the Sørensen fatigue test for trunk extensors.

Selected methodological issues and the results of the aforementioned studies are discussed in more detail in the sections 2.3, 2.4, and 2.5.

2.3 Assessment of the reliability of fitness tests

2.3.1 Different aspects of reliability and the components of measurement error

It is the reliability of a new measurement tool that should be tested first in since the toll will never be valid if it is not adequately consistent in whatever value it indicates from repeated measurements (Atkinson & Nevill 1998). When assessing HRF, the testers categorize subjects, make comparisons between individuals and monitor changes in fitness over time. Therefore, the applied fitness measurements need to be *objective*. In other words, 2 or more observers must obtain consistent results, while following the standard measurement procedure (inter-rater reliability). If trained testers cannot agree, the assessment procedure lacks objectivity and utility. *Stability* or consistency over time (test-retest reliability), another aspect of reliability, is critical for evaluating whether an observed change is real (Johnston et al. 1992). *Internal consistency* reliability is the variability between repeated trials within a day (Baumgartner 1989).

Irrespective of the aspect of reliability (objectivity, stability, internal consistency), there are 2 components of variability associated with each measurement error: systematic bias and random error (Atkinson & Nevill 1998). *Systematic bias* refers to a general trend for a measurement to be different in a particular direction between repeated tests (e.g., effects of learning and fatigue). *Random error* can occur due to inherent biological or mechanical variation or inconsistencies in the measurement protocol. Random error is usually larger than systematic bias.

In addition, 2 types of reliability have been identified: relative and absolute (Baumgartner 1989). *Relative reliability* is the degree to which people maintain their position in a sample with repeated measurements. Methods based on correlation coefficients and regression provide an indication of relative reliability (Atkinson & Nevill 1998). *Absolute reliability* is the degree to which repeated measurements vary for individual persons. The methods used to describe absolute reliability include the standard error of measurement (SEM), the coefficient of variation, and limits of agreement (Atkinson & Nevill 1998).

Most of the reliability studies on available fitness testing methods in sports medicine (Atkinson & Nevill 1998), physical education (Lamb 1998), and rehabilitation (Rankin & Stokes 1998) have used Pearson's correlation coefficient and paired t-test to assess reliability. The problems related to Pearson's correlation are that the correlation depends greatly on the range of values of the sample (high inter-individual variation increases the correlation) and the inference that a high correlation (say >0.80) between repeated scores can be equated to good agreement. A high correlation reflects well the stability of position or rank order within a particular sample. However, the absolute reliability can be poor (i.e., large individual variation). More recently, the intraclass correlation coefficient (ICC) has been used to assess relative reliability. It is also affected by sample heterogeneity to such a degree that a high correlation may still mean unacceptable measurement error for some "analytical goals" (Atkinson & Nevill 1998).

T-statistics are used to detect systematic bias, but they provide no indication of random variation between tests and are less likely to detect systematic bias if it is accompanied by large amounts of random error between tests. Therefore, *comparing the reliability results between studies* is not possible unless the size and attributes of the samples tested in each case are virtually identical (Rankin & Stokes 1998). Furthermore, a measurement that is reported to have good reliability for general application should be interpreted with caution unless the raters used in the study were a random sample from a larger population (Rankin & Stokes 1998).

The reliability results of many routinely used tests should be supplemented with the application of absolute indicators of reliability because they may have been erroneously concluded to be sufficiently reliable (Atkinson & Nevill 1998). Since some amount of error is always present with measurements, reliability can be considered the amount of measurement error that has been deemed acceptable for the effective practical use of a measurement tool. However, the results of reliability statistics in physical fitness testing have not yet been discussed with reference to any *"analytical goals" for acceptable reliability* (Atkinson & Nevill 1998). For example, such a goal for acceptable reliability in typical fitness testing would be that the SEM is smaller than the wideness of norm-referenced fitness categories. Future studies should include an examination of *how measurement error relates to the magnitude of the measured variables* irrespective of which type of absolute reliability statistics (SEM, coefficient of variation, limits of agreement) is employed (Atkinson & Nevill 1998).

2.3.2 Former reliability studies on field-based fitness assessment methods

Most of the former reliability studies that are relevant for the present study have used Pearson's correlation coefficient and the paired t-test as the only methods for assessing reliability (Frost et al. 1982, Stones and Kozma 1987, Mellin 1986a, Hyytiäinen et al. 1991, Alaranta et al. 1994a and 1994b, Ito et al. 1996, etc.). This also applies to the test batteries presented in section 2.2 with the exception of the work by Simmonds et al. (1998), who reported ICC and SEM values of the 8-item (see section 2.2) performance tests battery for patients with low-back pain.

The static trunk extension endurance test (Biering-Sørensen 1984) is the only single test for which several authors have reported ICC values and absolute measures of reliability. The results have varied. Biering-Sørensen (1984) reported the coefficients of variation to be 7% (mean 120 seconds).for one man who was tested 5 times in 10 days. Jørgensen and Nicolaisen (1986) assessed the 2-week test-retest reliability for 10 healthy male students aged 20-37 years. The coefficient of variation was 19% (mean 267 and 287 seconds).

Moffroid et al. (1994) reported 1-day test-retest reliability (ICC) to be 0.96 and 0.39 in active (n=7) and inactive (n=22) back pain patients, respectively. Moreland et al. (1997) assessed the inter-rater reliability of 3 raters over 3 days within a 1-week period. The subjects (n=39, mean age 35 years) had no history of back pain, and 77% of them were physically active. The ICC was 0.59 and the SEM 20 seconds (mean 101 seconds). In the same study the reliability of dynamic abdominal (ICC 0.89, SEM 8 repetitions, mean 31 repetitions) and dynamic extensor (ICC 0.78, SEM 9 repetitions, mean 46 repetitions) endurance tests was somewhat better.

Most recently Simmonds et al. (1998) studied patients with low-back pain (n=44) and a healthy control group (n=48). The 2-week test-retest reliability (stability) values were ICC 0.88 and SEM 8.7 seconds (mean 45.2 seconds) for the patients and ICC 0.68 and SEM 17.6 (mean 75.7) for the control group. The corresponding values for objectivity (6 different testers) were ICC 0.99 and SEM 1.4 (mean 45.9) and ICC 0.99 and SEM 1.2 (mean 77.8).

In regard to commonly used flexibility tests, Biering-Sørensen (1984) reported the coefficient of variation for 5 subjects tested 5 times in 10 days for the modified Schober test (4.8%, mean 67 mm). Rose (1991) studied physiotherapy students (n=18, mean age 20 years) and calculated the 3-week test-retest reliability for trunk lateral flexion in terms of the least significant difference, which was expressed in the units of measurement (right side 3 cm and mean 22.8 cm, left side 4 cm and mean 22.6). The value was the extent to which repeated measures must differ to be statistically significant.

2.4 Safety and the feasibility of assessing health-related fitness

Safety is a major concern in HRF testing and exercise prescription for adults. Knowledge of the current and former health status and physical activity level of the subjects, as well as the physiological exertion of fitness tests, are important factors affecting the safety and feasibility of such testing. It is not economically possible, and would be counterproductive to the goal of maximizing physical activity, to require physician-supervised exercise testing of all persons prior to participation in any form of fitness testing and physical activity [King & Senn 1996, ACSM & AHA (American College of Sports Medicine & American Heart Association) 1998]. However, some type of pretesting health screening should be an integral part of assessments of HRF (ACSM 1995, Oja & Tuxworth 1995).

2.4.1 Health risks of physical fitness testing

The assessment of HRF is indicated primarily for unfit and physically inactive adult populations. The health risks of heavy physical exertion are increased among this group (Mittleman et al. 1993). On the other hand, most of the health benefits of physical activity and exercise accrue at moderate levels of intensity (Pate et al. 1995), at which the risks are probably low. The potential health risks include (a) cardiovascular (Gibbons et al. 1989, Mittlemann et al. 1993, Gordon et al. 1995) and (b) musculoskeletal complications (Pollock et al. 1991, Jones et al. 1994, Shaw et al. 1995, Saxton et al. 1995).

Sudden cardiac death is the most serious cardiovascular complication during exercise testing and training. It is caused mainly by symptomatic or latent coronary heart disease (Vuori 1995). Habitually sedentary people have an increased risk, and regular physical activity protects against the cardiovascular complications related to physical exertion (Mittleman et al. 1993). The intensity of exercise, both absolute and relative, is probably the most important characteristic influencing the risk (Vuori 1995). A few serious cardiovascular complications have occurred during clinical exercise testing (Gibbons et al. 1989) or maximal strength testing (Gordon et al. 1995) despite thorough medical screening and supervision.

The safety of non-physician-supervised fitness testing among apparently healthy adults is less well documented. The Canadian experiences (Shephard 1991, Shephard et al. 1991) on exercise testing outside of the medical domain have been encouraging. In a representative population study (n=665) in Finland, no cardiovascular complications occurred during a submaximal walking test under field conditions, and most of the subjects were able to complete the 2-kilometer work adequately for fitness assessment (Laukkanen et al. 1992).

Only a few studies report information related to *musculoskeletal injuries during fitness testing*. Gordon et al. (1995) found no orthopedic complications during 1-repetition maximum and maximal isokinetic strength testing of apparently healthy men (n=5460) and women (n=1193) aged 20 to 69 years. Recently, the safety of 2 simple trunk muscle endurance tests of healthy adults (n=90) and chronic back pain patients (n=199) was reported (Ito et al. 1996). None of the patients experienced worsened low-back pain. Conversely (Moreland et al. 1997), among healthy workers (n=39), neck pain was reported by 2 subjects and low-back pain by 1 subject in an abdominal dynamic endurance test. In an extensor dynamic endurance test 2 subjects stopped because of low-back pain and 2 because of lower-extremity muscle cramps.

Among elderly persons, 11 (19%) of 57 subjects were injured during maximal strength testing with leg extension and chest presses, while no injuries occurred during treadmill testing (Pollock et al. 1991). Based on these findings, the authors stated that "1-repetition maximum strength testing is inappropriate for older men and women who have had previous joint problems specific to the muscle group being tested". In another study (Shaw et al. 1995), with a similar type of testing among 83 elderly subjects, only 2 subjects with no experience with weight training had a back injury and a rib fracture.

Although acute musculoskeletal injuries during fitness testing seem to be rare, musculoskeletal injuries occur frequently among participants in fitness programs, runners, athletes, military recruits, and others who engage in routine vigorous exercise (Jones et al. 1994). Furthermore, exertion injuries are common among the elderly, and they are often connected with degenerative aging processes (Kallinen & Alen 1995, Felson et al. 1997).

Delayed onset of muscle soreness (DOMS), occurring typically after unaccustomed or strenuous eccentric exercise and also after strenuous fitness testing, is usually a self-limiting condition (Kuipers 1994, MacIntyre et al. 1995). However, there is recent evidence that impaired neuromuscular function may affect the successful performance of certain motor tasks during recovery from exercise-induced muscle damage (Saxton et al. 1995). No studies were found which reported the occurrence of musculoskeletal injuries or functional consequences of DOMS in conjunction with adult HRF testing (Shephard 1991, Shaulis et al. 1994, Oja & Tuxworth 1995). DOMS and pain, as well as a high level of perceived exertion during the testing, may also have negative effects on exercise motivation and training adherence, especially among inactive persons (King et al. 1991, Dishman 1994).

2.4.2 Screening for health limitations to fitness testing

Preparticipation screening has been used to ensure the safety of testing and training (Shephard 1988, Laukkanen et al. 1992, ACSM 1995). The screening should identify persons at high risk and should be simple and easy to perform (ACSM 1995, King & Senn 1996, ACSM & AHA 1998). The Physical Activity Readiness Questionnaire (Chrisholm et al. 1975) has been successfully administered as a screening instrument in conjunction with the Canadian home fitness test (Shephard 1988, Shephard et al. 1991). It is sensitive in finding persons with potential health risks, but it excludes a high number of subjects (Shephard 1991, Shephard et al. 1991, Thomas et al. 1992). More recent recommendations for exercise testing and prescription (King & Senn 1996, Norton et al. 1998, ACSM & AHA 1998) apply screening procedures which involve classifying persons into 1 of the following 3 risk groups: apparently healthy persons, persons at higher risk, persons with known disease. The screening includes cardiovascular risk factors, health history, and physical activity level.

According to ACSM & AHA (1998), apparently healthy persons of all ages and asymptomatic persons at increased risk can participate in *moderate-intensity* exercise and non-diagnostic fitness assessment without first undergoing a medical examination or a medically supervised, symptom-limited exercise test. Younger persons (men younger than 45 and women younger than 55 years of age) may also participate in *vigorous* exercise. All other persons should undergo a medical examination and perform a maximal exercise test before participation in moderate or vigorous exercise or non-diagnostic fitness testing. [For details see King & Senn (1996), Norton et al. (1998), ACSM & AHA 1998).] This type of preparticipation screening can limit the number of unnecessary referrals for further medical evaluation and thus prevent undue expense and barriers to participation (ACSM & AHA 1998). However, fitness testing personnel should have the training and experience needed to ensure safe and effective testing. The level of education and experience needed varies with the health status of the client population (ACSM & AHA 1998).

2.5 Validity of health-related fitness tests

2.5.1 Methods used to validate health-related fitness tests

Ideally, a new instrument is compared with a "gold standard" [For example indirect methods of determining maximal oxygen uptake (VO_{2max}) are compared with the direct measurement of VO_{2max}] Unfortunately, for musculoskeletal and motor fitness tests, gold standards do not exist. Therefore, the new instrument is frequently compared with an established construct (Reuben et al. 1992) (i.e., the performance of an instrument is compared with other measures that might be related but are not identical.) For example, some correlation is expected between objectively measured physical performance and the self-report of function and health (Rejeski et al. 1995, Simmond et al. 1998). In the present study the term *content validity* is used when referring to this type of validity. Another method of validating an instrument is to measure its ability to predict health outcomes such as death, need for institutionalization, or incidence of selected disease or disorder (Reuben et al. 1992). In the present study *predictive validity* is used when referring to this type of validity. The content and predictive validity of HRF components and factors relating to low-back and mobility functioning, the main interest of the present study, are reviewed in sections 2.5.2 and 2.5.3, respectively.

A common means of validating HRF tests has been the comparison of results of subjects with and without (or between several groups) the outcomes of interest (Reuben et al. 1994, Guralnik et al. 1995, Simmonds et al. 1998, Thomas et al. 1998). Statistical methods of t-tests and analyses of variance have been used for continuous variables, and chi-square tests are common for categorical variables. Correlations between performance tests and selected outcome measures have also been reported in content validity analyses (Rejeski et al. 1995, Simmonds et al. 1998).

Multiple linear regression (Reuben et al. 1992, Guralnik et al. 1994) and logistic regression models (Guralnik et al. 1995, Ferrucci et. al. 1997) have been used to evaluate the content and predictive validity of performance tests for outcomes of interest. Multivariate analyses can gauge the independent contributions of fitness measures in predicting outcome measures (Reuben et al. 1992). Adjustments can be made for number of confounding variables (age, sex, education, smoking, etc.) in the models.

With reference to diagnostic tests, *sensitivity* (proportion of people for whom the test result is positive among those with the target disorder) and *specificity* (proportion of people for whom the test result is negative among those without the target disorder) are the old, and according to Jaeschke et al. (1994), less useful measures of accuracy of diagnostic tests. Recently, simpler and more efficient methods have been introduced (Jaeschke et al. 1994, Thomas et al. 1998).

In a study on back pain (Thomas et al. 1998) a *receiver-operator characteristics curve analysis* was used to determine the ability of 7 flexibility measures to discriminate between those with and those without back pain. The area under the curve was used as an indicator of discrimination of each method. On visual inspection of the curves, the point of maximal discrimination was derived and was taken as the cutoff in classifying a person's movement as "normal" or "abnormal". These cutoff points were then applied to the data set, and the sensitivity and specificity of each measure for detecting back pain was calculated together with the *likelihood ratios* for each measure. Likelihood ratios is a measure, that indicates the clinical usefulness of a diagnostic test by determining the accuracy with which it identifies its target disorder. [For details see Jaeschke, et al. (1994).]

As an example, based on the area under the curve, the angle of standing extension was the most discriminatory measure of 7 flexibility tests (Thomas et al.

1998). At the cutoff angle for maximal discrimination of 50 degrees, there was sensitivity of 71% (i.e., 71% of the individuals experiencing low-back pain could be identified by this cutoff). The likelihood ratio was 4.9 (i.e., those with this degree of restricted movement were 5 times more likely to have back pain than those without). In the future, these methods could be applied to *establish health-based threshold values* for adequate levels of HRF to promote positive health and the maintenance of physical functioning.

2.5.2 Physical activity, fitness and low-back trouble

Physical activity seems to have dual role as a positive and negative influence on the spine (Videman et al. 1995 and 1997). Investigations on the association between different components of fitness and low-back trouble (LBT) have also revealed contradictory results. There are considerable methodological difficulties involved in an attempt to study the effect of LTPA and fitness on LBT. Furthermore, genetics may play an important role in the degenerative processes (Jimenenz & Dharmavaram 1994, Battié et al. 1995).

Fergusson and Marras (1997) suggested that inconsistencies in the literature may be due to variations in the *outcome measures* and definitions of *risk factors* (independent measures). LBT may be viewed as a progression of events. It can begin with spinal loading and progress to discomfort, then symptoms, and then disorder (injury or illness), followed by the report of an incidence possibly leading to restricted work or disability (Fergusson & Marras 1997). LBT can be reported at any point in this progression of events, or it may never be reported.

The typical outcome measures in epidemiologic studies on LBT have been defined as a symptomatic state rather than as verifiable clinical outcomes, and the most non-specific, low-back pain, is the most common (Burdorf et al. 1997). It is also unclear how the structural changes and symptoms are related (Räty et al. 1997).

The proposed preventive function of physical activity and physical fitness may be *different for work-related and non-work-related LBT* (Barnekow-Bergkvist et al. 1998). The lifetime exposure to occupational physical loading is often many times greater than the loading during leisure-time, the former may obscure the effects of exercise-related physical loading (Videman et al. 1997). Many studies have shown an association between physical loading and a high incidence of LBT. Heavy physical work, in particular, has been correlated with back-related symptoms and degenerative changes of the spine (Videman et al. 1990, Riihimäki 1991). The results of studies of highly selected groups of former athletes with different lifetime loading patterns (Videman et al. 1995 and 1997, Räty et al. 1997) may not apply to less vigorous sports participation, which is more common among the general population (Videman et al. 1997).

Little is known of what aspects of exercise would be the most relevant to the back, and no established methods for its assessment exists (Leino 1993). Adams and Dolan (1997) recently proposed the hypothesis that *large and abrupt increases in a person's level of physical activity* may leave the lumbar discs as a weak link (due to the low metabolic rate) that prevents adequate adaptive remodeling changes. Discs adapting more slowly than bones could explain why former elite weight lifters have more bulging discs than former elite runners, while the number of end-plate defects does not notabley differ (Videman et al. 1995). Cumulative fatigue damage could also explain (Adams & Dolan 1997) the more common occurrence of back pain among student nurses after 9-12 months of active training on the wards (Moffet et al. 1993).

Despite a positive effect of LTPA on physical fitness, occupational physical activity does not improve fitness (Nygård et al. 1994) rather a decreased capacity is often seen among elderly doing physical work (Nygård et al. 1991). With reference to musculoskeletal fitness, physically heavy work seems to have a negative effect on the trunk and lower extremities, especially in women, but also a possible maintaining or training effect on the upper extremities (Torgén et al. 1999).

The common hypotheses on the associations of LTPA and HRF with LBT are presented in Table 2. To overcome some of the noted methodological problems, the criteria to include a study in the present review were as follows: first, the *subjects* had to be (a) a representative population sample, (b) a representative occupational group or several groups, or (c) a selected group in a properly matched case-control study, and, second, (d) a healthy reference group had to be defined and (e) the effect of age on the outcome measure had to be controlled.

TABLE 2. Hypotheses on the associations between leisure-time physical activity (LTPA) and components and factors of health-related fitness with low-back trouble (LBT).

LTPA or fitness component Fitness factor	Hypotheses
LTPA	Physical activity during leisure time is beneficial for the back (Nutter 1988).
Musculoskeletal fitness Trunk muscle strength	High trunk muscle strength minimizes pathological or functional change after an injurious event to the spine (Beimborn & Morrisey 1988, Lahad et al. 1994, Campello et al. 1996).
Trunk muscle endurance	High trunk muscular endurance decreases the loss of motor control, a risk factor for injury, due to the lower amount of fatigue in repeated submaximal trunk motion (Parnianpour et al. 1988).
Flexibility	"Stiff" people are at greater risk of injuring their back during bending and lifting activities because bending stresses are higher in people with poor mobility in the lumbar spine and hips (Dolan & Adams 1993).
Motor fitness	Impaired motor skill could be a cause or a consequence of certain musculoskeletal disorders, such as LBT (Alaranta et al. 1994c).
Aerobic fitness	Persons with a high level of aerobic fitness fatigue more slowly while performing repetitive tasks than those with lower levels of fitness, which decreases the risk of back injury (Nutter 1988).
Morphological fitness Body composition	Obesity is a risk factor for LBT (Deyo & Bass 1989).

Leisure-time physical activity and low-back trouble. In regard to *population-based* studies on the associations of LTPA with LBT, Gyntelberg (1974) found no consistent relationship between LTPA and 1-year incidence of low-back pain among male residents (n=4753) aged 40-59 years in Copenhagen. Those taking part in sport did, however, seem to have a little lower risk of low-back pain. No association was found in a 5-year follow-up study (n=262) with 4 age cohorts of men and women (25, 35, 45, 55 years) in Finland (Kujala et al. 1996). Similarly, no association between LTPA and back symptoms (n=238) was found in a 16-year follow-up study among Swedish men and women at the age of 34 years (Barnekow-Bergkvist et al. 1998). A case-control (n=459) epidemiologic study indicated that most recreational sports are not associated with an increased risk of lumbar herniation (Mundt et al. 1993).

Among metal industry *workers* (n=602) low level of LTPA was associated with worse clinical findings and predicted the development of the low-back symptoms in a 5-year follow-up of men but not that of women (Leino 1993). In occupation-specific analyses the incidence of sciatic pain in men (n=1149) was not related to frequent LPTA among office workers, but among blue-collar workers a tendency towards an increase in the risk was found (Riihimäki et al. 1994).

Physical fitness and low-back trouble. In the present review, 11 cross-sectional, 12 prospective studies, and 2 studies including results on both designs met the inclusion criteria. The *cross-sectional studies* investigated whether the current fitness level of subjects with LBT differs from the fitness level of healthy subjects. The *prospective studies* investigated whether former fitness level can predict back health status after a follow-up period. The main findings of the review on the selected hypotheses (Table 2) are presented in the following sections.

Musculoskeletal fitness: trunk muscle strength (see hypothesis in Table 2). The results of the *cross-sectional* studies on the associations between LBT and trunk strength in terms of maximal voluntary contraction (MVC) are extremely conflicting. The *MVC of the trunk extensors* did not differ between clinically positive and negative male (n=295 aged 19-65 years) construction workers (Holmström et al. 1992), but it did differ among male (n=383) and female (n=215) metal industry workers (Leino et al. 1987). With respect to *trunk flexion* strength no difference was found among men (Holmström et al. 1992, Leino et al. 1987); however, clinically positive women had lower strength (Leino et al. 1987). Among fairly young office and manual workers no difference in trunk extension strength was found between those with and without low-back pain; for trunk flexion the results were dependent on the strength assessment method (Suzuki & Endo 1983). The *extension/flexion strength ratio* was decreased among clinically positive male constructions workers (Holmström et al. 1992), the ratio was not assessed in the study by Leino et al. (1987), and the ratio did not differ between young workers with and those without LBT (Suzuki & Endo 1983). None of the cross-sectional studies meeting the selection criteria measured strength in lifting, trunk rotation or trunk lateral flexion.

In the *prospective* studies, MVC of *trunk extension and flexion* had no predictive value for low-back injury (Videman et al. 1989) among nurses (n=199, mean age 23 years), for first-time sciatic pain (Riihimäki et al. 1989) among concrete reinforcement workers and painters (n=228 aged 25-54 years), and for first-time back pain among a working age population (n=262) of men and women (Kujala et al. 1996) and male steel workers (n=215) under 40 years of age (Masset et al. 1998). Weak trunk muscles did predict recurrent back pain (Biering-Sørensen 1984) among Danish women (n=479) but not among Danish men (n=442).

Recently, Masset et al. (1998) assessed MVC, dynamic torque, and angular velocity (torque 25% and 50% of the MVC) of the trunk in all movement directions (flexion, extension, rotation, lateral flexion). *Rotation* was the only MVC measure that had predictive value for low-back pain incidence in the 2-year follow-up. In addition, dynamic rotation and extension torque were lower among those who developed back pain; however, individuals performing dynamic tests at higher velocities appeared to be at a greater risk of low-back pain. None of the performance measures were included in the final multivariate regression model.

Lifting strength had no predictive value for back injury 3-years later (Battié et al. 1989b) among workers in an aircraft plant (n=2178 aged 21-67 years), for the 1-year incidence of back pain (Luoto et al. 1995b) in an adult working population (n=126 aged 35-54 years), and for back symptoms (n=148) 16 years later among working men at the age of 34 years (Barnekow-Bergkvist et al. 1998). However, the 2-hand lift in adulthood predicted low-back symptoms among women (n=90)(Barnekow-Bergkvist et al. 1998).

Musculoskeletal fitness: trunk muscle endurance (see hypothesis in Table 2). *Trunk extensor endurance* was systematically decreased among subjects with LBT when compared with healthy subjects regardless of the population studied, method of testing or outcome measure in the reviewed *cross-sectional* studies. The results were consistent for static endurance (Holmström et al. 1992, Alaranta et al. 1994b, Barnekow-Bergkvist et al. 1998), repetitive dynamic extensions (Nummi et al. 1976, Leino et al. 1987, Alaranta et al. 1994b) and isokinetic extensions (Suzuki & Endo 1984) The results on *dynamic trunk flexion* (sit-ups, curl-ups) *endurance* show similar associations (Suzuki & Endo 1984, Alaranta et al. 1994b) although less consistently in some studies (Leino et al. 1987, Barnekow-Bergkvist et al. 1998). In a study among 185 chronic back pain patients with no reference group (Rissanen et al. 1994) dynamic performance of trunk muscles, when compared with isokinetic trunk tests, correlated more strongly with pain and the disability index of men, whereas for the women the tests correlated equally well.

In the *prospective* studies, s*tatic trunk extensor endurance* had a predictive value for first-time back pain in Danish men (Biering-Sørensen 1984) and Finnish men and women (Luoto et al. 1995b), but not for recurrent back pain (Biering-Sorensen 1984). *Dynamic trunk endurance* had predictive value for clinical back findings in a study among male metal industry workers (Leino et al. 1987) and for back injury among nurses (Videman et al. 1989), but not for first-time back pain (Luoto et al. 1995b) or other back pain (Kujala et al. 1996) among working age urban populations.

Musculoskeletal fitness: flexibility of the lumbar spine and hips (see hypothesis in Table 2). *Lumbar flexion* and *extension* has been measured with several different methods [flexicurve (Burton et al. 1989b), gravity inclinometer (Alaranta et al. 1994a), the modified Schober test (Biering-Sørensen 1984) and fingertip-to-floor distance (Biering-Sørensen 1984)], and the results are conflicting.

With reference to *cross-sectional* studies, Burton et al. (1989b) found a reduced extension range of motion (ROM) for men (n=274) and a reduced total lumbar sagittal ROM for women (n=271) with history of LBT working in variety of occupations. Alaranta et al. (1994a) found no difference in flexion or extension with reference to previous low-back pain among workers of the Helsinki city council (n=508).

Only measures of *lumbar flexion* were used in the *prospective* studies. High mobility in the Schober test increased the risk of first-time low-back pain among men, but not among women (Biering-Sørensen 1984), and it had no predictive value for back injury (n=3020) among workers in an aircraft plant (Battié et al. 1990).

One *cross-sectional* study (Alaranta et al.1994a) measured *trunk lateral flexion*, which was reduced in subjects with LBT compared with those never having experienced it. In a *prospective* design lateral mobility had some predictive value for low-back symptoms among nurses (Videman et al. 1989) but not for back injury in an aircraft plant (Battié et al. 1990). *Trunk rotation* ROM did not differ between subjects with LBT and those never having experienced it (Alaranta et al. 1994a). None of the prospective studies included measures of trunk rotation.

Poor (lower than the median) *hamstring muscle extensibility* was not associated with self-reported back symptoms within 12 months (Barneko-Bergkvist et al. 1998) in a *cross-sectional* design. In a *prospective* design *short hamstring muscles* increased the risk of recurrent low-back pain among women but not among men (Biering-Sørensen 1984), and it had no predictive value for back injury when measured by the sit-and-reach test (Battié et al. 1990).

Motor fitness. It has recently been suggested that impaired motor skill could be a cause or a consequence of certain musculoskeletal disorders (Alaranta et al. 1994c). There is some evidence that low-back pain patients have impaired postural control (Byl & Sinnot 1991, Luoto et al. 1996 and 1998), long reaction time (Taimela et al. 1993, Venna et al. 1994, Luoto et al. 1995a) and deficits of motor control of the trunk muscles (Hides et al. 1996, Hodges & Richardson 1996). However, none of these studies met the inclusion criteria for the present review. Three other studies were accepted (Videman et al. 1989, Takala et al. 1997, Barnekow-Bergkvist et al. 1998).

High-fitness (equal or higher than the median) in the one-leg *balance* test was associated with a decreased risk of low-back symptoms in working men (n=148) but not in working women (n=90) at the age of 34 years (Barnekow-Bergkvist et al. 1998). Postural sway, measured by force-plate, had a wide variation among forest industry workers (n=508), the sway being slightly lower in the non-symptomatic group than in the symptomatic group (Takala et al. 1997). Prospectively, poor patient handling skill was associated with the 1-year incidence for job-related back injury (Videman et al. 1989) among nurses (n=199).

Aerobic fitness (see Table 2 for hypothesis). In the present review, 3 studies on the associations between aerobic fitness and LBT, all *prospective* in design, were found. Among urban populations, aerobic fitness had no predictive value for 1-year (Gyntelberg 1974, n=3894) or 5-year (Kujala et al. 1996, n=262) incidence of low-back pain symptoms. Among workers in an aircraft plant (n=2434) it had no predictive value for back injury (Battié et al 1989a).

Morfological fitness (see Table 2 for hypothesis). In the United States *obesity*, measured in terms of *body mass index* (BMI) was an independent risk factor for low-back pain among 10 404 adults aged 25 years and older (Deyo & Bass, 1989).The prevalence rose substantially only among the most obese 20% of the subjects. Accordingly, obesity (BMI >30 kg/m^2) based on self-reported weight and height was associated with increased prevalence of severe back pain in nationally representative (n=26 402) Swedish population (Wolk & Rössner 1996). In a British population (n=9003, aged ≥18 years) women, but not men, with the BMI in the highest 40% had increased prevalence of back pain in the month before interview (Croft & Gigby 1994). Similarly, 5 consecutive cross-sectional population surveys of adults (n= 29 043, aged 30-59 years) in eastern Finland between 1972 and 1992 showed that BMI was directly proportional to the prevalence of back pain among women, but not among men (Heistaro et al. 1998). However, in 1987 and 1992, the BMI category differences seemed to become obvious among the men as well. *Prospectively*, in a representative Finnish population sample (n=31 111, aged 25-64) overweight predicted work disability due to diseases of the back (Rissanen et al. 1990).

2.5.3 Physical activity, fitness and mobility-related disability

Physical activity and mobility-related disability. There is increasing concern that too much physical activity may lead to *osteoarthritis* (OA) of the lower extremities (Saxon et al. 1999), a major cause of mobility-related disability. Mobility-related disability seems to be more common among middle-aged and older women than among men in the same age range (Davis et al. 1991, McAlindon et al. 1993, Launer et al. 1994). Some studies provide evidence that a decline in mobility function precedes changes in activities in daily living (Guralnik et al. 1993, Dunlop et al. 1997).

Interpreting the available information on the proposed risk factors for OA is difficult because the selection of outcome measures influences the results (Davis et al. 1991, Panush 1994), and there is variability in the predisposition to joint degeneration, exercise patterns, and previous joint injuries among populations (Buckwalter & Lane 1997, Saxon et al. 1999). A combination of sports participation and occupational workloads can affect the development of OA (Vingård et al. 1993 and 1996).

Population-based studies. Most recently, the relationship between sports activities and knee OA in the general Swedish population (n=1173) was investigated in case-referent study (Sandmark & Vingård 1999). For men between 55 and 65 years of age, the risk of severe knee OA was increased among those highly exposed to all kinds of sports, and the highest risk estimates were found for those who reported exposure to cross-country skiing and soccer. No association was found for the women; however, only few had participated in sports activities. Moderate daily general physical activity was not found to be a risk factor. Accordingly, Hannan et al. (1993) reported that general habitual physical activity during middle age was not associated with knee OA in later years (n= 1415, mean age 73 years). However, the analysis of the same cohort 10 years later (n=598) indicated that LTPA (highest quartile) increased the risk of radiographic OA (Felson et al. 1997). In a case-control study (Imeokparia et al. 1994), highly active women (n=308) aged 55 to 64 years were at increased risk of knee OA, but no such association was found for men (n=170).

Despite of the possible negative effects of LTPA on OA, several large epidemiological studies on *functional disability*, focused on populations over 65 years of age (Huang et al. 1998), consistently indicate that moderate and higher levels of physical activity are associated with the maintenance of basic physical performance, as well as with higher-order goal functions, and general *mobility* (LaCroix et al. 1993, Nelson et al. 1994, DiPietro 1996, Schroll et al. 1997, Huang et al. 1998, Laukkanen et al. 1998).

With regard to *athletes*, Saxon et al. (1999) stated in their recent review article that, despite the methodological problems, results have shown that the sports characteristics that appear to increase the risk of developing OA include torsional loading, fast acceleration and deceleration, repetitive high impact training, and high levels of sports participation. High risk sports include power and mixed sports such as track and field, racket sports, soccer, rock climbing and throwing activities. Among the athletes one of the major problems in estimating the relationship between sports participation and OA is the fact that athletes who develop OA may stop participating in sports. This change results in selection bias and an underestimation of the risk of developing sports-related OA. Well-conducted, long-term prospective cohort studies are needed to provide the best estimate of the risk of sports-related OA (Saxon et al. 1999).

Physical fitness and mobility-related disability. Increasingly, physical fitness (or performance) tests have been used in epidemiological studies as independent factors for disability. However, very few studies have included middle-aged populations (Launer et al. 1994, Huang et al. 1998, Rantanen et al. 1999b). Performance measures reflect functional limitations, whereas self-reported functional status refers more to disabilities (Hoyemans et al. 1996, Kivinen et al. 1998). Self-report scales are likely to be insensitive to change, particularly early in the course of functional decline (Reuben et al. 1992).

The results of cross-sectional and prospective studies on the association between objectively measured fitness (or performance) tests and self-reported mobility-related disability are presented as follows. The criteria to include a study into the review were that (a) the subjects were a representative sample of an independently living elderly population and (b) only subjects with no mobility limitations at baseline had been included in the follow-up studies.

Cross-sectional studies, investigating whether the current fitness level of elderly subjects with mobility-related disability differed from the fitness level of non-disabled subjects, clearly indicated that objectively measured walking speed (Avlund et al. 1994, Ensrud et al. 1994, Fried et al. 1994, Guralnik et al. 1994, Hirsch et al. 1997), ability to rise from a chair or climb stairs (Avlund et al. 1994, Guralnik et al. 1994, Hirsch et al. 1997), standing balance (Ensrud et al. 1994, Guralnik et al. 1994), knee extension strength (Avlund et al. 1994, Ensrud et al. 1994, Rantanen 1994, Ferrucci et al. 1997) and trunk extension strength (Avlund et al. 1994, Rantanen et al. 1994) were associated with mobility-related disability. Contrary to simple performance tests of balance, measures on a force platform were not associated with mobility disability (Era et al. 1997).

The MVC of the knee extensors was also associated with objectively measured walking speed, stair mounting, ability to climb stairs or rise from a chair, and standing balance (Rantanen et al. 1994 and 1996, Ferrucci et al. 1997). Several of the studies discussed the possible threshold levels of knee extension strength for lower extremity functioning (Sonn et al. 1995, Rantanen et al. 1996, Ferrucci et al. 1997). In addition, hip strength (Ensrud et al. 1994, Ferrucci et al. 1997) and hand-grip strength (Ensrud et al. 1994, Fried et al. 1994, Hirch et al. 1997) were associated with both self-reported mobility disability and objectively measured functioning.

Flexibility of the trunk and lower extremities and aerobic capacity were not assessed in any of the studies that met the inclusion criteria. In a more selected elderly group (Morey et al. 1998a and 1998b) maximal aerobic power (VO_{2max}) and flexibility showed direct associations with functional limitations and disability.

Prospective studies among the elderly, investigating whether the subjects' former fitness level could predict mobility-related disability after a follow-up period, confirm the findings of the cross-sectional studies. The studies consistently indicate that objectively measured walking speed (Guralnik et al. 1995, Schroll et al. 1997), ability to rise from a chair (Guralnik et al. 1995) or climb stairs (Schroll et al. 1997), and standing balance (Guralnik et al. 1995) have predictive value for mobility-related disability. Furthermore, the MVC of the trunk extensors proved to be a predictor of independence in mobility, and there was a dose-response relationship between the MVC of the knee extensors and stair mounting height (Schroll et al. 1997).

Hand-grip strength was a strong predictor of mobility-related functional limitations (objective measures of slow walking speed, inability to rise from a chair) and self-reported disability at old age (Rantanen et al. 1999b). In addition, prospective data from Launer et al. (1994) suggested that a high BMI is a strong predictor of long-term risk for mobility disability among women aged 60 to 65 years.

2.6 Summary

Assessing and monitoring relevant aspects of fitness may have an important role in the promotion of physical activity for health. The Toronto model for physical activity, fitness and health, including the concept of HRF, offers a theoretical frame with which to assess the interrelationships between physical activity, fitness and health. Positive aspects of health, such as good function and well-being, are emphasized in the model.

The majority of the currently available, scientifically validated methods for assessing the musculoskeletal and motor components of HRF have been designed for elderly persons. Thus reliable, safe, and feasible tests need to be developed for middle-aged populations. Furthermore, the validity of the methods has to be evaluated with reference to important health outcomes.

Different aspects of reliability and components of the measurement error of any new measurement tool need to be assessed and the results should be interpreted with reference to the intended purpose and use. There is an obvious need to reassess the measurement errors of many existing fitness tests, especially in terms of absolute reliability (i.e., agreement).

Safety is a major concern in HRF testing for adults. The potential health risks include cardiovascular and musculoskeletal complications. The risks are increased among unfit and physically inactive adults. Preparticipation screening, including cardiovascular risk factors, health history and physical activity level, is an integral part of assessments of HRF aimed at ensuring safe testing.

The health-related validity of fitness tests can be evaluated by comparing the test results with selected health outcomes. In regard to LBT, a major cause of work disability and activity restrictions among middle-aged persons, most of the hypotheses on the relationships between fitness and LBT need further investigation in methodologically well-designed studies.

Trunk extensor endurance is the only musculoskeletal fitness factor that has been systematically associated with LBT in cross-sectional studies, and it has been shown to have predictive value for first-time back pain in 2 population studies (see page 30). The methods for assessing flexibility in the reviewed studies varied and therefore made it difficult to evaluate the study results. With reference to motor fitness, postural control and muscular coordination have been shown to be impaired in patients with low-back pain, but population-based studies are needed to confirm the findings and to evaluate their predictive role in back health. The results on aerobic fitness, measured in 3 of the reviewed prospective studies (see page 31), indicated that it has no predictive value for LBT. Obesity, a morphological fitness factor, has been systematically associated with current LBT in several cross-sectional population studies. Only 1 prospective study was found for the review (see page 32); its results were consistent with the cross-sectional findings.

OA of the lower limbs is a major cause of mobility-disability among the elderly. Although intensive LTPA may have adverse effects on the development of lower extremity OA, both cross-sectional and prospective studies consistently show the benefits of regular physical activity and good performance capacity in maintaining mobility. The important fitness factors for mobility functioning are walking speed, lower extremity strength, and balance. Muscle strength may have a mediating role between physical activity and mobility-related disability. Threshold values for the strength of knee extensor muscles have been discussed, but no such values have been reported for practical application. Measures of aerobic capacity and flexibility were seldom included in the studies.

The review of the literature reveals that the inherent characteristics of the reliability, safety, feasibility, and validity of existing fitness tests for middle-aged adults have been incompletely studied and documented. Therefore, there is definite need for a systematic evaluation of such characteristics of potential tests in order to design a scientifically sound HRF test battery.

3 PURPOSE AND DESIGN OF THE STUDY

The purpose of this study was to develop a field-based HRF test battery for middle-aged adults. The emphasis was on musculoskeletal and motor fitness. More specifically the aims were to:

(1) (a) assess the *reliability* (I) in terms of *objectivity* (inter-rater reliability) and *stability* (1-week test-retest repeatability) of the preliminary selected musculoskeletal and motor fitness tests and to (b) propose a reliable HRF test battery for further evaluations among a representative population sample,

(2) assess (a) the *safety* in terms of acute cardiovascular and muscular complications, heart rate and DOMS and (b) the *feasibility* in terms of subject suitability and practicality of the whole testing procedure and single fitness test items (II),

(3) evaluate the *health-related content validity* of the proposed test battery by studying the cross-sectional associations of fitness test results with *self-rated* perceived health, mobility function, and back functioning and pain (III),

(4) evaluate the *health-related predictive validity* of baseline fitness test results with 3-year changes in *self-rated* perceived health, mobility function, and back functioning and pain (IV),

(5) evaluate the *physical-activity-related content validity* of the proposed test battery by studying the cross-sectional associations between physical activity patterns and fitness test results (V).

The development procedure and study design for the proposed HRF test battery are presented in Figure 2. Before the reliability study (I), a literature study for the selection of a preliminary test battery and a pilot study for feasibility assessment were conducted. A list of the tests that were tentatively selected but later excluded is presented in Appendix 1. The proposed test items are presented in Figure 4 in the Methods section.

The criteria for accepting a test into the battery were that it had to be objective and stable over time, a majority of the middle-aged subjects could safely participate and the test could be associated with a person's current or future self-rated health status.

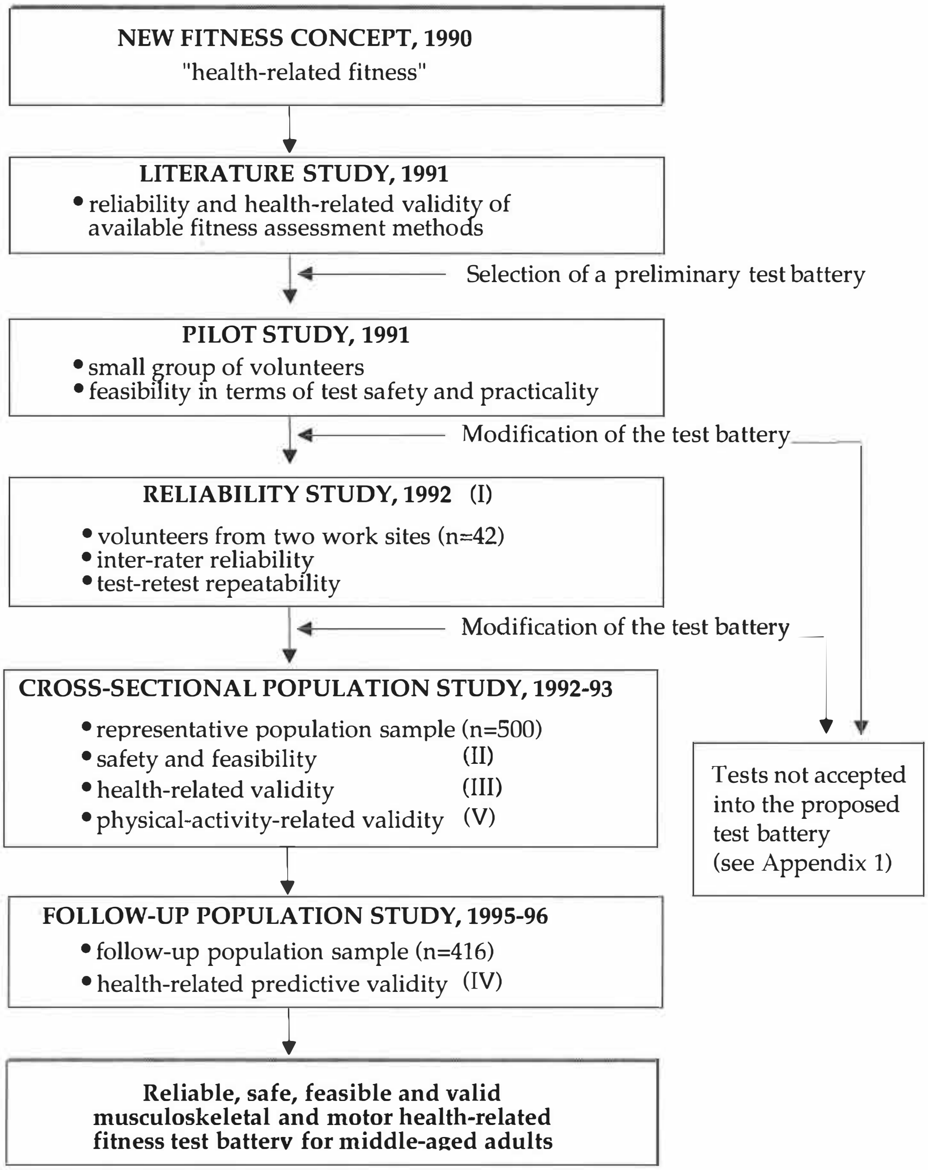


FIGURE 2. Study procedure and design for the development of musculoskeletal and motor health-related fitness test battery for middle-aged adults.

4 MATERIAL AND METHODS

4.1 Subjects

4.1.1 Reliability study (I)

The subjects were volunteers from two work sites. Twenty-two men and 20 women were chosen from different age (mean 41.9, range 25-59 years) and occupational groups.

4.1.2 Cross-sectional population study (II, III, V)

The study subjects represented middle-aged residents of the city of Tampere (II). The study sample was drawn from persons who had previously attended preventive health examinations for all city residents. The selection of the subjects is presented in Figure 3. Altogether 437 men and 389 women, evenly selected from 5 age groups by systematic sampling, were invited to participate in the study, and 56% of the men and 65% of the women agreed. The background characteristics of the subjects are given in Table 3. The non-participants had a somewhat lower level of education, rated their health as lower, used prescribed medication more often, were smokers more often, and exercised briskly less often than the participants (II).

4.1.3 Follow-up population study (IV)

Of the subjects in the cross-sectional study, 87% of the men and 80% of the women participated in the 3-year follow-up study. The age-specific participation rates are shown in Figure 3. A short telephone interview concerning the reasons for not participating in the follow-up was conducted among the subjects who dropped out (n=82). Of them, 79% answered the interview, 12% were not reached, 6% were not willing to answer, and 3% had died. The detailed reasons for not participating are given in original article IV.

POPULATION OF CITY OF TAMPERE

Preventive health examinations: 80% of selected age brackets
- 80% gave consent to use data for research

↓

TARGET POPULATION

Random sample: 437 men, 389 women
Five age cohorts: subjects born in 1935, 1940, 1945, 1950, 1955

↓

did not participate: 316
interrupted: 10

CROSS-SECTIONAL STUDY POPULATION

Age-specific (years) participation rates

	37	42	47	52	57	All
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Men:	57 (50)	58 (46)	52 (50)	59 (50)	57 (50)	56 (246)
Women:	62 (53)	63 (50)	68 (54)	66 (47)	68 (50)	65 (254)

↓

did not participate: 82
were excluded: 2

THREE-YEAR FOLLOW-UP STUDY POPULATION

Age-specific (years) participation rates

	40	45	50	55	60	All
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Men:	88 (44)	80 (37)	84 (42)	90 (45)	92 (46)	87 (214)
Women:	77 (41)	80 (40)	72 (39)	85 (40)	84 (42)	80 (202)

FIGURE 3 Selection of the study population.

TABLE 3 Background characteristics of the cross-sectional study population.

Factor	Men n=245	Women n=253
Marital status	%	%
Married/cohabiting	85	75
Single, divorced, separated, widowed	15	25
Education level		
Secondary school	33	40
High school/vocational training	54	46
University degree	13	14
Occupational physical activity		
Not working	20	21
Sitting work	35	36
Light to moderate movement at work	31	40
Heavy physical work	14	3
Smoking status		
Never smoked	36	67
Past smoker	30	16
Current smoker	34	17
Body mass index		
≥ 30 (mass/height^2)	15	16
Blood pressure		
systolic ≥ 160 mmHg	5	2
diastolic ≥ 100 mmHg	12	5

4.2 Measurements

The procedures and measurements described in this report apply to the cross-sectional (II, III, V) and follow-up (IV) studies. Detailed information on the measurements are given in the original articles. The procedures concerning the reliability study are given in original article I.

4.2.1 Procedures and pretesting health screening (II)

In both the cross-sectional and follow-up studies the subjects attended 2 measurement sessions at the UKK Institute. At the first visit the participants answered a questionnaire on self-rated health, LTPA, and other living habits. In addition, a standard pretesting health screening, including measures of BMI, blood pressure and a modified physical activity readiness questionnaire (Thomas et al. 1992), was conducted. According to prepared safety instructions the fitness testers used the health screening results to refer subjects with severe diseases or symptoms to a physician for a health examination or to exclude subjects with minor limitations from selected fitness tests. At the second visit the assessment of HRF was conducted by 3 testers.

4.2.2 Health-related fitness assessment (II, III, IV,V)

The components and items in the HRF test battery are given in Figure 4, along with a short description of the methods. Each subject was tested individually. All the tests followed a standard sequence, balance being measured first, followed by flexibility, leg power, leg strength, modified push-up strength, and static trunk extensor endurance. After these tests the subjects rested for about 10 minutes before performing the Walk Test (Oja et al. 1991, Laukkanen et al. 1993, Laukkanen 1993). The subjects were given verbal encouragement in a consistent manner to achieve their best performance. Relations between the fitness test items were assessed by age-adjusted partial correlations, and some test items were found to be moderately interrelated. Detailed information is presented in original paper III.

For the statistical analysis the subjects were grouped into fitness categories (Blair et al. 1996) for all the test scores based on age- and sex-specific cut points, and they were classified as "low-fit" (lowest 20% or 40%), "mid-fit" (mid 20% or 40%) or "high-fit" (highest 40%). The distributions of the one-leg balance and squat test scores were skewed. Details of the fitness categories are presented in original papers III, IV, and V.

4.2.3 Self-rated perceived health, mobility function, and back functioning and pain (III, IV)

Five simple measures of self-rated health were used to evaluate the health-related validity of the proposed HRF test battery (III, IV). A questionnaire (III) measuring 3 aspects of *self-rated health* (global health perceptions, physical function, bodily pain) within the classification scheme of health-related quality of life (Caspersen et al. 1994, Wilson & Cleary 1995) was applied. The questions with original ratings and categories used in the cross-sectional and follow-up study are presented in Table 4 and the descriptive results in Table 5. The rationale for the selection of the particular measures was as follows.

Perceived health is a measure of global health perception (see Table 4) which represents an individual integration of many aspects of the health concept (Wilson & Cleary 1995). There is strong evidence of the validity of the perceived health measure among middle-aged populations in terms of the relationship with health as measured by other indicators (Kaplan et al. 1996, Miilunpalo et al. 1997). Recently, the short-term (mean 22 days) test-retest reliability of perceived health (good, bad, something between) was shown to be good (weighted kappa 0.67) in all population groups among 2 representative samples of the Swedish population (Lundberg & Manderbacka 1996). Among a representative sample in north-eastern Finland (Miilunpalo et al. 1997) the test-retest reliability over 13 months of a 5-alternative question (good, fairly good, average, rather poor, poor) was somewhat poorer (weighted kappa range 0.42-0.47, depending on age and sex).

Questions on physical functional activities reveal the persons' perceptions of their ability to perform particular tasks (Wilson & Cleary 1995). The question selected to assess *mobility function* was the *ability to climb several flights of stairs* (see

MUSCULOSKELETAL FITNESS
Muscular strength and endurance

ONE-LEG SQUAT with increasing weights
Purpose: To assess functional leg extensor strength.
Method: The subject takes a short step forward on the mat, first with the right leg, squats down until lightly touching the mat with the left knee, rises immediately and steps backward to the starting position; she then repeats the squat with the left leg.
Outcome: The load limit for a successful squat task measured as the maximum weight relative to the subject's body weight (BW) up to 140%. The test starts with BW (i.e., no added weight) and 10% increments of BW are added at 4 successive steps of 10%, 20%, 30% and 40% using a weight belt system.

VERTICAL JUMP
Purpose: To measure leg extensor power.
Method: The subject is instructed to jump as high as possible and to bend the knees and swing the arms to enhance performance. Before jumping, one arm is raised to mark the standing height. During the jump the subject touches a board with the middle finger at the highest position.
Outcome: The difference between the reach height and the jump height is measured with a tape to the nearest centimeter.

STATIC TRUNK EXTENSION ENDURANCE
Purpose: To measure the endurance capacity of the trunk extensor muscles.
Method: The subject lies prone with the lower body (from the level of the spina iliaca anterior superior) resting on a low bench and crosses the hands behind the neck. The bench is placed on one gymnastic mat and another mat is placed on the bench. The tester stabilizes the subject by sitting on the subject's ankles. The subject is asked to rise the upper body to a horizontal level and hold the position as long as possible for up to 4 minutes.
Outcome: Endurance time in seconds.

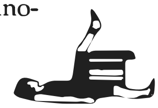
MODIFIED PUSH-UPS
Purpose: To measure dynamic muscular endurance of the upper-extremity extensor muscles and the ability to stabilize the trunk.
Method: The subject lies prone. The push-up cycle begins as the subject claps the hands behind the back. Next a normal straight-leg push-up to straight elbows is performed, followed by a touch of one hand to the top of the supporting hand. The cycle ends in the prone lying position.
Outcome: The number of push-ups completed in 40 seconds.

FIGURE 4 Components, test items, and description of the methods of the proposed musculoskeletal and motor health-related fitness test battery for middle-aged adults.

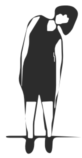
FIGURE 4 continued

Flexibility

HAMSTRING MUSCLE EXTENSIBILITY
Purpose: To measure the active knee extension range of motion in order to assess hamstring muscle extensibility.
Method: The subject lies supine. The hip and the knee of the limb to be measured is flexed to 90 degrees. The opposite leg rests extended. The inclinometer (Vinkelmätare "Myrin", LIC, Rehab Vårdum,Solna, Sweden) is attached to the medial side of the ankle.
Outcome: End point range of motion angle in degrees at maximal extension.



TRUNK SIDE-BENDING to the right and left
Purpose: To measure the total range of movement of lateral flexion of the thoracic and lumbar spine and pelvis.
Method: The subject stands on marked lines (15 cm apart) with the back against the wall. Arms are kept straight at the sides of the body. The subject bends to the right and then to the left as far as possible; the middle finger slides laterally down along the thigh.
Outcome: The distance the fingertip moves down the leg during maximal bending measured with a cloth tape in millimeters. The average value of the right and left sides is calculated.



MOTOR FITNESS
Static balance

ONE-LEG STANDING
Purpose: To assess static postural control while the area of support is reduced.
Method: The subject wears sport shoes. He or she places one foot at knee level along the inner side of the supporting leg and rotates the thigh outwards. The subject is advised to stand as still as possible.
Outcome: Duration of the balance task up to 1 minute as measured with a stopwatch in seconds.

CARDIORESPIRATORY FITNESS

UKK 2-KM WALK TEST
Purpose: To predict maximal oxygen uptake (VO_{2max}) on the basis of time, heart rate at the end, body mass index, age and sex.
Method: Subject walks as fast as possible on a flat surface using a normal walking style.
Outcome: Predicted VO_{2max} ($ml/min^{-1}/kg^{-1}$) and test time (min).

MORPHOLOGY

BODY MASS INDEX (BMI)
Purpose: To assess obesity.
Method: Standard measures of height and weight.
Outcome: BMI as weight/height^2 (kg/m^2).

Table 4). It was chosen because it is a fundamental activity of everyday life, 22% of Finnish men and 31% of Finnish women aged 30 to 64 years reported at least some difficulties in the particular task in the Mini Finland Survey (Aromaa et al. 1989), and decline in mobility function precedes decline in activities of daily life in later life (Guralnik et al. 1994, Dunlop et al. 1997). The test-retest reliability (kappa 0.57) of a similar type of dichotomic question was acceptable among the Swedish population (Lundberg & Manderbacka 1996). The validity of the questions on mobility status has been established among the elderly (Guralnik et al. 1993).

Back functioning and pain were assessed by 3 questions representing (a) general functioning, (b) task-specific functional limitation (stooping), and (c) pain symptoms (see Table 4). General back functioning is a positively defined health measure (Caspersen et al. 1994) that reflects well-being (Ware et al. 1981). Positive measures are applicable when general population is studied, while only some 10% to 20% of the population will have chronic physical limitations (Ware et al. 1981). Intolerance of a stooping posture is common in everyday life regardless of the occupation of the person. The task-specific question about stooping is therefore applicable to the general population.

Symptom status is one important determinant of physical functional status (Harper et al. 1992, Wilson & Cleary 1995). Questions about bodily pain and other symptoms address the interface between physical and mental health (Ware et al. 1981) (i.e., pain is only partially determined by biological or physiological factors) (Wilson & Cleary 1995). Back pain is common among the general population. Over 50% of Finns over 30 years of age have experienced at least 5 spells of low-back pain (Aromaa et al. 1999). Among middle aged Finnish men and women, almost 50% reported back pain during the preceding month (Heistaro et al. 1998).

There is limited evidence on both the reliability and validity of simple (typically yes or no) questions on back symptoms and disease. The test-retest reliability (15-day interval) of the Nordic questionnaire about low-back trouble was good in terms of non-identical answers, which varied from 0% to 4% (Kuorinka et al. 1987). The weighted kappa coefficient of the test-retest (mean 22 days) reliability for backache during the last 12 months was high (0.75) when rated "No", "Yes, mild" and "Yes, severe"(Lundberg & Manderbacka 1996).

In regard to validity, the prevalence of chronic low-back disorder was underestimated in population-based interview data (Heliövaara et al. 1993) when compared with definite diagnoses of a health examination survey, and the agreement between the methods was relatively poor (kappa 0.43). In the pretesting health screening of the present study only 5% of the subjects (n=25/498) reported a low back disorder (yes or no) which had been or might be aggravated by physical activity (II). This rate is less than the Finnish population estimate of 17% of reduced capacity for leisure time activities attributable to the low-back syndrome (Heliövaara et al. 1989). However, the proportion of subjects with poor back function (17%), frequent back problems while stooping (18%) and frequent back pain (18%) in the present study matched well this estimate (see Table 5).

The self-rated health outcome measures in the cross-sectional study were good and poor general health, no difficulty in stair climbing, good and poor general back functioning, seldom and often problems with the back while stooping, and seldom of often back pain (see Tables 4 and 5).

For the follow-up study, the health outcomes were dichotomized (see Table 4) as either "low" or "high" at baseline (IV). The 3-year changes were assessed separately for the subjects with "low" and "high" status at baseline. The outcome was "improved status" among the subjects with "low status "at baseline and "high status" in the follow-up. The outcome was "deteriorated status" among the subjects with "high status" at baseline and "low status" in the follow-up. The outcome for the other subjects was "no change"(see Table 5).

TABLE 4 Questionnaire ratings and categories of the health outcome measures used in the cross-sectional and follow-up studies.

Questionnaire rating		Cross-sectional study categories		Follow-up study categories Baseline		Change	
How would you describe your state of health in comparison with people of your own age?							
1	very poor	1/2	poor	1/2/3	low	4/5	improved
2	poor					1/2/3	no change
3	average	3	average				
4	good			4/5	high	1/2/3	deteriorated
5	very good	4/5	good			4/5	no change
How well can you manage climbing several flights of stairs without resting?							
1	cannot	1/2/3	some difficulty	1/2/3	low	4	improved
2	much difficulty					1/2/3	no change
3	some difficulty						
4	cannot	4	no difficulty	4	high	1/2/3/	deteriorated
						4	no change
How would you describe the functioning of your back?							
1	very poor	1/2	poor	1/2/3	low	4/5	improved
2	poor					1/2/3	no change
3	average	3	average				
4	good			4/5	high	1/2/3	deteriorated
5	very good	4/5	good			4/5	no change
How often do you have problems with your back while functioning in a stooped position?							
1	constantly	1/2	often	1/2/3	low	4/5	improved
2	often					1/2/3	no change
3	now and then	3	now and then				
4	seldom			4/5	high	1/2/3	deteriorated
5	never	4/5	seldom			4/5	no change
How often do you have back pain?							
1	constantly	1/2	often	1/2/3	low	4/5	improved
2	often					1/2/3	no change
3	now and then	3	now and then				
4	seldom			4/5	high	1/2/3	deteriorated
5	never	4/5	seldom			4/5	no change

TABLE 5 Descriptive results of the health outcomes in the cross-sectional and follow-up studies.

HEALTH OUTCOME categories (baseline for follow-up)	Age group (years at baseline) 37 %	42 %	47 %	52 %	57 %	Men %	Women %	All n
PERCEIVED HEALTH								
Poor* (low)	9	10	13	26	26	16	17	83
Average* (low)	76	64	59	51	53	58	64	301
Good* (high)	15	26	28	23	21	26	19	113
No change** (low)	67	61	60	65	68	62	66	267
No change** (high)	10	22	24	17	11	19	14	68
Improved** (low)	15	12	6	9	13	10	12	46
Deteriorated** (high)	8	5	10	9	8	9	8	34
ABILITY TO CLIMB STAIRS								
Some difficulty* (low)	11	22	24	35	39	14	38	130
No difficulty* (high)	89	78	76	65	61	86	62	367
No change** (low)	7	8	13	25	31	9	24	70
No change**(high)	80	73	75	54	57	80	55	280
Improved** (low)	4	11	7	8	7	4	11	31
Deteriorated** (high)	9	8	5	13	5	7	10	34
BACK FUNCTIONING AND PAIN								
General back functioning								
Poor* (low)	15	6	15	21	30	14	20	86
Average* (low)	47	58	55	45	47	48	53	250
Good* (high)	39	35	30	34	23	38	27	161
No change** (low)	55	53	49	57	71	53	61	237
No change** (high)	32	24	28	21	13	26	21	97
Improved** (low)	5	14	17	6	6	9	10	39
Deteriorated** (high)	8	9	6	16	10	12	8	42
Back problems while stooping								
Often* (low)	10	11	20	18	31	14	21	89
Now and then* (low)	32	27	28	37	36	28	36	158
Seldom* (high)	58	62	52	45	33	58	43	248
No change** (low)	28	22	31	34	49	28	39	138
No change** (high)	48	43	46	38	20	43	35	161
Improved** (low)	11	17	14	21	17	15	17	66
Deteriorated** (high)	13	18	9	7	14	14	9	50
Back pain frequency								
Often* (low)	13	8	14	20	34	16	19	89
Now and then* (low)	39	46	35	35	37	37	40	191
Seldom* (high)	48	46	51	45	29	47	41	217
No change** (low)	37	34	36	47	56	38	46	175
No change** (high)	35	31	43	32	20	36	28	133
Improved** (low)	14	23	12	7	17	15	15	61
Deteriorated** (high)	14	12	9	14	7	11	11	46

*Cross-sectional study **Follow-up study

4.2.4 Assessment of leisure-time physical activity (IV, V)

LTPA was assessed using a self-administered questionnaire (Oja et al. 1994, Haapanen et al. 1996). Participation in leisure-time *exercise* primarily for keeping fit and healthy was asked in terms of frequency, intensity, duration and mode with average weekly time (in minutes). *Active transportation* in terms of daily walking distance (in kilometers) and cycling time (in minutes) was assessed, as well as the average weekly time (in minutes) spent in *other physical leisure-time activities* categorized as light, brisk or strenuous (V). The data of the LTPA questions were used to construct an overall physical activity level (inactive, moderately active, active) and exercise type (aerobic, muscular). Descriptive results are given in original article V. The continuity of physical activity since leaving school (yes/no) was also questioned. The inactive group was the reference group in all the statistical analyses.

4.3 Assessment of reliability

Inter-rater reliability and test-retest reproducibility was studied by administering the test battery twice to a small sample (n=42) of volunteers (I). The interval between the sessions was from 6 to 8 days. Two trained testers evaluated the subjects during 2 testing days according to a preplanned protocol. In addition, the inter-rater reliability between 3 pairs of testers of the one-leg balance test was reassessed with 48 randomly selected subjects in the cross-sectional population study. Detailed information is given in original report I.

4.4 Assessment of safety and feasibility

An assessment of safety and feasibility was conducted during the cross-sectional population study. In addition, a safety model for non-physician fitness testing was developed. Detailed information is given in original report II.

4.4.1 Assessment of safety (II)

The testers recorded all acute musculoskeletal injuries and symptoms, and also any cardiovascular complications during the cross-sectional fitness testing. DOMS was assessed with a questionnaire that was completed 4 to 6 days after the testing. Cardiovascular exertion in the fitness tests was evaluated by recording the heart rate immediately after each test, and it was expressed as the percentage of age-predicted maximum heart rate (%HR_{max}) calculated according to Arstila et al. (1984). The subjects wore a heart rate monitor (Polar Sport Tester, Polar Electro, Kempele, Finland) continuously during the cross-sectional fitness testing.

4.4.2 Assessment of feasibility (II)

The subject exclusion rate from each fitness test due to health limitations and the reasons for the test interruption or refusal were recorded. In addition, 3 testers evaluated the average time required to perform the complete test battery and rated the feasibility of each musculoskeletal and motor test on a 5-point scale on the basis of the time required to prepare, administer and score the test.

4.5 Assessment of validity: concept and hypothesis

4.5.1 Concept and variables (III, IV, V)

The conceptual model on physical activity, fitness and health (see Figure 1), presented by Bouchard and Shephard (1994), was the basis for studying the content and predictive validity of the proposed HRF test battery. Accordingly, the associations between single fitness items (univariate analysis) and self-rated perceived health, mobility function, and back functioning and pain were studied first in a cross-sectional design (III). Later, new univariate and multivariate analysis on the associations of both fitness and physical activity with self-rated health were conducted. Finally, the predictive value of baseline fitness and LTPA with the 3-year changes in self-rated perceived health, mobility function, and back functioning were assessed (IV).

The later cross-sectional (content validity) and follow-up (predictive validity) analyses were conducted in a similar manner in order to evaluate the results more systematically. These results are presented as the main findings on health-related validity. In addition, cross-sectional associations of HRF with self-reported LTPA patterns were assessed (V). A summary of the variables included in the validity studies is presented in Table 6.

4.5.2 Hypotheses (III, IV, V)

The main hypothesis was that the effects of regular physical activity on health are mediated through physiological responses that affect specific components of fitness. More specifically, the hypotheses concerning musculoskeletal and motor fitness with self-reported health (III, IV) were that baseline

(a) musculoskeletal fitness in lower extremity strength (one-leg squat) and power (vertical jump), motor fitness in one-leg balance, cardiorespiratory fitness in the Walk Test, and muscular-type exercise are associated with baseline (content validity) ability to climb stairs and its 3-year changes (predictive validity) and

(b) musculoskeletal fitness in trunk and upper-body functions (static trunk extension endurance, modified push-ups, trunk side-bending flexibility) and in lower-limb flexibility (hamstring muscle extensibility), motor fitness (one-leg balance), and muscular-type exercise are associated with baseline (content validity) back functioning and back pain and their 3-year changes (predictive validity).

The hypothesis concerning physical activity-related associations (V) was that baseline

(c) muscular exercise is the type of activity that is the most strongly associated with baseline musculoskeletal and motor fitness (content validity).

TABLE 6 Summary of the variables included in the cross-sectional and follow-up population studies evaluating the content validity of the proposed health-related fitness (HRF) test battery.

LEISURE-TIME PHYSICAL ACTIVITY - questionnaire	HEALTH-RELATED FITNESS - standard field test	SELF-RATED HEALTH - questionnaire
Overall leisure-time physical activity level - inactive - moderately active - active	MUSCULOSKELETAL *Muscular strength and endurance* - leg strength in one-leg squat - leg power in vertical jump - static trunk extension endurance - modified push-up strength	Perceived general health Mobility function - ability to climb stairs
Exercise type - aerobic - muscular	*Flexibility* - hamstring muscle extensibility - trunk side-bending flexibility	Back functioning & pain - general back functioning - frequency of back problems while stooping - frequency of back pain
Continuity of physical activity since school - no - yes	MOTOR *Static balance* - one-leg standing	
	CARDIOVASCULAR - Walk Test: -predicted maximal oxygen uptake -test time	
	MORPHOLOGICAL *Body composition* - body mass index	

4.6 Statistical analyses

All the descriptive results are presented as percentages, means with standard deviations, or ranges. In the *reliability study* (I), the ICC of the repeated interval scale measures was used as the measure of inter-rater reliability (Baumgartner 1989). A one-way analysis of variance was used for the ICC taking into account the score variability between the subjects and raters and the difference in the measurement levels between the raters. The degree of measurement error was expressed as the SEM (Baumgartner 1989, Roebroeck et al. 1993). For test-retest reproducibility the mean difference with the 95% confidence intervals (CI) (Altman & Gardner 1989, Roebroeck et al. 1993) and the coefficient of variation [(standard deviation/mean) x 100%] between the testing days were calculated.

Univariate and multivariate stepwise logistic regression analyses were conducted to assess the *content and predictive validity* of the proposed HRF test battery. Odds ratios (OR) with 95% CI values were used to estimate the strength of the associations of HRF and LTPA with health (III, IV), and the strength of the association of LTPA with HRF (V). When the 95% CI of the OR did not include 1.00, the result was considered *statistically significant* ($p<0.05$). The association was considered *positive* when the OR was higher than 1.00 and *negative* when the OR was less than 1.00. Age, sex (only when men and women were combined in the analysis), occupational physical activity, smoking, level of education, and marital status were included in all the models as possible confounders. All the HRF and LTPA variables that showed a statistical significance of $p \leq 0.15$ in the univariate analysis were included in the corresponding multivariate stepwise logistic regression analysis. A summary of the design and methods of the validity studies is presented in Table 7.

TABLE 7 Summary of the design and methods used to assess the content and predictive validity of the proposed health-related fitness (HRF) test battery.

Study design Predictor variables	Fitness categories	Outcome measure(s)	Type of LR analysis	Original paper
Cross-sectional HRF	Low-fit 20% Mid-fit 40% High-fit 40%	Poor & good perceived health and musculoskeletal functioning	Univariate M/W	III
Cross-sectional HRF & LTPA	Low-fit 40% Mid-fit 20% High-fit 40%	Poor & good perceived health and musculoskeletal functioning	Univariate & multivariate M+W	
Follow-up HRF & LTPA	Low-fit 40% Mid-fit 20% High-fit 40%	Improved & deteriorated perceived health and musculoskeletal functioning	Univariate & multivariate M+W	IV
Cross-sectional LTPA	Low-fit 40% High-fit 60%	Low-fitness in HRF	Univariate M/W	V

LR=logistic regression, LTPA=leisure-time physical activity, M/W= analyses conducted separately for the men and women, M+W=men and women combined in the analyses

5 RESULTS

5.1 Reliability

Descriptive data for the reliability assessments of the fitness tests is presented in original article I. The results of the reliability assessments are presented in Table 8.

TABLE 8 Inter-rater and test-retest reliability of the musculoskeletal and motor fitness tests.

FITNESS COMPONENT *Fitness factor* Test item	Inter-rater reliability			Test-retest reliability		
	N	ICC	SEM	Mean of two tests	CV (%)	Mean diff. (95%CI)
MUSCULOSKELETAL						
Muscular strength and endurance						
Strength in one-leg squat (points)	20	0.86	0.9	8.2	12.1	-0.1(-0.7 to 0.5)
Power in vertical jump (cm)	20	0.98	3.0	32.2	2.4	1.7(-0.2 to 3.6)
Static trunk endurance (time)		NA		NA		
Modified push-ups (repetitions)	19	0.88	2.6	12.1	0.6	3.0 (2.1 to 3.9)
Flexibility						
Hamstring muscle extensibility		NA		NA		
Side-bending (cm)						
Right	40	0.90	1.6	21.6	4.7	-0.5(-1.3 to 0.3)
Left	39	0.90	1.7	21.1	6.2	-0.5(-1.4 to 0.5)
Average	39	0.92	1.4	21.3	4.7	-0.5(-1.3 to 0.3)
MOTOR						
Static balance	40	0.76	13.3	43.7	5.0	3.7(-2.2 to 9.6)
One-leg standing (time)	48*	1.00	0.3	44.1	NA	NA

*Reassesessment in the cross-sectional population study, 2 testers took measurements from 1 performance simultaneously. CI=confidence interval, CV=coefficient of variation, diff.= difference, ICC=intraclass correlation coefficient, N=number of subjects, NA=not assessed in this study, SEM=standard error of measurement

The *inter-rater ICC values* were high (ICC ≥ 0.90) for the vertical jump and trunk side-bending, good (ICC ≥ 0.80 and < 0.90) for the modified push-ups and one-leg squat, and fair (ICC ≥ 0.70 and < 0.80) for the one-leg standing in the reliability study. High values were obtained for the one-leg balance test when it was reassessed in the cross-sectional study (ICC=1.0). The SEM in relation to the mean of 2 tests was lowest for the average side-bending (7%) followed by vertical jump (9%), one leg squat (11%), modified push-ups (21%) and one-leg standing (30%). The SEM in relation to the mean of 2 simultaneous test recordings of one-leg balance was very low (1%) in the cross-sectional study.

The *test-retest reproducibility* over 1 week, as measured with the coefficient of variation, ranged from 0.6% to 12.1% (see Table 8). Small test-retest mean differences with narrow confidence intervals compared with the mean of 2 tests were obtained for vertical jump, one-leg squat and average side-bending; a somewhat larger variability was found for the one-leg balance test. The results of the modified push-up test improved from the first measurement day to the second. No other systematic changes were observed in the test results between the first and second measurement day.

5.2 Safety and feasibility

5.2.1 Safety (II)

Acute health problems. No major complications occurred during the testing. Two subjects interrupted the modified push-ups due to back pain and 2 due to arm pain. The tester interrupted the static trunk extension endurance test of 2 subjects with a history of elevated blood pressure because their heart rate increased dramatically during the test. Three subjects interrupted the Walk Test due to pain in their lower limbs and 1 due to symptoms of flu.

Delayed onset of muscle soreness (DOMS). The response rate of the DOMS questionnaire was 95% (n=477). Sixteen men and 24 women indicated that DOMS was severe. For 16 men and 28 women the DOMS caused difficulties in daily activities, especially in stair climbing, squatting and walking. Most (n=30) of the subjects experiencing severe DOMS reported that the location of pain was in the thigh and gluteal muscles, and most of them assumed that the one-leg squat test was the cause. Severe DOMS is presented in Figure 5 by age, sex and LTPA intensity. Different trends were found for the men and women, as described in original article II.

Cardiovascular exertion. In general, the range of heart rates was large after all the tests. The mean $\%HR_{max}$ after each musculoskeletal and motor test, presented in Figure 6, did not differ more than 5% between the age groups. The highest levels were recorded after the cardiorespiratory Walk Test, followed by the modified push-up and static trunk extension muscular endurance tests. Heart rates higher than 85% of the predicted maximum were detected in 43% of the men and in 37% of the women after the Walk Test, in 19% of the men and 24% of the women after modified push-ups, in 2% of the men and 4% of the women after the trunk extension endurance test, and in 1% of the women after the one-leg squat test. In all the other tests the mean $\%HR_{max}$ was ≤60%, and no one reach the 85% level.

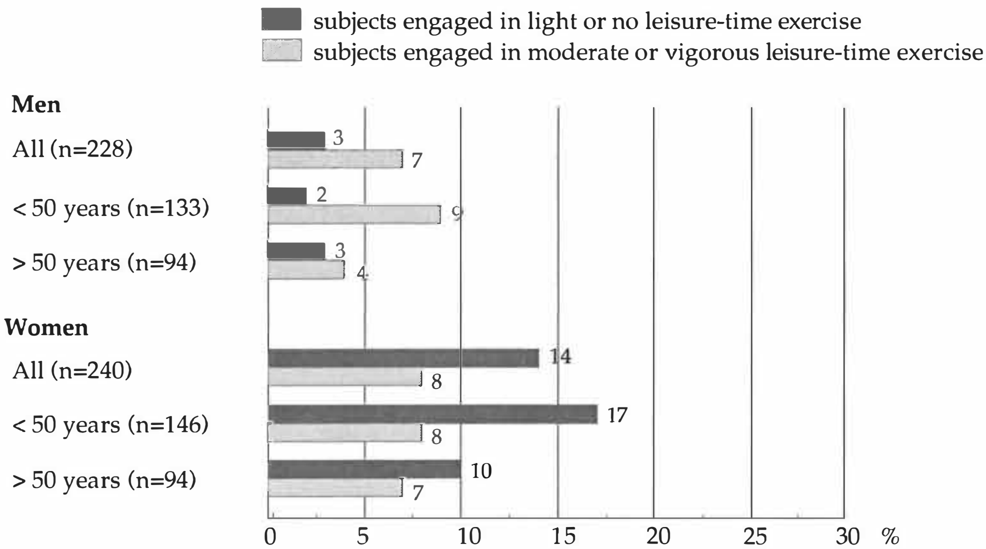


FIGURE 5 Reported severe delayed onset of muscular soreness by habitual exercise intensity.

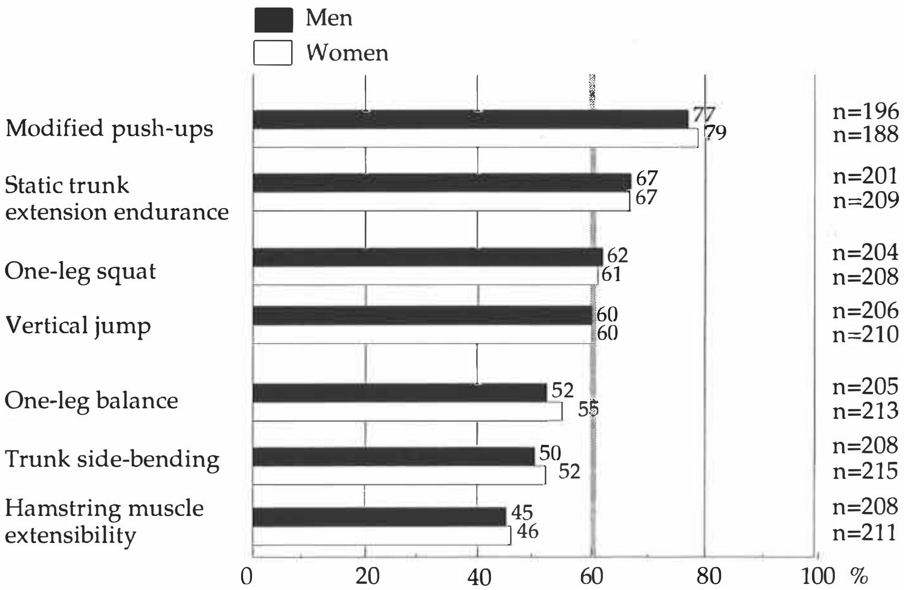


FIGURE 6 Mean percentages of the age-predicted maximum heart rate after each musculoskeletal and motor fitness test. The 60% line indicates the upper exertion level that is considered safe for most apparently healthy adults.

A safety model for HRF assessment conducted by non-physician testing personnel was introduced on the basis of the acceptable results on safety. The model is presented in Figure 7. The model includes (a) standard screening for health limitations, (b) standard instructions to refer subjects with severe health limitations to a physician for further examination, and (c) standard instructions to exclude subjects with minor health limitations from selected fitness tests. In the present study 53% of the subjects (n=500) had no health limitations to fitness testing, the fitness testers excluded 45% from selected tests, and 2% were referred to a physician for a more thorough health examination.

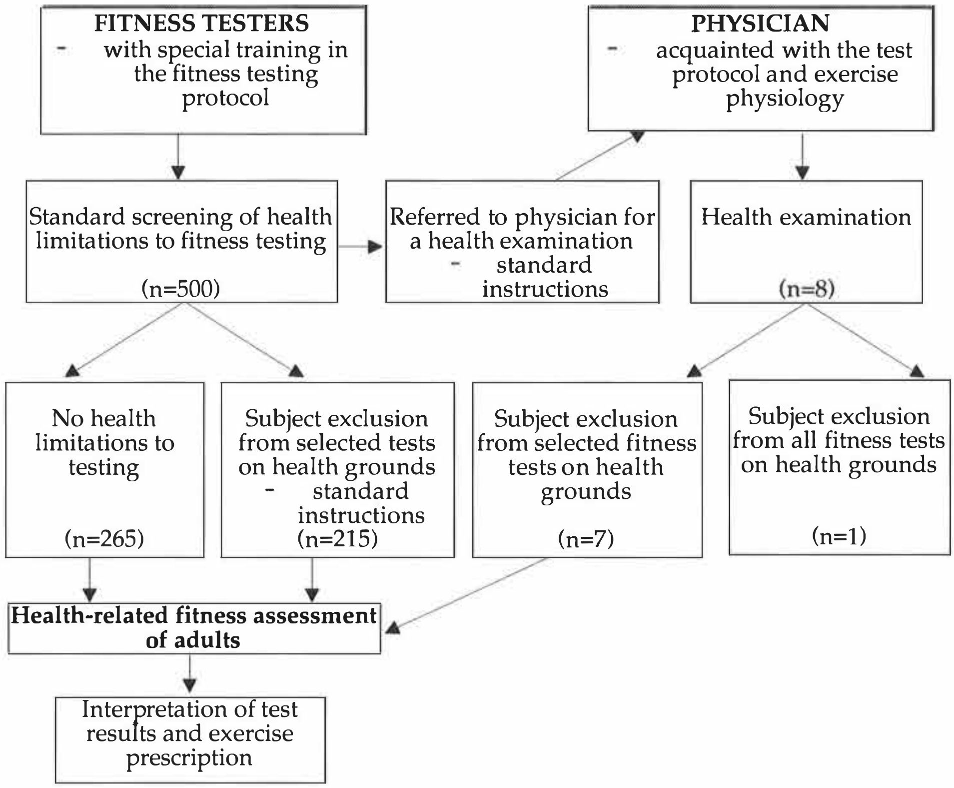


FIGURE 7 Safety model of the health-related fitness assessment of adults. The number of subjects (n) corresponds to the cross-sectional study procedure.

5.2.2 Feasibility (II)

Subject exclusion and limitations to fitness testing. The physician excluded 1 subject from all the tests; all the others were able to participate in selected tests. The overall percentages of the subjects in each age group excluded from, interrupting, or refusing one or more of the tests are presented in Figure 8, and the test-specific exclusion percentages are shown in Figure 9. The overall exclusion rate increased with age. The test-specific exclusion rates varied. The greatest number of subjects was excluded from the modified push-up (n=60) and static trunk extension endurance (n=50) tests. Heart disease and high blood pressure were the main reasons, followed by low-back disorders.

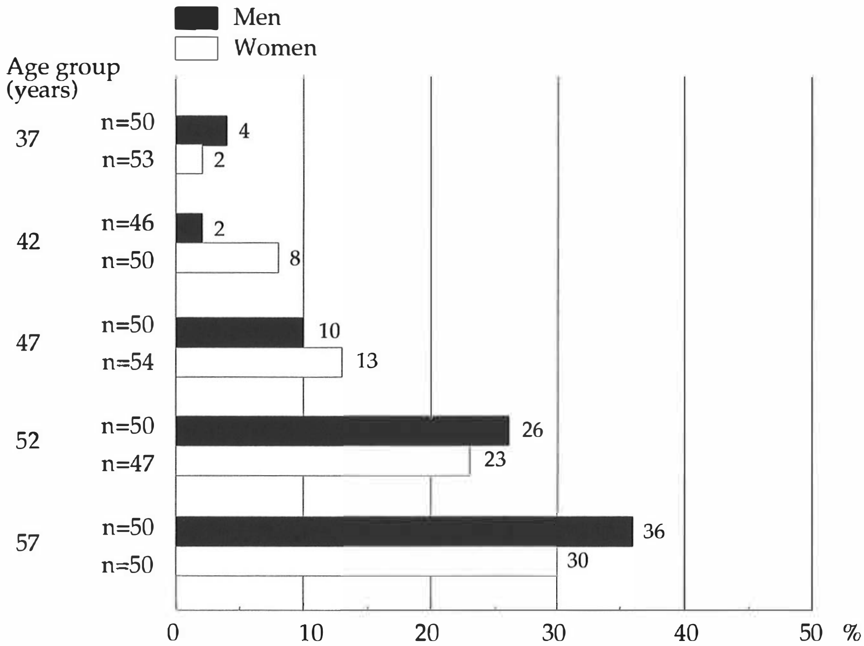


FIGURE 8 Overall percentage of subjects excluded from, interrupting or refusing to participate in one or more of the health-related fitness tests.

No more than 10% of the subjects in any age group were excluded from the vertical jump and one-leg squat tests. The main limitations were low-back and lower limb pain and severe heart disease with respect to the squat test. Less than 5% of the subjects were excluded from the one-leg standing balance, trunk side-bending flexibility, and hamstring muscle extensibility tests. Severe dizziness limited the balance testing, and musculoskeletal problems limited the flexibility testing. Over 95% of the subjects (n=481) completed the Walk Test; however, the VO_{2max} could not be predicted for 83 of them due to medication affecting heart rate.

Feasibility in terms of time requirements. The average time needed to perform all the musculoskeletal and motor fitness tests was 40 to 45 minutes, the Walk Test took another 20 to 25 minutes. The fitness testers rated the vertical jump as the most feasible (mean 5.0 points) of the musculoskeletal and motor tests, and the knee extension ROM measurements of the hamstring muscle extensibility test was as the least feasible (mean 2.7 points). The mean scores for the other tests ranged from 3.7 to 4.7 points.

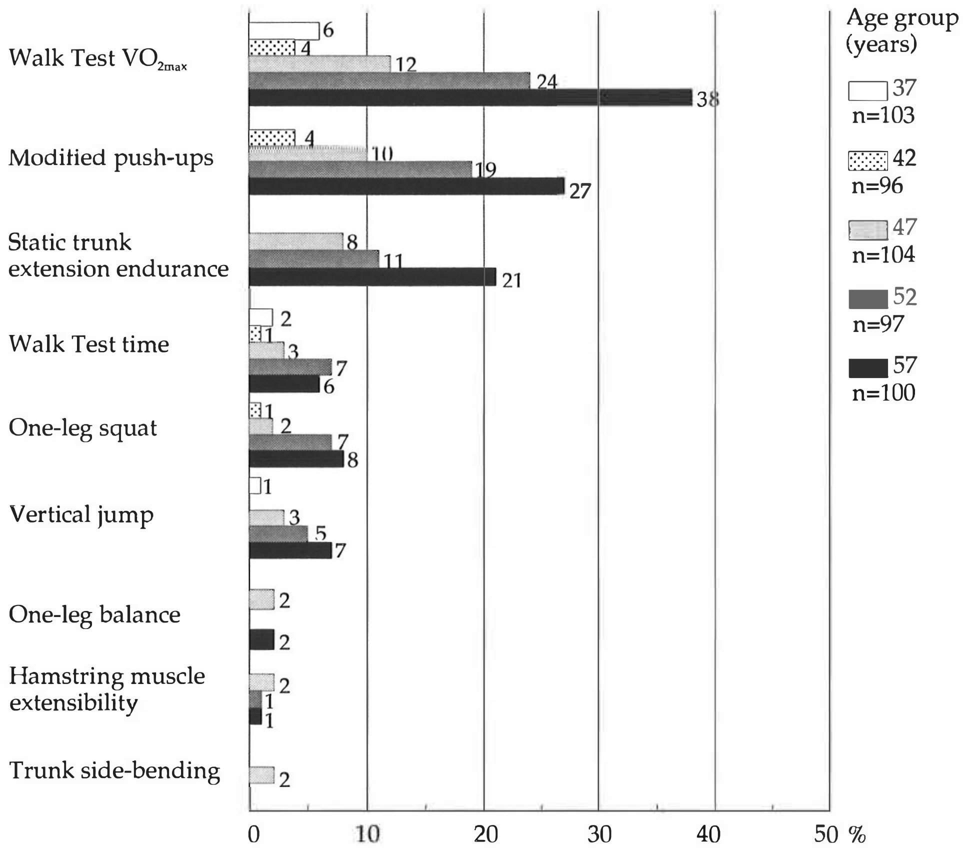


FIGURE 9 Test-specific exclusion rates for each health-related fitness test.

5.3 Validity of the proposed health-related fitness test battery

Content and predictive validity was evaluated on the basis of the hypotheses described in section 4.5.2. A summary of the study design and methods was presented in section 4.6 in Table 7.

5.3.1 Associations of physical fitness and activity with self-rated perceived health and its 3-year changes

Cross-sectional associations. In the univariate logistic regression analyses, all the fitness variables (see Table 6) were positively (OR>1.00) associated with *good perceived health* when the high-fitness group was compared with the low-fitness group. The highest OR values were found for muscular-type exercise (OR 7.62), high overall LTPA (4.56), high cardiorespiratory fitness in predicted VO_{2max}, (4.86) and the Walk Test time (3.73). The results of the multivariate stepwise logistic regression analysis for *good perceived health* are presented in Table 9.

In the univariate analysis, all the fitness variables, except one-leg balance, were negatively (OR<1.00) associated with *poor perceived health* when the high-fitness group was compared with the low-fitness group. The lowest OR values were found for the mid-level of musculoskeletal fitness in trunk extension endurance (OR 0.11) and for high-fitness in the modified push-up test (0.25), high cardiorespiratory fitness according to the Walk Test time (0.14) and predicted VO_{2max}, (0.22), muscular-type exercise (0.15), and a high level of overall LTPA (0.24). The results of the multivariate analysis for *poor perceived health* are presented in Table 9.

When the data of the men and women were analyzed separately (III), one-leg standing was not associated with perceived health among the men; vertical jump was not associated with perceived health among the women.

Prospective associations (IV). In the univariate analyses muscular-type exercise (OR 3.95), high overall level of LTPA (3.00), high cardiorespiratory fitness according to predicted VO_{2max} (3.80) and the Walk Test time (2.66), and continuous LTPA since school (2.45) were positively associated with *improved perceived health*. The corresponding results of the multivariate analysis are presented in Table 9. For *deteriorated perceived health*, in the univariate analysis, high cardiorespiratory fitness according to VO_{2max} (0.21) and the Walk Test time (0.25) showed a negative association. The variable selected in the multivariate analysis, presented in Table 9, was predicted VO_{2max}.

5.3.2 Associations of physical fitness and activity with self-rated ability to climb stairs and its 3-year changes

Cross-sectional associations. In the univariate analyses all the fitness variables (see Table 7), except the flexibility measures, were positively (OR>1.00) associated with *no difficulty in stair climbing* when the high-fitness group was compared with the low-fitness group. The highest OR values were found for high cardiorespiratory fitness according to the Walk Test time (OR 6.03) and predicted VO_{2max} (4.50), high musculoskeletal fitness in the modified push-up (3.62) and trunk extension endurance (3.42) tests, high overall level of LTPA (3.57), muscular exercise (3.42), and low BMI (3.28). When the data of the men and women were analyzed separately (III), vertical jumps were associated with no difficulty in stair climbing only among the men, and one-leg squats and BMI only among the women. The results of the multivariate stepwise logistic regression analysis for *no difficulty in stair climbing* are presented in Table 10.

Prospective associations (IV). In the univariate analyses high cardiorespiratory fitness according to the Walk Test time (OR 8.57), muscular-type exercise (8.36), moderate overall level of LTPA (5.41), and high musculoskeletal fitness in trunk side-bending flexibility (3.52) were positively associated with *improved ability to climb stairs*. The results of the corresponding multivariate analyses are presented in Table 10. For overall LTPA the "moderately active" group, but not the "active" group, had an increased OR when compared with the "inactive" group.

For *deteriorated ability to climb stairs*, in the univariate analyses, high musculoskeletal fitness for the one-leg squat test (0.29) and mid-fitness for the modified push-up test (0.27), high overall level of LTPA (0.32), and high cardiorespiratory fitness in predicted VO_{2max} (0.39) showed a negative association.

TABLE 9 Associations of baseline health-related fitness (HRF) and leisure-time physical activity (LTPA) with baseline perceived health and its 3-year changes [multiple stepwise logistic regression analysis adjusted for sex, age, marital status, education level, smoking and occupational physical activity: odds ratios (OR) with 95% confidence intervals (CI)].

Baseline HRF & LTPA variables selected into the multivariate stepwise logistic regression models		CROSS-SECTIONAL ASSOCIATIONS: Good perceived health (n=389) OR	(95% CI)	Poor perceived health (n=422) OR	(95% CI)	3-YEAR FOLLOW-UP ASSOCIATIONS: Improved perceived health (poor at baseline) (n=243) OR	(95% CI)	Deteriorated health (good at baseline) (n=89) OR	(95% CI)
HRF	Categories		*p-value*		*p-value*		*p-value*		*p-value*
• One-leg squat	Low 40%	1.00	*0.001*						
	High 60%	3.57	(1.55 to 8.19)						
• Static trunk extension endurance	Low 40%	1.00	0.027	1.00	*0.001*				
	Mid 20%	1.57	(0.71 to 3.47)	0.14	(0.04 to 0.49)				
	High 40%	2.35	(1.24 to 4.44)	0.75	(0.37 to 1.52)				
• Walk Test - Predicted VO_{2max}	Low 40%	1.00	*0.001*			1.00	*0.002*	1.00	*0.017*
	Mid 20%	1.11	(0.49 to 2.47)			2.25	(0.80 to 6.34)	0.90	(0.19 to 4.31)
	High 40%	2.90*	(1.51 to 5.59)			3.47	(1.39 to 8.70)	0.21	(0.06 to 0.81)
- Test time	Low 40%			1.00	*0.002*				
	Mid 20%			0.51	(0.22 to 1.16)				
	High 40%			0.23	(0.10 to 0.56)				
LTPA									
• Type of exercise	Inactive	1.00	*<0.001*	1.00	*0.015*				
	Aerobic	0.79	(0.37 to 1.71)	0.55	(0.27 to 1.13)				
	Muscular	3.34**	(1.52 to 7.33)	0.21	(0.06 to 0.68)				
• Continuity of LTPA since school	Non-contin.					1.00	*0.032*		
	Contin.					2.30	(1.07 to 4.94)		

*Differs significantly also from mid-fitness group. **Differs significantly also from aerobic exercise group.
contin.=continuous, VO_{2max}=maximal oxygen uptake

TABLE 10 Associations of baseline health-related fitness (HRF) and leisure-time physical activity (LTPA) with baseline mobility (ability to climb several flights of stairs) and 3-year changes in mobility [multiple stepwise logistic regression analysis adjusted for sex, age, marital status, education level, smoking and occupational physical activity: odds ratios (OR) with 95% confidence intervals (CI)].

Baseline HRF & LTPA variables selected into the multivariate stepwise logistic regression models		CROSS-SECTIONAL ASSOCIATIONS		3-YEAR FOLLOW-UP ASSOCIATIONS			
		No difficulty in climbing stairs (n=422)		Improved ability to climb stairs (some difficulty at baseline) (n=115)		Deteriorated ability to climb stairs (no difficulty at baseline) (n=316)	
		OR	(95% CI)	OR	(95% CI)	OR	(95% CI)
HRF	Categories		*p-value*		*p-value*		*p-value*
• One-leg squat	Low 40%	1.00	*0.022*			1.00	*0.004*
	High 60%	1.96	(1.10 to 3.47)			0.29	(0.13 to 0.65)
• Static trunk extension endurance	Low 40%	1.00	*0.016*				
	Mid 20%	1.93	(0.96 to 3.91)				
	High 40%	2.34	(1.25 to 4.36)				
• Walk Test - Test time	Low 40%	1.00	*<0.001*	1.00	*0.010*		
	Mid 20%	2.98	(1.46 to 6.08)	1.37	(0.41 to 4.63)		
	High 40%	4.48	(2.54 to 9.22)	6.06	(1.77 to 20.8)		
LTPA	Inactive			1.00	*0.030*		
• Overall LTPA	Mod. active			3.81	(1.25 to 11.6)		
	Active			1.35	(0.37 to 4.99)		

Mod.=moderately

The only selected variable in the corresponding multivariate stepwise logistic regression analysis, presented in Table 10, was the one-leg squat test.

5.3.3 Associations of physical fitness and activity with self-rated back functioning and pain

High cardiorespiratory fitness according to the predicted VO_{2max} (OR 2.68) and mid-fitness according to the Walk Test time (1.91), high musculoskeletal fitness in trunk extensor endurance (2.56), muscular-type exercise (2.42), high overall level of LTPA (2.20), and low BMI (2.01) showed positive (OR<1.00) associations with *good general back functioning* when the high-fitness groups was compared with the low-fitness or inactive group in the univariate analyses. The corresponding results of the multivariate stepwise logistic regression analysis are presented in Table 11.

In the univariate analyses for *good back functioning while stooping*, the mid-level of cardiorespiratory fitness according to the Walk Test time (OR 2.20) and high musculoskeletal fitness in the trunk extension endurance test (2.02) showed positive associations. The results of the corresponding multivariate analysis are presented in Table 11.

In the univariate analyses for *seldom back pain*, high and mid levels of musculoskeletal fitness in trunk extension endurance (OR 2.39 and 2.02, respectively) and hamstring muscle extensibility (1.64 and 2.18) were positively associated when compared with low-fitness. The results of the corresponding multivariate analysis are presented in Table 11.

For *poor general back functioning*, in the univariate analyses, high-fitness in all the musculoskeletal tests (range of OR 0.30-0.53), except vertical jump, showed a negative (OR<1.00) association, as did high motor fitness in one-leg balance (OR 0.48), muscular-type exercise (0.21), high overall level of LTPA (0.39), and continuous LTPA since school (0.48). The results of the corresponding multivariate analysis are presented in Table 12.

In the univariate analyses for *poor back function while stooping* high-fitness in all the test variables, except leg strength (one-leg squat) and leg power (vertical jump), showed a negative association. The OR values were the lowest for a mid-level of musculoskeletal fitness in hamstring muscle extensibility (OR 0.28) and high-fitness in modified push-up test (0.28), low BMI (0.29), mid- (0.36) and high- (0.40) levels of cardiorespiratory fitness according to predicted $VO2_{max}$, and muscular-type exercise (0.39). The results of the multivariate analysis for *poor back function while stooping* are presented in Table 12.

High musculoskeletal fitness in back extension endurance (OR 0.30), the modified push-up test (0.40) and the one-leg squat test (0.45), high motor fitness in one-leg balance (0.44), low BMI (0.40), muscular-type exercise (0.43), and overall level of LTPA (0.43) were negatively associated with *frequent back pain* in the univariate analyses. The results of the corresponding multivariate analysis are presented in Table 12.

When the data of the men and women were analyzed separately (III), one-leg balance and BMI were associated with back functioning and pain among the women but not among the men.

TABLE 11 Associations of baseline health-related fitness (HRF) and leisure-time physical activity (LTPA) with good baseline back function and pain and with their positive 3-year changes [multiple stepwise logistic regression analysis adjusted for sex, age, marital status, education level, smoking and occupational physical activity: odds ratios (OR) with 95% confidence intervals (CI)].

Baseline HRF & LTPA variables selected into the multivariate stepwise logistic regression models		CROSS-SECTIONAL ASSOCIATIONS						3-YEAR FOLLOW-UP ASSOCIATIONS			
		Good general back function (n=406)		Seldom problems while stooping (n=428)		Seldom back pain (n=450)		Improved general back function (n=242)		Improved back function while stooping (n=190)	
		OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)
HRF	Categories		*p-value*		*p-value*		*p-value*		*p-value*		*p-value*
• One-leg squat	Low 40%					1.00	*0.003*				
	High 60%					0.50	(0.31 to 0.80)				
• Static trunk extension endurance	Low 40%	1.00	*0.023*	1.00	*0.009*	1.00	*<0.001*				
	Mid 20%	1.44	(0.79 to 2.62)	1.31	(0.76 to 2.25)	2.00	(1.16 to 3.43)				
	High 40%	2.03	(1.22 to 3.37)	2.04	(1.29 to 3.24)	2.66	(1.69 to 4.20)				
• Modified push-ups	Low 40%							1.00	*0.009*		
	Mid 20%							4.13	(1.57 to 10.8)		
	High 40%							1.39	(0.54 to 3.57)		
• Hamstring muscle extensibility	Low 40%					1.00	*0.001*			1.00	*0.002*
	Mid 20%					2.64	(1.52 to 4.60)			4.25**	(1.82 to 9.97)
	High 40%					1.77	(1.12 to 2.79)			1.33	(0.62 to 2.83)
• Trunk side-bending flexibility	Low 40%							1.00	*0.028*		
	Mid 20%							4.09	(1.46 to 11.5)		
	High 40%							1.73	(0.70 to 4.24)		
• Walk Test, - Predicted VO_{2max}	low 40%	1.00	*0.001*								
	mid 20%	0.98	(0.52 to 1.81)								
	high 40%	2.31*	(1.40 to 3.81)								
- Test time	low 40%			1.00	0.041						
	mid 20%			2.03	(1.16 to 3.56)						
	high 40%			1.19	(0.75 to 1.88)						

*Differs significantly also from the mid-fitness group. **Differs significantly also from the high-fitness group. VO_{2max}=maximal oxygen uptake

TABLE 12 Associations of baseline health-related fitness (HRF) and leisure-time physical activity (LTPA) with poor baseline back function and pain and with their negative 3-year changes [multiple stepwise logistic regression analysis adjusted for sex, age, marital status, education level, smoking and occupational physical activity: odds ratios (OR) with 95% confidence intervals (CI)].

Baseline HRF & LTPA variables selected into the multivariate stepwise logistic regression models		CROSS-SECTIONAL ASSOCIATIONS: Poor general back function (n=432) OR (95% CI)	CROSS-SECTIONAL ASSOCIATIONS: Frequent problems while stooping (n=393) OR (95% CI)	CROSS-SECTIONAL ASSOCIATIONS: Frequent back pain (n=429) OR (95% CI)	3-YEAR FOLLOW-UP ASSOCIATIONS: Deteriorated general back function (n=137) OR (95% CI)	3-YEAR FOLLOW-UP ASSOCIATIONS: Deteriorated back function while stooping (n=205) OR (95% CI)
HRF	Categories	*p-value*	*p-value*	*p-value*	*p-value*	*p-value*
• Trunk extension endurance	Low 40%			1.00 *0.002*		
	Mid 20%			0.96 (0.48 to 1.90)		
	High 40%			0.32*(0.15 to 0.66)		
• Modified push-ups	Low 40%	1.00 *0.003*	1.00 *0.012*			
	Mid 20%	0.53 (0.26 to 1.06)	1.06 (0.51 to 2.22)			
	High 40%	0.34 (0.17 to 0.66)	0.37 (0.17 to 0.79)			
• Hamstring muscle extensibility	Low 40%		1.00 *0.005*			
	Mid 20%		0.22 (0.08 to 0.61)			
	High 40%		0.61 (0.31 to 1.18)			
• Trunk side-bending flexibility	Low 40%				1.00 *0.019*	
	Mid 20%				0.29 (0.10 to 0.86)	
	High 40%				0.31 (0.12 to 0.76)	
• One-leg standing balance	Low 40%	1.00 *0.026*	1.00 *0.009*	1.00 *0.026*		
	High 60%	0.53 (0.31 to 0.93)	0.45 (0.25 to 0.82)	0.52 (0.29 to 0.92)		
LTPA	Inactive				1.00 *0.035*	
• Overall LTPA	Mod. active				0.58**(0.18 to 1.86)	
	Active				1.95 (0.62 to 6.17)	
• Continuity of LTPA since school	Non-contin.	1.00 *0.020*				1.00 *0.011*
	Contin.	0.50 (0.28 to 0.91)				2.40 (1.21 to 4.77)

*Differs significantly from the mid-fitness group. **Differs significantly from the high-fitness group. contin.=continuous, Mod.=moderately

5.3.4 Associations of physical fitness and activity with 3-year changes in self-rated back functioning and pain

In the univariate analyses *for improved general back functioning*, mid-level of musculoskeletal fitness in the modified push-up (OR 4.32) and trunk side-bending (3.48) tests, and high-fitness in hamstring muscle extensibility (2.85), trunk extension endurance (2.62), and low BMI (2.28) showed positive (OR>1.00) association when compared with low-fitness. The results of the corresponding multivariate analysis are presented in Table 11.

Mid-level of musculoskeletal fitness in hamstring muscle extensibility (OR 3.61) and high cardiorespiratory fitness according to the Walk Test time (2.34) were positively associated with *improved back function while stooping* in the univariate analyses. In the corresponding multivariate analysis, presented in Table 11, the only selected variable was hamstring muscle extensibility.

For *deteriorated general back functioning*, a high-level of musuloskeletal fitness in trunk side-bending (0.34) and high- (0.36) and mid- (0.36) levels of hamstring muscle extensibility were negatively (OR<1.00) associated in the univariate analyses when compared with low-fitness. The results of the corresponding multivariate analysis are presented in Table 12.

In the univariate analyses for *deteriorated back function while stooping* continuous LTPA since school (OR 2.29) was the only variable showing a significant association, and it was the only variable selected in the multivariate analysis, as presented in Table 12.

There were no associations of HRF and LTPA with 3-year *changes in back pain frequency*.

5.3.5 Leisure-time physical activity patterns in relation to health-related fitness

Patterns of leisure-time physical activity. Of the men 30% and of the women 35% were physically "active" 3 or more times a week. Of the men 46% and of the women 43% were "moderately active"(i.e., engaged in LTPA 1 to 2 times a week). Of the men 24% and of the women 23 % were categorized as physically "inactive". Seventeen percent of the subjects participated only in "aerobic exercise", whereas 24% of the men and 25% of the women participated in "muscular exercise". Of the muscular exercise group 74% were also engaged in some aerobic exercise. Detailed information on the LTPA patterns of this adult population has been given in original article V.

Cross-sectional a*ssociations between baseline physical activity and health-related fitness* (V). The results of the analyses studying the associations between LTPA and HRF are presented in Table 13. LTPA was not systematically associated with *lower limb musculoskeletal fitness* in muscular strength (one-leg squat), power (vertical jump), or flexibility (hamstring extensibility, trunk side-bending) for either sex. LTPA was systematically associated with *musculoskeletal fitness* in respect to modified push-up strength for both sexes, and there were fewer men and women with low trunk extension endurance among the active subjects with respect to overall LTPA and muscular exercise, but not among the aerobic exercisers, when they were compared with the inactive subjects. There were fewer women with low *motor fitness* in the one-leg balance test among the active subjects with respect to overall LTPA and muscular exercise, but not among the aerobic exercisers, when they were compared with the inactive subjects. No associations between LTPA and balance were found for the men. LTPA was the most strongly and systematically associated with *cardiorespiratory fitness* in terms of the Walk Test results. LTPA was systematically associated with *body composition* in terms of the BMI for the women but not for men.

TABLE 13 Associations between leisure-time physical activity (LTPA) and health-related fitness (HRF): Odds ratios (OR) of low-fitness among active subjects with reference to inactive subjects adjusted for age, marital status, educational level, smoking and occupational physical activity.

HRF COMPONENT Test item, lowest 40%	LEISURE-TIME PHYSICAL ACTIVITY GROUPS Sex	Overall LTPA OR	 95% CI	Aerobic exercise OR	 95% CI	Muscular exercise OR	 95% CI
MUSCULOSKELETAL							
One-leg squat	M	0.61	0.26-1.42	0.32*	0.10-0.98	0.65	0.25-1.73
	W	0.77	0.34-1.73	1.07	0.39-2.93	0.52	0.21-1.30
Vertical jump	M	1.56	0.74-3.28	1.50	0.60-3.77	0.66	0.29-1.53
	W	0.49	0.23-1.02	0.39	0.15-1.01	0.50	0.23-1.09
Static trunk extension	M	0.29*	0.13-0.66	0.72	0.28-1.81	0.41*	0.18-0.93
	W	0.44*	0.20-0.93	0.71	0.29-1.71	0.29*	0.12-0.68
Modified push-up	M	0.43*	0.20-0.96	0.31*	0.11-0.85	0.28*	0.12-0.67
	W	0.35*	0.16-0.74	0.30*	0.12-0.76	0.24*	0.10-0.58
Hamstring extensibility	M	1.04	0.49-2.23	1.27	0.51-3.18	0.66	0.30-1.47
	W	0.66	0.33-1.34	0.53	0.22-1.29	0.55	0.25-1-17
Trunk side-bending	M	0.85	0.41-1.75	0.58	0.24-1.42	0.87	0.41-1.88
	W	1.30	0.61-2.74	1.46	0.57-3.77	0.88	0.39-1.97
MOTOR							
One leg standing	M	1.06	0.49-2.29	0.63	0.24-1.67	0.83	0.35-1.98
	W	0.40*	0.19-0.82	0.70	0.29-1.68	0.27*	0.12-0.60
CARDIORESPIRATORY							
Walk Test:	M	0.43*	0.19-0.97	0.44	0.16-1.21	0.38*	0.16-0.92
Predicted VO_{2max}	W	0.20*	0.09-0.46	0.18*	0.07-0.49	0.11*	0.04-0.29
Walk Test time	M	0.25*	0.11-0.53	0.17*	0.06-0.45	0.18*	0.08-0.41
	W	0.21*	0.10-0.46	0.12*	0.04-0.35	0.21*	0.09-0.46
MORPHOLOGICAL							
Body mass index	M	0.55	0.26-1.15	0.42	0.17-1.04	0.83	0.39-1.75
	W	0.32*	0.15-0.67	0.40*	0.16-0.98	0.28*	0.12-0.62

*Significantly different ($p<0.05$)
CI= confidence interval, M=men, W=women, VO_{2max}=maximal oxygen uptake

6 DISCUSSION

The reliability, safety, feasibility, content, and predictive validity of a proposed HRF test battery for middle-aged adults was evaluated with special interest in musculoskeletal and motor tests. The criteria for accepting a test into the battery were that the (a) measurement had to be objective and stable over time, (b) a majority of the middle-aged subjects could safely participate in the test, and (c) the test result could be associated with a person's current or future perceived health or musculoskeletal functioning.

Furthermore, the relationship between current LTPA patterns and the fitness test results were evaluated. However, these results were not used as criteria for accepting a test as part of the battery due to the shortcomings in former population studies on the relations between LTPA and fitness (van Heuvelen et al. 1998). In most studies, the subjects have not been representative of the general population (Sandler et al. 1991, Knapik et al. 1993 and 1996), and the studies on fitness components other than cardiorespiratory one are underrepresented, especially those concerning balance and flexibility (Sandler et al. 1991, Viljanen et al. 1991, Era et al. 1992, Rantanen et al. 1992, Kujala et al. 1994, Knapik et al. 1996). In the future, an exercise intervention study is needed to evaluate the sensitivity of the proposed HRF tests to fitness changes, a final criterion for acceptance as a test for the battery.

6.1 Subjects

The subjects were initially drawn from a representative population, but were, due to multistage sampling (see Figure 3), selected to some extent. They had a higher education level, rated their health as better, and were physically more active than the non-participating part of the study sample. Nevertheless, there were subjects with health limitations (II, III), and the subjects reported a variety of physical activity patterns in terms of exercise frequency, intensity, duration, and mode (V).

The differences made it possible for the safety and feasibility to be investigated, along with the health-related validity of the test battery for middle-aged adults with certain limitations. The selection bias of the population may (a) raise the question of whether the test battery is safe and feasible for the less healthy and sedentary part of the middle-aged population. In addition, (b) the population estimates of fitness, if used as norm reference values, may be too high. The bias can be bigger for the tests and age groups from which a large proportion of the subjects were excluded. The fact that 71% of the men in the one-leg squat test (n=238) and 58% of the men in the one-leg balance test (n=243) were able to perform the respective test to the highest possible level limits the discriminatory power of the these tests (III, IV,V). Furthermore, (c) the variation in the study sample was decreased, and the contrast in different groups was lower than in the general population. This situation is likely to lead to an underestimation of the strength of the associations of HRF with health and LTPA (Blair et al. 1996).

6.2 Reliability of the proposed musculoskeletal and motor fitness tests

The degree of reliability necessary for a fitness test depends on the purpose of the measurement. In epidemiological-type studies aimed at establishing relations between fitness and other constructs, the fitness test results of selected population groups are typically compared or the test results are used to categorize subjects into different fitness levels. In these types of studies good relative reliability in terms of objectivity and stability is desirable.

In the present study, the ICC values describe the relative inter-rater reliability (the degree to which people maintain their position in a sample with repeated measurements by two testers). In this respect, all the musculoskeletal and motor fitness tests that were studied (see Table 8), except for the one-leg balance test in the first study, proved to have acceptable inter-rater reliability (ICC ≥ 0.88). Most of the former reliability studies, referred to in original article I, have used Pearson's correlation coefficient as the method for assessing relative reliability. The drawbacks of their use were discussed in section 2.4 (Atkinson & Nevill 1998, Lamb 1998, Rankin & Stokes 1998).

The mean differences of the 2 measurement sessions with the 95% CI (including the variation between days 1 and 2 and between the two testers in the present study) can be interpreted as the minimum difference between the test results that indicate a real change in the participants fitness level over time at the group level (I). It is of importance when the fitness level of population samples or other groups are being monitored. The test-retest reliability analysis also reveals systematic bias.

In the present study (see Table 8), mean test-retest difference was also acceptable (≤ 8%) for the vertical jump, one-leg squat, average side-bending and one-leg balance (reassessment) measures when compared with the mean of the two tests. The results of the modified push-up test improved significantly (25% compared with the mean). The absolute test-retest reliability of all the items was also acceptable as measured with the coefficients of variation (range from 0.6 to 12.1%).

In practice, fitness testers categorize subjects, make comparisons between persons, and monitor individual fitness changes over time. In individual fitness testing, the SEM value for inter-rater reliability can be interpreted as the minimum difference in the test results that indicates a real difference between the fitness levels of 2 subjects measured by different testers (I). It should be noted that the SEM covers about 68% of the variability as opposed to the 95% CI.

When the SEM values of the present study (see Table 8) are compared with the mean of 2 tests, average side-bending (7%), vertical jump (9%), one-leg squat (11%) and one-leg standing (1%), as reassessed, seem to have acceptable reliability. The SEM for the modified push-up test is difficult to interpret, since the test-retest analyses revealed a large systematic bias (3 repetitions), proposed to be caused by learning (I); however, the range of the 95% CI was only 1.8 repetitions.

An "analytical goal" for acceptable reliability in the present study could be that the SEM should not be larger than the range of the test results included in the "mid-fit"(mid 20%) category (based on the mean of age- and sex-specific cut points). The values for this range were as follows: 3.0 centimetres for vertical jump, 25 seconds for back extensor endurance, 1.6 repetitions for modified push-up, 6 degrees for hamstring muscle extensibility, and 1.6 centimetres for trunk side-bending. According to this analytical goal the vertical jump and trunk side-bending tests had an acceptable level of reliability (see Table 8). The one-leg squat and one-leg balance tests had only 2 fitness categories, and the proposed analytical goal was not relevant.

Two of the musculoskeletal fitness tests proposed for the HRF test battery were not included in the reliability study: static trunk extension endurance (due to safety problems in the pilot studies when conducted on a high couch) and hamstring muscle extensibility test (due to the complicated procedure). Former reliability studies assessing the ICC and SEM for static trunk extensor endurance (see section 2.3) have revealed varying results for different population groups (Moffroid et al. 1994, Moreland et al. 1997, Simmonds et al. 1998). As for the hamstring muscle extensibility test, high Pearson correlation coefficients (r=0.94-0.99) have been reported (Gajdosik & Lusin 1983, Kane & Bernasconi 1992). However, measurement errors of more than 6 degrees, the proposed analytical goal, are likely, and therefore some problems with the reliability exist.

6.3 Safety and feasibility of the proposed health-related fitness test battery

When HRF assessments are conducted to promote physical activity for health, major issues are how to ensure safe testing without (a) expensive referrals for medical evaluation or (b) excluding large numbers of subjects who would actually benefit from regular physical activity despite of some health problems.

In the present study a standard *health screening* procedure (II), applied to a middle-aged population, was successful in ensuring safe testing with minor physician participation and minor subject exclusion (see Figure 7). Only 1 subject was excluded from all the fitness tests due to severe health limitations revealed by the physician. All the others participated in one or more tests. Ninety percent or more of each age group was qualified to perform the balance, flexibility, muscular power (vertical jump), and strength (one leg squat) tests, and the Walk Test. The low exclusion rates are in accordance with those reported for the Allied Dunbar National Fitness Survey (1992), but they are somewhat lower than reported in an earlier population study in Finland (Mälkiä 1983). Of the subjects 31% were on medication affecting heart rate, and therefore the prediction of VO_{2max} was limited. These exclusion rates are similar to those reported in surveys assessing the cardiorespiratory fitness of adults through the use of submaximal tests (Mälkiä 1983, Shephard 1991, Laukkanen et al. 1992).

The major concerns related to the *occurrence of DOMS* were severe symptoms (see Figure 5) among inactive subjects and impaired function (II). Accordingly, inactive women were the most prone to severe DOMS; this result agrees with the findings that training acts in a preventive fashion to reduce muscular damage and soreness (MacIntyre et al. 1995). The one-leg squat test seemed to be the major cause for both severe DOMS and reported difficulties with stair climbing, squatting and walking. No other major problems were discovered. Despite the risk of DOMS, the inclusion of the leg strength test in the HRF test battery is warranted because it is an important indicator of the mobility and functional independence of older adults (Rantanen et al. 1994 and 1996, Ferrucci et al. 1997, Schroll et al. 1997).

Physiological *exertion in individual fitness tests* is an important factor affecting the safety of HRF testing. Heart rate is a good indicator of cardiovascular volume load in tests requiring dynamic movements of large muscles. In musculoskeletal and motor fitness testing involving isometric types of muscular work, heart rate reflects primarily the cardiovascular pressure load and was accepted as a relevant measure of safety-related exertion. However, the proposed risks of isometric types of muscular work have to be carefully considered before subjects are allowed to perform modified push-ups and static trunk extension endurance, both of which require a substantial amount of isometric muscular work and, in the present study, had the highest mean values of $\%HR_{max}$.

Accordingly, in the present study, a substantial proportion of the subjects over 50 years of age *was not qualified* for the modified push-up (22%) and trunk extension endurance (16%) tests due to elevated blood pressure and self-reported heart disease, although most of them (93%) were allowed to perform the Walk Test. Blood pressure measurements during muscular testing would provide further information on the physiological exertion and related cardiovascular risks (Arstila et al. 1984). Another, more accessible possibility during field testing would be the ratings of perceived exertion, which indicate how close the subject is to maximal exertion (Dishman 1994, ACSM 1995). Visual analog scales of pain and discomfort have also been used among back pain patients (Simmonds et al. 1998).

Most of the musculoskeletal and motor fitness tests were quick and easy to administer. The only test that required a considerable amount of time was the knee extension ROM measurement (hamstring muscle extensibility) with an inclinometer. (For details see original article II.) This time requirement may limit the use of the test in larger populations. The average time required to perform the musculoskeletal and motor fitness tests was 40 minutes.

6.4 Validity of the proposed health-related fitness test battery

6.4.1 Concept and design of the study

The content and predictive validity of the proposed HRF test battery was evaluated with respect to a theoretical model (see Figure 1) on physical activity, fitness and health (Bouchard & Shephard 1994). The cross-sectional health-related evaluations revealed how the current level of HRF was related to the current state (content validity) of self-rated perceived health, mobility function, and back functioning and pain. The prospective evaluations revealed whether the baseline HRF level had value in predicting the 3-year changes in the selected health outcomes. Furthermore, the cross-sectional evaluations made in relation to physical activity revealed how current LTPA patterns were associated with current HRF levels.

The main interets were the associations of objectively measured musculoskeletal and motor fitness test results with the self-rated level of mobility and back functioning. It has been proposed that the objective performance test can reflect a wide range of functional levels, including the positive aspects, whereas self-measures mainly reflect disability (Guralnik et al. 1994, Kivinen et al. 1998). The new International Classification of Function and Disability (WHO 1999), aimed at providing a scientific basis for understanding and studying functional states associated with health conditions, also states that positive aspects of activity are reflected by functioning and negative aspects by disability. Furthermore, this new process model of functioning and disability includes all the essential components of the Toronto model (see Figure 1) on physical activity, fitness and health.

The major limitation of the present study is that simple self-ratings of health were used as outcome measures to validate the field-based HRF test battery among middle-aged adults. The quality of the selected measures was discussed in Methods section 4.2.3. With respect to mobility and back functioning the other alternatives would have been a clinical examination (Viikari-Juntura et al. 1998) or use of established disease-specific questionnaires (Delitto 1994, Rejeski et al. 1995). However, these alternatives were not considered feasible due to the significant human resource and time requirements.

The present results of the health-related validity studies provide some new data on the role of physical fitness in musculoskeletal functioning among middle-aged adults. In general, the findings suggest that several factors of musculoskeletal and motor fitness, as well as aerobic fitness in terms of walking, are independent factors for self-reported mobility and back functioning. Similar findings on functional decline were recently reported by Huang et al. (1998) and Morey et al. (1998b).

A more advanced statistical model is needed to study further the possible mediating role of physical fitness between physical activity and health, as suggested by Rantanen et al. (1999a). Ultimately, randomized controlled trials on the effects of specific types of fitness training are needed to confirm the role of physical activity and HRF in the prevention of mobility and back-related disability. At present, results from randomized controlled trials aimed at prevention are rare (Gundewall et al. 1993, Lahad et al. 1994, Ettinger et al. 1997).

There are two other limitations of the study. First, the impact of genetic factors on the relationships between HRF and the selected health outcomes cannot be estimated (Bouchard & Pérusse 1994, Jimenenz & Dharmavaram 1994, Battié et al. 1995, Thomis et al. 1998), nor can that between LTPA and HRF (Bouchard & Preusse 1994, Launderdale et al. 1997, Simonen et al. 1998, Thomis et al. 1998). Second, the LTPA and HRF data of the study were collected at one point in time and the intra-individual changes in them during the follow-up were not identified (van Heuvelen et al. 1998). The use of this procedure could have affected the relationships of the baseline LTPA and HRF with the 3-year changes in health status.

6.4.2 Musculoskeletal fitness: strength and power of the lower extremities

The one-leg squat (strength) and vertical jump (power) tests were expected to have health-related validity for the ability to climb stairs. (See the hypotheses presented in section 4.5.2.) The *one-leg squat* test showed stronger health-related validity than the vertical jump test. High-fitness in leg strength was positively associated with good current mobility function in stair climbing, and it showed negative predictive validity for deteriorated ability to climb stairs (see Table 10). These findings are in agreement with the results of several cross-sectional (Avlund et al. 1994, Guralnik et al. 1994, Rantanen et al. 1996) and prospective (Hoeymans et al. 1994, Guranic et al. 1995, Schroll et al. 1997) studies on elderly populations. High-fitness in leg strength was also positively related to good perceived health (see Table 9), and this result agrees with the findings of Era et al. (1992). The contradictory associations of the one-leg squat test with back pain (see Table 11 and section 5.3.3) may be partly due to the ceiling effect of the test among the men (III).

Leg extensor power in the *vertical jump* test showed weak associations with current perceived health status. (See section 5.3.1.) In the analyses conducted separately for the men and women (III), low leg power was an indicator of poor perceived health and difficulties in stair climbing among the men but not among the women. Leg power had no predictive value for changes in self-rated health. In a former prospective study (Fujita et al. 1995) poor leg power among men, but not among women, was associated with an excess risk of death from cardiovascular disease and all causes.

The *physical-activity-related validity* (V) of lower-limb strength and power tests was poor (see Table 13). This finding disagrees with the results of former studies (Allied Dunbar National Fitness Survey 1992, Kujala et al. 1994). Physical activities of daily life (Loy et al. 1994) may have been sufficient to maintain muscular fitness of the lower limbs among the study population.

6.4.3 Musculoskeletal fitness: trunk and upper-body muscular endurance

The trunk and upper-body function tests were expected to have health-related validity for back functioning and pain. (See the hypotheses presented in section 4.5.2.) Accordingly, high-fitness in *trunk extension endurance* showed strong positive associations with good current back health (see Table 11). Several former studies agree with these findings (Holmström et al. 1992, Alaranta et al. 1994b, Barnekow-Bergkvist et al. 1998). In the present study, trunk endurance had no predictive value for 3-year changes in the frequency of low-back pain. This finding is in disagreement with former study results (Biering-Sørensen 1984, Luoto et al. 1995b). The difference in the outcome measures for back pain may be an explanation for the finding. The outcome in the former studies was first-time episode of back pain. Trunk extension endurance showed some predictive validity for improved general back functioning in the univariate analysis (OR 2.6, 95% CI 1.2 to 5.9), but it was not selected into the corresponding multivariate model. Unexpectedly, high trunk extension endurance was positively associated with good current perceived health and mobility status (see Tables 9 and 10). Trunk extensor strength, but not endurance, has earlier been associated with the mobility function of the elderly (Avlund et al. 1994).

Modified push-up assesses dynamic endurance strength of the upper body and the ability to stabilize the trunk. High-fitness according to push-up ability was negatively associated with poor current back health. (See Table 12 and section 5.3.3.) The results may indicate that subjects with low-back dysfunction had difficulties in stabilizing their trunks, which is a sign of weakened postural and motor control (Byl & Sinnot 1992, Luoto et al. 1996 and 1998, Hodges & Richardson 1996, Hides et al. 1996). This result agrees with the other findings of the present study on the associations of the one-leg balance test results with back dysfunction and pain (see Table 12). Prospectively, the "mid-fit" but not the "high-fit" group was more likely to have improved general back functioning than the "low-fit" group (see Table 11). This result may be due to the strong cross-sectional association in that there were not many "high-fit" subjects among the subjects grouped into the "low-fit" in the baseline examination. Unexpectedly, the modified push-up test was also associated with the current status of mobility function (see section 5.3.2), but it was not selected into the multivariate logistic regression model (see Table 10), possibly because of the high correlations of modified push-up test with Walk Test time (r= -0.42) and the one-leg squat test (r=0.43).

In support of the hypotheses (see section 4.5.2) concerning *physical-activity-related validity* (V), muscular-type exercise was associated with trunk extensor endurance and modified push-up strength among both sexes (see Table 13). No former studies were found that related exercise type with trunk and upper-body muscular endurance among middle-aged adults. A more-detailed discussion is presented in original article V.

6.4.4 Musculoskeletal fitness: flexibility

The *hamstring muscle extensibility* test was expected to *show health-related validity* for back functioning and pain (see section 4.5.2). The "mid-fit" group was less likely than the "low-fit" group to have frequent back problems while stooping, and it was more likely to experience back pain seldom in the cross-sectional study (see Tables 12 and 11, respectively). Prospectively, the "mid-fit" group was more likely to improve back function while stooping (see Table 11). These results may indicate a non-linear relationship between hamstring extensibility and low-back dysfunction, a finding not presented in former studies (Biering-Sørensen 1984, Thomas et al. 1998). No former studies have assessed the relationship between hamstring extensibility and task-specific functional problems of stooping, but it has been related to total hip and trunk flexion movement among low-back patients (Esola et al. 1996).

Typically reduced spinal flexiblity has been found to be a residual sign of persistent LBT (Mellin 1986b, Battié et al. 1990). In the present study, the cross-sectional associations of *trunk side-bending* with back functioning and pain were weak. Prospectively, the "mid-fit" subjects were more likely to improve their general back functioning than the "low-fit" and "high-fit" subjects (see Table 11). This result agrees with the cross-sectional findings on lumbar flexion and extension mobility by Burton et al. (1989a). Both the "mid-fit" and "high-fit" subjects were less likely to have deteriorated general back functioning (see Table 12). Results from the two former prospective studies are conflicting. These studies showed that trunk side-bending had no predictive value for back injury (Battié et al. 1990), but poor lateral mobility to the right increased the risk for low-back pain (Videman et al. 1989).

The flexibility measures showed no *physical-activity-related validity* in the present study (V). This finding is in contrast with the results of former studies among adult (Knapik et al. 1993, Kozma et al. 1991) and elderly (Allied Dunbar National Fitness Survey 1992, van Heuvelen et al. 1998) populations. The absence of associations may be a result of a relatively small contribution of specific activities relying heavily on flexibility, or some activities may even have a negative effect on flexibility (van Heuvelen et al. 1998).

6.4.5 Motor fitness

The *one-leg standing balance* test was expected to show health-related validity for mobility (stair climbing), back dysfunction and back pain. In contrast to the expectation and former cross-sectional (Ensrud et al. 1994, Guralnik et al. 1994) and longitudinal (Hoeymans et al. 1994, Guralnik et al. 1995) studies among the elderly, balance was not associated with stair climbing ability among this middle-aged study population. However, balance was systematically associated with poor general and task-specific back function and frequent back pain in the cross-sectional design (see Table 12), but it had no predictive value (see Tables 11 and 12). The cross-sectional findings on LBT agree with the results of Byl & Sinnot (1992), Luoto et al. (1996 and 1998) and Takala et al. (1997). Former prospective studies on postural control employed follow-up intervention (Luoto et al. 1996 and 1998).

The hypothesis that *muscular type of exercise* is associated with balance was based on exercise studies indicating that exercise training of the elderly reduced fall rates (Province et al. 1995). The hypothesis was supported by the present results for the women, but not for the men (see Table 13). The lack of association among the men may be due to the ceiling effect of the one-leg balance test.

6.4.6 Cardiorespiratory fitness

Predicted VO_{2max} and test time according to the 2 kilometer Walk Test showed a strong and consistent association with current and future status of self-rated perceived health and mobility (see Tables 9 and 10). Predicted VO_{2max} was more strongly associated with perceived health, while test time was related to mobility function. Similar associations with cardiorespiratory fitness and perceived health were reported by Kaplan et al. (1996). The results on mobility function agree with the findings of numerous studies among the elderly (see section 2.5.3). In addition, both predicted VO_{2max} and test time showed cross-sectional validity for back functioning, but not for pain. The Walk Test time also showed some predictive value for improved back function while stooping in the univariate analyses.

The associations of LTPA with cardiorespiratory fitness in the Walk Test were the strongest and most consistent of all the test results regardless of the type of activity. This finding indicates good *physical-activity-related validity* for the test (see Table 13). A more-detailed discussion has been given in original article V.

6.4.7 Morphological fitness

BMI was associated with current status of perceived health (see section 5.3.1), mobility (see section 5.3.2), and back functioning and pain (see section 5.3.3) in the univariate anlayses, but it was not selected into the multivariate logistic regression models, possibly because it is included in the prediction model for VO_{2max} of the Walk Test and has a high correlation with the Walk Test time (r=0.27 for men and r= 0.48 for women). In contrast with the findings of former studies on functional disability (Rissanen et al. 1990, Launer et al. 1994), in the present study, BMI showed only a weak predictive value for improved general back functioning in the univariate analyses (OR 2.28, 95% CI 1.03 to 5.05).

Several cross-sectional epidemiological studies have reported an inverse association between self-reported LTPA and weight or BMI (Allied Dunbar National Fitness Survey 1992, U.S. Department of Health and Human Services 1996). Prospectively, in another study physical inactivity was a risk factor for body mass gain and obesity among Finnish adult men and women (Haapanen et al. 1997). In the present study current LTPA was significantly associated with the BMI of the women but not with that of the men. A more-detailed discussion has been given in original article V.

6.5 Leisure-time physical activity patterns as determinants of self-rated health status

The main hypothesis of the study was that the effects of regular physical activity on health are mediated through physiological responses that affect specific components of fitness. However, the Toronto model, presented in Figure 1, shows that there are direct, two-way associations between physical activity and health.

Subjects engaged in muscular types of exercise were more likely to have good current perceived health than both the aerobic exercise and inactive groups, and they were less likely to have poor current health when compared with an inactive group. Continuous physical activity since school was associated with improved perceived health in the follow-up. These results agree with the recent health-enhancing physical activity recommendations emphasizing versatile and regular physical activity (Pate et al. 1995, U.S. Department of Health and Human Services 1996, ACSM 1998).

The physically active subjects were more likely to experience no difficulty in stair climbing at baseline than the inactive subjects (see section 5.3.2). However, only fitness variables were selected into the multivariate logistic regression model (see Table 10). This finding agrees with the results of Huang et al. (1998), who reported that, in relation to functional limitations, fitness models had a steeper gradient of dose-response and were more accurate in quantifying the association than self-reported physical activity was.

As in former studies (Gyntelberg 1974, Leino 1993, Mundt et al. 1993, Riihimäki et al. 1994, Battié et al. 1995, Kujala et al. 1996, Barnekow-Bergwist et al. 1998), the findings on the associations of LTPA with back functioning and pain were inconsistent or showed no association in the present study. The subjects who had been continuously physically active since school were less likely to have poor general back functioning than the non-continuous group, but they were more likely to deteriorate in back function while stooping in the follow-up (see Table 12).

Not the most active (LTPA 3 times a week or more) but the moderately active (LTPA 1-2 time a week) subjects with difficulties in stair climbing at baseline were more likely to improve their mobility in the follow-up when they were compared with the inactive subjects (see Table 10). Similarly, moderately active subjects with good general back functioning in the baseline examination were less likely to have a deteriorated back function status in the follow-up when compared with both the active and inactive subjects (see Table 12). These findings reveal the problems and open questions related to the proper type and dose of physical activity needed to enhance mobility (see section 2.5.3) and back functioning (see section 2.5.2).

7 SUMMARY AND CONCLUSIONS

A summary of the results on the proposed test battery is presented in Table 14.

(1) The following musuloskeletal and motor tests appeared to provide acceptable reliability for the field testing of HRF: one-leg squat, vertical jump, trunk side-bending and one-leg balance. The reliability of the static trunk extensor endurance test is acceptable on the basis of former studies. There was a learning effect in the modified push-up test and its inter-rater reliability needs further assessment. New studies using proper statistical methods are needed to elaborate the available knowledge on the reliabiity of the proposed test battery.

(2) The present results highlighted the importance of standard health screening and showed that HRF testing can be safely and effectively performed with minor physician participation. The proposed battery offer safe and feasible methods for the HRF testing of middle-aged adults with some reservations. Most people in all age groups qualify for the majority of the tests. For older subjects, cardiovascular diseases limit their participation in trunk extension endurance and modified push-up tests considerably, and to some extent in one-leg squat strength testing. Inactive women are prone to DOMS in the one-leg squat test near maximal strength level.

(3 & 4) In regard to content and predictive validity, all of the items, except the vertical jump test and BMI, showed meaningful associations with self-rated health: the one-leg squat test has validity for current perceived health and current and future mobility (stair climbing); trunk extension endurance for current perceived health, mobility, and back functioning and pain; modified push-up, hamstring extensibility and side-bending for current and future back functioning and current back pain; one-leg balance for poor current back functioning and pain; and the Walk Test for current and future perceived health, mobility and back functioning.

(5) The Walk Test and the trunk extensor endurance and modified push-up tests showed physical-activity-related validity among both sexes, and the one-leg balance test and the BMI showed corresponding results for the women.

The proposed test battery seems to be a promising field-based method for the assessment of HRF among apparently healthy middle-aged adults. With respect to health-related validity, the development is an ongoing process aiming at assessing the validity for different and more objective health outcomes.

TABLE 14. Summary of the results concerning the reliability, safety, feasibility, and validity of the proposed health-related fitness (HRF) test battery for apparently healthy middle-aged adults.

FITNESS COMPONENT	RELIABILITY[R]		SAFETY[R]		FEASIBILITY[R]		VALIDITY								
							Health-related						*LTPA related**		
Test item	Inter rater	Test-retest	CV risk	DOMS	Excl. rate	Time req.	Content validity			Predictive validity			Current status		
MUSCULOSKELETAL							PH	M	BFP	PH	M	BF	ME	AE	OPA
Muscular strength & endur.															
One-leg squat	A,S	A,S	A	AL	A	A	●●	●●	?		●●			°^{M}	
Vertical jump	A	A	A	A	A	A	°^{M}	°^{M}							
Static trunk extensor endur.	NA,S	NA,S	AL	A	AL	A	●●	●●	●●			°	°		°
Modified push-ups	S	AL,S	AL	A	AL	A	°	°	●●			●●	°	°	°
Flexibility															
Hamstring muscle ext.	NA,S	NA,S	A	A	A	AL	°		●●			●●			
Trunk side-bending	A	A	A	A	A	A	°		°		°	●●			
MOTOR															
Static balance															
One-leg standing	A,S	A,S	A	A	A	A	°^{W}		●●				°^{W}		°^{W}
CARDIORESPIRATORY															
UKK Walk Test															
Predicted VO_{2max}	NA	NA	A	A	AL	A	●●	°	●●	●●			°	°^{W}	°
Test time	NA	NA	A	A	A	A	●●	●●	●●	°	●●	°	°	°	°
MORPHOLOGY															
Body mass index	NA	NA	NA	NA	NA	NA	°	°^{W}	°^{W}				°^{W}	°^{W}	°^{W}

[R]=rating: A=acceptable, AL=acceptable with limitations, S=further study needed, NA=not assessed in the present study.
*only univariate analyses were conducted, •• selected into the multivariate stepwise logistic regression model, °univariate association (odds ratio ≥ 2.0 or ≤ 0.50), ?=contradictory results, ^{M}=men only, ^{W}=women only, AE=aerobic type of exercise, BF=back functioning, BFP=back functioning and pain, CV=cardiovascular, DOMS=delayed onset muscle soreness, excl.=exclusion, endur.=enduranse, extens.= extensibility, LTPA=leisure-time physical activity, M=mobility, ME=muscular type of exercise, OPA=overall leisure-time physical activity, PH=perceived health, req.=requirements

8 RECOMMENDATIONS FOR ASSESSING THE HEALTH-RELATED FITNESS OF ADULTS

The HRF test battery described in this paper was *designed to be used in the context of physical activity promotion for health*. The emphasis is on enhancing people's physical functional capacity for everyday life rather than preventing specific diseases (Breslow 1999). It is a measure that can help people to develop a physically active life-style that enhances their state of well-being. More specifically, it is aimed at assessing the level of HRF and monitoring its changes over time among middle-aged people and populations.

Professionals in health care and physical education, such as physical therapists and exercise leaders, have optimal *qualifications to conduct HRF testing*. However, the need to provide specific training for testers should be emphasized to ensure reliable, safe and useful HRF testing with proper interpretations of the results and subsequent exercise recommendations for health promotion. The fitness testers need to be well acquainted with the testing procedures. They should also appreciate the strict standardization of the procedures, be aware of the potential errors in their measurement, and understand the rationale for each test and for the interpretation of the results according to the HRF concept. Furthermore, they must be able to screen the health limitations of the subjects and refer them to a physician when necessary. The context of the health screening procedure, the role of the physician, and the criteria to exclude subjects need to be reconsidered according to the target population and legal or organizational norms and quality requirements.

The *interpretation of the test* results from the point of view of health and physical activity *prescription* are the most important parts of HRF assessment. The recent recommendations for physical activity (ACSM 1998, Feigenbaum & Pollock 1999) provide general criteria of the adequacy of a person's level and type of physical activity with regard to health. The present study provides some elements of the health criteria for fitness (Oja & Tuxworth 1995). The UKK Institute's HRF test battery provides an individual fitness profile based on age- and sex- specific norm reference values (quintiles 1-5) derived from the population sample described in this report. The general interpretation of the test results is that the components of fitness that are lower than the "mid-fit" level (lowest 40%) need to be enhanced. For the flexibility measures, the suggested target is from extremely low and high values towards the "mid-fit" level. More prospective studies with representative population samples are needed to establish more definite health criteria for different components of fitness. The interpretation of the HRF test results is used as the *individual basis for health-enhancing physical activity prescription*. A feasible and effective prescription is a result of discussion between the tester and the client, so that the expert knowledge of the "optimal program" can be balanced to match the opportunities, motivation, and aims of the client (Laitakari & Asikainen 1998). A special aim is to create an individual health target for physical activity.

The direction of the change in a person's fitness level seems to be an equally or even more important factor for health than is the actual level of fitness (Blair et al. 1995, Gill et al. 1997, Erikssen et al. 1988, McMurray et al. 1998). Thus *monitoring changes* in a person's fitness levels is more valuable than interpreting the HRF results at one point in time. Recommendations and comments for the selection of test items for HRF test battery for practical fitness assessment among middle-aged adults are given in Table 15.

TABLE 15 Test-specific recommendations for the practical assessment of health-related fitness among middle-aged adults.

FITNESS COMPONENT Test item	Recommendations and comments
MUSCULOSKELETAL One-leg squat	Good indicator of mobility function, especially among women. Inactive persons need to be informed about the possibility of DOMS.
Vertical jump	Most men score at the highest possible level in the squat test, and therefore the jumping test can be chosen as an alternative test.
Static trunk extension endurance	Good indicator of current functioning and pain of the back. Unsuitable test for persons with hypertension or other CVD. Motivation may have a marked effect on the test result.
Modified push-up	Good indicator of current and future functioning of the back. Unsuitable test for persons with hypertension or other CVD.
Hamstring muscle extensibility	Difficult and time consuming test to obtain reliable results. Good indicator of current and future functioning of the back. Applicable to most middle-aged persons.
Trunk side-bending	Reliable results can be obtained quickly. Good indicator of current and future functioning of the back. Applicable to most middle-aged persons.
MOTOR One-leg standing	Most men score at the highest possible level in the test, and therefore a more challenging test should be selected. Good indicator of current back functioning and pain.
CARDIORESPIRATORY The Walk Test	A versatile indicator of several aspects of health, including musculoskeletal functioning. Large groups of people can be safely and reliably tested within a reasonable length of time.
MORPHOLOGY Body mass index	Good indicator of metabolic aspects of health.

DOMS=delayed oneset of mucle soreness, CVD=cardiovascular diseases

9 YHTEENVETO

Uusi tutkimustieto osoittaa, että kohtuullisesti kuormittavan liikunnan lisäämisellä voidaan merkittävästi edistää väestön terveyttä. Terveyskunnon testauksen yksi tavoite on motivoida ihmisiä säännölliseen liikuntaan. Testituloksia voidaan myös hyödyntää liikuntaohjelmien suunnittelussa ja liikuntatottumusten muutosten vaikutusten arvioinnissa. Liikkumiskyvyn ja selän toimintakyvyn osalta tutkimustieto terveyttä edistävästä liikunnasta ja kunnosta on toistaiseksi puutteellista ja osin ristiriitaista. Keski-ikäiselle väestölle ei myöskään ole olemassa toistettavuudeltaan luotettaviksi ja sisällöltään päteviksi osoitettuja kenttätestausmenetelmiä tuki- ja liikuntaelimistön ja motorisen kunnon arviointiin. Uusia menetelmiä tarvitaankin sekä tutkimukseen että käytännön liikunnan edistämistyöhön.

Tämän tutkimuksen tarkoitus oli kehittää luotettava, turvallinen, toteutuskelpoinen ja pätevä terveyskuntotestistö mittaamaan keski-ikäisen väestön tuki- ja liikuntaelimistön ja motorista kuntoa kenttäolosuhteissa. Poikkileikkaustutkimuksen koehenkilöt olivat edustava otos 37-57-vuotiaita tamperelaisia miehiä (n=246) ja naisia (n=254). Heistä 83% osallistui myös 3-vuoden seurantatutkimukseen. Vapaa-ajan fyysinen aktiivisuus, itse arvioitu koettu terveydentila, liikkumiskyky (portaidennousu), selän toimintakyky ja selkäkipu selvitettiin kyselylomakkeella.

Terveyskuntoa mitattiin yhdeksällä kenttätestillä, joiden suoritustapa oli tarkoin vakioitu. Testistön luotettavuutta (mittaajien välinen, toistettu mittaus) tutkittiin 42 vapaaehtoisen henkilön ryhmässä. Testistön turvallisuus arvioitiin poikkileikkaustutkimuksessa kirjaamalla testauksen aiheuttamat akuutit terveydelliset ongelmat, rekisteröimällä sydämen syke kunkin testin jälkeen ja selvittämällä jälkeenpäin ilmenneen lihaskivun yleisyys ja aste. Testistön sopivuus aikuisille selvitettiin laskemalla kustakin testistä poissuljettujen osuus eri ikäryhmissä. Lisäksi 3 testaajaa arvioi testien toteuttamisen helppoutta ja niihin tarvittavaa aikaa. Testistön pätevyyttä terveyteen liittyvän kunnon mittarina selvitettiin tutkimalla kuntotulosten yhteyksiä itse arvioituun koettuun terveyteen, portaidennousukyyn, selän toimintakykyyn ja selkäkipuihin (poikkileikkaustutkimus) sekä näissä 3-vuoden aikana tapahtuneisiin muutoksiin (seurantatutkimus). Lisäksi tutkittiin fyysisen aktiivisuuden ja terveyskunnon yhteyksiä poikkileikkausasetelmassa. Seuraavat testit todettiin luotettaviksi tuki- ja liikuntaelimistön ja motorisen kunnon mittausmenetelmiksi kenttäolosuhteissa: askelkyykistys, ponnistushyppy, selän sivutaivutus ja yhdellä jalalla seisominen. Muunnellun punnerrustestin tulokset paranivat viikon jälkeen toistetussa mittauksessa. Lisätutkimusta tarvitaan varmentamaan punnerrustestin, vartalon ojentajien staattisen kestävyystestin ja reiden takaosan lihasten venyvyystestin luotettavuus.

Arvioitavina olleet testit osoittautuivat yleisesti turvallisiksi ja toteutuskelpoisiksi keski-ikäisten terveyskunnon mittaamiseen. Lisäksi ennalta laaditun terveydentilan seulontamallin käyttö osoitti, että terveyskunnon testaukset voidaan toteuttaa suurelle osalle väestöä turvallisesti ja tehokkaasti ilman lääkärintarkastusta. Valtaosa koehenkilöistä kaikissa ikäryhmissä pystyi osallistumaan useimpiin testeihin. Vanhimmissa ikäryhmissä sydän- ja verisuonisairaudet rajoittivat kuitenkin noin neljäsosan osallistumista punnerrus- ja vartalon ojentajien staattiseen kestävyystestiin, ja jossain määrin askelkyykistystestiin. Tuki- ja liikuntaelinvaivat rajoittivat osallistumista yksittäisiin testeihin. Lisäksi vähän liikuntaa harrastaneet naiset olivat alttiita askelkyykistystestin aiheuttamalle viivästyneelle lihaskivulle.

Ponnistushyppyä ja kehon painoindeksiä lukuun ottamatta kaikilla testeillä oli yhteyksiä itse arvioituun koettuun terveyteen, liikkumiskykyyn, ja selän toimintakykyyn ja selkäkipuihin (terveyteen liittyvä pätevyys). Askelkyykistys oli yhteydessä portaidennousukykyyn ja sen muutoksiin. Vartalon ojentajien staattinen kestävyys oli yhteydessä koettuun terveyteen, portaidennousukykyyn, selän toimintakykyyn ja selkäkipuihin. Muunneltu punnerrus, reiden takaosan lihasten venyvyys ja selän sivutaivutus olivat yhteydessä selän toimintakykyyn ja niiden muutoksiin sekä selkäkipuihin. Yhdellä jalalla seisominen oli yhteydessä huonoon selän toimintakykyyn ja toistuviin selkäkipuihin. Kävelytesti oli yhteydessä koettuun terveyteen, portaiden nousukykyyn ja selän toimintakykyyn sekä niiden muutoksiin.

Kävelytestin, vartalon ojentajien kestävyys- ja muunnellun punnerrustestin tulokset olivat yhteydessä vapaa-ajan fyysiseen aktiivisuuteen (liikunta-aktiivisuuteen liittyvä pätevyys) miehillä ja naisilla. Lisäksi yhdellä jalalla seisominen ja kehon suhteellinen paino olivat naisilla yhteydessä fyysiseen aktiivisuuteen.

Tässä tutkimuksessa arvioitu tuki- ja liikuntaelimistön kuntoa mittaava testistö on lupaava menetelmä mitata keski-ikäisen väestön terveyskuntoa kenttäolosuhteissa. Testistön pätevyyden jatkoarviointi ja sisällön kehittäminen on tarpeen. Tämä tarkoittaa esimerkiksi toisenlaisten terveydentilan indikaattoreiden käyttöä ja interventiotutkimusten tekemistä. Uusien terveydellä tärkeitä kuntotekijöitä mittaavien testien kehittäminen edellyttää tuki- ja liikuntaelimistön toimintakykykyyn liittyvän uuden epidemiologisen, biomekaanisen ja kliinisen tutkimustietoa seuraamista ja soveltamista.

REFERENCES

ACSM (American College of Sports Medicine) 1995. ACSM's guidelines for exercise testing and prescription. Fifth edition. Pennsylvania: Williams and Wilkins.

ACSM (American College of Sports Medicine) 1998. Position stand on the recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness and flexibility in healthy adults. Medicine and Science in Sports and Exercise 30, 975-991.

ACSM/AHA (American College of Sports Medicine and American Heart Association) 1998. Recommendations for cardiovascular screening, staffing, and emergency policies at health/fitness facilities. Joint position statement. Medicine and Science in Sports and Exercise 30, 1009-1018.

Adams, M.A. & Dolan, P. 1997. Could sudden increases in physical activity cause degeneration of intervertebral discs? Lancet 350, 734-735.

Alaranta, H., Hurri H., Heliövaara, M., Soukka, A. & Harju, R. 1994a. Flexibility of the spine: normative values of goniometric and tape measurements. Scandinavian Journal of Rehabilitation Medicine 26, 147-154.

Alaranta, H., Hurri H., Heliövaara, M., Soukka, A. & Harju, R. 1994b. Non-dynamometric trunk performance tests: reliability and normative data. Scandinavian Journal of Rehabilitation Medicine 26, 211-215.

Alaranta, H., Moffroid, M., Elmqvist, L.-G., Held, J., Pope, M. & Renström, P. 1994c. Postural control of adults with musculoskeletal impairment. Critical Reviews in Physical & Rehabilitation Medicine 6, 337-370.

Allied Dunbar National Fitness Survey 1992. The Sports Council and the Health Education Authority. A report on activity patterns and fitness levels [main findings] Northampton, England: Belmont Press, 110-112.

Altman, D.G. & Gardner, M.J. 1989. Calculating confidence intervals for means and their differences. In M.J. Gardner & D.G. Altman (Eds.) Statistics with confidence. London: British Medical Journal, 20-27.

Aromaa, A., Heliövaara, M., Impivaara, O., Knekt, P., Maatela, J., Joukamaa, M., Klaukka, T., Lehtinen, V., Melkas, Y., Mälkiä, E., Nyman, K., Paunio, I., Reunanen, A., Sievers, K., Kalimo, E. & Kallio. V. 1989. Terveys, toimintakyky ja hoidontarve Suomessa. [English summary: Health, functional limitations and need for care in Finland] Helsinki & Turku: Social Insurance Institution Finland, AL 23, 372.

Aromaa, A., Koskinen, S. & Huttunen, J. 1999. Health in Finland. Helsinki: Edita Ltd, 71-85, 98-103.

Arstila, M., Kallio, V. & Seppänen, A. (Eds.) 1984. Suositus kliinisen rasituskokeen suorittamisesta ja tulkinnasta [Clinical exercise testing. Standards for procedures and recommendations for the interpretation]. Turku, Finland: Social Insurance Institution, ML 39.

Atkinson, G. & Nevill, A.M. 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine 26, 217-238.

Avlund, K., Schroll, M., Davidsen, M., Løvborg, R. & Rantanen, T. 1994. Maximal isometric muscle strength and functional ability in daily activities among 75-year-old men and women. Scandinavian Journal of Medicine and Science in Sports 4, 32-40.

Barnekow-Bergkvist, M., Hedberg, G.E., Janlert, U. & Jansson E. 1998. Determinants of self-reported neck-shoulder and low back symptoms in a general population. Spine 123, 235-243.

Battié, M.C., Bigos, S.J., Fisher, L.D., Hansson, T.H., Nachemson, A.L., Spengler, D.M., Wortley, M.D. & Zeh, J.1989a. A prospective study of the role of cardiovascular risk factors and fitness in industrial back pain complaints. Spine 14, 141-147.

Battié, M.C., Bigos, S.J., Fisher, L.D., Hansson, T.H., Nachemson, A.L., Jones, M.E. & Wortely M.D. 1989b. Isometric lifting strength as a predictor of industrial back pain reports. Spine 14, 851-856.

Battié, M.C., Bigos, S.J., Fisher, L.D., Spengler, D.M., Hansson, T.H., Nachemson, A.L. & Wortely M.D. 1990. The role of spinal flexibility in back pain complaints within industry. Spine 15, 768-773.

Battié, M.C., Videman, T., Gibbons, L.E., Fisher, L.D., Manninen, H. & Gill, K. 1995. Determinants of lumbar disc degeneration. A study relating lifetime exposures and magnetic resonance imaging findings in identical twins. Spine 20, 2601-2612.

Baumgartner, T.A. 1989. Norm-referenced measurement: reliability. In: M.J. Safrit &T.M. Wood (Eds). Measurement concepts in physical education and exercise science. Champaign, IL: Human Kinetics Books, 45-72.

Beimborn, D. & Morrissey, M. 1988. A review of the literature related to trunk muscle performance. Spine 13, 655-660.

Biering-Sørensen, F. 1984. Physical measurements as risk indicators for low-back trouble over a one-year period. Spine 9, 106-119.

Blair, S.N., Kohl, H.W., Barlow, C.E., Paffenbarger, R.S., Gibbons, L.W. & Macera, C.A. 1995. Changes in physical fitness and all-cause mortality. Journal of American Medical Association 273, 1093-1098.

Blair, S.N., Kampert, J.B., Kohl, H.W., Barlow, C.E., Macera, C.A., Paffenbarger, R.S. & Gibbons, L.W. 1996. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. Journal of American Medical Association 276, 205-210.

Bouchard, C., Shephard, R.J., Stephens, T., Sutton, J.R. & McPherson B.D. 1990. Exercise, fitness, and health: the consensus statement. In: C. Bouchard, R.J. Shephard, T. Stephens, J.R. Sutton & B.D. McPherson (Eds.) Exercise, fitness, and health. Champaign, IL: Human Kinetics Books, 3-26.

Bouchard, C. & Shephard, R.J. 1994. Physical activity, fitness, and health: the model and key concepts. In: C. Bouchard, R.J. Shephard & T. Stephens (Eds.) Physical activity, fitness, and health. Champaign, IL: Human Kinetics Books, 77-88.

Bouchard, C. & Pérusse, L. 1994. Heredity, activity level, fitness and health. In: C. Bouchard, R.J. Shephard & T. Stephens (Eds.) Physical activity, fitness, and health. Champaign, IL: Human Kinetics Books, 106-118.

Bravo, G., Gauthier, P., Roy, P.-M., Tessier, D., Gaulin, P., Dubois, M.-F. & Péloquin L. 1994. The functional fitness assessment battery: reliability and validity data for elderly women. Journal of Aging and Physical Activity 2, 67-79.

Breslow, L. 1999. From disease prevention to health promotion. Journal of American Medical Association 281, 1030-1033.

Buckwalter, J.A. & Lane, N.E. 1997. Athletics and osteoarthritis. American Journal of Sports Medicine 25; 873-881.

Burdorf, A., Rossignol, M., Fathallah, F.A., Snook, S.H. & Herrick, R.F. 1997. Challenges in assessing risk factors in epidemiologic studies on back disorders. American Journal of Industrial Medicine 32, 142-152.

Burton, A.K., Tillotson, K.M., Troup, J.D.G. 1989a. Prediction of low-back trouble frequency in a working population. Spine 14, 939-946.

Burton, A.K., Tillotson, K.M. & Troup, J.D.G. 1989b. Variations in lumbar sagittal mobility with low-back trouble. Spine 14, 584-590.

Byl, N.N. & Sinnot, P.L. 1991. Variations in balance and body sway in middle-aged adults: subjects with healthy backs compared with subjects with low-back dysfunction. Spine 16, 325-330.

Campello, M., Nordin, M. & Weiser, W. 1996. Physical exercise and low back pain. Scandinavian Journal of Medicine and Science in Sports 6, 63-72.

Caspersen, C.J., Powell K.E. & Christenson, G.M. 1985. Physical activity, exercise and physical fitness: definitions and distinctions for health-related research. Public Health Reports 100, 126-130.

Caspersen, C.J., Powell, K.E. & Merritt, R.K. 1994. Measurement of health status and well-being. In: C. Bouchard, R.J. Shephard & T. Stephens (Eds.) Physical activity, fitness, and health. Champaign, IL: Human Kinetics Books, 180-202.

Chrisholm, D.M., Collins, M.L., Kulak, L.L., Davenport, W. & Gruber, N. 1975. Physical activity readiness. British Columbia Medical Journal 17, 375-378.

Croft, P.R. & Rigby, A.S. 1994. Socioeconomic influences on back problems in the community in Britain. Journal of Epidemiology and Cummunity Health 48, 166-170.

Davis, M.A., Ettinger, W.H., Neuhaus, J.M. & Mallon, K.P. 1991. Knee osteoarthritis and physical functioning: evidence from the NHANES I epidemiologic follow-up study. Journal of Rheumatology 18, 591-598.

Delitto, A. 1994. Are measures of function and disability important in low back care? Physical Therapy 74, 452-462.

Deyo, R.A. & Bass J.E. 1989. Lifestyle and low-back pain. The influence of smoking and obesity. Spine 14, 501-506.

DiPietro, L.1996. The epidemiology of physical activity and physical function in older people. Medicine and Science in Sports and Exercise 28, 596-600.

Dishman, R.K. 1994. Prescribing exercise intensity for healthy adults using perceived exertion. Medicine and Science in Sports and Exercise 26, 1087-1094.

Dishman, R.K. & Sallis, J. 1994. Determinants and interventions for physical activity and exercise. In: C. Bouchard, R.J. Shephard & T. Stephens (Eds.) Physical activity, fitness, and health. Champaign, IL: Human Kinetics, 214-238.

Dolan, P. & Adams, M.D. 1993. Influence of lumbar and hip mobility on the bending stresses acting on the lumbar spine. Clinical Biomechanics 8, 185-92.

Dunlop, D.D., Hughes, S.L & Manheim, L.M. 1997. Disability in activities of daily living: patterns of change and a hierarchy of disability. American Journal of Public Health 87, 378-383.

Ensrud, K.E., Nevitt, M.C., Yunis, C., Cauley, J.A., Seeley, D.G., Fox, K.M. & Cummings, S.R. 1994. Correlates of impaired function in older women. Journal of American Geriatric Society 42, 481-489.

Era, P., Lyyra, A.L., Viitasalo, J.T. & Heikkinen, E. 1992. Determinants of isometric muscle strength in men of different ages. European Journal of Applied Physiology 64, 84-91.

Era, P., Avlund, K., Jokela, J., Gause-Nilson, I., Heikkinen, E., Steen, B. & Schroll, M. 1997. Postural balance and self-reported functional ability in 75-year-old men and women: a cross-national comparative study. Journal of American Geriatric Society 45, 21-29.

Erikssen, G., Liestol, K., Bjornholt, J., Thaulow, E., Sandvik, L. & Erikssen, J. 1998. Changes in physical fitness and changes in mortality. Lancet 352, 759-762.

Esola, M.A., McClure, P.W., Kelley, F.G. & Siegler, S. 1996. Analysis of lumbar spine and hip motion during forward bending in subjects with and without a history of low back pain. Spine 21, 71-78.

Ettinger, W.H., Burns, R., Messier, S.P., Applegate, W., Rjeski, J., Morgan, T., Berry, M.J., O'Toole, M., Monu, J. & Craven, T. 1997. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. Journal of American Medical Association 277, 25-31.

Feigenbaum, M.S. & Pollock, M.L. 1999. Prescription of resistance training for health and disease. Medicine and Science in Sports and Exercise 31, 38-45.

Felson, D.T., Zhang, Y., Hannan, M.T., Naimark, A., Weissman, B., Aliabadi, P. & Levy, D. 1997. Risk factors for incident radiographic knee osteoarthrosis in the elderly. Arthritis and Rheumatism 40, 728-733.

Fergusson, S.A. & Marras, W.S. 1997. A literature review of low back disorder surveillance measures and risk factors. Clinical Biomechanics 12, 211-226.

Ferrucci, L., Guranick, J.M., Buchner, D., Kasper, J., Lamb, S.E., Simonsick, E.M., Corti, M.C., Bandeen-Roche, K. & Fried, L.P. 1997. Departures form linearity in the relationship between measures of muscular strength and physical performance of the lower extremities: the women's health and aging study. Journals of Gerontology: *Medical Sciences* 52, M275-M285.

Fried, L.P., Ettinger, W.H., Lind, B., Newman, A.B. & Gardin, J. 1994. Physical disability in older adults: a physiological approach. Journal of Clinical Epidemiology 47, 747-760.

Frost, M., Stuckey, S., Smalley, L. & Dorman, G. 1982. Reliability of measuring trunk motions in centimetres. Physical Therapy 62, 1431-1437.

Fujita, Y., Nakamura, Y., Hiraoka, J., Kobayashi, K., Sakata, K., Nagai, M. & Yanagawa, H. 1995. Physical strength tests and mortality among visitors to health-promotion centers in Japan. Journal of Clinical Epidemiology 48, 1349-1359.

Gajdosik, R. & Lusin, G. 1983. Hamstring muscle tightness: reliability of an active knee-extension test. Physical Therapy 63, 1085-1088.

Gibbons, L., Blair, S.N., Kohl, H.W. & Cooper, K. 1989. The safety of maximal exercise testing. Circulation 80, 846-852.

Gill, T.M., Williams, C.S., Mendes de Leon, C.F. & Tinetti, M. 1997. The role of change in physical performance in determining risk for dependence in activities of daily living among nondisabled community-living elderly persons. Journal of Clinical Epidemiology 50, 765-772.

Gordon, N.F., Kohl, H.W., Pollock, M.L., Vaandrager, H., Gibbons, L.W. & Blair S.N. 1995. Cardiovascular safety of maximal strength testing in healthy adults. American Journal of Cardiology 76, 851-853.

Grimby, G. 1995. Physical performance, physical activity and quality of life in the elderly people. Scandinavian Journal of Medicine and Science in Sports 5, 127-128.

Gundewall, B., Liljeqvist, M. & Hansson, T. 1993. Primary prevention of back symptoms and absence from work: a prospective randomized study among hospital employees. Spine 18, 587-594.

Guralnik, J.M., LaCroix, A.Z., Abbott, R.D., Berkman, L.F., Satterfield, S., Evans, D.A. & Wallace, R.B. 1993. Maintaining mobility in late life. I. Demographic characteristics and chronic conditions. American Journal of Epidemiology 137, 845-857.

Guralnik, J.M., Simonsick, E.M., Ferrucci, L., Glynn, R. J., Berkman, L. F., Blazer, D.G., Scherr, P. A. & Wallace, R.B. 1994. A short physical performance battery assessing lower-extremity function: association with self-reported disability and preiditon of mortality and nursing home admission. Journals of Gerontology: *Medical Sciences* 49, M85-M94.

Guralnik, J.M., Ferrucci, L., Simonsick, E.M., Salive, M.E. & Wallace, R.B. 1995. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. New England Journal of Medicine 332, 556-561.

Gyntelberg, F. 1974. One-year incidence of low back pain among male residents of Copenhagen aged 40-59. Danish Medical Bulletin 21, 30-36.

Haapanen, N., Miilunpalo, S., Vuori I., Oja P., Pasanen, M. 1996. The characteristics of leisure time physical activity associated with decreased risk of premature all-cause and CVD-mortality in middle-aged men. American Journal of Epidemiology143, 870-880.

Haapanen, N., Miilunpalo, S., Pasanen, M., Oja, P. & Vuori I. 1997. Association between leisure time physical activity and 10-year body mass change among working-aged men and women. International Journal of Obesity 21, 288-296.

Hagan, R.E., Parrish, G. & Licciardone, J.C. 1991. Physical fitness is inversely related to heart disease risk: a factor analytic study. American Journal of Preventive Medicine 7, 237-243.

Hagen, K. & Thune O. 1998. Work incapacity from low back pain the general population. Spine 23, 2091-2095.

Hannan, M.T., Felson, D.T., Anderson, J.J. & Naimark, A. 1993. Habitual physical activity is not associated with knee osteoarthritis: the Framingham study. Journal of Rheumatology 20, 704-709.

Harper, A.C., Harper, D.A, Lamber, L.J., Andrews, H.B., Kai Lo S, Ross, F.M.& Straker, L.M. 1992. Symptoms of impairment, disability and handicap in low back pain: a taxonomy. Pain 50, 189-195.

Heistaro, S., Heliövaara, M. & Puska, P. 1998. Trends of back pain in eastern Finland, 1972-1992, in relation to socioeconomic status and behavioral risk factors. American Journal of Epidemiology 148, 671-682.

Heliövaara, M., Sievers, K., Impivaara, O., Maatela, J., Knekt, P., Mäkelä, M. & Aromaa, A. 1989. Descriptive epidemiology and public health aspects of low back pain. Annals of Medicine 21,327-333.

Heliövaara, M., Aromaa, A., Klaukka, T., Knekt, P., Joukamaa, M. & Impivaara, O. 1993. Reliability and validity of interview data on chronic diseases. Journal of Clinical Epidemiology 46, 181-191.

Hides, J.A., Richardson, C.A.& Jull, G.A. 1996. Multifidus muscle recovery is not automatic after resolution of acute, first-episode low back pain. Spine 21, 2763-2769.

Hirsch, C.H., Fried, L.P., Harris, T., Fitzpatrick, A., Enright, P. & Schulz, R. 1997. Correlates of performance-based measures of muscle function in the elderly. The cardiovascular health study. Journals of Gerontology: *Medical Sciences* 52, M192-M200.

Hodges, P.W. & Richardson, C.A. 1996. Inefficient muscular stabilisation of the lumbar spine associated with low back pain. A motor control evaluation of transversus abdominis. Spine 21, 2640-2650.

Hoeymans, N., Feskens, E.J.M., van den Bos, G.A.M. & Kromhout, D. 1996. Measuring functional status. Cross-sectional and longitudinal associations between performance and self-report. Journal of Clinical Epidemiology 49, 1103-1110.

Holmström, E., Moritz, U. & Andersson, M. 1992. Trunk muscle strength and back muscle endurance in construction workers with and without low back disorders. Scandinavian Journal of Rehabilitation Medicine 24, 3-10.

Huang, Y., Macera, C.A., Blair, S.N., Brill, P.A., Kohl III, H.W. & Kronenfeld, J.J. 1998. Physical fitness, physical activity, and functional limitations in adults aged 40 and older. Medicine and Science in Sports and Exercise 30, 1430-1435.

Hyytiäinen. K., Salminen, J., Suvitie, S., Wickström, G. & Pentti J. 1991. Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. Scandinavian Journal of Rehabilitation Medicine 23, 3-10

Imeokparia, R.L., Barret J.P., Arrieta, M.I., Leaverton P.E., Wilson, A.A., Hall, B.J. & Marlowe S.M. 1994. Physical activity as a risk factor for osteorathritis of the knee. Annals of Epidemiology 4, 221-230.

Ito, T., Shirado, O., Suzuki, H., Takahashi, M., Kaneda, K. & Strax, T.E. 1996. Lumbar trunk muscle endurance testing: an inexpensive alternative to a machine for evaluation. Archives of Physical Medicine and Rehabilitation 77, 75-79.

Jaeschke, R., Guyatt, G.H. & Sackett, D.L. 1994. Users' guides to the medical literature III: how to use an article about a diagnostic test. Journal of American Medical Association 271, 703-707.

Jette, M., Quenneville, J. & Sideny, K. 1992. Fitness testing and counselling in health promotion. Canadian Journal of Sport Sciences 17, 193-198.

Jette, A.M. 1994. Physical disablement concepts for physical therapy research and practice. Physical Therapy 74, 380-386.

Jimenenz, S.A. & Dharmavaram, R.M. 1994. Genetic aspects of familial osteoarthrosis. Annals of Rheumatic Diseases 53, 789-797.

Jones, B.H., Cowan, K.N. & Knapik, J.J. 1994. Exercise, training and injuries. Sports Medicine 18, 202-2014.

Jørgensen, K. & Nicolaisen, T. 1986. Two methods for determining trunk extensor endurance. European Journal of Applied Physiology 55, 639-644.

Johnston, M.V., Keith, R.A. & Hinderer, S.R. 1992. Measurement standards for interdisciplinary medical rehabilitation. Archives Physical Medicine and Rehabilitation 73 (Suppl), 3S-23S.

Kallinen, M. & Alen, M. 1995. Aging, physical activity and sports injuries. Sports Medicine 20, 41-52.

Kane, Y. & Bernasconi, J. 1992. Analysis of a modified active knee extension test. Journal of Orthopaedic and Sports Physical Therapy 15, 141-146.

Kannus, P., Jozsa, L., Renström, P., Järvinen, M., Kvist, M., Lehto, M., Oja, P. & Vuori, I. 1992. The effects of training, immobilization and remobilization on musculoskeletal tissue. Scandinavian Journal of Medicine and Science in Sports 2, 100-118.

Kaplan, G.A., Goldberg, D.E., Everson, S.A., Cohen, R.D., Salonen, R., Tuomilehto, J. & Salonen, J. 1996. Perceived health-status and morbidity and mortality: evidence from the Kuopio heart disease risk factor study. International Journal of Epidemiology 25, 259-265.

King, A.S., Haskell, W.L., Taylor, C.B., Kraemer, H.C. & DeBusk R.F. 1991. Group- vs home based exercise training in healthy older men and women. Journal of American Medical Association 266, 1535-1542.

King, C.N. & Senn, M.D. 1996. Exercise testing and prescription. Practical recommendations for the sedentary. Sports Medicine 21, 326-336.

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. Journal of Clinical Epidemiology 51, 1243-1252.

Knapik, J., Zoltick, J., Rottner, H.C., Phillips, J., Bielenda, C., Jones, B. & Drews F. 1993. Relationships between self-reported physical activity and physical fitness in active men. American Journal of Preventive Medicne 9, 203-208.

Knapik, J.J., Banderet, L.E., Voger, J.A., Bahrke, M.S. & O'Connor, J.S. 1996. Influence of age and physical training on measures of cardiorespiratory and muscle endurance. European Jorunal of Applied Physiology 72, 490-495.

Koch, M., Gottschack, M., Baker, D.I., Palumbo, S. & Tinetti, M. 1994. An impairment and disability assessment and treatment protocol for community living elderly persons. Physical Therapy 74, 286-294.

Kozma, A., Stones, M.J. & Hannah, T.E. 1991. Age, activity, and physical performance: an evaluation of performance models. Psychology and Aging 6, 43-49.

Kuipers, H. 1994. Exercise-induced muscle damage. International Journal of Sports Medicine 15, 132-135.

Kujala, U.M., Viljanen, T., Taimela. S. & Viitasalo J.T. 1994. Physical activity, VO_{2max}, and jumping height in an urban population. Medicine and Science in Sports and Exercise 26, 889-895.
Kujala, U.M., Taimela, S., Viljanen, T., Jutila, H., Viitasalo, J.T., Videman, T. & Battie, M.C. 1996. Physical loading and performance as predictors of back pain in healthy adults. European Journal of Applied Physiology 73, 452-458.
Kujala, U.M., Kaprio, J., Sarna, S. & Koskenvuo, M. 1998. Relationship of leisure-time physical activity and mortality. Journal of American Medical Association 279, 440-444.
Kuorinka, I., Jonsson, B., Kilbom, Å., Vinterberg, H., Biering-Sørensen, F., Andersson, G. & Jørgensen K. Standized Nordic questionnaires for the analysis of musculoskeletal symptoms. Applied Ergonomics 18, 233-237.
LaCroix, A.Z., Guralnik, J.M., Berkman, L.F., Wallace, R.B. & Satterfield, S. 1993. Maintaining mobility in late life. II Smoking, alcohol consumption, physical activity, and body mass index. American Journal of Epidemiology 137, 858-869.
Lahad, A., Malter, A.D., Berg, A.O. & Deyo, R.A. 1994. The effectiveness of four interventions for the prevention of low back pain. Journal of American Medical Association 272, 1268-1291.
Laitakari, J. & Asikainen, T.-M. 1998. How to promote physical activity through individual counselling – A proposal for a practical model of counselling on health-related physical activity. Patient Education and Counselling 33, S13-S24.
Lamb, K. 1998. Test-retest reliability in quantitative physical education research: a commentary. European Physical Education Review 4, 145-152.
Laukkanen, R.M.T., Oja, P., Ojala, K.H., Pasanen, M.E. & Vuori, I.M.1992. Feasibility of a 2-km walking test for fitness assessment in a population study. Scandinavian Journal of Social Medicine 20, 119-126.
Laukkanen, R.M.T., Oja, P., Pasanen, M.E. & Vuori, I.M. 1993. Two-kilometre walking test: effect of walking speed on the prediction of maximal oxygen uptake. Scandinavian Journal of Medicine and Science in Sports 3, 263-266.
Laukkanen, R. 1993. Development and evaluation of a 2-km walking test for assessing maximal aerobic power of adults in field conditions, Kuopio University Publications D. Medical Sciences 23. Dissertation.
Laukkanen, P., Kauppinen, M. & Heikkinen, E. 1998. Physical activity as a predictor of health and disability in 75- and 80-year-old men and women. Journal of Aging and Physical Activity 6, 141-156.
Launderdale, D.S., Fabsitz, R., Meyer, J.M., Sholinsky, P., Ramakrishnan, V. & Goldberg, J. 1997. Familial determinants of moderate and intense physical activity: a twin study. Medicine and Science in Sports and Exercise 29, 1062-1068.
Launer, L.J., Harris, T., Rumpel, C. & Madans, J. 1994. Body mass index, weight change, and risk of mobility disability in middle-aged and older women. The epidemiologic follow-up study of NHANES I. Journal of American Medical Association 271, 1093-1098.
Lawrence, R.H. & Jette, A.M. 1996. Disentangling the disablement process. Journals of Gerontology: *Social Sciences* 51B, S173-S182.

Leino, P., Aro, S. & Hasan, J. 1987. Trunk muscle function and low-back disorders: a ten-year follow-up study. J Chronic Dis 40, 289-296.

Leino, P.I. 1993. Does leisure time physical activity prevent low back disorders. Spine 18, 863-871.

Loy, S.F., Conley, L.M., Sacco, E.R., Vincent, W.J., Holland, G.J., Sletten, E.G. & Trueblood P.R. 1994. Effect of stair climbing on VO_{2max} and quadriceps strength in middle-aged females. Medicine and Science in Sports and Exercise 26, 241-247.

Lundberg, O. & Manderbacka, K. 1996. Assessing reliability of a measure of self-rated health. Scandinavian Journal of Social Medicine 24, 218-224.

Luoto, S., Hurri, H. & Alaranta, H. 1995a. Reaction time in patients with chronic low-back pain. European Journal of Physical Medicine and Rehabilitation 5, 47-50.

Luoto, S., Heliövaara, M., Hurri, H. & Alaranta, H. 1995b. Static back endurance and the risk of low back pain. Clinical Biomechanics 10, 323-324.

Luoto, S., Taimela, S., Hurri, H., Aalto, H., Pyykkö, I. & Alaranta, H. 1996. Psychomotor speed and postural control in chronic low back pain patients. Spine 21, 2621-2627.

Luoto, S., Aalto, H., Taimela, S., Hurri, H., Pyykkö, I. & Alaranta, H. 1998. One-footed and externally disturbed two-footed postural control in patients with chronic low back pain and healthy control subjects. Spine, 23, 2081-2090.

MacIntyre, D.L., Reid, W.D. & McKenzie, D.C. 1995. Delayed muscle soreness: the inflammatory response to muscle injury and its clinical implications. Sports Medicine 20, 24-40.

Mäkelä, M., Heliövaara, M., Sievers, K., Knekt, P., Maatela, J. & Aromaa, A. 1993. Musculoskeletal disorders as determinants of disability in Finns aged 30 years or more. Journal of Clinical Epidemiology 46, 549-559.

Mälkiä, E. 1983. Muscular performance as a determinant of physical ability in Finnish adult population. Turku, Finland: Social Insurance Institution Finland, Turku, AC, 23.

Marsh, H.S. 1993. The multidimensional structure of physical fitness: invariance over gender and age. Research Quarterly for Exercise and Sport 64, 256-273.

Masset, D.F., Piette, A.G. & Malchaire, J.B. 1998. Relation between functional characteristics of the trunk and the occurence of low back pain. Spine 23, 359-365.

McAlindon, T.E., Cooper, C., Kirwan, J.R. & Deipe, P.A. 1993 Determinants of disability in osteoarthritis of the knee. Annals of Rheumatic Diseases. 52, 258-262.

McMurray, R.G., Ainsworth, B.E., Harrell, J.S., Griggs, T.R. & Williams, O.D. 1998. Is physical activity or aerobic power more influential on reducing cardiovascular disease risk factors? Medicine and Science in Sports and Exercise 30, 1521-1529.

Mellin, G.P. 1996a. Accuracy of measuring lateral flexion of the spine with a tape. Clinical Biomechanics 1, 85-89.

Mellin, G. 1996b. Chronic low back pain in men 54-63 years of age. Correlations of physical measurements with the degree of trouble and progress after treatment. Spine 11, 421-426.

Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M. & Urponen, H. 1997. Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. Journal of Clinical Epidemiology 50, 517-528.

Mittlemann, M.A., Maclure, M., Tofler, G.H., Sherwood, J.B., Goldberg, R.J. & Muller, J.E.1993. Triggering of acute myocardial infarction by heavy physical exertion. Protection against triggering by regular exertion. New England Journal of Medicine 329, 1677-1683.

Moffet, J.A.K., Hughes, G.I. & Griffiths, P. 1993. A longitudinal study of low back pain in student nurses. International Journal of Nursing Studies 30, 197-212.

Moffroid, M., Reid, S., Henry, S.M., Haugh, L.D. & Ricamato, A. 1994. Some endurance measures in persons with chronic low back pain. Journal of Orthopaedic and Sports Physical Therpay 20, 81-87.

Moreland, J., Finch, E., Stratford, P., Balsor B. & Gill C. 1997. Interrater reliability of six tests of trunk muscle function and endurance. Journal of Orthopaedic and Sports Physical Therapy 26, 200-208.

Morey, M.C., Pieper, C.F.& Cornoni-Huntley, J. 1998a. Is there a threshold between peak oxygen uptake and self-reported physical functioning in older adults. Medicine and Science in Sports and Exercise 30, 1223-1229.

Morey, M.C., Pieper, C.F. & Cornoni-Huntley, J. 1998b. Physical fitness and functional limitations in community-dwelling older adults. Medicine and Science in Sports and Exercise 30, 715-723.

Mundt, D.J., Kelsey, J.L., Golden, A.L., Panjabi, M.M., Pastides, H., Berg, A., Sklar, J. & Hosea, T. 1993. An epidemiologic study of sports and weight lifting as possible risk factors for herniated lumbar and cervical discs. American Journal of Sports Medicine 21, 854-860.

Nelson, H.D., Nevitt, M.C., Scott, J.C., Stone, K.L. & Cummings, S.R. 1994. Smoking, alcohol, and neuromuscular and physical function of older women. Journal of American Medical Association 272, 1825-1831.

Norton, K., Olds, T., Bowes, D., van Ly, V. & Gore, C. 1998. Applying the sports medicine Australia pre-exercise screening procedures: who will be excluded? Journal of Science and Medicine in Sports 1, 38-51.

Nummi, J., Järvinen, T., Stambej, U. & Wickström, G. 1978. Diminished dynamic performance capacity of back and abdominal muscles in concrete reinforcement workers. Scandinavian Journal of Work Environment & Health 4 (suppl 1), 39-46.

Nutter, P. 1988. Aerobic exercise in the treatment and prevention of low back pain. Occupational Medicine 3, 137-145.

Nygård, C.-H., Luopajärvi, T. & Ilmarinen, J. 1991. Musculoskeletal capacity and its changes among aging municipal employees in different work categories. Scandinavian Journal of Work, Environment & Health 17 (suppl 1), 110-117.

Nygård, C.-H., Kilbom, Å., Hjelm E.W. & Winkel, J. 1994. Life-time occupational exposure to heavy work and individual physical capacity. International Journal of Industrial Ergonomics 14, 365-372.

Oja, P., Laukkanen, R., Pasanen, M., Tyry, T. & Vuori, I. 1991. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. International Journal of Sports Medicine 12, 356-362.

Oja, P., Miilunpalo, S., Vuori, I., Pasanen, M. & Urponen, H. 1994. Trends of health-related physical activity in Finland: 10 year follow-up of and adult cohort in eastern Finland. Scandinavian Journal of Medicine and Science in Sports 4, 75-81.

Oja, P. & Tuxworth, W. 1995 (Eds.) Eurofit for adults. Assessment of health-related fitness. Finland: Council of Europe Publishing, 27-35.

Panush, R.S. 1994. Physical activity, fitness and osteoarthritis. In: C. Bouchard, R.J. Shephard & T. Stephens (Eds.) Physical activity, fitness, and fealth. Champaign, IL: Human Kinetics, 712-723.

Parnianpour, M., Nordin, M., Kahanovitz, N. & Frankel, V. 1988. The triaxial coupling of torque generation of trunk muscles during isometric exertions and the effect of fatiguing isoinertial movements on the motor output and movement patterns. Spine 9, 982-992.

Pate, R.R., Pratt, M., Blair, S.N., Haskell, W.L., Macera, C.A., Bouchard, C., Buchner, D., Ettinger, W., Heath, G.W., King, A.C., Kriska, A., Leon, A.S., Marcus, B.H., Morris, J., Paffenbarger, R.S., Patrick, K., Pollock, M.L., Rippe, J.M., Sallis, J. & Wilmore, J. 1995. Physical activity and public health. A recommendation from the Centres for Disease Control and Prevention and the American College of Sports Medicine. Journal of American Medical Association 273, 402-407.

Phillips, W.T. & Haskell, W.L. 1995. "Muscular fitness" easing the burden of disability for elderly adults. Journal of Aging and Physical Activity 3, 261-289.

Pollock, M.L., Carroll, J.F., Graves, J.E., Legget, S.H., Braith, R.W., Limacher, M. & Hagberg, J.M. 1991. Injuries and adherence to walk/jog and resistance training programs in the elderly. Medicine and Science in Sports and Exercise 23, 1194-1200.

Province, M.A., Hadley, E.C., Hornbrook, M.C., Lipsitz, L.A., Miller, J.P., Mulrow, C.D., Ory, M.G., Sattin, R.W., Tinetti, M.E. & Wolf, W.L. 1995. The effect of exercise on falls in elderly patients. A preplanned meta-analysis of the FICSIT Trials. Journal of Aging and Physical Activity 273, 1341-1347.

Rankin, G. & Stokes, M. 1998. Reliability of assessment tools in rehabilitation. An illustration of appropriate statistical analyses. Clinical Rehabilitation 12, 187-199.

Rantanen, T., Parkatti, T. & Heikkinen, E. 1992. Muscle strength according to level of physical exercise and educational background in middle-aged women in Finland. European Journal of Applied Physiollgy 65, 507-512.

Rantanen, T., Era, P. & Heikkinen, E. 1994. Maximal isometric strength and mobility among 75 year-old men and women. Age and Aging 23, 132-137.

Rantanen, T., Era, P. & Heikkinen, E. 1996. Maximal isometric knee extension strength and stair-mounting ability in 75- and 80-year-old men and women. Scandinavian Journal of Rehabilitation Medicine 28, 89-93.

Rantanen, T., Guralnik, J.M., Sakari-Rantala, R., Leveille, S., Simonsick, E.M., Ling. S. & Fried, L.P. 1999a. Disability, physical activity, and muscle strength in older women: the women's health and aging study. Archives of Physical Medicine and Rehabilitation 80, 130-135.

Rantanen, T., Guralnik, J.M., Foley, D., Maskai, K., Leveille, S., Curb, J.D. & White L. 1999b. Midlife hand grip strength as a predictor of old age disability. Journal of American Medical Association 281, 558-560.

Räty, H.P., Kujala, U.M., Videman, T., Impivaara, O., Battie, M.C. & Sarna, S. 1997. Lifetime musculoskeletal symptoms and injuries among former elite male athletes. International Journal of Sports Medicine18, 625-632.

Rejeski W.J., Ettinger, W.H., Schumaker, S., James, P., Burns, R. & Elam, J.T. 1995. Assessing performance-related disability in patients with knee osteoarthritis. Osteoarthritis and Cartilage 3, 157-167.

Reuben, D.B., Siu, A.L. & Kimpau, S. 1992. The predictive validity of self-report and performance based measures of function and health. Journals of Gerontology: *Medical Sciences* 47, M106-M110.

Riihimäki, H.,Wickström, G., Hänninen, K. & Luopajärvi, T. 1989. Predictors of sciatic pain among concrete reinforcement workers and house painters – five year follow-up. Scandinavian Journal of Work Environment & Health 15, 415-423.

Riihimäki, H. 1991. Low-back pain, its origin and risk indicators. Scandinavian Journal of Work Environment and Health 17, 81-90.

Riihimäki, H., Viikari-Juntura, E., Moneta, G., Kuha, J., Videman, T. & Tola, S. 1994. Incidence of sciatic pain among men in machine operating, dynamic physical work, and sedentary work. Spine 19, 138-142.

Rissanen, A., Heliövaara, M., Knekt, P., Reunanen, A., Aromaa, A. & Maatela, J. 1990. Risk of disability and mortality due to overweight in a Finnish population. British Medical Jorunal 301, 835-837.

Rissanen, A., Alaranta, H., Sainio, P. & Härkönen H. 1994. Isokinetic and non-dynamometric tests in low back patients related to pain and disability index. Spine 19, 1963-1967.

Roebroeck, M.E., Harlaar, J. & Lankhorst, G.F. 1993. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Physical Therapy 73, 386-401.

Rose, M.J. 1991. The statistical analysis of the intra-observer repeatability of four clinical measurement techniques. Physiotherapy 77, 89-91.

Sandler, R.B., Burdett, R., Zaleskiewitch, M., Sprowls-Repcheck, C. & Harwell, M. 1991. Muscle strength as an indicator of the habitual level of physical activity. Medicine and Science in Sports and Exercise 23, 1375-1381.

Sandmark, H. & Vingård, E. 1999. Sports and risk for severe osteoarthrosis of the knee. Scandinavian Journal of Medicine and Science in Sports 9, 279-284.

Saxon, L., Finch, C. & Bass, S. 1999. Sports participation, sports injuries and osteoarthritis. Sports Medicine 28, 123-135.

Saxton, J.M., Clarkson, P.M., James, R., Miles, M., Westerfer, M., Clark, S. & Donelly, A.E. 1995. Neuromuscular dysfunction following eccentric exercise. Medicine and Science in Sports and Exercise 27, 1185-1193.

Schroll, M. 1994. The main pathway to musculoskeletal disability. Scandinavian Journal of Medicine and Science in Sports 4, 3-12.

Schroll, M., Avlund, K. & Davidsen M. 1997. Predictors of five-year functional ability in a longitudinal survey of men and women aged 75 to 80. The 1914-population in Glostrup, Denmark. Aging in Clinical and Experimental Research 9, 143-152.

Shaw, C.E., McCully, K.K. & Posner JD. 1995. Injuries during the one repetition maximum assessment in the elderly. Journal of Cardiopulmonary Rehabilitation 15, 283-287.

Shaulis, D., Golding, L.A. & Tandy, R.D. 1994. Reliability of the AAHPERD functional fitness assessment across multiple practise sessions in older men and women. Journal of Aging and Physical Activity 2, 272-279.

Shephard, R.J. 1988. PAR-Q, Canadian home fitness test and exercise screening alternatives. Sports Medicine 5, 185-195.

Shephard, R.J. 1991. Measurement of fitness. The Canadian experience. Journal of Sports Medicine and Physical Fitness 31, 470-480.

Shephard, R.J., Thomas, S. & Weller, I. 1991. The Canadian home fitness test. 1991 update. Sports Medicine. 11, 358-366.

Shephard, R.J. & Bouchard, C. 1994. Principal components of fitness: relatioship to physical activity and lifestyle. Canadian Journal of Applied Physiology 19, 200-214.

Simmonds, M.J., Olson, S.L., Jones, S., Hussein, T., Novy, D. & Radwan H. 1998. Psychometric characteristics and clinical usefulness of physical performance tests in patients with low back pain. Spine 23, 2412-2421.

Simonen, R., Videman, T., Battié, M.C. & Gibbons, L.E. 1998. The effects of lifelong exercise on psychomotor reaction time. A study of 38 pairs of male monozygotic twins. Medicine and Science in Sports and Exercise 30, 1445-1450.

Skinner, J. & Oja, P. 1994. Laboratory and field tests for assessing health-related fitness. In: C. Bouchard, R.J. Shephard & T. Stephens (Eds.) Physical activity, fitness, and health. Champaign, IL: Human Kinetics, 160-179.

Sonn, U., Frändin, K. & Grimby, G. 1995. Instrumental activities of daily living related to impairments and functional limitations in 70-year-olds and changes between 70 and 76 years of age. Scandinavian Journal of Rehabilitation Medicine 27, 119-128.

Stones, M.J. & Kozma, A.K. 1987. Balance and age in the sighted and blind. Archives of Physical Medicine and Rehabilitation 68, 85-89.

Suzuki, N. & Endo, S. 1983. A quantitative study of trunk muscle strength and fatigability in the low-back pain syndrome. Spine 8, 69-74.

Taimela, S., Österman, K., Alaranta, H., Soukka, A. & Kujala, U.M. 1993. Long psychomotor reaction time in patients with chronic low-back pain. Archives of Physical Medicine and Rehabilitation 74, 1161-1164.

Takala, E.-P., Korhonen, I. & Viikari-Juntura, E. 1997. Postural sway and stepping response among working population: reproducibility, long term stability, and associations with symptoms of the low back. Clinical Biomechanics 12, 429-437.

Thomas, S., Reading, J. & Shephard, R.J. 1992. Revision of the physical activity readiness questionnaire (PAR-Q). Canadian Journal of Sports Science 17, 338-345.

Thomas, E., Silman, A.J., Papageorgiou, A.C., Macfarlane, G.J. & Croft, P.R. 1998. Association between measures of spinal mobility and low back pain. Spine 23, 343-347.

Thomis, M.A.I., Beunen, G.P., Maes, H.H., Blimkie, C.J., Leemputte, M.van, Claessens, A.L., Marchal, G., Willems, E. & Vlietinck R.F. 1998. Strength training: importance of genetic factors. Medicine and Science in Sports and Exercise 30, 724-731.

Tinetti, M.E., Baker, D.I., McAvay, G., Claus, E.B., Garret, P., Gottschalk, M., Koch M.L., Trainor, K. & Horwitz, R.I. 1994. A multifactorial intervention to reduce the risk of falling among elderly people living in the community. New England Journal of Medicine 331, 821-827.

Torgén, M., Punnet, L., Alfredsson, L. & Kilbom, Å. 1999. Physical capacity in relation to present and past physical load at work: a study of 484 men and women aged 41 to 58 years. American Journal of Industrial Medicine 36, 388-400.

Troup, J.D.G. & Videman, T. 1989. Inactivity and the aetiopathogenesis of musculoskeletal disorders. Clinical Biomechanics 4, 173-178.

U.S. Department of Health and Human Services. Physical activity and health: A report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, 28-29, 33-37, 43-45, 72, 133-134.

van Heuvelen, M.J.G., Kempen, G.I.J.M., Ormel, J. & Rispens, P.1998. Physical fitness related to age and physical activity in older persons. Medicine and Science in Sports and Exercise 30, 434-441.

Venna, S., Hurri, H. & Alaranta, H. 1994. Correlation between neurological leg deficits and reaction time of upper limbs among low-back pain patients. Scandinavian Journal of Rehabilitation Medicine 26, 87-90.

Videman, T., Rauhala, H., Asp, S., Lindström, K., Cedercreutz, G., Kämppi, M., Tola, S. & Troup J.D.G. 1989. Patient-handling skill, back injuries and back pain: an intervention study in nursing. Spine 14, 148-56.

Videman, T., Nurminen, T. & Troup, J.D.G. 1990. Lumbar spine pathology in cadaveric meterial in relation to history of back pain, occupation, and physical loading. Spine 15, 728-737.

Videman, T., Sarna, S., Battie, M.C., Koskinen, S., Gill, K., Paananen, H. & Gibbons L. 1995. The long term effects of physical loading and exercise lifestyle on back-related symptoms, disability, and spinal pathology among men. Spine 20, 699-709.

Videman, T., Battie, M.C., Gibbons, L.E., Manninen, H., Gill, K., Fisher, L.D. & Koskenvuo, M. 1997. Lifetime exercise and disk degeneration: an MRI study of monozygotic twins. Medicine and Science in Sports and Exercise 29, 1350-1356.

Viikari-Juntura, E., Takala, E.-P., Riihimäki, H., Malmivaara, A., Martikainen, R. & Jäppinen P. 1998. Standardized physical examination protocol for low back disorders: feasibility and of use and validity for symptoms and signs. Journal of Clinical Epidemiology 51, 245-255.

Viljanen, T., Viitasalo, J.T. & Kujala, U.M. 1991. Strength characteristics of a healthy urban population. European Journal of Applied Physiology 63, 43-47.
Vingård, E., Alfredsson, L., Goldie, I. & Hogstedt, C. 1993. Sports and osteoarhtrosis of the hip. An epidemiologic study. American Journal of Sports Medicine 21, 195-200.
Vingård, E., Alfredsson, L. & Malchau, H. 1998. Osteoarthrosis of the hip and its relationship to physical load from sports activities. American Journal of Sports Medicine 26, 78-82.
Vuori, I.M. 1995. Sudden death and exercise: effects of age and type of activity. Sport Science Review 4, 46-84.
Ware, J.E., Brook, R.H., Davies. A.R. & Lohr K.N. 1981. Choosing measures of health status for individuals in general populations. American Journal of Public Health 71, 620-625.
WHO, 1999. International classification of functioning and disability: beta-2 draft short version July 1999. Assessment, Classification and Epidemiology Group. World Health Organization. Switzerland, Geneva: http://www.who.int/icidh.
Wilson, I.B. & Cleary, P.D. 1995. Linking the clinical variables with health-related quality of life: a conceptual model of patient outcomes. Journal of American Medical Association 273, 59-65.
Wolff, I., van Croonenborg, J.J., Kemper, H.C.G., Kostense, P.J. & Twisk, J.W.R. 1999. The effect of exercise training programs on bone mass: a meta-analysis of published controlled trials in pre- and postmenopausal women. Osteoporosis International 9, 1-12.
Wolk, A. & Rössner, S. 1996. Obesity and self-perceived health in Sweden. International Journal of Obesity 20, 369-372.

Appendix 1. Evaluated test items that were excluded from the proposed health-related fitness test battery.

FITNESS COMPONENT Test item	Reason (s) for exclusion
MUSCULOSKELETAL FITNESS	
• **Flexibility**	
Sit-and-reach	Soft-tissue low-back pain during testing
Lower extremity muscular tightness: iliopsoas, hamstring, rectus femoris	Subjective evaluations of degree of tightness had poor inter-rater reliability
• **Muscular strength and endurance**	
Push-ups in 30 seconds	Difficult to define improper technique
Partial sit-ups (maximum repetitions)	High number of zero scores among the women
Rotated sit-up (can/cannot)	Reasons for failure difficult to define
Isometric sit-up (up to 4 minutes)	Test position biomechanically "impossible "for many subjects
Maximal repetitions of two-leg squatting with extra weight	Caused severe delayed muscle soreness
MOTOR FITNESS	
• **Coordination**	
Jumping jack Throw and catch the ball Swinging eight Embrace ball	Subjective evaluations of the quality of motor performance had poor inter-rater reliability
• **Balance**	
One-leg standing, eyes open	Poor inter-rater reliability
One-leg standing, head turns	Poor inter-rater reliability

I

Health-related fitness test battery for adults: aspects of reliability

by

Jaana Suni, Pekka Oja, Raija Laukkanen, Seppo Miilunpalo, Matti Pasanen,
Ilkka Vuori, Klaus Bös

Archives of Physical Medicine and Rehabilitation 1996; 77:399-405

https://doi.org/10.1016/S0003-9993(96)90092-1

Health-Related Fitness Test Battery for Adults: Aspects of Reliability

Jaana H. Suni, MSc, Pekka Oja, PhD, Raija T. Laukkanen, PhD, Seppo I. Miilunpalo, MD, Matti E. Pasanen, MSc, Ilkka M. Vuori, MD, Tuula-Marja Vartiainen, MD, Klaus Bös

ABSTRACT. Suni JH, Oja P, Laukkanen RT, Miilunpalo SI, Pasanen ME, Vuori IM, et al. Health-related fitness test battery for adults: aspects of reliability. Arch Phys Med Rehabil 1996;77:399-405.

Objective: In two studies, the reliability of 3 balance, 2 flexibility, and 4 muscular strength tests proposed as test items were investigated in a health-related fitness (HRF) test battery for adults.

Design: Methodological study.

Setting: A health promotion research institute.

Subjects: In study A, volunteers ($n = 42$) from two worksites participated. In study B, a population sample ($n = 510$) of 37- to 57-year-old men and women was selected.

Main Outcome Measures: Intraclass correlation coefficient of repeated measures was used to assess inter rater reliability. The degree of measurement error was expressed as the standard error of measurement. The mean difference with 95% confidence intervals between the testing days or test trials was used to assess test-retest or trial-to-trial reproducibility. The coefficient of variation (CV = {SD/mean} × 100%) from day to day was also calculated.

Results: The following tests appeared to provide acceptable reliability as methods for field assessment of HRF: standing on one leg with eyes open for balance, side-bending of the trunk for spinal flexibility, modified push-ups for upper body muscular function, and jump and reach and one leg squat for leg muscular function.

Conclusions: This reliability assessment provided useful information on the characteristics of potential test items in a HRF test battery for adults and on the limitations of its practical use. Testers must be properly trained to ensure reliable assessment of HRF of adults.

PROMOTING PHYSICAL ACTIVITY is considered to be an important public health measure in Western societies. To some extent, physical fitness is an objective marker for physical activity, and fitness can be measured more accurately than physical activity.[1,2] Health-related fitness (HRF) has recently been recognized as a distinct type of fitness with special relevance to the health potential of physical activity.[3,4] When fitness testing is used as a tool to promote health-related exercise in large populations, the methods need to be simple, practical, and safe under conditions available in ordinary communities.[5] In addition, the methods must be reliable to obtain valid and useful information about the fitness level of individuals or populations.

Several field test batteries of HRF have been developed in recent years: AAHPERD,[6] Fitness Canada,[7] Eurofit.[8] The reliability of subtests of AAHPERD and Eurofit, as well as an entire test battery[9] designed to test school children, have been thoroughly studied. The reliability of the AAHPERD functional fitness assessment in older men and women was recently reported.[10] In adults, the reliability of selected single tests has been evaluated (see Skinner and Oja[5]), but to our knowledge, no studies of the reliability of several potential items in a HRF test battery for adults have been published.

To achieve inter-rater reliability, 2 observers, following the standard measurement procedure, bave to obtain consistent results. If trained testers cannot agree, the assessment procedure lacks objectivity and utility. Test-retest reproducibility or consistency over time is another aspect of reliability, and is critical for evaluating whether an observed change is real. Research is required to distinguish the instability caused by unreliable measurements from the instability caused by the phenomena being measured.[11] Trial-to-trial reproducibility is determined to reveal the immediate effects of test repetition (eg, learning) on results.

Our Institute for Health Promotion Research is developing a fitness test battery for adults designed to be a tool for health-related exercise promotion. The components of the health-related fitness test battery are body composition, cardiorespiratory fitness, motor fitness (coordination, balance), and musculoskeletal fitness (flexibility, muscular strength and endurance). This article presents the results of the reliability assessments of motor and musculoskeletal fitness tests proposed to this test battery. In 2 consecutive studies, we investigated the inter-rater reliability, test-retest and trial-to-trial reproducibility of 3 balance, 2 flexibility, and 4 muscular strength and endurance tests.

METHODS

Subjects

In study A, the subjects were volunteers from 2 worksites. Twenty-two men and 20 women were chosen from different age (mean 41.9 years, range 25 to 59 years) and occupational groups. In study B, a sample of an urban population, 250 men and 260 women, was selected from residents 37, 42, 47, 52, and 57 years of age in Tampere, Finland. About 50 men and 50 women of each 5 age cohorts were selected. All together, 499 subjects participated in fitness testing.

The ethical committee of our Institute approved the study. An informed consent was obtained from each subject prior to their inclusion in the study.

Fitness Assessment

The selection of motor and musculoskeletal test items to the health-related fitness test battery was based on a literature review on the method for fitness assessment with special reference

From the UKK Institute for Health Promotion Research (Ms. Suni, Dr. Oja, Dr. Laukkanen, Dr. Miilunpalo, Dr. Pasanen, Dr. Vuori, Dr. Vartiainen), Tampere, Finland, and the Sports Research Institute of the University of Frankfurt (Dr. Bös), Germany.

Submitted for publication June 15, 1995. Accepted in revised form September 20, 1995.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to Jaana Suni, MSc, UKK Institute, PO Box 30, FIN-35501 Tampere, Finland.

0003 9993/96/7704-3561$3.00/0

to reliability and health-related validity. The reliability of the cardiorespiratory fitness test in the battery, the UKK 2-km walking test,[12] developed by the UKK Institute, has been previously established[13,14] and was not further evaluated in this study.

The testing procedures of the motor and musculoskeletal tests were as follows:

Motor Fitness: Balance. *Standing on one leg with eyes open (1), eyes closed (2), head turns (3).* Three tasks of one leg stand[15-18] were used to measure the efficiency of postural control under different sensory conditions (visual, vestibular, proprioceptive). The subjects wore sport shoes. They placed the heel of the opposite foot at the level of knee joint against the inner side of the supporting leg and rotated the thigh outwards. Arms hung relaxed at the subject's sides. In the eyes closed condition, the test position was taken before the eyes were closed, and in the head turns condition, the test position was taken before subjects started turning their head from side to side in sequence with the metronome (50 times per minute to one side).

The subjects were first instructed to familiarize themselves with the balance position with both legs and to choose the leg they felt to be better able to perform the task. Subjects were advised to stand in the position as long as possible, as quietly as possible, and to use arm movements for balance only when necessary.

The balance time of standing on one leg in each task was measured in seconds with a stopwatch. Changing the position of the supporting leg or losing knee contact with the heel were the end points for the task, as well as opening the eyes, or interrupting the head turns. Sixty seconds was the upper limit for the task with the eyes open. Thirty seconds was the upper limit for the eyes closed and head turns tasks. Two trials were recorded with the subject's eyes open, unless the time limit of 60 seconds was reached in the first trial. Two to three trials were recorded with the eyes closed and head turns unless the time limit of 30 seconds was reached in the first trial.

Musculoskeletal Fitness: Flexibility, Upper Body. *Shoulder-neck mobility.* Functional shoulder-neck mobility restrictions were estimated by a visual observation method. The subjects stood with their backs against the wall; the feet were placed at a distance of $1\frac{1}{2}$ foot lengths from the wall. The buttocks, back and shoulders rested against the wall.

The subjects were instructed to raise their hands above their head as far as possible while keeping their upper arms close to the ears. The backs of the hands were turned against the wall and the elbows were kept straight.

The testers made an ordinal scale estimation of the restrictions on functional movement by observing the final position of the hands against the wall. Results were separately scored for the right and left sides. The classification criteria were as follows: 0 = severe restriction of range of motion (ROM), no hand contact with the wall; 1 = moderate restriction of ROM, only the fingers reach the wall; 2 = no restriction of ROM, the whole dorsal side of the hand is in contact with the wall.

Musculoskeletal Fitness: Flexibility, Trunk. *Side-bending* to the right and left was used to measure the total range of movement of lateral flexion of the thoracic and lumbar spine and pelvis.[19-21] Two parallel lines, 15cm apart, were marked on the floor right in front of a wall. The subjects stood on the marked lines with their backs against the wall, the scapula and buttocks resting against the wall. Arms were kept straight at the sides of the body. The position of the tip of the middle finger for both sides was marked with a horizontal line on the lateral thigh, first in the upright position, and then at the end point of the lateral flexion movement. No rotation of the trunk or movement of the pelvic area was allowed, and the heels had to stay in contact with the floor.

The subjects were instructed to bend to the right and then to the left as far as possible while keeping their shoulder blades and buttocks in contact with the wall. The middle finger slid down along their lateral thigh.

The test score was the distance the fingertip moved down the leg during maximum lateral bending. A cloth tape was used to measure the distance between the markings on the thigh. The results were recorded for both sides, added together and averaged for the mean side-bending score.

Musculoskeletal Fitness: Muscular Strength, Upper Body. *A modified push-up test* was developed to measure the short term endurance capacity of the upper extremity extensor muscles, and the ability to stabilize the trunk. The subjects lay prone on a mat. The push-up cycle began by clapping the hands behind the back to standardize the beginning of each cycle. Next the subject performed a normal straight-leg push-up with the elbows completely straight in the up position, followed by a touch of one hand on the top of the supporting hand to standardize the up position. The cycle ended in the prone lying position. The subjects practiced one push-up cycle.

The subjects were instructed to do as many push-ups as they could during 40 seconds. The number of push-ups completed in 40 seconds was counted.

Musculoskeletal Fitness: Muscular Endurance, Trunk. *Isometric sit-up test* was used to measure the isometric endurance capacity of the trunk flexor muscles and the ability to stabilize the trunk, modified after Hyytiäinen and coworkers.[22] The test was performed in an unsupported, straight back sit-up position. The subjects were carefully guided into a straight-back sit-up position with bent knees (about 90°) and the feet placed flat on the floor, arms at the sides, fingertips lightly touching the mat. Flexion of the lumbar spine was not allowed. The correct thigh-trunk angle was 90°, and it was controlled with the aid of a "cardboard model" during the study.

The subjects were instructed to stay in the test position as long as possible for up to 240 seconds. They were informed about the time every 30 seconds.

Endurance time was measured as the number of seconds the subjects were able to keep the position. Zero was recorded if the subjects were unable to reach the correct test position or were not able to keep it stable for 1 second.

Musculoskeletal Fitness: Muscular Strength, Legs. *Jump and reach test* was used to measure the leg extensor power.[23] The task was to jump as high as possible. Subjects stood beside the jump-and-reach board facing forward. They were allowed to flex their knees, but not move their feet while preparing for the jump. Subjects were advised to touch the board with their middle finger while at the highest position. One pretrial of the jump was allowed for practice. Before jumping, the dominant arm was raised straight up as high as possible to mark the standing height with the magnesium-powdered middle finger.

The subjects were instructed to jump as high as possible and to swing their arms to enhance the performance. The difference between the reach height and the jump height was measured in centimeters with a tape measure. Two test trials were performed and the best result was recorded.

One-leg squat test was developed to assess the functional restrictions in the lower extremity extensor strength. The purpose was to determine the load limit for a successful one-leg squat task for the right and left leg measured as maximum weight relative to the subject's body weight up to 130%. A two-leg squat down to 90° knee flexion was performed before the one-leg trials. The test squats began at body weight. Next, 10% of body weight was added at each successive step, up to

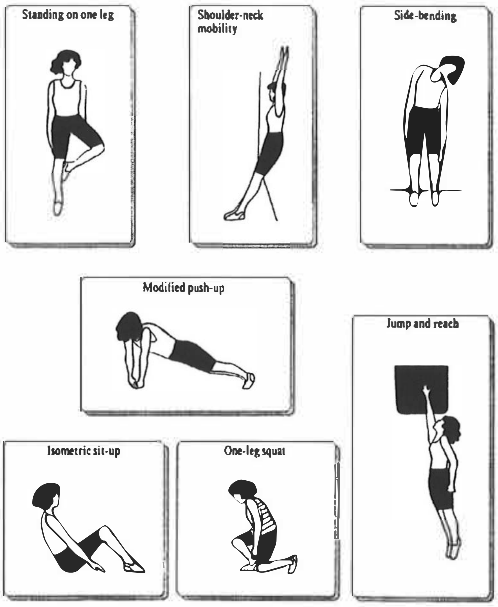


Fig 1. The motor and musculoskeletal fitness tests.

30%. The test ended when the subjects felt they could not step with any more weight. A weight belt was used to add extra weights.

The subjects were instructed to take a short step forward on the mat, first with their right leg, and then squat down with a straight back until their left knee lightly touched the mat, and raise up immediately to the starting position. The squat was then repeated with the left leg.

Results for the right and left sides were added together. The load limit for a successful one-leg squat task for the right and left legs was rated as follows: 0 = unable to perform squat with two legs; 1 = able to perform squat with two legs; 2 = able to perform one-leg squat with body weight; 3 = able to perform one-leg squat with an extra load of 10% of body weight; 4 = able to perform one-leg squat with an extra load of 20% of body weight; 5 = able to perform one-leg squat with an extra load of 30% of body weight.

The fitness tests are presented in figure 1. All tests followed a standard sequence: motor fitness was tested first, followed by flexibility and muscular strength tests. No warm-up was used and the subjects were not allowed to perform preliminary trials of the test unless explicitly instructed to do so. Each subject was tested individually and specific instructions were given for each test. The subjects were encouraged by the tester for best performance in a consistent manner.

Assessment of Reliability

We investigated the inter-rater reliability and test-retest reproducibility by administering the test battery twice to the subjects in study A. The time interval between the sessions was from 6 to 8 days. Of the 42 subjects, 40 participated in both testing sessions. The subjects were randomly divided into 2 groups (n = 20 + 20), each group consisting of 10 men and 10 women. Every subject participated in balance and flexibility tests on both testing days. Half of the subjects in both groups (n = 10 + 10) took part in muscular strength and endurance tests on both testing days, while the other half participated in parallel forms of strength tests in laboratory conditions during the second testing day. The results concerning the criterion-related validity have been reported elsewhere.[24] Two trained testers evaluated the subjects during the two testing days, according to the protocol presented in table 1.

The inter-rater reliability of the eyes open balance test was reassessed with 48 randomly selected subjects in study B. Three trained testers were grouped into three pairs of raters. In simultaneous evaluations, each pair rated the test performance of 16 subjects. For 8 of the 16 subjects other tester gave the instructions while both rated the performance. The testers reversed their roles for the second group of 8 subjects. Trial to trial reproducibility of the more difficult balance tests (standing on one leg with eyes closed and head turns) during one measurement session were studied for the entire cross-sectional sample.

Statistical Methods

The mean and standard deviation (SD) or the frequency distribution is presented as descriptive statistics. The intraclass correlation coefficient (ICC) of repeated interval scale measures was determined as the measure of inter-rater reliability.[25] A one-way analysis of variance (ANOVA) was used to ICC, taking into account the score variability between subjects and raters, and the difference in measurement levels between the raters. For individual measurements, the degree of measurement error was expressed as the standard error of measurement (SEM).[25,26] The inter-rater reliability for ordinal scale measures was determined with a weighted Kappa (κ_w) coefficient, which has been used to examine the error patterns of ratings; the more serious the error the greater the weight.[27]

The mean difference with a 95% confidence interval (CI) between the testing days was determined as a measure of test-retest reproducibility.[28,26] As an additional measure of test-retest reproducibility, the coefficient of variation (CV = [SD/mean] × 100%), which represents the relative magnitude of variation as a percentage of the score variability from day to day, was calculated. The test-retest reproducibility of the ordinal scale measurements was evaluated by calculating the ratio of discordant pairs with a 95%CI between the test days. These measures reveal the pattern of the effect of test repetition.

Trial-to-trial reproducibility during one measurement session was determined with the mean difference as described with the measures of test-retest reproducibility.

For evaluation of the correlation coefficients as measures of reliability we used previously reviewed scales based on our ICC[29] and κ_w values[27]: ICC values .90-.99, high; .80-.89, good; .70-.79, fair, ≤.69, poor; κ_w values ≥ 75, excellent; .41-.74, fair to good; ≤.40, poor.

RESULTS

Descriptive data for the reliability assessments of fitness tests is presented in table 2. The results of the inter-rater reliability

Table 1: Test Protocol for Reliability Assessment, Study A

Fitness Component	Testing Day	Rater 1	Rater 2
Motor Fitness			
Balance	1st testing day	Subjects 1-20	Subjects 21-40
	2nd testing day	Subjects 21-40	Subjects 1-20
Musculoskeletal Fitness			
Flexibility	1st testing day	Subjects 1-20	Subjects 21-40
	2nd testing day	Subjects 21-40	Subjects 1-20
Muscular Strength	1st testing day	Subjects 1-10	Subjects 21-30
and Endurance	2nd testing day	Subjects 21-30	Subjects 1-10

Table 2: Descriptive Data for Reliability Assessment of the Motor and Musculoskeletal Fitness Tests

Test Item	Study A								Study B							
	1st Day				2nd Day				1st Rater*/Trial†				2nd Rater*/Trial†			
Standing on one leg (s)	Mean	SD	*n*		Mean	SD	*n*			Mean	SD	*n*		Mean	SD	*n*
Eyes open	42.7	20.8	42		45.6	21.5	40		Rater 1	44.2	21.0	48	Rater 2	44.0	21.1	48
Eyes closed	5.6	4.7	41		6.4	5.2	40		Trial 1	4.6	4.6	494	Trial 2	5.3	5.3	491
Head turns	6.3	6.0	41		8.7	8.9	40		Trial 1	5.5	5.8	494	Trial 2	6.0	6.0	489
Shoulder-neck mobility	S‡	M§	N‖	(*n*)	S	M	N	(*n*)								
Right	21%	29%	50%	(42)	18%	27%	55%	(40)								
Left	21%	31%	48%	(42)	20%	25%	55%	(40)								
Side-bending (cm)	Mean	SD	*n*		Mean	SD	*n*									
Right	22.1	4.4	42		21.4	4.9	40									
Left	21.4	4.9	41		20.8	5.0	39									
Average	21.7	4.4	41		21.0	4.8	39									
Modified push-ups (repetitions)	11.1	5.1	21		13.6	5.5	19									
Isometric sit-up (sec)	34.7	31.4	22		41.3	33.4	20									
Jump and reach (cm)	32.0	13.4	22		33.1	13.7	20									
One-leg squat (points 0-10)	8.3	1.3	22		8.2	2.0	20									

* Inter-rater comparisons.
† Trial-totrial comparisons.
‡ Severe restrictions of range of motion (ROM).
§ Moderate restrictions of ROM.
‖ No restrictions of ROM.

and test-retest reproducibility assessments are presented in table 3.

Inter-Rater Reliability

The inter-rater ICC values were high for the trunk side-bending (.90 and .92, with SEM ranging from 1.4 to 1.7cm) and for the jump and reach (.98, with SEM 3.0cm), and good for the modified push-ups (.88, with SEM 2.6 repetitions) and for the one-leg squat tests (.86, with SEM 0.9 points). Fair ICC values were obtained for the isometric sit-up test (.76, with SEM 20.1 seconds) and the standing on one leg with eyes open test in study A (.76, SEM 13.3 seconds). Better results were obtained for this balance test in study B (ICC = 1.0, SEM 0.3 seconds). Inter-rater ICC values were poor for the balance tasks with eyes closed and head turns (.18 and .28, respectively). κ_w coefficients for the shoulder-neck mobility test were good (.61 and .62).

Test-Retest Reproducibility

The reproducibility (CV) of the tests over one week ranged from 0.6% to 12.1% (table 3). In general, the values of the mean differences were small (table 3), but when the 95% confidence intervals of the mean differences were compared to the descriptive mean values of the first measurement day (table 2) large variability was discovered for standing on one leg with eyes closed (95% CI from −1.6 to 2.8 seconds), head turns (from −0.7 to 5.9 seconds), and isometric sit-up test (from −5.7 to 21.5). A moderate amount of variability was discovered for the one leg standing with eyes open test (mean difference between testing days 3.7 seconds with a 95% CI from −2.2 to 9.6). Very small mean differences with narrow confidence intervals were obtained for the average side-bending (mean difference −0.5cm with a 95% CI from −1.3 to 0.3), jump and reach (mean difference 1.7cm with a 95% CI from −0.2 to 3.6), and one leg squat test (mean difference −0.1 points with a 95% CI from −0.7 to 0.5).

The modified push-up test results improved from the first measurement day to the second measurement day (mean difference 3.0 with a 95% CI from 2.1 to 3.9). Less restriction in shoulder-neck mobility was recorded during the second testing

Table 3: Inter-Rater Reliability and Test-Retest Reproducibility of the Motor and Musculoskeletal Fitness Tests

Fitness Test	Inter-Rater Reliability*†			Test-Retest Reproducibility*		
Balance						
Standing on one leg (sec)	*n*	ICC	SEM	*n*	Mean difference (95% CI)	CV (%)
Eyes open	40	.76*	13.3	40	3.7 (−2.2 to 9.6)	5.0
	48	1.00†	0.3			
Eyes closed	39	.18*	4.6	39	0.6 (−1.6 to 2.8)	10.9
Head turns	39	.28*	7.5	39	2.6 (−0.7 to 5.9)	4.0
Flexibility						
Upper body: Shoulder-neck mobility	*n*	κ_w		*n*	Ratio of discordant pairs (95% CI)	
Right	40	.61*		40	2.3 (0.7 to 7.1)	
Left	40	.62*		40	2.3 (0.7 to 7.1)	
Trunk: Side-bending (cm)	*n*	ICC	SEM	*n*	Mean difference (95% CI)	CV (%)
Right	40	.90*	1.6	40	−0.5 (−1.3 to 0.3)	4.7
Left	39	.90*	1.7	39	−0.5 (−1.4 to 0.5)	6.2
Average	39	.92*	1.4	40	−0.5 (−1.3 to 0.3)	4.7
Muscular Strength and Endurance						
Upper body: Modified push-up (repetitions)	19	.88*	2.6	19	3.0 (2.1 to 3.9)	0.6
Trunk: Isometric sit-up, (sec)	20	.76*	20.1	20	7.9 (−5.7 to 21.5)	3.7
Legs: Jump and reach (cm)	20	.98*	3.0	20	1.7 (−0.2 to 3.6)	2.4
Legs: One leg squat (points)	20	.86*	0.9	20	0.1 (−0.7 to 0.5)	12.1

Abbreviations: ICC, intraclass correlation coefficient; SEM, standard error of measurement; CI, confidence interval; CV, coefficient of variation; κ_w, weighted Kappa coefficient.
* Study A.
† Study B.

day, the ratio of discordant pairs being 2.3 with a 95% CI from 0.7 to 7.1. No other systematic changes in the test results between the first and second measurement day were observed.

Trial-to-Trial Reproducibility

The balance times of standing on one leg with eyes closed and with head turns from trial to trial were slightly longer for the second trial. The mean difference was 0.9 seconds with a 95% CI from 0.4 to 1.4 for the eyes closed task ($n = 491$) and 0.6 seconds with a 95% CI from 0.0 to 1.1 for the head turn task ($n = 489$).

DISCUSSION

General Aspects

We studied different aspects of reliability of several motor and musculoskeletal fitness tests proposed to form a health-related fitness test battery for adults. The amount of reliability that is necessary for a fitness test depends on the purpose and use of the measurement. We had two goals in mind while performing our reliability assessments. For scientific purposes we needed objective fitness measures that are reproducible over time to investigate the interrelationships between physical activity, fitness, and health. On the other hand, this health-related fitness test battery was also designed to be a practical tool for promoting health-enhancing exercise in adult populations. For this purpose, the reliability assessment provides useful information on the limitations of the practical use of the tests and proper interpretation of the test results. In practice, fitness testers make comparisons between individuals or they monitor the changes in fitness over time. They need to be educated to be aware of the errors in their measurement in order to ensure the quality of fitness testing and proper interpretation of the results.

In individual fitness testing, the SEM values of our study could be interpreted as the minimum difference in the test results between two subjects measured by different testers that indicate a real difference between their fitness levels. In a similar manner, the 95% confidence intervals between the mean differences of the two measurement sessions could be interpreted as a minimum difference between the results of individuals that indicate a real change in their fitness level over time.

Single Fitness Test Items of the HRF Test Battery

There is no single test that could serve as a global measure of postural control or balance.[30] Standing on one leg tests are proposed to have validity in relation to falls in the elderly[31] and possibly to back pain and injury in middle-aged adults.[32-34] The inter-rater reliability of the standing on one leg with eyes open test was fair in study A, but improved to high in study B. The fitness testers reevaluated the possible standardization problems of all of the three balance tasks after the first study (A), and learned to take the measurements in a more standardized way in study B. We did not reassess inter-rater reliability of the eyes closed and head turn tasks in study B, even though they showed very poor reliability in study A. We believe that the difficulty in determining the exact moment the test begins and the need for repeated trials were the two main factors that decreased the reliability of the results in study A. To ensure the best possible performance, subjects in the study B were always instructed to repeat the trial if they lost their concentration or were disturbed by outside noises. This was not allowed in study A. As other researches have pointed out, it is useful to determine the need for repeated trials and whether the subject's performance improves over trials.[35]

Stones and Kozma[16] reported a 1-year test-retest correlation of .68 for the standing on one leg with eyes open test, and .32 for the eyes closed test. We used the same test forms as they used, but used different statistics to study test-retest reproducibility. Only the eyes open task had acceptable reproducibility in the first study (A). In study B, the reproducibility from trial to trial for eyes closed and head turns tasks showed small mean differences and narrow 95% confidence intervals, indicating small learning effects from trial to trial and good reproducibility within one rater.

In older adults, mobility restriction of shoulder joint often cause limitation of daily activities. We chose to estimate functional shoulder-neck mobility restrictions by observation method with an ordinal scale. The test was developed to be a simple screening test for subjects with severe functional mobility restrictions in shoulder joints or cervicothoracic spine. The test showed acceptable inter-rater reliability. Viikari-Juntura[36] reported fair inter-rater κ_w coefficients for cervical range of motion estimations with a similar type of ordinal scale. Test repetition within a week seemed to improve the shoulder-neck mobility results. This limits the use of the test in the assessment of change since the test only has a 3-point scale.

Restricted spinal mobility has been proposed as a risk factor for low back trouble,[37-39] but the protective role of mobility against back problems has not been proven.[40,41] We selected side-bending to the test battery, because evidence has indicated that among the simple clinically used tests it gives the best association with low back-pain.[42] In a study by Frost and coworkers,[43] side-bending had the next highest single measurement (rater × day × repetition) reliability ($r = .70$) after forward bending. Mellin[19] and Hyytiäinen and coworkers[22] reported almost identical inter-rater reliability using Pearson correlation coefficients ($r = .84$-.88). Rose[44] reported the least significant difference (LSD) for intra-observer test repetition to be quite large (right, 3.0cm; left, 4.0cm). In our study the inter-rater ICC value was very high (.92) with a low SEM of 1.4cm. We used ICC values to indicate inter-rater reliability, because ICC is generally considered a more proper method than the Pearson correlation coefficient. The reproducibility of the test was very good (CV 4.7%, mean difference −0.5cm with 95% CI from −1.3 to 0.3).

Push-up tests measure the muscular endurance capacity of the upper body. We used a modified form of the push-up test to improve the test standardization. Other researchers have reported difficulties in differentiating between correct and incorrect performances of conventional push-ups in a straight-leg position for men and a bent knee position for women.[45] Our results showed acceptable inter-rater reliability and small test-retest variation in the CV measure. However, there seemed to be a clear learning effect between the first and the second testing day. This may be due to the complexity of the test performance. The subjects need to practice the performance before testing, but this might also affect the results by causing undue fatigue before actual testing, especially in subjects with poor strength.

Trunk muscle endurance capacity is the best documented factor of fitness in relation to back health.[46-49] Our method for testing the endurance capacity of the trunk flexion muscles was adapted from the test used by Hyytiäinen and coworkers.[22] They reported an interobserver correlation coefficient of .90 for this isometric sit-up test, and concluded that the reproducibility of the test was good. Our results showed only fair inter-rater reliability with a large SEM. The variability of test-retest results was also large. The subjects in the Hyytiäinen were all middle-aged men, so the better reliability results they obtained could be a function of a more homogeneous study population. Another reason might be the different statistics used to assess reliability. In our experience, the standardization of the test position was difficult, and motivational factors had a substantial effect on the results.

The jump and reach test had the best reliability of all the tests studied. Even though good coordination was necessary for successful performance, the reproducibility of the test was not a problem. We believe this was because of the natural movement pattern of the performance. The subjects flexed their knees to prestretch their leg extensors, and swung their arms during the jump. The jump and reach test was the only test in our battery in which speed of movement was required.

We developed the one-leg squat test for the assessment of functional leg extensor strength. This test was thought to correlate with the everyday needs of physical exertion, including stair climbing or lifting loads. The test proved to be reliable.

In summary, we evaluated the reliability of simple motor and musculoskeletal fitness tests proposed to a HRF test battery for adults. On the basis of these results, the following tests appeared to provide acceptable reliability for field testing of fitness: standing on one leg with eyes open for balance, side-bending of the trunk for spinal flexibility, modified push-ups for upper body muscular function, and jump and reach or one leg squat for leg muscular function. In addition to the use of reliable assessment methods, we emphasize the need for proper training of testers to ensure reliable and useful HRF testing and interpretation of the results in the context of exercise promotion for health.

References

1. Blair SN, Kohl HW, Paffenbarger RS, Barlow CE. Physical activity, physical fitness, and all-cause mortality in women. J Am Coll Nutr 1993;12:368-71.
2. Blair SN. 1993 C.H. McCloy research lecture: physical activity, physical fitness, and health. Res Quart Exerc Sport 1993;64(4):365-76.
3. Bouchard C, Shephard RJ, Stephens T, Sutton JR, McPherson BD. Exercise, fitness, and health: the consensus statement. In: Bouchard C, Shephard RJ, Stephens T, Sutton JR, McPherson BD, editors. Exercise, fitness, and health. Champaign, IL: Human Kinetics Books, 1990:3-26.
4. Bouchard C, Shephard RJ. Physical activity, fitness, and health: the model and key concepts. In: Bouchard C, Shephard RJ, Stephens T, editors. Physical activity, fitness, and health: international proceedings and consensus statement. Champaign, IL: Human Kinetics Books, 1994:77-88.
5. Skinner J, Oja P. Laboratory and field tests for assessing health-related fitness. In: Bouchard C, Shephard RJ, Stephens T, editors. Physical activity, fitness, and health: international proceedings and consensus statement. Champaign, IL: Human Kinetics Books, 1994:160-79.
6. AAHPERD. Technical manual. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance, 1984.
7. Fitness Canada. Canadian standardized test of fitness operations manual, fitness and amateur sport. Ottawa: Fitness Canada, 1986.
8. Eurofit. Handbook for the EUROFIT tests of physical fitness. Rome, Italy: Council of Europe, committee for the development of sport. Edigraf editoriale grafica, 1988.
9. Safrit MJ, Wood TM. The test battery reliability of the health-related physical fitness test. Res Quart Exer Sport 1987;58:160-7.
10. Shaulis D, Golding LA, Tandy RD. Reliability of the AAHPERD functional fitness assessment across multiple practise sessions in older men and women. J Aging Phys Act 1994;2:2729.
11. Johnston MV, Keith RA, Hinderer SR. Measurement standards for interdisciplinary medical rehabilitation. Arch Phys Med Rehabil 1992;73 Suppl:3S-23S.
12. Laukkanen R, Hynninen E. Guide for the UKK 2-km walking test. Tampere, Finland: President Urho Kaleva Kekkonen Institute for Health Promotion Research, 1993.
13. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. Int J Sports Med 1991;12:356-62.
14. Laukkanen R. Development and evaluation of a 2-km walking test for assessing maximal aerobic power of adults in field conditions [dissertation]. Tampere, Finland: Kuopio University Publications D. Medical Sciences 23, 1993.
15. Bohannon RW, Larkin PA, Cook AC, Gear J, Singer J. Decrease in timed balance test scores with aging. Phys Ther 1984;64:1067-70.
16. Stones M, Kozma A. Balance and age in the sighted and blind. Arch Phys Med Rehabil 1987;68:85-9.
17. Briggs RC, Gossman MR, Birch R, Drews J, Shaddeau S. Balance performance among noninstitutionalized elderly women. Phys Ther 1989;69:748-56.
18. Johansson G, Jarnlo G-B. Balance training in 70-year-old women. Physiother Theory Pract 1991;7:121-5.
19. Mellin G. Accuracy of measuring lateral flexion of the spine with a tape. Clin Biomech 1986;1:85-9.
20. Battie MC, Bigos SJ, Sheehy A, Wortley M. Spinal flexibility and individual factors that influence it. Phys Ther 1987;67:653-8.
21. Alaranta H, Soukka A, Harju R, Heliövaara M. Selän ja niskahartiaseudun suoritustestistö työterveyshuollon terveystarkastuksiin. Työsuojelurahaston julkaisuja C21. Helsinki: Invalidisäätiö, 1990.
22. Hyytiäinen K, Salminen J, Suvitie S, Wickström G, Pentti J. Reproducibility of nine tests to measure spinal mobility and trunk muscle strength. Scand J Rehabil Med 1991;23:3-10.
23. Beuker F. Leistungsprüfungen im freizeit und erholungssport. Leipzig, Germany, 1976.
24. Suni J, Oja P, Laukkanen R, Miilunpalo S, Pasanen M, Vuori I. Evaluation of a fitness test battery for assessment of health-related fitness in field studies. In: Claessens AL, Lefevre J, Vanden Eynde B, editors. World-wide variation in physical fitness. Proceedings of the 1992 symposium of the international council for physical fitness research: 1992 July 10-13; Leuven, Belgium. Leuven, Belgium, 1993:160-8.
25. Baumgartner TA. Norm-referenced measurement: reliability. In: Safrit MJ Wood TM, editors. Measurement concepts in physical education and exercise science. Champaign, IL: Human Kinetics Books, 1989:45-72.
26. Roebroeck ME, Harlaar J, Lankhorst GF. The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther 1993;73:386-401.
27. Haley SM, Osberg JS. Kappa coefficient calculation using multiple ratings per subject: a special communication. Phys Ther 1989;69:970-4.
28. Altman DG, Gardner MJ. Calculating confidence intervals for means and their differences. In: Gardner MJ, Altman DG, editors. Statistics with confidence. London: British Medical Journal, 1989:20-7.
29. Youdas JW, Carey JR, Garret TR. Reliability of measurements of cervical spine range of motion—comparison of three methods. Phys Ther 1991;71:98-106.
30. Horak FB. Clinical measurement of postural control in adults. Phys Ther 1987;67:1881-5.
31. Gehlsen BM, Whaley MH. Falls in the elderly: part II, balance, strength, and flexibility. Arch Phys Med Rehabil 1990;71:739-41.
32. Byl NN, Sinnot PL. Variations in balance and body sway in middle-aged adults: subjects with healthy backs compared with subjects with low-back dysfunction. Spine 1991;16:325-30.
33. Videman T, Rauhala H, Asp S, Lindström K, Cedercreutz G, Kämppi M, et al. Patient-handling skill, back injuries and back pain: an intervention study in nursing. Spine 1989;14:148-56.
34. Oddsson LI. Control of voluntary trunk movements in man: mechanism for postural equilibrium during standing. Acta Physiol Scand 1990;140 Suppl:595.
35. Cohen H, Blatchly CA, Gombash LL. A study of the clinical test of sensory interaction and balance. Phys Ther 1993;73:346-54.
36. Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. Phys Ther 1987;67:1526-32.
37. Frymoyer JW, Cats-Baril W. Predictors of low-back pain disability. Clin Orthop 1987;221:89-98.
38. Burton AK, Tillotson KM, Troup JDG. Prediction of low-back trouble frequency in a working population. Spine 1989a;14:939-46.

39. Riihimäki H. Low-back pain, its origin and risk indicators. Scand J Work Environ Health 1991;17:81-90.
40. Burton AK, Tillotson KM, Troup JDG. Variation in lumbar sagittal mobility with low-back trouble. Spine 1989b;14:584-90.
41. Battie MC, Bigos SJ, Fisher LD, Spengler DM, Hansson TH, Nachemson AL, et al. The role of spinal flexibility in back pain complaints within industry. Spine 1990;15:768-73.
42. Mellin G. Correlations of spinal mobility with degree of chronic low-back pain after correction for age and anthropometric factors. Spine 1987;12:464-8.
43. Frost M, Stuckey S. Smalley L, Dorman G. Reliability of measuring trunk motions in centimetres. Phys Ther 1982;62:1431-7.
44. Rose MJ. The statistical analysis of the intra-observer repeatability of four clinical measurement techniques. Physiotherapy 1991; 77:89-91.
45. Invergo JJ, Ball TE, Looney M. Relationship of push-ups and absolute muscular endurance to bench press strength. J Appl Sport Sci Res 1991;5:121-5.
46. Biering-Sorensen F. Physical measurements as risk indicators for low back trouble over a one-year period. Spine 1984;9:106-18.
47. Leino P, Aro S, Hasan J. Trunk muscle function and low-back disorders: a ten-year follow-up study. J Chronic Dis 1987;40:289-96.
48. Holmström E, Moritz U, Andersson M. Trunk muscle strength and back muscle endurance in construction workers with and without low back disorders. Scand J Rehabil 1992;24:3-10.
49. Moffet JAK, Hughes GI, Griffiths P. A longitudinal study of low back pain in student nurses. Int J Nurs Stud 1993;30:197-212.

II

Safety and feasibility of a health-related fitness test battery for adults

by
Jaana Suni, Seppo Miilunpalo, Tuula-Maria Asikainen, Raija Laukkanen, Pekka Oja
Matti Pasanen, Klaus Bös, Ilkka Vuori

Physical Therapy 1998; 78:134-148

https://doi.org/10.1093/ptj/78.2.134

III

Health-related fitness test battery for adults: associations with perceived health, mobility, and back function and symptoms

by

Jaana Suni, Pekka Oja, Seppo Miilunpalo, Matti Pasanen, Ilkka Vuori, Klaus Bös

Archives of Physical Medicine and Rehabilitation 1998; 79:559-569

https://doi.org/10.1016/S0003-9993(98)90073-9

Health-Related Fitness Test Battery for Adults: Associations With Perceived Health, Mobility, and Back Function and Symptoms

Jaana H. Suni, MSc, Pekka Oja, PhD, Seppo I. Miilunpalo, MD, Matti E. Pasanen, MSc, Ilkka M. Vuori, MD, Klaus Bös, Prof

ABSTRACT. Suni JH, Oja P, Miilunpalo SI, Pasanen ME, Vuori IM, Bös K. Health-related fitness test battery for adults: associations with perceived health, mobility, and back function and symptoms. Arch Phys Med Rehabil 1998;79:559-569.

Objective: To evaluate the health-related content validity of nine fitness tests by studying how low, mid, and high levels of fitness are associated with perceived health and musculoskeletal functioning.

Design: Cross-sectional methodological study.

Setting: A research institute for health promotion.

Participants: Middle-aged (37 to 57 years) men (n = 245) and women (n = 253), evenly selected from five age cohorts of a random population sample.

Main Outcome Measures: The odds ratios (ORs) of selected health outcomes for low (least fit 20%), mid (next 40%), and high (most fit 40%) fitness categories in the different tests adjusted for several possible confounders.

Results: Cardiorespiratory fitness, as measured by 2-km walk test, was strongly and consistently associated with perceived health and mobility (stair climbing) in both genders (range of ORs, 2.4 to 17.6), and a somewhat weaker relationship was found with leg power and with leg strength (ORs, 2.5 to 7.2). Low fitness in back muscular endurance and upper-body strength were associated with mobility disability (ORs, 2.8 to 8.5) and with back dysfunction and pain (ORs, 2.9 to 6.1). High fitness in back endurance in men and in balance in women were related to positive back health (ORs, 2.5 to 3.7). Body mass index was associated with musculoskeletal disability in women (ORs, 2.4 to 5.3). Balance, leg strength, and leg flexibility in men; and leg power, trunk and leg flexibility in women were not associated with health outcomes.

Conclusions: Among a middle-aged population, the majority of the evaluated fitness tests demonstrated health-related validity by strong associations with perceived health and musculoskeletal functioning, and by weaker associations with back symptoms.

From the Urho Kaleva Kekkonen Institute for Health Promotion Research, Tampere, Finland (Ms. Suni, Dr. Oja, Dr. Miilunpalo, Mr. Pasanen, Dr. Vuori), and the Sports Research Institute of the University of Frankfurt, Frankfurt, Germany (Prof. Bös).

Submitted for publication May 30, 1997. Accepted in revised form October 23, 1997.

Supported by the Finnish Ministry of Education.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

Reprint requests to Jaana Suni, MSc, UKK Institute, PO Box 30, FIN-35501 Tampere, Finland.

0003-9993/98/7905-4523$3.00/0

THE PROMOTION OF health-related physical activity has gained importance as a preventive public health measure[1,2] and as a means to maintain work ability[3] and functional capabilities.[4] Increasing scientific evidence, although inconclusive,[5,6] indicates that in addition to promoting aerobic exercise and fitness for cardiovascular[1,2] and metabolic[1,7] health, the enhancement of motor and musculoskeletal fitness maintains functional capability,[1,4] may prevent falls[1,8] and reduce the risk for osteoporotic fractures among the elderly[1,9] and decreases the burden of back-related disability among the working age population.[10,11]

Based on the known relationships between physical activity, fitness, and health, a new fitness concept—health-related fitness (HRF)—has been introduced.[12] The components of HRF are cardiorespiratory, motor, and musculoskeletal fitness; body composition; and metabolism.[12,13] According to this conceptual model[12] several fitness components are associated with various health outcomes. For the measurement of cardiorespiratory fitness, body composition, and metabolism, standard laboratory methods are available. To evaluate and monitor the fitness level of the general population, however, field measures of HRF that are reliable, safe, economic, and easy to administer need to be developed.[13] Most importantly, the validity of the tests should be established by demonstrating significant and meaningful relationships with health. This is necessary to determine the contributionof of a particular fitness component to important health outcomes,[4] as well as to interpret the HRF scores in terms of adequacy of fitness with regard to health.[14]

Several field test batteries of HRF have been developed in recent years. Guralnic and coworkers[15] showed that among nondisabled older persons, objective performance measures of balance, walking speed, and ability to rise from a chair were highly predictive of subsequent disability. The reliability of all fitness parameters and the criterion-related validity of the cardiorespiratory endurance test of the American Alliance for Health, Physical Education, Recreation, and Dance functional fitness assessment battery in elderly women have been reported.[16] However, we have not found studies concerning the health-related validity of test batteries designed to assess the HRF of middle-aged populations.[14]

Our institute is developing a field-based health-related fitness test battery (HRFTB) for adults. In addition to the well-accepted health-related measurements of cardiorespiratory fitness and body composition, measures of motor and musculoskeletal fitness were considered. The selection of motor and musculoskeletal tests for the HRFTB was first based on a literature review of available methods with special reference to reliability and health-related validity. Secondly, their reliability,[17] safety, and feasibility[18] were established in two consecutive studies. The reliability[19,20] and feasibility[21] of the cardiorespiratory Walk Test had been previously established.

As a first step towards establishing the health-related content validity of the proposed tests we studied in cross-sectional design the associations of objective measures of HRF with self-assessed health in a population sample of middle-aged adults. More specifically, we studied the question of whether people with low, mid, and high levels of physical fitness differ in their perceived health, mobility, and back function and symptoms. Based on the review of literature regarding the relations between fitness and health by Skinner and Oja[13] we formulated the following working hypothesis: (1) cardiorespiratory fitness is primarily associated with perceived health and lower extremity function (ie, mobility); (2) musculoskeletal fitness of the lower extremities is associated with mobility; (3) musculoskeletal fitness of the upper body and trunk muscular endurance is associated with back function and pain; (4) motor fitness is associated with mobility and back pain; and (5) body mass index (BMI) is associated with perceived health and mobility.

METHODS

Subjects

The study sample was selected from the residents of the city of Tampere who had previously attended preventive health examinations provided by the municipal primary health care center for selected age brackets of the whole population. On average 80% of the residents have attended the examinations annually. About 80% of the participants had given consent to allow their personal data to be used for research purposes. Of these individuals, five age cohorts—subjects born in 1955 (37 years old at the time of the study), 1950 (42yrs), 1945 (47yrs), 1940 (52yrs) and 1935 (57yrs)—formed the present study population. All together 437 men and 389 women, evenly selected from each age cohort by systematic sampling, were invited to participate in the study. About 50 persons (range from 46 to 54) were included in each age and sex group. Fifty-six percent ($n = 246$) of the invited men and 65% ($n = 253$) of the women participated in the fitness testing. All subjects signed the informed consent, which contained detailed information about the study and the terms for participation.

Nonparticipant rates from the youngest to the oldest group were 40%, 40%, 41%, 38%, and 38%. A standard questionnaire, accessible to two age groups (47- and 52-year-olds), was used to compare the participants (62%) and the nonparticipants (38%) of the sample.[18] The nonparticipants had somewhat lower education level ("no vocational education," 22% vs 13%), rated their health lower ("good health," 65% vs 79%), used prescribed medication more often (35% vs 29%), were smokers much more often (40% vs 13%), and exercised briskly less often (43% vs 55%) than the participants.

Procedure

The subjects attended two measurement sessions at the Institute. At the first visit a standard pretesting health screening, including sociodemographic background factors, a modified Physical Activity Readiness Questionnaire (MPAR-Q),[22] measurements of body mass and height for BMI (kg/m^2), and systolic and diastolic blood pressure (BP, in mmHg, auscultation method after 5 minutes rest in a sitting position) were conducted by laboratory technicians and fitness testers. At the second visit the assessment of HRF was conducted by three testers, all of whom had a masters degree in sport or health sciences. The study team has substantial experience in research projects and is trained to conduct all measurements according to rigorous procedures and standards. In case of an emergency during fitness testing, a physician and nurses with skills and equipment for cardiopulmonary resuscitation were available.

The fitness testers used the pretesting screening results according to safety instructions determined in advance to refer subjects with severe diseases or symptoms to a physician for a health examination or to exclude subjects with minor limitations from selected fitness tests.[18] Accordingly, 8 subjects of 500 (2%) were referred to a physician. One woman with severe physical and mental limitations was excluded from the study population ($n = 499$); the 7 others were able to participate in some tests. Fifty-three percent of the subjects had no health limitations to testing, and the fitness testers excluded 45% from selected tests. A detailed description of the exclusion criteria for fitness testing has been reported elsewhere.[18] With these prescreening procedures, no acute complications occurred during the testing.[18]

Health-Related Fitness Assessment

Each subject was tested individually. All tests followed a standard sequence: balance (one-leg standing) was measured first, followed by flexibility (trunk side-bending, knee extension range of motion), leg muscular power (jump and reach), leg strength (one-leg squat), upper-body strength (modified push-ups) and trunk muscular endurance (static back extension) assessment, after which the subjects rested for about 10 minutes before performing the Urho Kaleva Kekkonen Institute (UKK) 2-km Walk Test. The subjects were given verbal encouragement in a consistent manner to achieve their best performance. The components, factors, and items in the HRFTB with brief description of methods are presented in figure 1. Detailed descriptions of the methods,[17,19,23] their reliability,[17,19,20] and feasibility[18,21] have been reported elsewhere. In summary, the interrater intraclass correlation coefficients (ICCs) for one-leg balance, trunk side-bending, push-up strength, leg power, and leg strength ranged from .86 to 1.00, and the test-retest coefficients of variation from 0.6% to 12.1%.[17] The test-retest reliability, as measured with the Pearson correlation coefficient (r), has been reported to be .89 for the Sorensen test of back muscular endurance,[24] and .99 for active knee extension range of motion.[25] In the Walk Test, the test-retest correlation coefficients (r) of men and women, respectively, were .98 and .94 for predicted maximal oxygen uptake ($\dot{V}O_2max$).[20] Concerning feasibility, the overall exclusion rate increased with age. Up to 27% of subjects aged 52 and 57 years were excluded from muscular endurance tests, mainly because of elevated BP and self-reported heart disease. However, over 90% qualified for balance, flexibility, muscular strength, and Walk tests. Thirty-eight percent of the subjects in the oldest age group and 24% of the 52-year-olds were on medication affecting heart rate, which limited the use of the Walk Test for the prediction of $\dot{V}O_2max$. Specific musculoskeletal symptoms limited the participation in the tests selectively.[18]

Interrelations between the fitness test items were assessed with age-adjusted partial correlations for men and women. The majority of the correlation coefficients (r) were below .40. In men, modified push-ups correlated strongly with jump and reach ($r = .51$) and with Walk Test time (WTT) ($r = -.43$). In women, WTT and BMI were the most strongly interrelated variables ($r = .48$), while one-leg squat correlated strongly with jump and reach (.43) and modified push-ups (.43), and WTT with modified push-ups ($r = -.41$). Based on this assessment we conclude that some of the test items are interrelated.

CARDIORESPIRATORY FITNESS

Test: ***UKK 2-km Walk Test*** for the prediction of maximal oxygen uptake (VO_{2max}) on the basis of time, heart rate at the end, BMI, age and sex.[19,20]
Method: Subject walks as fast as possible on flat surface using normal walking style.
Outcome: predicted VO_{2max} (ml/min/kg) and test time (min).

MUSCULOSKELETAL FITNESS

Lower extremity function

Test: ***Jump and reach*** for the leg extensor power.[17,18,37]
Method: Subject is instructed to jump as high as possible, and to bend her knees and swing her arms to enhance the performance. Before jumping, one arm is raised to mark the standing height. During the jump she touches a board with her middle finger at the highest position.
Outcome: The difference beween the reach height and the jumpheight is measured with a tape to the nearest cm.

Test: ***One-leg squat*** with increasing weights for the assessment of functional leg extensor strength.[17,18]
Method: Subject takes a short step forward on the mat, first with her right leg, squats down untill she lightly touches the mat with her left knee, rises up immediately and steps backward to the starting position.The squat is then repeated with the left leg.
Outcome: The load limit for a succesful squat task measured as maximum weight relative to the subjects body weight up to 140%. The test starts with body weight (BW) (ie. no added weight) and 10% increments of BW are added at four successive steps of 10%, 20%, 30% and 40% using a weight belt system.

Trunk muscular endurance

Test: ***Static back extension*** for the endurance capacity of the trunk extensor muscles.[43,24,18]
Method: Subject lies prone with the lower body (from the level of the spina iliaca anterior superior) resting on a low bench and crosses her hands behind the neck. The bench is placed on one gymnastic mat and another mat is placed on the bench. The tester stabilizes the subject by sitting on her ankles. The subject is asked to rise her upper-body to horizontal level and hold the position as long as possible for up to 4 minutes.
Outcome: endurance time in seconds.

Upper-body strength

Test: ***Modified push-ups*** for dynamic muscular endurance of the upper extremity extensor muscles and ability to stabilize the trunk.[17,18]
Method: The push-up cycle begins as the subject claps the hands behind her back. Next a normal straight leg push-up to straight elbows is performed, followed by a touch of one hand to the top of the supporting hand. The cycle ends in the prone lying position.
Outcome: The number of push-ups completed in 40 seconds.

Trunk flexibility

Test: ***Trunk side-bending*** to the right and left to measure the total range of movement of lateral flexion of the thoracic and lumbar spine and pelvis.[38,17,18]
Method: The subject stands on marked lines (15 cm apart) with her back against the wall. Arms are kept straight at the sides of the body. The subject bends to the right and then to the left as far as possible, the middle finger slides down along her lateral thigh.
Outcome: The distance the fingertip moves down the leg during maximal bending measured with a cloth tape in millimeters. Average value for the right and left sides is calculated.

Lower extremity flexibility

Test: Active ***knee extension range of motion*** (ROM) for the assessment of hamstring muscle extensiblity[23,18]
Method: The subject lies supine. The hip and the knee of the limb to be measured is flexed to 90 degrees. The opposite leg rests extended. The inclinometer (Vinkelmätare "Myrin", LIC, Rehab Vårdum, Solna, Sweden) is attached to the medial side of the ankle.
Outcome: End point ROM angle in degrees at maximal extension.

MOTOR FITNESS

Balance

Test: ***One-leg standing*** for the assessment of static postural control while the area of support is reduced.[37,17,18]
Method: The subject wears sport shoes. She places one foot at knee level along the inner side of the supporting leg and rotates the thigh outwards.Subject is advised to stand as still as possible.
Outcome: Duration of the balance task up to 1 minute measured with a stopwatch in seconds.

BODY COMPOSITION

Test: ***Body mass index*** (BMI) for the assessment of obesity.
Method: Standard measures of height and weight.
Outcome: BMI as weight/height^2 (kg/m^2).

Fig 1. Assessment methods of health-related fitness.

Assessment of Perceived Health, Mobility, and Back Function and Symptoms

A self-administered questionnaire was used to measure three aspects of health within the classification scheme of health-related quality of life according to Wilson and Cleary.[26] General health perception represents an individual integration of many aspects of the health concept. Questions on functional status reveal the individual's perception of ability to perform particular tasks. Functional status has been shown to be associated with general health perceptions.[26] Symptom status is one important determinant of functioning.[26,27]

A standard question on *perceived health* was used: "How would you describe your state of health in comparison with people of your age?" (five categories: very poor, poor, average, good, very good). Only 3% of the subjects rated themselves in the extreme categories, and in the analysis the two lowest and two highest categories were combined, given three ratings: poor, average, and good. In two Finnish population studies perceived health was consistently associated with use of outpatient physician services during 1 year, and with age-adjusted mortality over a 10-year follow-up period,[28] as well as with heart disease risk factors and disease indicators.[29]

We assessed *mobility* by the responses to a question about the degree of difficulty in stair climbing, a usual activity of daily living[30,31]: "How well can you manage climbing several blocks of stairs without resting?" (four categories: no difficulty, some difficulty, much difficulty, cannot). In the analysis the three lowest categories were combined because of the low proportion of subjects with severe limitations (rating: some difficulty, no difficulty). This question has been used earlier in the Mini–Finland Health Survey to assess the determinants of disability in Finns.[30,31] In that survey, with similar ratings, the proportion of men aged 30 to 64 years who had some difficulty was higher compared with our study (22% vs 14%), and the corresponding proportion of women was lower compared with our study (31% vs 38%).[30]

Back function and symptoms were assessed by three questions representing functional consequences (secondary impairments and disabilities) and common symptoms (primary impairments) of low back disorders.[27] *General back function* was elicited by the question, "How would you describe the functioning of your back?" (five categories: very poor, poor, average, good, very good). In the analysis the two lowest and two highest categories were combined (rating: poor, average, good). The question for *task-specific limitations* was, "How often do you have problems in your back while functioning in a stooped position?" (five categories: constantly, often, now and then, seldom, never). In the analysis the two lowest and two highest categories were combined (rating: often, now and then, seldom). *Back pain symptoms* were elicited by the question, "How often do you have back pain?" (five categories: constantly, often, now and then, seldom, never). In the analysis the two

Table 1: Distribution of Subjects in the Health Categories of Perceived Health, Mobility, Back Function, and Symptoms by Age and Sex

	Men (*n* = 245)						Women (*n* = 253)					
		37Yr (50)*	42Yr (46)	47Yr (50)	52Yr (49)	57Yr (50)		37Yr (53)	42Yr (50)	47Yr (53)	52Yr (47)	57Yr (50)
Health Outcome	(*n*)	%	%	%	%	%	*n*	%	%	%	%	%
Perceived health												
Poor	(40)	12	2	14	27	26	(44)	6	18	13	26	26
Average	(141)	70	65	50	46	56	(160)	81	62	66	55	50
Good	(64)	18	33	36	27	18	(49)	13	20	21	19	24
Mobility												
Ability to climb several blocks of stairs												
Some difficulty	(35)	4	9	12	22	24	(96)	17	34	38	49	54
No difficulty	(210)	96	91	88	78	76	(157)	83	66	62	51	46
Back function and symptoms												
General back function												
Poor	(35)	12	4	12	14	28	(51)	17	8	17	28	32
Average	(118)	46	59	62	41	34	(133)	47	58	49	49	60
Good	(92)	42	37	26	45	38	(69)	36	34	34	23	8
Frequency of problems in the back while functioning in a stopped posture												
Often	(35)	10	11	14†	12	24	(54)	9	12	25	24‡	38
Now and then	(69)	30	24	31†	33	24	(89)	34	30	25	41‡	48
Seldom	(140)	60	65	55†	55	52	(109)	57	58	50	35‡	14
Frequency of back pain												
Often	(40)	16	11	10	14	30	(49)	11	6	17	26	38
Now and then	(91)	38	48	40	33	28	(100)	40	44	30	38	46
Seldom	(114)	46	41	50	53	42	(104)	49	50	53	36	16

* Age (*n*).
† For this category, in the men's 47yr age group, *n* = 49.
‡ For this category, in the women's 52yr age group, *n* = 46.

lowest and two highest categories were combined (rating: often, now and then, seldom). The distribution of subjects in the health categories of the five self-assessed outcome variables is presented in table 1.

Data Analysis

All analyses were conducted separately for men and women. For all fitness scores the subjects were grouped into fitness categories based on age- and sex-specific cut points. The least fit 20% of the participants in each age-sex group were classified as "low-fit," the next 40% as "mid-fit," and the remaining 40% as "high-fit."[2] Because the distributions of the balance and one-leg squat test scores were skewed, the men were assigned differently to two and women to three fitness categories as presented in table 2.

Logistic regression analysis (BMDP statistical software, program LR)[32] was used to estimate the *odds ratios* (ORs) with 95% confidence interval [95% CI]) of *positive and negative health outcomes* (perceived health, mobility, back function, and symptoms) for *HRF levels* (low, mid, high). Low-fit groups were used as the reference for positive health outcomes, and high-fit groups as the reference for negative health outcomes. When the 95% CI did not include the value 1.0 the results were considered statistically significant ($p < .05$). Age, marital status, level of school education, occupational physical activity, and smoking were included in all models as possible confounders.

RESULTS

Associations of Health-Related Fitness With Perceived Health

In men, mid and high fitness in predicted $\dot{V}O_2max$ (ORs, 4.3 and 5.9, respectively), WTT (ORs, 2.9 and 11.0), modified push-ups (ORs, 3.0 and 5.6), and BMI (ORs, 4.6 and 3.3); and high fitness in jump and reach (OR, 3.4) and one-leg squat (OR,

Table 2: Proportion of the Subjects in Fitness Categories of the Tests With Skew Distributions

	One-Leg Balance							One-Leg Squat						
	Men			Women				Men			Women			
	Low	Mid & High		Low	Mid	High		Low	Mid & High		Low	Mid	High	
	%	%	*n*	%	%	%	*n*	%	%	*n*	%	%	%	*n*
Age group (yrs)														
37	20	80	50	21	22	57	53	10	90	50	21	43	35	53
42	20	80	46	22	20	58	50	20	80	46	28	35	37	49
47	20	80	50	19	31	50	52	30	70	50	35	46	19	52
52	20	80	49	21	41	38	47	24	76	46	23	41	36	44
57	21	79	43	20	42	38	50	22	78	46	20	58	22	46

3.1) were associated with *good perceived health* in comparison with the low-fitness category (fig 2). Conversely, mid and low fitness in $\dot{V}O_2max$ (ORs, 3.0 and 6.0), WTT (ORs, 3.2 and 9.1), jump and reach (ORs, 2.5 and 7.2), static back extension (ORs, 3.5 and 5.7) and side-bending (ORs, 3.1 and 4.1) were associated with *poor perceived health* in comparison with the high-fitness category. A more modest association with good health was found for high fitness in knee extension range of motion. Balance was not related to perceived health in men.

In women, mid and high fitness in predicted $\dot{V}O_2max$ (ORs, 2.4 and 7.5), WTT (ORs, 2.4 and 4.2), and static back extension (ORs, 2.8 and 3.7); and high fitness in one-leg squat (OR, 5.4), side-bending (OR, 3.1) and BMI (OR, 4.1) were associated with *good perceived health* in comparison with the low-fitness category (fig 2). Conversely, mid and low fitness in $\dot{V}O_2max$ (ORs, 6.4 and 9.2), WTT (ORs, 2.8 and 14.0), static back extension (ORs, 4.7 and 3.2), and BMI (ORs, 2.5 and 4.3); and low fitness in modified push-ups (OR, 8.2) was associated with *poor perceived health* in comparison with the high-fitness category. More modest associations with perceived health were found for side-bending and balance. Jump and reach

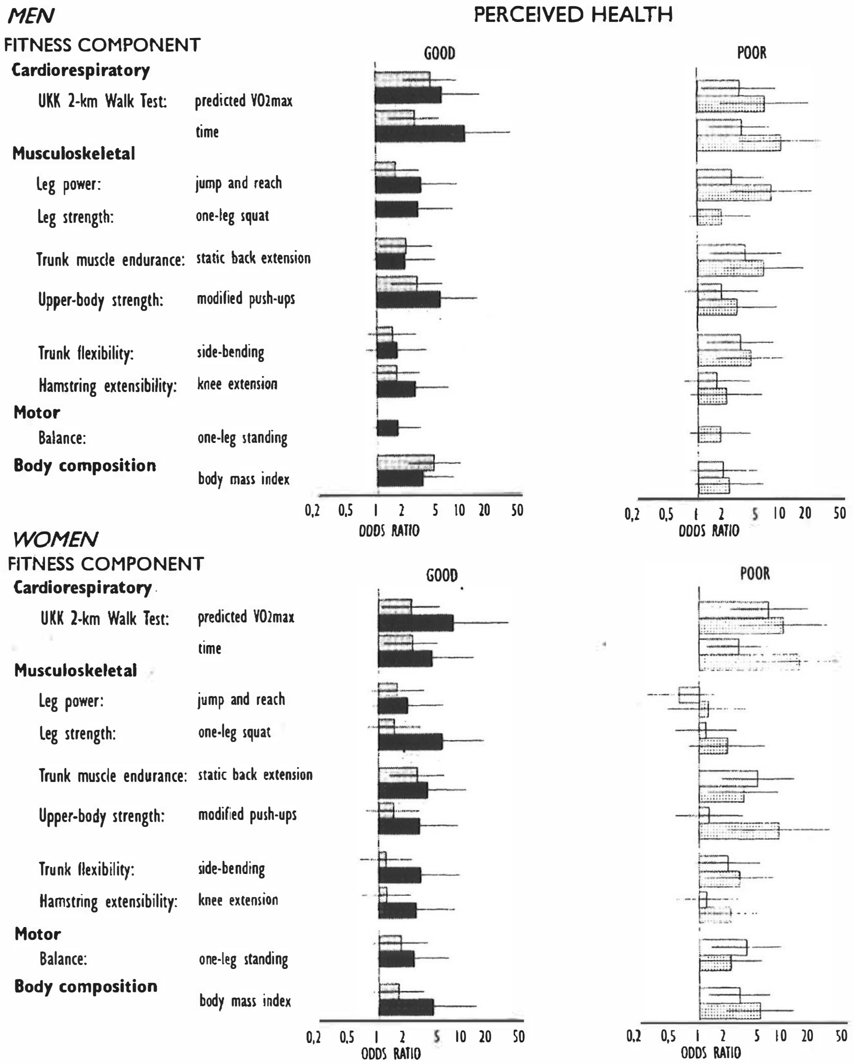


Fig 2. Associations between health-related fitness and perceived health (odds ratios with 95% confidence intervals adjusted for age, marital status, level of school education, occupational physical activity, and smoking) (⊡, low-fit group [20%]; □, mid-fit group [40%]; ■, high-fit group [40%]). Low-fit group was the reference for good health, and high-fit group for poor health.

and knee extension were not related to perceived health in women.

Associations of Health-Related Fitness With Mobility

In men, high fitness in predicted $\dot{V}O_2max$ (OR, 6.3), WTT (OR, 17.6), jump and reach (OR, 5.1), and static back extension (OR, 8.5) were associated with *no difficulties in climbing stairs* as compared with the low-fitness category (fig 3). Conversely, mid and low fitness in WTT (OR, 5.7 and 17.6, respectively) and static back extension (ORs, 3.3 and 8.5), and low fitness in $\dot{V}O_2max$ (OR, 6.3) and jump and reach (OR, 5.1) were associated with *some difficulties in climbing stairs* in comparison with the high-fitness category. One-leg squat, modified push-ups, side-bending, knee extension, balance, and BMI were not related to mobility in men.

In women, mid and high fitness in predicted $\dot{V}O_2max$ (ORs, 2.9 and 7.1), WTT (ORs, 3.4 and 5.8), one-leg squat (ORs, 2.3 and 4.8), static back extension (ORs, 2.3 and 2.8), modified push-ups (ORs, 2.5 and 5.9) and BMI (ORs, 3.0 and 5.1) were associated with *no difficulties in climbing stairs* in comparison with the low-fitness category (fig 3). Mid and low fitness in $\dot{V}O_2max$ (ORs, 2.5 and 7.1), one-leg squat (ORs, 2.1 and 4.8), and modified push-ups (ORs, 2.3 and 5.9); and low fitness in WTT (OR, 5.8), static back extension (OR, 2.8) and BMI (OR, 5.1) were associated with *some difficulties in climbing stairs* in comparison with the high-fitness category. Jump and reach,

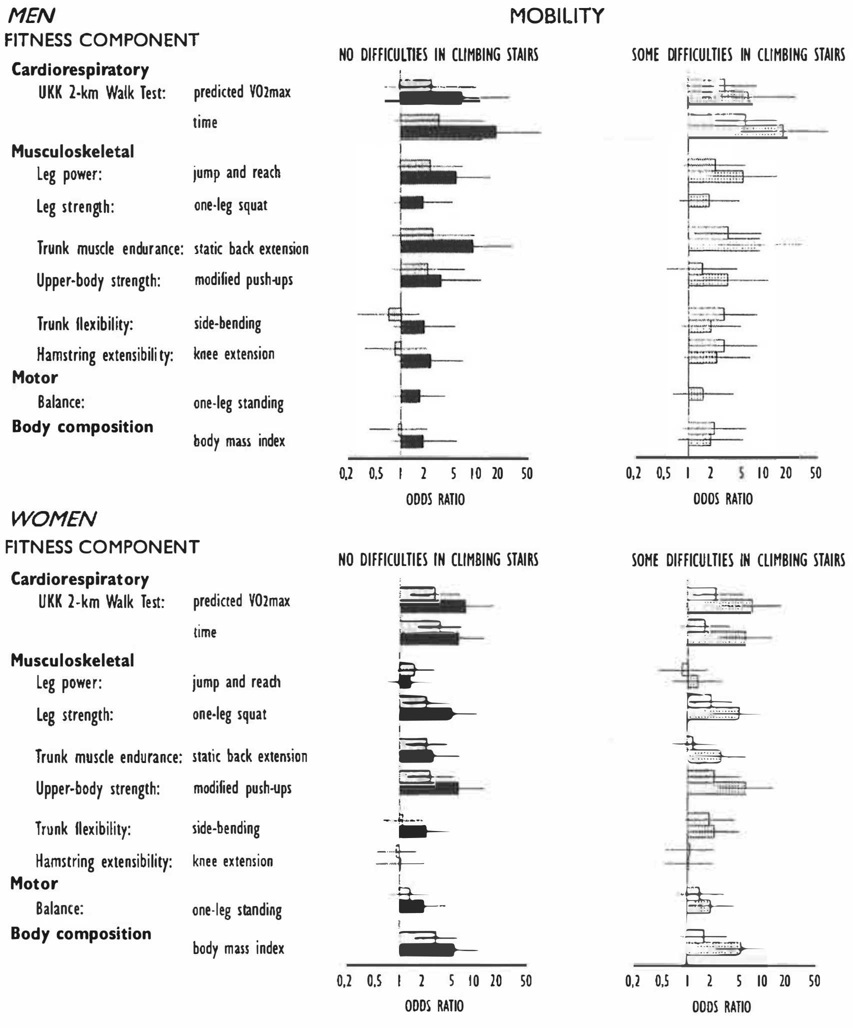


Fig 3. Associations between health-related fitness and mobility (odds ratios with 95% confidence intervals adjusted for age, marital status, level of school education, occupational physical activity, and smoking) (▒, low-fit group [20%]; □, mid-fit group [40%]; ■, high-fit group [40%]). Low-fit group was the reference for no difficulties, and high-fit group for some difficulties.

side-bending, knee extension, and balance were not related to mobility in women.

Associations of Health-Related Fitness With Back Function and Symptoms

In men, mid and high fitness in static back extension (ORs, 2.0 and 3.7, respectively) and mid fitness in predicted $\dot{V}o_2max$ (OR, 3.3) were associated with *good perceived back function* in comparison with the low fitness category (fig 4). Similarly, high fitness in jump and reach (OR, 2.9) and static back extension (OR, 2.6) were associated with *seldom problems in the back while stooping,* and mid and high fitness in static back endurance (ORs, 2.0 and 2.9) with *seldom back pain.* Conversely, mid and low fitness in jump and reach (ORs, 5.2 and 3.9) and side-bending (OR, 3.1 and 3.7), and low fitness in WTT (OR, 3.5), static back extension (OR, 4.8), and modified push-ups (OR, 6.1) were associated with *poor perceived back function* in comparison with the high fitness category. Similarly, mid and low fitness in WTT (ORs, 3.4 and 5.7), and low fitness in jump and reach (OR, 6.1), modified push-ups (OR, 5.5) and side-bending (OR, 3.4) were associated with *often problems in the back while stooping.* Low fitness in static back extension (OR, 3.9) and modified push-ups (OR, 3.6) were associated with *often back pain.* More modest associations with poor functioning and pain were found for one-leg squat, and with pain for WTT and jump and reach. Knee extension, balance, and BMI were not related to back function and symptoms in men.

In women, mid and high fitness in predicted $\dot{V}o_2max$ (ORs,

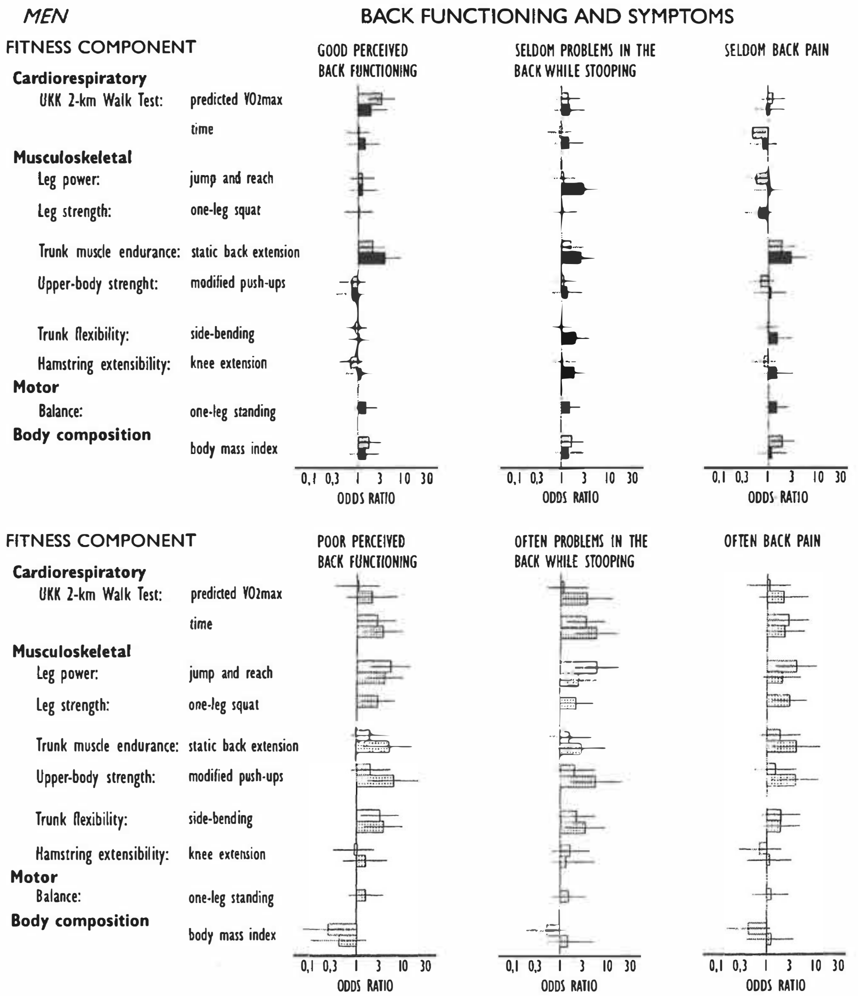


Fig 4. Associations between health-related fitness and back function and symptoms in men (odds ratios with 95% confidence intervals adjusted for age, marital status, level of school education, occupational physical activity, and smoking) (▨, low-fit group [20%]; □, mid-fit group [40%]; ■, high-fit group [40%]). Low-fit group was the reference for good back health, and high-fit group for poor back health.

2.3 and 3.1), and high fitness in one-leg balance (OR, 3.7) were associated with *good perceived back function*. Similarly, high fitness in modified push-ups test (OR, 2.7) and balance (OR, 2.5) were modestly associated *with seldom problems in the back while stooping*. There were no significant associations between HRF and *seldom back pain* in women. Mid and low fitness in static back extension (OR, 2.6 and 2.9) and BMI (OR, 2.9 and 2.4), and low fitness in modified push-ups (OR, 3.5), side-bending (OR, 3.1), knee extension range of motion (OR, 2.7) and balance (OR, 3.2) were associated with *poor perceived back function* in comparison with the high-fitness category. Similarly, mid and low fitness in predicted $\dot{V}O_2max$ (ORs, 2.8 and 2.7), and low fitness in static back extension (OR, 3.1), modified push-ups (OR, 4.4), knee extension (OR, 2.9), balance (OR, 2.4) and BMI (OR, 4.5) were associated with *often problems in the back while stooping*. Mid and low fitness in BMI (ORs, 2.6 and 4.2), and low fitness in static back extension (OR, 3.8) and balance (OR, 2.9) were associated with *often back pain*. WTT, jump and reach, and one-leg squat were not related to back function and symptoms in women.

DISCUSSION

General Aspects of the Study

We studied the associations of nine fitness tests with self-assessed health as a first step towards establishing their content validity in relation to selected aspects of health. Health-related validity is one criterion for selecting items to our HRFTB, the prior criteria having been reliability, safety, and feasibility of the field-based tests.[17,18] Because these cross-sectional associations cannot be interpreted as indicating causal relationships between fitness and health, we have conducted a 3-year follow-up study with the same population to assess the predictive validity of the tests, the final criterion for selection. These results will be reported later.

Our study sample was, to some extent, selected: the subjects had somewhat higher education, were healthier, and more physically active than the nonparticipating part of the population. This selection bias causes the population estimates of fitness to be too high in general, and the bias becomes even greater in those tests and age-groups from which a proportion of study subjects were excluded. However, the effect of this bias on the strength of the associations of fitness with health is more likely an underestimation than an overestimation.[2]

The HRFTB described in this article is aimed to assess the level of fitness of middle-aged individuals and populations, monitor changes in HRF, and offer a motivational tool for physical activity.[3,14] The ultimate purpose is to give feedback to the individual in terms of the adequacy of his or her fitness with regard to health criteria.[14] Although our results provide some elements for the development of such criteria, more prospective studies with representative population samples are needed to establish definite criteria. For the time being, this HRFTB provides an individual age- and sex-specific fitness profile based on norm-reference values derived from the present data.

Validity of the Test Items in the Health-Related Fitness Test Battery

Cardiorespiratory fitness. As hypothesised, the level of fitness in the 2-km Walk Test (predicted $\dot{V}O_2max$ or WTT) was most strongly and consistently associated with perceived health and mobility (ability to climb stairs) in both men and women (figs 2 and 3). There was a graded relationship between these variables. Similar patterns of association between cardiorespiratory fitness and perceived health,[29] and cardiorespiratory fitness and mortality[2] have been reported elsewhere. In our group of middle-aged men, high fitness in WTT was the best indicator of good mobility, while in women predicted $\dot{V}O_2max$ showed the strongest associations.

In our group of men, lower categories of WTT were associated with back dysfunction, but not with back pain (fig 4). In women, no consistent associations were found (fig 5). Other studies have shown that patients with chronic low back pain have slow walking speed,[33] and that fitness program can increase the walking speed.[34] Prospectively,[35] cardiorespiratory fitness was not predictive of future industrial back injury, but was related to development of chronic disabling pain.

Musculoskeletal fitness. Our hypothesis was that the tests of lower extremity function would be most strongly associated with mobility, and the trunk and upper-body function tests with back function and symptoms.

Lower extremity function. Leg power was not related to health status of the women in our study. In men, low leg power was an indicator of poor perceived health and difficulties in stair climbing (figs 2 and 3). In other studies perceived health has been associated with maximal isometric muscle strength[36] and with excess risk of death from cardiovascular diseases and from all causes.[37] Our findings on the associations between leg power and back health were inconsistent in men (fig 4). In another prospective study[38] isometric lifting strength was not a predictor of back pain, but cross-sectionally low back pain was related to low isokinetic leg strength.[39]

Leg strength was not related to the health status of the men in our study. This could be explained by the fact that 74% of them (table 2) were able to perform the squat with the highest load level (140% of body weight).[18] Among the women, high fitness in leg strength was associated with good perceived health and mobility, and low fitness with some disability in stair climbing. These findings are in line with former studies on aging subjects.[15,40,41]

Trunk muscular endurance. Static back extension was related to all measured health outcomes in both genders. Unexpectedly, although in line with the study on functional ability among healthy elderly,[42] the strongest, as well as a graded association found for mobility in men, both low and mid fitness were associated with some difficulties in stair climbing, and high fitness with no difficulties (fig 3). In women, static back extension was most strongly associated with perceived health (fig 2) and to a lesser extent with mobility.

As hypothesised, low fitness in static back extension was systematically associated with back dysfunctioning and pain (figs 4 and 5) in both genders, a finding consistent with other cross-sectional studies.[43,44] In addition, static back extension was the only test systematically associated with positive back health in men (fig 4); among the women no associations were found (fig 5). Our results support the prospective finding by Biering-Sorensen,[45] which suggested that good isometric endurance of back muscles may prevent first-time occurrence of low back pain in men. More recently, Gundewall and colleagues[11] showed that muscular exercises of back were efficient in preventing back symptoms and absence from work.

Upper-body muscular strength. The *modified push-ups* test is aimed to assess dynamic strength of the upper body and ability to stabilize the trunk.[17] As expected, low fitness was strongly associated with back dysfunction and pain (figs 4 and 5). This may indicate that subjects with back problems had difficulties in stabilizing their trunks, which is a sign of weakened postural control,[46] an impairment also detected among the women with back dysfunction in this study (fig 5). In addition, modified push-up was strongly related to good

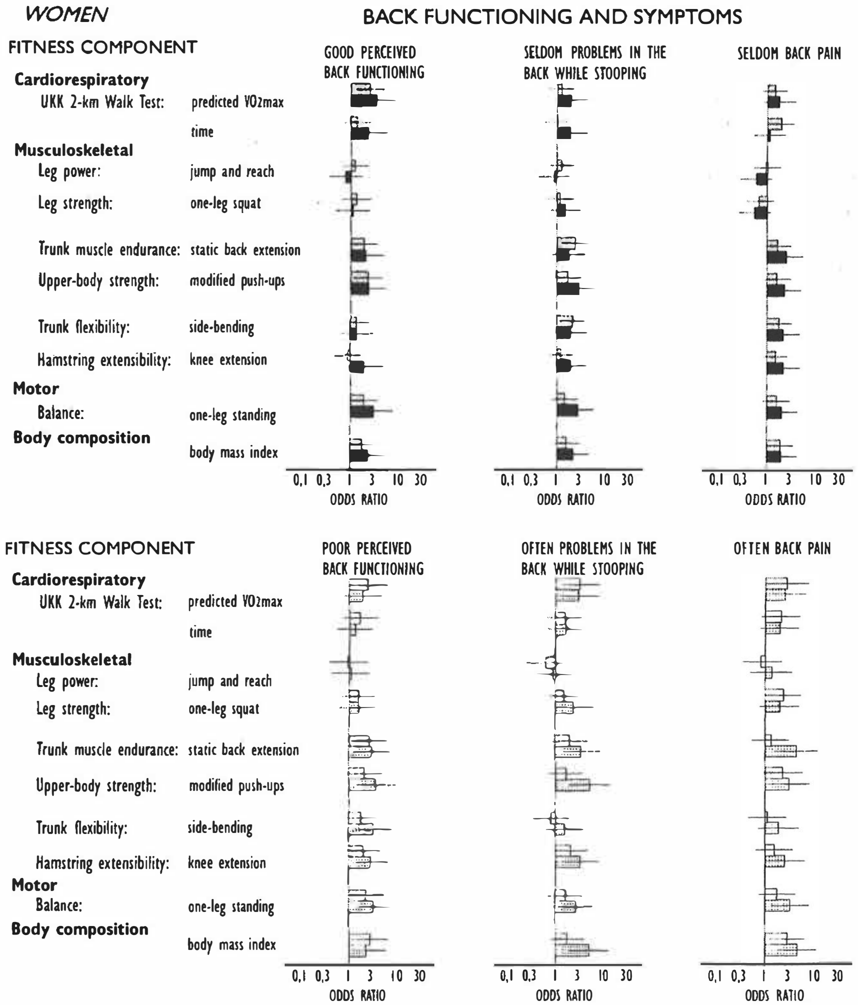


Fig 5. Associations between health-related fitness and back function and symptoms in women (odds ratios with 95% confidence intervals adjusted for age, marital status, level of school education, occupational physical activity, and smoking) (▨, low-fit group [20%]; □, mid-fit group [40%]; ■, high-fit group [40%]). Low-fit group was the reference for good back health, and high-fit group for poor back health.

perceived health in men (fig 2), and to poor perceived health and mobility disability in women (figs 2 and 3). Modified push-up was also interrelated with Walk Test performance and leg function tests in both men ($r = -.43$ and .51) and women ($r = -.41$ and .43). All these facts suggest that the modified push-up test is a highly functional test related to many important health outcomes.

Flexibility. Unexpectedly, low and mid fitness in trunk side bending were most strongly associated with poor perceived health in men; weaker trends were found in women. Typically, reduced spinal flexibility[47,48] and hamstring muscular extensibility[45,47] have been found as residual signs of persistent low back trouble. In the present study, trunk side bending was associated with back dysfunction, but not with back pain in men. Weaker associations in trunk side bending and knee extension were found in women. The practical significance of the small differences in flexibility between subjects with and without a history of back problems has been questioned by Battie and colleagues.[49] In the present study, the mean differences in trunk side bending between the high-fit and low-fit groups of the five age groups were 4.3cm (range 3.9 to 4.6) in men and 3.6 cm (range, 2.8 to 4.3) in women. While the standard error of our trunk side bending measurement is 1.4cm,[17] we consider the differences practically meaningful.

Motor fitness. In our group of middle-aged men there were no associations between the *one-leg standing balance* and health outcomes. This could be explained by the fact that the test did not differentiate the level of fitness very well (table 2). A more challenging balance task for men is needed for further studies.

Although balance is a risk factor for falls[9] and has been shown to have predictive value for mobility disability[15] among the elderly, in our group of middle-aged women there was only a trend for low fitness in balance to be associated with difficulties in climbing stairs. As hypothesised, however, balance was consistently associated with back function and symptoms in women. An association between low back pain and impaired postural control was first reported by Byl and Sinnot[46] and later by Luoto and colleagues.[50] Two prospective studies have indicated that training of postural control reduces certain types of back injuries,[11,51] but the changes in low back pain disability were not related to postural control after a restoration program.[50]

Body composition. Obesity is a known risk factor for non-insulin-dependent diabetes and coronary heart disease.[7] These two conditions have been related to poor perceived health.[29] As expected, BMI was related to good perceived health in men and to poor perceived health in women in our study.

In the present study BMI was not related to either mobility or to back function in men. Contrary to this, being overweight predicted functional disability of Finnish men and women[52] and was cross-sectionally associated with self-reported joint pain and reduced mobility in a Swedish population.[53] However, these associations were stronger for women than for men.

In our group of women, the expected and strongest relations were found between low BMI and no difficulty in climbing stairs. A similar type of relationship to the onset of mobility disability was found by Launer and colleagues.[54] In our study, high BMI in women was also strongly related to back pain and problems while stooping. A similar pattern of association with back pain was reported by Deyo and Bass.[55] In our study BMI correlated strongly with WTT (r = .48), indicating that the overweight women walked slowly. Our findings together with the other studies[52-55] suggest that weight control of women is of importance with respect to prevention of functional disability.

CONCLUSIONS

Based on their relations to measures of perceived health and musculoskeletal functioning the 2-km Walk Test, static back extension test, modified push-ups, and BMI proved to be valid measures of HRF in this cross-sectional study among middle-aged adults. Furthermore, the jump and reach and trunk side-bending tests demonstrated health-related validity among the men, as did the one-leg squat and one-leg balance tests among the women. Prospective studies are needed to assess the predictive health-related validity of these fitness tests.

References

1. U.S. Department of Health and Human Services. Physical activity and health: A report of the Surgeon General. Atlanta (GA): US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion; 1996.
2. Blair SN, Kampert JB, Kohl HW, Barlow CE, Macera CA, Paffenbarger RS, et al. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. JAMA 1996;276:205-10.
3. Shephard RJ. Worksite fitness and exercise programs: a review of methodology and health impact. Am J Health Promot 1996;10: 436-52.
4. Phillips WT, Haskell WL. "Muscular fitness" easing the burden of disability for elderly adults. J Aging Phys Act 1995;3:261-89.
5. Videman T, Sarna S, Battie MC, Koskinen S, Gill K, Paananen H, et al. The long-term effects of physical loading and exercise lifestyle on back-related symptoms, disability, and spinal pathology among men. Spine 1995;20:699-709.
6. Kujala UM, Kaprio J, Sarna S. Osteoarthritis of weight bearing joints of lower limbs in former elite male athletes. BMJ 1994;308: 231 4.
7. Perry IJ, Wannamethee SG, Walker MK, Thomson AG, Whincup CH, Shaper AG. Prospective study of risk factors for development of non-insulin-dependent diabetes in middle aged British men. BMJ 1995;310:560-4.
8. Province MA, Hadley EC, Hornbrook MC, Lipsitz LA, Miller JP, Mulrow CD, et al. The effect of exercise on falls in elderly patients. A preplanned meta-analysis of the FICSIT Trials. JAMA 1995;273: 1341-7.
9. Heinonen A, Kannus P, Sievänen H, Oja P, Rinne M, Uusi-Rasi K, et al. Randomized controlled trial of the effect of high-impact exercise on selected risk factors for osteoporotic fractures. Lancet 1996;348:1343-7.
10. Lahad A, Malter AD, Berg AO, Deyo RA. The effectiveness of four interventions for the prevention of low back pain. JAMA 1994;272:1268-91.
11. Gundewall B, Liljeqvist M, Hansson T. Primary prevention of back symptoms and absence from work: a prospective randomized study among hospital employees. Spine 1993;18:587-94.
12. Bouchard C, Shephard RJ. Physical activity, fitness, and health: the model and key concepts. In: Bouchard C, Shephard RJ, Stephens T, editors. Physical activity, fitness, and health. Champaign (IL): Human Kinetics; 1994. p. 77-88.
13. Skinner J, Oja P. Laboratory and field tests for assessing health-related fitness. In: Bouchard C, Shephard RJ, Stephens T, editors. Physical activity, fitness, and health. Champaign (IL): Human Kinetics; 1994. p. 160-79.
14. Oja P, Tuxworth W, editors. Eurofit for adults. Assessment of health-related fitness. Tampere, Finland: Council of Europe Publishing; 1995.
15. Guralnic JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. New Engl J Med 1995;332: 556-61.
16. Bravo G, Gauthier P, Roy P-M, Tessier D, Gaulin P, Dubois M-F, et al. The functional fitness assessment battery: reliability and validity data for elderly women. J Aging Phys Act 1994;2:67-79.
17. Suni JH, Oja P, Laukkanen RT, Miilunpalo SI, Pasanen ME, Vuori IM, et al. Development of a health-related fitness test battery for adults: aspects of reliability. Arch Phys Med Rehabil 1996;77:399-405.
18. Suni JH, Miilunpalo SI, Asikainen T-M, Laukkanen RT, Oja P, Pasanen ME, et al. Safety and feasibility of a health-related fitness test battery for adults. Phys Ther (In Press)
19. Oja P, Laukkanen R, Pasanen M, Tyry T, Vuori I. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. Int J Sports Med 1991;12:356-62.
20. Laukkanen RMT, Oja P, Pasanen ME, Vuori IM. Two-kilometre walking test: effect of walking speed on the prediction of maximal oxygen uptake. Scand J Med Sci Sports 1993;3:263-6.
21. Laukkanen RMT, Oja P, Ojala KH, Pasanen ME, Vuori IM. Feasibility of a 2-km walking test for fitness assessment in a population study. Scand J Soc Med 1992;20:119-26.
22. Thomas S, Reading J, Shephard RJ. Revision of the physical activity readiness questionnaire (PAR-Q). Can J Sports Sci 1992; 17:338-45.
23. Laukkanen R, Hynninen E, editors. Guide for the UKK 2-km walking test. Tampere (Finland): President Urho Kaleva Kekkonen Institute for Health Promotion Research; 1993.
24. Jörgensen K, Nicolaisen T. Two methods for determining trunk muscle endurance: a comparative study. Eur J Appl Physiol 1986;55:639-44.
25. Gajdosik R, Lusin G. Hamstring muscle tightness: reliability of an active knee-extension test. Phys Ther 1983;63:1085-8.
26. Wilson IB, Cleary PD. Linking the clinical variables with health-related quality of life: a conceptual model of patient outcomes. JAMA 1995;273:59-65.

27. Harper AC, Harper DA, Lamber LJ, Andrews HB, Kai Lo S, Ross FM, et al. Symptoms of impairment, disability and handicap in low back pain: a taxonomy. Pain 1992;50:189-95.
28. Miilunpalo S, Vuori I, Oja P, Pasanen M, Urponen H. Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. J Clin Epidemiol 1997;50: 517-28.
29. Kaplan GA, Goldberg DE, Everson SA, Cohen RD, Salonen R, Tuomilehto J, et al. Perceived health-status and morbidity and mortality: evidence from the Kuopio heart disease risk factor study. Int J Epidemiol 1996;25:259-65.
30. Aromaa A, Heliövaara M, Impivaara O, Knekt P, Maatela J, Joukamaa M, et al. Health, functional limitations and need for care in Finland. Basic results from the Mini-Finland Health Survey. Helsinki and Turku (Finland): Publications of the Social Insurance Institution (AL.32); 1989.
31. Mäkelä M, Heliövaara M, Sievers K, Knekt P, Maatela J, Aromaa A. Musculoskeletal disorders as determinants of disability in Finns aged 30 years or more. J Clin Epidemiol 1993;46:549-59.
32. Dixon WJ, editor. BMDP statistical software manual. Vol 2. Berkeley (CA): University Press; 1992.
33. Khodadadeh S, Eisenstein SM. Gait analysis of patients with low back pain before and after surgery. Spine 1993;18:1451-5.
34. Frost H, Moffet JAK, JS Moser, Fairbank JCT. Randomized control trial for evaluation of fitness programme for patients with chronic low back pain. BMJ 1995;310:151-4.
35. Battie MC, Bigos SJ, Fisher LD, Hansson TH, Nachemson AL, Spengler DM, et al. A prospective study of the role of cardiovascular risk factors and fitness in industrial back pain complaints. Spine 1989;14:141-7.
36. Era P, Lyyra AL, Viitasalo JT, Heikkinen E. Determinants of isometric muscle strength in men of different ages. Eur J Appl Physiol 1992;64:84-91.
37. Fujita Y, Nakamura Y, Hiraoka J, Kobayashi K, Sakata K, Nagai M, et al. Physical strength tests and mortality among visitors to health-promotion centers in Japan. J Clin Epidemiol 1995;48: 1349-59.
38. Battie MC, Bigos SJ, Fisher LD, Hansson TH, Nachemson AL, Jones ME, et al. Isometric lifting strength as a predictor of industrial back pain reports. Spine 1989;14:851-6.
39. Lee JH, Ooi Y, Nakamura K. Measurement of muscle strength of the trunk and the lower extremities in subjects withhistory of back pain. Spine 1995;20:1994-6.
40. McAlindon TE, Cooper C, Kirwan JR, Dieppe PA. Determinants of disability in osteoarthritis of the knee. Ann Rheum Dis 1993;52:258-62.
41. Rantanen T, Era P, Heikkinen E. Maximal isometric knee extension strength and stair-mounting ability in 75- and 80-year-old men and women. Scand J Rehabil Med 1996;28:89-93.
42. Avlund K, Schroll M, Davidsen M, Lowvborg R, Rantanen T. Maximal isometric muscle strength and functional ability in daily activities among 75-year-old men and women. Scand J Med Sci Sports 1994;4:32-40.
43. Jorgensen K, Nicolaisen T. Trunk extensor endurance: determination and relation to low-back trouble. Ergonomics 1987;30: 259-67.
44. Holmström E, Moritz U, Andersson M. Trunk muscle strength and back muscle endurance in construction workers with and without low back disorders. Scand J Rehabil Med 1992;24:3-10.
45. Biering-Sorensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. Spine 1984;9:106-19.
46. Byl NN, Sinnot PL. Variations in balance and body sway in middle-aged adults: subjects with healthy backs compared with subjects with low-back dysfunction. Spine 1991;16:325-30.
47. Mellin G. Chronic low back pain in men 54-63 years of age. Correlations of physical measurements with the degree of trouble and progress after treatment. Spine 1986;11:421-6.
48. Burton AK, Tillotson KM, Troup JDG. Variations in lumbar sagittal mobility with low-back trouble. Spine 1989;14:584-90.
49. Battie MC, Bigos SJ, Fisher LD, Spengler DM, Hansson TH, Nachemson AL, et al. The role of spinal flexibility in back pain complaints within industry. Spine 1990;15:768-73.
50. Luoto S, Taimela S, Hurri H, Aalto H, Pyykkö I, Alaranta H. Psychomotor speed and postural control in chronic low back pain patients. Spine 1996;21:2621-7.
51. Videman T, Rauhala H, Asp S, Lindström K, Cedercreutz G, Kämppi M, et al. Patient-handling skill, back injuries and back pain: an intervention study in nursing. Spine 1989;14:148-56.
52. Rissanen A, Heliövaara M, Knekt P, Reunanen A, Aromaa A, Maatela J. Risk of disability and mortality due to overweight in a Finnish population. BMJ 1990;301:835-7.
53. Wolk A, Rössner S. Obesity and self-perceived health in Sweden. Int J Obesity 1996;20:369-72.
54. Launer LJ, Harris T, Rumpel C, Madans J. Body mass index, weight change, and risk of mobility disability in middle-aged and older women. The epidemiologic follow-up study of NHANES I. JAMA 1994;271:1093-8.
55. Deyo RA, Bass JE. Lifestyle and low-back pain. The influence of smoking and obesity. Spine 1989;14:501-6.
56. Mellin G. Accuracy of measuring lateral flexion of the spine with a tape. Clin Biomech 1986;1:85-9.
57. Stones M, Kozma A. Balance and age in the sighted and blind. Arch Phys Med Rehabil 1987;68:85-9.

IV

Predictive value of fitness for 3-year changes in musculoskeletal functioning

by

Jaana Suni, Pekka Oja, Seppo Miilunpalo, Matti Pasanen, Ilkka Vuori, Klaus Bös

Manuscript

PREDICTIVE VALUE OF FITNESS FOR 3-YEAR CHANGES IN MUSCULOSKELETAL FUNCTIONING.

Jaana H. Suni, M.Sc., Pekka Oja, Ph.D., Seppo I. Miilunpalo, M.D., Matti E. Pasanen, M.Sc. Ilkka M. Vuori, M.D., Prof., Klaus Bös, Prof.

ABSTRACT

Purpose: To evaluate the predictive validity of 9 field-based health-related fitness (HRF) tests by studying the associations of baseline fitness with 3-year changes in self-rated perceived health, mobility function, and back functioning and back pain. In addition, the relations between leisure-time physical activity (LTPA) and 3-year changes in the selected health outcomes status were assessed. **Methods:** The subjects were a random sample of five age groups of middle-aged (37-57-years) adults, 246 men and 254 women at baseline. Of the men 87% (n=214) and of the women 80% (n=202) participated in the follow-up. Logistic regression models, adjusted for several confounding variables, were used to analyse the associations of HRF and LTPA with the changes in health and functional status. **Results:** Maximal oxygen uptake, estimated according to the Walk Test, predicted both negative and positive changes in perceived health. The Walk Test time predicted positive and the one-leg squat strength negative changes in ability to climb stairs. For modified push-ups, trunk side-bending flexibility and hamstring muscle extensibility tests mid-level of fitness (mid 20%), but not high-level (highest 40%), compared with low-level (lowest 40%) predicted positive changes in back functioning. Low-fitness in side-bending was associated with negative changes in back functioning while stooping. None of the tests predicted changes in back pain. Associations between LTPA and the changes in musculosketal functioning were contradictory. The vertical jump, trunk extension endurance and one-leg balance tests, and body mass index showed no predictive value for health. **Conclusions:** Five of the 9 evaluated HRF tests showed predictive validity for the used indices of perceived health. New follow-up and intervention studies are needed to investigate the significance of different components of fitness and LTPA for the maintenance of musculoskeletal functioning and health.

INTRODUCTION

Scientific knowledge on the significance of physical activity and physical fitness for health and functional capability has increased and became more specified during the 1990's [Bouchard & Shephard 1994, U.S. Department of Health and Human Services 1996, ACSM (American College of Sports Medicine) 1998]. Based on the known relationships between physical activity, fitness and health, a new fitness concept, health-related fitness (HRF), has been introduced (Bouchard and Shephard 1994). HRF includes components of fitness that are related to health and can be influenced by regular physical activity. The components of HRF are body composition, metabolic, aerobic, musculoskeletal and motor fitness (Bouchard and Shephard 1994, Skinner and Oja 1994). Assessing and monitoring relevant aspects of fitness may have an important role in the promotion of health-enhancing physical activity (HEPA) (King & Senn 1996).

The role of endurance-type physical activity and the respective fitness component, aerobic fitness, in preventing common chronic diseases such as coronary heart disease, hypertension and non-insulin dependent diabetes is scientifically well established (Pate et al. 1995, U.S. Department of Health and Human Services 1996). Among adult populations, degenerative musculoskeletal disorders such as chronic low back pain, and osteoarhtritis of the knee and hip are major courses of mobility limitations and disability in activities needed for everyday chores (Davis et al. 1991, Mäkelä et al. 1993, Hagen and Thune 1998). Among the elderly mobility-related problems may limit substantially their ability to remain independent (Davis et al. 1991, Launer et al. 1994, Guralnik et al 1995). Falls leading to bone fractures are the most serious health consequences of mobility-related limitations in old age (Province et al. 1995). Unfortunately, the knowledge to promote physical activity, exercise and fitness that enhance musculoskeletal health and prevent functional disability is inconclusive and inconsistent (Videman et al. 1995, Buckwalter and Lane 1997, Tulder et al. 1997). Furthermore, physical fitness has not yet been extensively examined with respect to musculoskeletal functional disability, especially in middle-aged populations. (Huang et al. 1998, Morey et al. 1998).

Scientifically sound population-based assessment methods of musucloskeletal and motor fitness for middle-aged adults are needed for both promotional and research purposes. The majority of the currently available, scientifically validated methods for assessing musculoskeletal and motor fitness were designed for the elderly (Shaulis et al. 1994, Bravo et al. 1994, Koch et al. 1994, Guralnik et al. 1995, Rejeski et al. 1995). Recently, an expert group of the Council of Europe proposed the test battery "Eurofit for adults", which is aimed at assessing HRF of middle-aged and older adults (Oja & Tuxworth 1995). However, there are no systematic studies concerning the reliability, safety and validity of the Eurofit test battery. To our knowledge the UKK Institute's HRF Test battery, designed for HEPA promotion of middle-aged adults, is the only battery that has been systematically evaluated for its reliability (Suni et al. 1996), as well as for safety and feasibility (Suni et al. 1998b). Furthermore, the health-related (Suni et al. 1998a) and physical activity-related (Suni et al. 1999) content validity of the Test battery has been evaluated in a cross-sectional population study.

The purpose of the present study was to evaluate the predictive health-related validity of the proposed HRF test battery by investigating the associations of baseline fitness with positive and negative changes in perceived health, mobility, and back functioning and back pain assessed in a three year follow-up study of an adult population to. In addition, the associations of leisure time physical activity (LTPA) with the selected health outcomes was assessed. The main hypotheses concerning musculoskeletal and motor fitness, the special interest of the study, were that baseline (i) musculoskeletal fitness in lower extremity strength and power, motor fitness, cardiorespiratory fitness, and muscular-type exercise are associated with three year changes in ability to climb stairs; and (ii) musculoskeletal fitness in trunk and upper-body functions and in lower limb flexibility, motor fitness, and muscular-type exercise are associated with three year changes in back functioning and back pain.

METHODS

The study subjects represent the middle-aged residents of the city of Tampere. The study sample was drawn from individuals, who had previously attended preventive health examinations for all inhabitants, organized by the municipal primary health care centre. Selection of the subjects is presented in Figure 1. On average 80% of the invited residents attended the examinations. About 80% of the participants had given a consent to allow their personal data to be used for research purposes. Of these individuals, five age cohorts formed the present study population. Altogether 437 men and 389 women, evenly selected from each age cohort by systematic sampling, were invited to participate in the cross-sectional study, and 56% of the men and 65% of the women participated (see Figure 1). The non-participants had somewhat lower education level, rated their health lower, used prescribed medication more often, were smokers more often, and exercised briskly less often than the participants [for details see Suni et al. (1998b)].

Of the cross-sectional study subjects 87% of the men and 80% of the women attended the three-year follow-up study. The age-specific participation rates are shown in Figure 1. A short phone interview of the reasons for not participating in the follow-up study was conducted among the subjects who dropped out (n=82). Of them 79% answered the interview, 12% were not reached, six percent were not willing to answer and three percent had died. The main reasons for not participating, presented in Table 1, were health (27%) and work (24%) related hindrances. The subjects who dropped out of the follow-up study had somewhat lower education level (secondary school or lower 46% vs. 35%), rated their health lower (good health 14% vs. 25%), were smokers more often (34% vs. 25%), and exercised briskly less often (36% vs. 44%) than the participants at baseline. In both cross-sectional and follow-up studies, all subjects signed an informed consent which contained detailed information about the study and the terms for participation.

POPULATION OF CITY OF TAMPERE

Preventive health examinations: 80% of selected age brackets
- 80% gave consent to use data for research

↓

TARGET POPULATION

Random sample: 437 men, 389 women
Five age cohorts: subjects born in 1935, 1940, 1945, 1950, 1955

↓ **did not participate: 316**
interrupted: 10

CROSS-SECTIONAL STUDY POPULATION

Age-specific (years) participation rates

	37	42	47	52	57	All
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Men:	57 (50)	58 (46)	52 (50)	59 (50)	57 (50)	56 (246)
Women:	62 (53)	63 (50)	68 (54)	66 (47)	68 (50)	65 (254)

↓ **did not participate: 82**
were excluded: 2

THREE-YEAR FOLLOW-UP STUDY POPULATION

Age-specific (years) participation rates

	40	45	50	55	60	All
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Men:	88 (44)	80 (37)	84 (42)	90 (45)	92 (46)	87 (214)
Women:	77 (41)	80 (40)	72 (39)	85 (40)	84 (42)	80 (202)

FIGURE 1 Selection of the study population.

In both cross-sectional (baseline) and follow-up studies the subjects attended two measurement sessions at the Institute. At the first visit a standard pre-testing health screening (Suni et al. 1998) was conducted by laboratory technicians and fitness testers. The fitness testers used the pre-testing health screening results, according to safety instructions made in advance, to refer subjects with severe diseases or symptoms to a physician for a health examination, or to exclude subjects with minor limitations from selected fitness tests. Accordingly, eight subjects (2%) out of 500 in the cross-sectional study were referred to a physician, seven of them participated in selected tests and one was excluded form the study population. The fitness tester excluded 45% from selected tests and 53% of the subjects had no limitations to any of the fitness tests. A detailed description of the screening procedure and exclusion criteria for fitness testing has been reported elsewhere (Suni et al. 1998b). The participants also answered a questionnaire on perceived health and musculoskeletal functioning, and physical activity and other living habits. An additional subject was excluded from the cross-sectional study population (n=498) because he did not complete the physical activity questionnaire. At the second visit assessment of HRF was conducted by three testers all of whom had a masters degree in sport or health sciences. The study team had substantial experience in research projects and was thoroughly trained to conduct all measurements according to standard procedures. In case of emergency during fitness testing a physician and nurses with skills and equipment for cardiopulmonary resuscitation were available on site.

TABLE 1. Reasons for not participating in the follow-up study among subjects who dropped out after the cross-sectional study.

Reason for not participating (n=82)	%	(n)
Subjects who answered the phone interview	79.3	(65)
• Health-related hindrance	26.8	(22)
-worsened musculoskeletal symptoms	9.8	(8)
-infant baby or pregnancy or recent miscarriage	6.1	(5)
-new cardiovascular disease	2.4	(2)
-other	8.5	(7)
• Work-related hindrance	24.4	(20)
• Not interested	9.8	(8)
• Has no time	8.5	(7)
• Negative experience from the cross-sectional study	4.9	(4)
• Too tired to participate	2.4	(2)
• Moved out of town or changed address	2.4	(2)
Subjects who did not answer the phone interview	20.7	(17)
• Not reached at all	12.2	(10)
• Reached, but not willing to answer the phone interview	6.1	(5)
• Deceased	2.4	(2)

Assessment of health-related fitness at baseline. All tests followed a standard sequence: balance (one-leg standing) was measured first, followed by flexibility (trunk side-bending, hamstring muscle extensibility), muscular power (vertical jump) and strength (one-leg squat) of lower extremities, upper body (modified push-ups) and trunk muscular endurance (static trunk extension). After these tests the subjects rested for about 10 minutes before performing the UKK 2-km Walk Test. The subjects were given verbal encouragement in a consistent manner to achieve their best performance. The components and items in the HRF test battery with short description of the methods are given in Figure 2. The reliability (Oja et al. 1991, Suni et al. 1996), safety, and feasibility of the methods have been reported elsewhere (Laukkanen et al. 1992, Suni et al. 1998b).

For all test scores the subjects were grouped into fitness categories based on age and sex specific cut points. The least fit 40% of the participants were classified as 'low-fit', the mid 20% as 'mid-fit' and the most fit 40% as 'high-fit'.

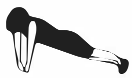
MUSCULOSKELETAL FITNESS
Muscular strength and endurance

ONE-LEG SQUAT with increasing weights
Purpose: To assess functional leg extensor strength.
Method: The subject takes a short step forward on the mat, first with the right leg, squats down until lightly touching the mat with the left knee, rises immediately and steps backward to the starting position; she then repeats the squat with the left leg.
Outcome: The load limit for a successful squat task measured as the maximum weight relative to the subject's body weight (BW) up to 140%. The test starts with BW (i.e., no added weight) and 10% increments of BW are added at 4 successive steps of 10%, 20%, 30% and 40% using a weight belt system.

VERTICAL JUMP
Purpose: To measure leg extensor power.
Method: The subject is instructed to jump as high as possible and to bend the knees and swing the arms to enhance performance. Before jumping, one arm is raised to mark the standing height. During the jump the subject touches a board with the middle finger at the highest position.
Outcome: The difference between the reach height and the jump height is measured with a tape to the nearest centimeter.

STATIC TRUNK EXTENSION ENDURANCE
Purpose: To measure the endurance capacity of the trunk extensor muscles.
Method: The subject lies prone with the lower body (from the level of the spina iliaca anterior superior) resting on a low bench and crosses the hands behind the neck. The bench is placed on one gymnastic mat and another mat is placed on the bench. The tester stabilizes the subject by sitting on the subject's ankles. The subject is asked to rise the upper body to a horizontal level and hold the position as long as possible for up to 4 minutes.
Outcome: Endurance time in seconds.

MODIFIED PUSH-UPS
Purpose: To measure dynamic muscular endurance of the upper-extremity extensor muscles and the ability to stabilize the trunk.
Method: The subject lies prone. The push-up cycle begins as the subject claps the hands behind the back. Next a normal straight-leg push-up to straight elbows is performed, followed by a touch of one hand to the top of the supporting hand. The cycle ends in the prone lying position.
Outcome: The number of push-ups completed in 40 seconds.



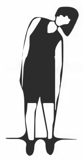
Flexibility

HAMSTRING MUSCLE EXTENSIBILITY
Purpose: To measure the active knee extension range of motion in order to assess hamstring muscle extensibility.
Method: The subject lies supine. The hip and the knee of the limb to be measured is flexed to 90 degrees. The opposite leg rests extended. The inclinometer (Vinkelmätare "Myrin", LIC, Rehab Vårdum,Solna, Sweden) is attached to the medial side of the ankle.
Outcome: End point range of motion angle in degrees at maximal extension.

Figure 2 Health-related fitness assessment methods.

Figure 2 continued

TRUNK SIDE-BENDING to the right and left
Purpose: To measure the total range of movement of lateral flexion of the thoracic and lumbar spine and pelvis.
Method: The subject stands on marked lines (15 cm apart) with the back against the wall. Arms are kept straight at the sides of the body. The subject bends to the right and then to the left as far as possible; the middle finger slides laterally down along the thigh.
Outcome: The distance the fingertip moves down the leg during maximal bending measured with a cloth tape in millimeters. The average value of the right and left sides is calculated.



MOTOR FITNESS
Static balance

ONE-LEG STANDING
Purpose: To assess static postural control while the area of support is reduced.
Method: The subject wears sport shoes. He or she places one foot at knee level along the inner side of the supporting leg and rotates the thigh outwards. The subject is advised to stand as still as possible.
Outcome: Duration of the balance task up to 1 minute as measured with a stopwatch in seconds.

CARDIORESPIRATORY FITNESS

UKK 2-KM WALK TEST
Purpose: To predict maximal oxygen uptake (VO_{2max}) on the basis of time, heart rate at the end, body mass index, age and sex.
Method: Subject walks as fast as possible on a flat surface using a normal walking style.
Outcome: Predicted VO_{2max} ($ml/min^{-1}/kg^{-1}$) and test time (min).

MORPHOLOGY

BODY MASS INDEX (BMI)
Purpose: To assess obesity.
Method: Standard measures of height and weight.
Outcome: BMI as weight/height^2 (kg/m^2).

The distributions of the balance and one-leg squat test scores were skewed, and only two fitness categories were formed. Accordingly, from 28% to 41% of the men in the five age groups were assigned to 'low-fit' group in the balance test, and from 10% to 50% in the squat test. Corresponding percentages of the women were from 40% to 43% and from 21% to 39%, respectively.

Assessment of leisure time physical activity (LTPA) at baseline. LTPA was assessed using a self-administered questionnaire. In short, the following aspects of baseline LTPA were questioned (Oja et al. 1994): (i) participation in *leisure-time exercise* primarily for keeping fit and healthy in terms of frequency, intensity, duration, mode of exercise and average weekly exercise time (min), (ii) *active transportation* including daily walking distance (km) and daily cycling time (min), (iii) the average weekly time (min) spent in *other physical leisure time activities*, categorised as light, brisk or strenuous, modified after the study by Haapanen et al. (1996). A detailed description of the methods has been given elsewhere (Suni et al. 1999). In addition, the maintenance of LTPA since leaving the school was asked.

The data of current LTPA questions was used to construct two variables: overall LTPA level and exercise type. For overall LTPA, subjects engaged in any type of brisk physical activity three times a week or more were categorised as 'active'. In the present study 'brisk' referred to activity that causes some perspiration and shortness of breath. It is comparable to 'moderate' intensity level as presented by American College of Sports Medicine (1). Subjects engaged in LTPA once or twice a week were categorised as 'somewhat active', and the rest as 'inactive' (for details see Suni et al. 1999). Exercise modes were characterised as aerobic (typical aerobic training including repetitive movement patterns for neuromuscular system e.g. walking, jogging, cycling, skiing, swimming) or muscular (team or individual sport including various types of movement patterns for neuromuscular system e.g. ball games, racquet games, aerobics, home gymnastics, strength training, downhill skiing) according to the principle presented by Kujala et al. (1994). Subjects exercising briskly or vigorously at least two times a week and engaged in one or more muscular-type exercises at least an hour a week were categorised as 'muscular exercise' group. The rest of the 'active' or 'somewhat active' subjects were categorised as 'aerobic exercise' group. The 'inactive' group was the same as for overall HEPA.

Assessment of three year changes in perceived health, mobility, and back functioning and pain. A self-administered questionnaire was used to measure three aspects of health within the classification scheme of health-related quality of life (Wilson & Cleary 1995). Accordingly, general health perception represents an individual integration of many aspects of the health concept. Questions on functional status reveal the individuals' perception of their ability to perform particular tasks and are associated with general health perceptions. Symptom status is one important determinant of functional status (Wilson & Cleary 1995, Harper et al. 1992).

The health questions were as follows: (a) 'how would you describe your state of health in comparison to people of your age?', (b) 'how well can you manage climbing several blocks of stairs without resting?, (c) 'how would you describe the functioning of your back?, (d) 'how often do you have problems in your back while functioning in a stooped position? and (e) 'how often do you have back pain? Categories of health questions, classification of the subjects and descriptive results at baseline were presented in the former article by Suni et al. (1998a).

For the follow-up study, the perceived health and musculoskeletal functional status of the subjects was dichotomised as either 'low or 'high' at baseline. The three-year changes were assessed separately among subjects with 'low' and 'high' status at baseline. The outcome was 'improved status' among subjects with 'low status' at baseline and 'high status' at follow up. The outcome was 'deteriorated status' among subjects with 'high status' at baseline and 'low status' at follow up. The outcome for other subjects was 'no change'. Descriptive results of the health outcomes are presented in Table 2.

TABLE 2. Descriptive results of the changes in health outcome measures at follow-up study.

HEALTH OUTCOME categories (baseline status)	Age-group (years at baseline) 37	42	47	52	57	Men	Women	All
PERCEIVED HEALTH	%	%	%	%	%	%	%	n
no change (low)	67	61	60	65	68	62	66	267
no change (high)	10	22	24	17	11	19	14	68
improved (low)	15	12	6	9	13	10	12	46
deteriorated (high)	8	5	10	9	8	9	8	34
ABILITY TO CLIMB STAIRS								
no change (low)	7	8	13	25	31	9	24	70
no change (high)	80	73	75	54	57	80	55	280
improved (low)	4	11	7	8	7	4	11	31
deteriorated (high)	9	8	5	13	5	7	10	34
BACK FUNCTIONING AND PAIN								
General back functioning								
no change (low)	55	53	49	57	71	53	61	237
no change (high)	32	24	28	21	13	26	21	97
improved (low)	5	14	17	6	6	9	10	39
deteriorated (high)	8	9	6	16	10	12	8	42
Back problems while stooping								
no change (low)	28	22	31	34	49	28	39	138
no change (high)	48	43	46	38	20	43	35	161
improved (low)	11	17	14	21	17	15	17	66
deteriorated (high)	13	18	9	7	14	14	9	50
Back pain frequency								
no change (low)	37	34	36	47	56	38	46	175
no change (high)	35	31	43	32	20	36	28	133
improved (low)	14	23	12	7	17	15	15	61
deteriorated (high)	14	12	9	14	7	11	11	46

Statistical analysis. Logistic regression models were used to assess the associations of HRF and LTPA with the changes in perceived health and musculoskeletal functioning. The low-fit groups (least fit 40%) and inactive group were the reference groups in the models. Sex, age, occupational physical activity (heavy physical work vs. other groups), smoking (current smoker vs. other groups), level of education (secondary school vs. higher levels) and marital status (single vs. married or cohabiting) were included in all models as possible confounders. Series of univariate analysis using the 10 fitness (see Figure 1) and three physical activity variables (overall LTPA, exercise type, continuity of LTPA since school) as independent variables and each health change outcome as dependent variable were conducted first. When the 95% confidence interval (CI) of odds ratio (OR) did not include the value of 1.00 the result was considered statistically significant ($p<0.05$). Secondly, HRF and LTPA variables that had significance level of $p<0.15$ in univariate analysis were entered into multivariate logistic regression analysis conducted in a stepwise manner ($p<0.01$).

RESULTS

Associations of baseline HRF and LTPA with three year changes in perceived health. In the univariate analyses muscular-type exercise (OR 3.95, CI 1.62 to 8.94), high overall level of LTPA (3.00, 1.08 to 8.33), high cardiorespiratory fitness according to predicted VO_{2max} (3.80, 1.62 to 8.94) and Walk Test time (2.66, 1.22 to 5.81), and sustained LTPA since school (2.45, 1.26 to 4.76) were positively associated with *improved perceived health* when the active or high-fitness group was compared with the inactive or low-fitness group. The variables selected into the multivariate model (Table 3) were predicted VO_{2max} and continuity of LTPA. For *deteriorated perceived health,* in the univariate analysis, high cardiorespiratory fitness according to predicted VO_{2max} (0.21, 0.06 to 0.81) and Walk Test time (0.25, 0.08 to 0.79) showed negative association when the high-fitness group was compared with the low-fitness group. The selected variable in the multivariate analysis, (Table 3), was predicted VO_{2max}. Musculoskeletal and motor fitness, and body composition were not significantly associated with changes in perceived health.

TABLE 3. Associations of baseline health-related fitness (HRF) and leisure-time physical activity (LTPA) with 3-year changes in perceived health and ability to climb several flights of stairs [multivariate stepwise logistic regression analysis adjusted for sex, age, marital status, education level, smoking and occupational physical activity: odds ratios (OR) with 95% confidence intervals (CI)].

• Baseline HRF & LTPA variables selected into the regression models		CHANGE IN HEALTH AND FUNCTIONAL STATUS					
		Positive change 'low status' at baseline			**Negative change** 'high status' at baseline		
	categories	OR	95% CI	*p-value*	OR	95% CI	*p-value*
HRF		**Improved perceived health** (n=243)			**Deteriorated perceived health** (n=89)		
• Walk Test,	low 40%	1.0		*0.002*	1.0		*0.017*
predicted	mid 20%	2.25	0.80-6.34		0.90	0.19-4.31	
VO_{2max}	high 40%	3.47	1.39-8.70		0.21	0.06-0.81	
LTPA							
• Continuity	non-cont.	1.0		*0.032*			
since school	cont.	2.30	1.07-4.94				
HRF		**Improved ability to climb stairs** (n=95)			**Deteriorated ability to climb stairs** (n=265)		
• Walk Test	low 40%	1.0		*0.002*			
time	mid 20%	2.84	0.67-12.1				
	high 40%	12.4	2.61-58.7				
• One-leg	low 40%				1.0		*0.005*
squat	high 60%				0.28	0.12-0.66	
LTPA							
• Overall	inactive	1.0		*0.004*			
level	mod. active	8.17	1.77-37.7				
	active	1.72	0.29-10.4				

cont.=continuos, mod.=moderately

Associations of baseline HRF and LTPA with three year changes in self-reported ability to climb stairs. In the univariate analyses high cardiorespiratory fitness according to Walk Test time (OR 8.57, CI 2.34 to 31.4), muscular-type exercise type (8.36, 1.86 to 37.5), moderate overall level of LTPA (5.41, 1.50 to 19.6) and high musculoskeletal fitness in trunk side-bending flexibility (3.52, 1.17 to 10.5) were positively associated with *improved ability to climb stairs*. when compared with the low-fitness or inactive group. The variables selected to the multivariate model (Table 3) were the Walk Test time and overall LTPA. The results of the corresponding multivariate analyses are presented in table 3. For overall LTPA the 'moderately active' group but not the 'active' group had increased OR when compared with the 'inactive' group. For *deteriorated ability to climb stairs*, in the univariate analyses, high musculoskeletal fitness in one-leg squat (0.29, 0.14 to 0.64) and mid-fitness in modified push-ups (0.27, 0.07 to 0.99), high overall level of LTPA (0.32, 0.12 to 0.86), and high cardiorespiratory fitness according to predicted VO_{2max} (0.39, 0.16 to 0.97) were negatively associated. The only variable selected into multivariate stepwise logistic regression analysis (Table 3) was one-leg squat. Musculoskeletal fitness in lower extremity power, hamstring muscle extensibility and trunk extensor muscular endurance, motor fitness in static balance, and body composition were not associated with self-reported changes in ability to climb stairs.

Associations of baseline HRF and LTPA with three year changes in self-reported back functioning and pain. In the univariate analyses *for improved general back functioning*, musculoskeletal fitness in terms of mid-level in modified push-ups (OR 4.32, CI 1.69 to 11.0) and trunk side-bending (3.48, 1.34 to 9.07), and high level in hamstring muscle extensibility (2.85, 1.23 to 6.60) and trunk extension endurance (2.62, 1.17 to 5.89), as well as body composition in terms of low BMI (2.28, 1.03 to 5.05) were positively associated when compared with the low-fitness group. The variables selected to the multivariate model (Table 4) were modified push-ups and trunk side-bending. For both of the selected variables the 'moderately active' group but not the 'active' group had increased OR compared with the 'inactive' group. In the univariate analyses for *improved back functioning while stooping* mid-level of musculoskeletal fitness in hamstring muscle extensibility (3.61, 1.60 to 8.15) and high cardiorespiratory fitness according to Walk Test time (2.34, 1.17 to 4.68) were positively associated when compared with the low-fitness group. In the corresponding multivariate analysis (Table 4) the only selected variable was hamstring muscle extensibility and the 'moderately active' group but not the 'active' group had increased OR compared with the 'inactive' group. There were no associations of HRF and LTPA with *positive changes in back pain frequency*.

For deteriorated general back functioning, in the univariate analyses, high level of musuloskeletal fitness in trunk side-bending (OR 0.34, CI 0.14 to 0.84) and high- (0.36, 0.14 to 0.92) and mid- (0.36, 0.13 to 0.97) levels of hamstring muscle extensibility were negatively associated when compared with the low-fitness group. The variables selected into the multivariate analysis (Table 4) were trunk side-bending and overall level of LTPA. Both the 'mid-fit' and 'high-fit' groups in side-bending had decreased OR. Concerning LTPA the 'moderately active' group had decreased OR of deteriorated general back functioning compared with the 'active' group, but did not differ from the 'inactive' group. In the univariate analyses for deteriorated back functioning while stooping continuous LTPA since school (2.29, 1.16 to 4.52) was the only variable showing significant association, and was selected into multivariate model (Table 4). Continuos LTPA since school increased the OR of deteriorated back functioning while stooping when compared with non-continuous LTPA. There were no associations of HRF and LTPA with negative changes in back pain frequency.

TABLE 4. Associations of baseline health-related fitness (HRF) and leisure-time physical activity (LTPA) with 3-year changes in general and task specific (stooping) back functioning [multivariate stepwise logistic regression analysis adjusted for sex, age, marital status, education level, smoking and occupational physical activity: odds ratios (OR) with 95% confidence intervals (CI)].

• Baseline HRF & LTPA variables selected into the regression models	categories	CHANGE IN HEALTH AND FUNCTIONAL STATUS Positive change 'low status' at baseline OR	95% CI	*p-value*	Negative change 'high status' at baseline OR	95% CI	*p-value*
HRF		**Improved general back functioning** (n=242)			**Deteriorated general back functioning** (n=137)		
• Modified push-ups	low 40%	1.0		*0.009*			
	mid 20%	4.13	1.57-10.8*				
	high 40%	1.39	0.53-3.57				
• Trunk side-bending flexibility	low 40%	1.0		*0.028*	1.0		*0.019*
	mid 20%	4.09	1.46-11.5		0.29	0.10-0.86	
	high 40%	1.73	0.70-4.24		0.31	0.12-0.78	
LTPA	inacitve				1.0		*0.035*
• Overall level	mod. active				0.58	0.18-1.86**	
	active				1.95	0.62-6.17	
HRF		**Improved back functioning while stooping** (n=190)			**Deteriorated back functioning while stooping** (n=205)		
• Hamstring muscle extensibility	low 40%			*0.002*			
	mid 20%	4.25	1.82-9.97*				
	high 40%	1.33	0.62-2.83				
LTPA							
• continuity since school	non-cont.				1.0		*0.011*
	cont.				2.40	1.21-4.77	

*Differs significantly also from the high-fitness group **Differs significantly from the high-fitness group. cont.=continuos, mod.=moderately

Musculoskeletal fitness in lower extremity strength (one-leg squat) and power (vertical jump), motor fitness in one-leg balance, cardiorespiratory fitness in terms of predicted VO_{2max}, and exercise type showed no associations with the changes in back functioning and pain.

DISCUSSION

In order to evaluate and monitor the fitness level of large groups and populations field-based measures of HRF that are reliable, safe, economic, and easy to administer are needed. Most importantly, the validity of the tests should be established by demonstrating significant and meaningful relationship with health and functional status. This is necessary to determine the contribution of a particular fitness component to important health outcomes (Phillips and Haskell 1994), as well as to interpret the fitness scores in terms of adequacy of fitness with regard to health (Oja and Tuxworth 1995). The prospective evaluations of the present study tested whether baseline HRF level and LTPA had value in predicting the three-year changes in perceived health and musculoskeletal functioning.

These results, together with the former findings on the health-related validity of this test battery (Suni et al. 1998a), increase our understanding of its possibilities and limitations in the assessment of HRF within the context of health promotion.

The study subjects were initially drawn from a representative population, but were, due to multistage sampling (see Figure 1), to some extend, selected. They had higher education level, rated their health better and were more physically active than the non-participating part of the study sample. Despite of this, there were subjects with health-limitations (Suni et al. 1998a and 1998b) and the subjects reported a variety of LTPA patterns in terms of exercise frequency, intensity, duration and mode of activity (Suni et al. 1999). This enabled us to investigate the safety and feasibility, as well as the content validity of the test battery among middle-aged adults with reasonable variance. However, the contrast in different groups was less than in the general population. This is likely to lead into underestimation of the strength of the associations of HRF and LTPA with health (Blair et al. 1996). In addition, we cannot estimate the impact of genetic factors on the relationships between HRF and the health outcomes or between LTPA and the health outcomes (Bouchard & Perusse 1994). Furthermore, the LTPA and HRF data of the present study were collected at one point in time so that the intra-individual changes in them during the follow-up time were not identified (van Heuvelen et al. 1998). This could have affected the associations between baseline LTPA and HRF with the three year changes in perceived health and musculoskeletal functioning.

The major limitation of the present study is that simple self-ratings of health were used as outcome measures to validate the field-based HRF test battery among middle-aged adults. With respect to mobility and back functioning the other alternatives would have been a clinical examination or use of established disease-specific questionnaires. However, these alternatives were not considered feasible due to the significant human resource and time requirements.

The results provide some new views of the role of physical fitness in the development of musculoskeletal disability among middle-aged adults. In general, the findings suggest that several factors of musculoskeletal and motor fitness, as well as aerobic fitness in terms of walking, are independent risk factors for mobility- and back-related disability. Similar findings on functional decline were recently reported by Huang et al.(1998) and Morey et al. (1998). A more advanced statistical model is needed to further study the possible mediating role of physical fitness between physical activity and health as suggested by Rantanen et al. (1999). Ultimately, randomised controlled trials on the effects of specific type fitness training are needed to confirm the role of physical activity and HRF in the prevention of musculoskeletal disability.

Predictive health-related validity of the single fitness test items. *Musculoskeletal fitness: strength and power of the lower extremities*. The one-leg squat and vertical jump tests were expected to have predictive health-related validity for ability to climb stairs. Accordingly, one-leg squat strength test was negatively associated with self-rated deterioration in ability to climb stairs among those with no functional problems at baseline (see Table 3). These finding are in agreement with several other prospective studies among the elderly (Hoeymans et al. 1994, Guranik et al. 1995, Schroll et al. 1997). Vertical jump had no predictive value for changes in perceived health and self-rated musculoskeletal functioning. In a former prospective study (Fujita et al. 1995) poor leg power in men, but not in women, was associated with an excess risk of death from cardiovascular and all causes.

Musculoskeletal fitness: trunk and upper-body muscular endurance. Trunk and upper-body function tests were expected to have health-related validity for back functioning and pain. Trunk extensor endurance showed predictive value for self-rated improvement in general back functioning among those with 'low' functional status at baseline, but only in univariate analysis. This is somewhat contradictory to the former studies (Biering-Sørensen 1984, Luoto et al. 1995) indicating predictive value for first-time back pain. The *modified push-up* test assesses dynamic endurance strength of the upper-body and ability to stabilize the trunk. The 'mid-fit' but not the 'high-fit' group was more likely to have improved general back functioning compared with 'low-fit' (see Table 4). This may be due to the strong baseline association: there were not many 'high-fit' subjects among those who's back functioning at baseline was categorised as 'low'. Unexpectedly, modified push-up test was also associated with future mobility function in univariate analysis, but was not selected into the stepwise multivariate logistic regression models (see Table 3). This may be due to the high correlations (Suni et al. 1998a) of modified push-up test with Walk Test time (r= -0.42) and one-leg squat test (r=0.43).

Musculoskeletal fitness: flexibility. The hamstring muscle extensibility test was expected to show health-related validity to self-rated back functioning and pain. The 'mid-fit' group was more likely to improve functioning while stooping (see Table 4). These results may indicate a non-linear relationship between hamstring extensibility and low back dysfunction, a finding not presented in former studies (Biering-Sorensen 1984, Nicolaisen & Jørgensen 1985, Esola et al. all 1996, Thomas et al. 1998). No former studies have assessed the relationship of hamstring extensibility to task-specific functional problems of stooping but it has been related to total hip and trunk flexion movement in low back patients (Esola et al. 1996).

Reduced spinal flexibility has been found as residual sign of persistent low back disorder The results from two former prospective studies are conflicting: trunk side-bending had no predictive value for back injury (Battie et al. 1990), but poor lateral mobility to the right increased the risk for low back pain (Videman et al. 1989). In the present study the 'mid-fit' subjects with low functional status at baseline were more likely to improve their general back functioning compared with the 'low-fit' and 'high-fit' subjects (see Table 4). Among subjects with high function status at baseline both the 'mid-fit' and 'high-fit' subjects were less likely to have deteriorated general back functioning than the 'low-fit' (see Table 4). The former finding is in agreement with cross-sectional findings on lumbar flexion and extension mobility by Burton et al. (1989).

Motor fitness. The *one-leg standing balance* test was expected to show predictive health-related validity to self-rated changes in mobility (stair climbing), and back functioning and pain. Contrary to the expectations and former longitudinal (Hoeymans et al. 1994, Guralnik et al. 1995) studies among the elderly, balance was not associated with stair climbing ability among this middle aged study population. No associations with self-rated changes in back functioning and pain were found either. Other former prospective studies on postural control have been follow-up interventions (Luoto et al. 1996 & 1998). In cross-sectional studies balance has been associated with low back disorders (Byl & Sinnot 1992, Luoto et al.1996 & 1998).

Cardiorespiratory fitness. Predicted VO_{2max} of the 2-km Walk Test showed strong associations with changes in perceived health. (see Table3). High-fitness in predicted VO_{2max} was also associated with decreased OR of negative changes in mobility function (among subjects with no difficulty at baseline, but only in univariate analysis. Among those with 'low' functional status at baseline, high-fitness according to Walk Test time had strong predictive value for self-rated improvement in mobility (see Table 3), and some predicted value for improved back functioning while stooping in univariate analysis. The former finding is in line with several prospective studies among the elderly (Hoeymans et al. 1994, Guralnik et al. 1995, Schroll et al. 1997).

Body composition. There was only one significant univariate association of low BMI with improved back functioning among those with 'low status' at baseline, which disagrees with former prospective studies on functional disability (Rissanen et al. 1990, Launer et al. 1994).

Associations of LTPA with three year changes in perceived health and musculoskeletal functioning. Muscular-type exercise at baseline was expected to be associated with the changes in self-rated musculoskeletal functioning. The only significant univariate association, although strong, with improved ability to climb stairs was for overall LTPA (OR 8.36), and it was selected into the multivariate model. Relating to overall LTPA level, moderately active (LTPA one to two time a week) rather than active (LTPA three times a week or more) subjects with self-rated difficulties in stair climbing at baseline ('low status') were more likely to improve their mobility at the follow-up when compared with inactive subjects (see Table 3). Similarly, moderately active subjects with good self-rated general back functioning at baseline were less likely to have deteriorated back functioning at the follow-up compared with both active and inactive subjects (see Table 4). Furthermore, sustained activity since school increased the risk of deteriorated back function while stooping among those with 'high status' at baseline. These findings bring out the problems and open questions related to the proper type and dose of physical activity needed to enhance musculoskeletal functioning and health. Sustained (continuous) physical activity since school was associated with improved perceived health at the follow-up. This is in line with the recent HEPA recommendations emphasising regular (or continuous) physical activity (Pate et al. 1995, Surgeon General 1996, ACSM 1998).

CONCLUSIONS

Based on their relations to changes in self-rated musculoskeletal functioning the following fitness tests proved to have predictive validity for health: One-leg squat test of lower extremity strength, hamstring muscle extensibility test of lower limb flexibility, modified push-up test of upper-body and trunk muscular function, and trunk side-bending test of spinal flexibility. The 2-km Walk Test of aerobic fitness proved to have predictive validity for both changes in self-rated health and musculoskeletal functioning. Trunk extension test of muscular endurance and BMI showed weak predictive validity for self-rated changes in back functioning. Vertical jump test of lower extremity power and one-leg standing balance test had no predictive validity for perceived health and musculoskeletal functioning. These findings give further support to the HRF testing concept and indicate validity for several test items of the proposed test battery. However, there is a continuous need to assess the validity of former and new fitness factors for different and more objective health outcomes in long-term prospective studies. Furthermore, randomised controlled fitness training trials with scientifically sound fitness assessment methods are needed to confirm the role of physical activity and HRF in the prevention of musculoskeletal disability.

Acknowledgement

The Finnish Ministry of Education granted financial support for this study.

References

ACSM (American College of Sports Medicine). Position stand on the recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness and flexibility in healthy adults. *Med. Sci. Sports Exerc.* 30:975-991, 1998.

Battie, M. C., S. J. Bigos, L. D. Fisher, D. M. Spengler, T. H. Hansson, A. L. Nachemson, and M. D. Wortley. The role of spinal flexibility in back pain complaints within industry. *Spine* 15:768-73, 1990.

Biering-Sorensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine* 9:106-119, 1984.

Blair, S.N., J. B. Kampert, H. W. Kohl, C. E. Barlow, C. A. Macera, R. S. Paffenbarger, and L. W. Gibbons. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *JAMA* 276:205-210, 1996.

Bouchard, C. and L. Pérusse. Heredity, activity level, fitness and health. In: *Physical Activity, Fitness, and Health*. C. Bouchard, R. J. Shephard, T. Stephens (Eds.) Champaign, IL: Human Kinetics, 1994, pp. 106-118.

Bouchard, C., and R. J. Shephard. Physical activity, fitness, and health: the model and key concept. In: *Physical Activity, Fitness, and Health.* C. Bouchard, R. J. Shephard, T. Stephens (Eds.) Champaign, IL: Human Kinetics, 1994, pp. 77-88.

Bravo, G., P. Gauthier, P.-M. Roy, D. Tessier, P. Gaulin, M.-F. Dubois, and L. Péloquin. 1994. The functional fitness assessment battery: reliability and validity data for elderly women. *J. Aging Phys. Act.* 2:67-79, 1994.

Buckwalter, J.A., and N. E. Lane. Athletic and osteoarthritis. *Am. J. Sports Med.* 25: 873-881, 1997.

Byl, N. N., and P. L. Sinnot. Variations in balance and body sway in middle-aged adults: subjects with healthy backs compared with subjects with low-back dysfunction. *Spine* 16:325-330, 1991.

Davis, M. A., W. H. Ettinger, J. M. Neuhaus, and K. P. Mallon. Knee osteoarthritis and physical functioning: evidence from the NHANES I epidemiologic follow-up study. *J. Rheum.* 18: 591-598, 1991.

Esola, M.A., P. W. McClure, F. G. Kelley, and S. Siegler. Analysis of lumbar spine and hip motion during forward bending in subjects with and without a history of low back pain. *Spine* 21:71-78, 1996.

Fujita, Y., Y. Nakamura, J. Hiraoka, K. Kobayashi, K. Sakata, M. Nagai, and H. Yanagawa. Physical strength tests and mortality among visitors to health-promotion centers in Japan. *J. Clin. Epidemiol.* 48:1349-1359, 1995..

Guralnik, J.M., L. Ferrucci, E. M. Simonsick, M. E. Salive, and R. B. Wallace. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *New Engl. J. Med.* 332:556-561, 1995.

Hagen, K., and O. Thune. Work incapacity form low back pain the general population. *Spine* 23:2091-2095, 1998.

Harper, A.C., D. A. Harper, L. J. Lamber, H. B. Andrews, S. Kai Lo, F. M. Ross, and L. M. Straker. Symptoms of impairment, disability and handicap in low back pain: a taxonomy. *Pain* 50:189-95, 1992.

Hoeymans, N., E. J. M Feskens, G. A. M. van den Bos, and D. Kromhout. Measuring functional status: cross-sectional and longitudinal associations between performance and self-report. *J. Clin. Epidemiol.* 49:1103-1110, 1996.

Huang, Y., C. A. Macera, S. N. Blair, P. A. Brill, H. W. Kohl III, and J. J. Kronenfeld. Physical fitness, physical activity, and functional limitations in adults aged 40 and older. *Med. Sci. Sports Exerc.* 30:1430-1435, 1998.

JKing C. N., and M. D. Senn. Exercise testing and prescription. Practical recommendations for the sedentary. *Sports Med.* 21:326-336, 1996.

Koch, M., M. Gottschack, D. I. Baker, S. Palumbo, and M. Tinetti. In impairment and disability assessment and treatment protocol for community living elderly persons. *Phys. Ther.* 74:286-294, 1994.

Kujala, U. M., T. Viljanen, S. Taimela, and J. T. Viitasalo. Physical activity, VO_{2max}, and jumping height in an urban population. *Med. Sci. Sports Exerc.* 26:889-895, 1994.

Laukkanen, R.M.T., P. Oja, K. H. Ojala, M. E. Pasanen, and I. M. Vuori. Feasibility of a 2-km walking test for fitness assessment in a population study. *Scand. J. Soc. Med.* 20:119-126, 1992.

Launer, L.J., T. Harris, C. Rumpel, and J. Madans. Body mass index, weight change, and risk of mobility disability in middle-aged and older women. The epidemiologic follow-up study of NHANES I. *JAMA* 271:1093-1098, 1994.

Luoto, S., M. Heliövaara, H. Hurri, and H. Alaranta. Static back endurance and the risk of low back pain. *Clin. Biomec.* 10:323-324, 1995.

Luoto, S., S. Taimela, H. Hurri, H. Aalto, I. Pyykkö, and H. Alaranta. Psychomotor speed and postural control in chronic low back pain patients. *Spine* 21:2621-2627, 1996.

Luoto, S., H. Aalto, S. Taimela, H. Hurri, I. Pyykkö, and H. Alaranta. One-footed and externally disturbed two-footed postural control in patients with chronic low back pain and healthy control subjects. *Spine* 23:2081-2090, 1998.

Mäkelä, M., M. Heliövaara, K. Sievers, P. Knekt, J. Maatela, and A. Aromaa. Musculoskeletal disorders as determinants of disability in Finns aged 30 years or more. J Clin Epidemiol 46:549-59, 1993.

Morey, M. C., C. F. Pieper, and J. Cornoni-Huntley. Physical fitness and functional limitations in community-dwelling older adults. *Med. Sci. Sports Exerc.* 30:715-723, 1998.

Nicolaisen, T., and K. Jørgensen. Trunk strength, back muscle endurance and low-back trouble. *Scand. J. Rehabil. Med.* 17:212-227, 1985.

Oja P., R. Laukkanen, M. Pasanen, T. Tyry, and I. Vuori. A 2-km walking test for assessing the cardiorespiratory fitness of healthy adults. *Int. J. Sports Med.* 12:356-362, 1991.

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

Oja, P., and W. Tuxworth (Eds.) Eurofit for adults. Assessment of health-related fitness. Tampere Finland: Council of Europe Publishing, 1995, pp. 37-83.

Pate, R.R., M. Pratt, S. N. Blair, et al. Physical activity and public health. A recommendation from the Centres for Disease Control and Prevention and the American College of Sports Medicine. *JAMA* 273:402-407, 1995.

Phillips, W.T., and W. L. Haskell. 'Muscular fitness' easing the burden of disability for elderly adults. *J. Aging Phys. Act.* 3:261-289, 1994.

Province M. A., E. C. Hadley, L. A. Hornbrook, et al. The effect of exercise on falls in elderly patients. A preplanned meta-analysis of the FICSIT Trials. *JAMA* 273: 1341-1347, 1995.

Rantanen, T., J. M. Guralnick, R. Sakari-Rantala, S. Leveille, M. E. Simonsick, S. Ling. S., and L. P. Fried. Disability, physical activity, and muscle strength in older women: the women's health and aging study. *Arch. Phys. Med. Rehabil.* 80:130-135, 1999.

Rejeski, W. .J, W. H. Ettinger, S. Schumaker, P. James, R. Burns, and J . T. Elam. Assessing performance-related disability in patients with knee osteoarthritis. *Osteoarthritis and Cartilage* 3: 157-167, 1995.

Rissanen, A., M. Heliövaara, P. Knekt, A. Reunanen, A. Aromaa, and J. Maatela. Risk of disability and mortality due to overweight in a Finnish population. *BMJ* 301:835-837, 1990.

Schroll, M., K. Avlund, and M. Davidsen M. Predictors of five-year functional ability in a longitudinal survey of men and women aged 75 to 80. The 1914-population in Glostrup, Denmark. *Aging Clin. Exp. Res.* 9:143-152, 1997.

Shaulis, D., L. A. Golding, and R. D. Tandy. Reliability of the AAHPERD functional fitness assessment across multiple practise sessions in older men and women. *J Aging Phys. Act.* 2:272-279, 1994.

Skinner, J., and P. Oja. Laboratory and field tests for assessing health-related fitness. In: *Physical Activity, Fitness, and Health.* C. Bouchard, R. J. Shephard, T. Stephens (Eds.) Champaign, IL: Human Kinetics, 1994, pp. 160-179.

Suni, J. H., P. Oja, R. T. Laukkanen, S. I. Miilunpalo, M. E. Pasanen, I. M. Vuori , and K. Bös. Development of a health-related fitness test battery for adults: aspects of reliability. *Arch. Phys. Med. Rehabil.* 77:399-405, 1996.

Suni, J. H., S. I. Miilunpalo, T.-M. Asikainen, et al. Safety and feasibility of a health-related fitness test battery for adults. *Phys. Ther.* 78:134-148, 1998.

Suni, J. H., P. Oja P, S. I. Miilunpalo, M. E. Pasanen, I. M. Vuori, and K. Bös. Health-related fitness test battery for adults: associations with perceived health, mobility, and back function and symptoms: *Arch. Phys. Med. Rehabil.* 79:559-569, 1998.

Suni, J. H., P. Oja, S. I. Miilunpalo, M. E. Pasanen, Vuori, I.M. Health-related fitness test battery for middle-aged adults: associations with physical activity patterns. *Int. J. Sports. Med.* 20:183-191, 1998.

Thomas, E., A. J. Silman, A. C. Papageorgiou, G. J. Macfarlane, and P. R. Croft, P.R. Association between measures of spinal mobility and low back pain. *Spine* 23:343-347, 1998.

Tulder, M. W., B. W. Koes, and L. M. Bouter. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. *Spine* 22:2128-2156, 1997.

U.S. Department of Health and Human Services. Physical activity and health: A Report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centres for Disease Control and Prevention, National Centre for Chronic Disease Prevention and Health Promotion, 1996, pp. 28-29, 33-37, 43-45, 72, 133-134.

van Heuvelen, M.J.G., G. I. J. M Kempen, J. Ormel, and P. Rispens. Physical fitness related to age and physical activity in older persons. *Med. Sci. Sports Exerc.* 30:434-441, 1998.

Videman, T., H. Rauhala, H., S. Asp, S., et al. 1989. Patient-handling skill, back injuries and back pain: an intervention study in nursing. *Spine* 14:148-56, 1989.

Videman, T., S. Sarna, M. C. Battie, S. Koskinen, K. Gill, H. Paananen, L. Gibbons. The long term effects of physical loading and exercise lifestyle on back-related symptoms, disability, and spinal pathology among men. *Spine* 20:699-709,1995.

Wilson, I.B., and P. D. Cleary. Linking the clinical variables with health-related quality of life: a conceptual model of patient outcomes. *JAMA* 273:59-65, 1995.

V

Health-related fitness test battery for middle-aged adults: associations with physical activity patterns

by

Jaana Suni, Pekka Oja, Seppo Miilunpalo, Matti Pasanen, Ilkka Vuori, Klaus Bös

International Journal of Sports Medicine 1999; 20:183-191

https://doi.org/10.1055/s-1999-970287