This article was downloaded by: *[European College of Sport Science]*
On: *8 October 2009*
Access details: *Access Details: [subscription number 909201977]*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

European Journal of Sport Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t714592354

Position statement: Testing physical condition in a population - how good are the methods?

Torben Jørgensen [a]; Lars B. Andersen [bc]; Karsten Froberg [c]; Urs Maeder [d]; Lisa von Huth Smith [a]; Mette Aadahl [a]

[a] Research Centre for Prevention and Health, Glostrup Hospital, Glostrup, Denmark [b] Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway [c] Institute of Sport Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark [d] Swiss Federal Institute of Sport, Magglingen, Switzerland

Online Publication Date: 01 September 2009

To cite this Article Jørgensen, Torben, Andersen, Lars B., Froberg, Karsten, Maeder, Urs, von Huth Smith, Lisa and Aadahl, Mette(2009)'Position statement: Testing physical condition in a population - how good are the methods?',European Journal of Sport Science,9:5,257 — 267

To link to this Article: DOI: 10.1080/17461390902862664

URL: http://dx.doi.org/10.1080/17461390902862664

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

European Journal of Sport Science, September 2009; 9(5): 257–267

POSITION STATEMENT

Position statement: Testing physical condition in a population – how good are the methods?

TORBEN JØRGENSEN[1], LARS B. ANDERSEN[2,3], KARSTEN FROBERG[3], URS MAEDER[4], LISA VON HUTH SMITH[1], & METTE AADAHL[1]

[1]*Research Centre for Prevention and Health, Glostrup Hospital, Glostrup, Denmark,* [2]*Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway,* [3]*Institute of Sport Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark, and* [4]*Swiss Federal Institute of Sport, Magglingen, Switzerland*

Abstract

A poor physical condition – expressed as physical inactivity and poor physical fitness – is associated with the development of chronic diseases and premature death. Our aim was to evaluate the methods currently available for measuring physical activity and physical fitness in the general population.

Physical activity is determined by duration, frequency, and intensity and derives from many different domains, making it difficult to assess over long periods and no feasible general criterion measure exists. Both objective and subjective methods are available. Of the objective methods, accelerometry is the most attractive technology, and is well enough developed for general use in large populations. The advantage of accelerometry is that it is not dependent on the memory of the individual, but its main disadvantage is that it grossly underestimates energy expenditure, due to the lack of registration of certain activities. This may be overcome to a certain extent by combining accelerometry with heart rate monitoring, although this still does not measure activity in different domains. Of the subjective methods, self-report questionnaires are inexpensive and easy to administer. Many questionnaires have been developed, but we require (1) consensus on which measures to use for validation and (2) further development of internationally standardized questionnaires for use in different settings and to address different scientific questions. Many questionnaires correlate well with biological markers and development of chronic diseases, but subjective measurement will always entail a certain degree of misclassification. Furthermore, unstructured physical activity such as housework and gardening may be subject to recall bias. No method appears better to any other, and the choice of instrument will depend on the research question being asked. Future research should combine information from both objective and subjective methods.

Physical fitness comprises several components, including cardiorespiratory endurance and muscle strength and endurance. Direct measurement of oxygen consumption is the criterion measure for cardiorespiratory endurance. As regards muscle strength and endurance, only test–retest reliability is available. Hand-held dynamometers greatly facilitate field testing for maximal isometric muscle strength assessment, while force plate measurements can be used for the lower extremities. For endurance, simple tests such as push-ups and sit-ups are reliable.

Keywords: *Physical activity, physical fitness, epidemiology*

Introduction

A sedentary lifestyle has become prevalent in modern society and physical inactivity (Berlin & Colditz, 1990; Blair, Cheng, & Holder, 2001) together with poor physical fitness (Blair *et al.*, 1996, 2001) are associated with the development of chronic diseases and premature death. Our knowledge of the specific dose–response relationship between physical activity and different health outcomes is still limited, especially in children. From a public health perspective, we need to be able to facilitate the surveillance and assess the effect of preventive understandings. Therefore, we require sensitive, valid, and reliable instruments for measuring the physical condition of large populations.

Physical condition comprises physical activity and physical fitness. Physical fitness and physical activity should be regarded as different but complementary

Correspondence: T. Jørgensen, Research Centre for Prevention and Health, Glostrup Hospital, Ndr. Ringvej 57, Building 84/85, DK-2600 Glostrup, Denmark. E-mail: tojo@glo.regionh.dk

ISSN 1746-1391 print/ISSN 1536-7290 online
DOI: 10.1080/17461390902862664

aspects of physical condition. Although physical activity is an important determinant of physical fitness (Bouchard, Blair, & Haskell, 2007; De Backer *et al.*, 2003), genetics plays an even more important role (Bouchard *et al.*, 1999). However, genetics may be more important for the highest level of fitness an individual can achieve, and less important in determining the fitness of a sedentary individual (i.e. overweight), and bed rest may decrease fitness substantially regardless of genotype. This distinction is important because health problems may increase exponentially at very low levels of fitness. The ability to improve individual physical fitness through physical activity appears to be genetically determined. When physical activity and fitness are included as explanatory variables in the same statistical model, only the latter predicts cardiovascular mortality (Blair *et al.*, 2001). This could, however, be due to the objective measurement of fitness, as opposed to the subjective self-report measurement of physical activity by questionnaire.

Our aims here are to evaluate current methods for measuring the physical condition of a population and to recommend further steps to improve such measurements to increase our understanding of the implications of physical condition at a population level.

Physical activity

Physical activity is defined as any bodily movement produced by contraction of skeletal muscle that substantially increases energy expenditure (Caspersen, Powell, & Christenson, 1985). In the general population, 60–70% of total energy expenditure is derived from resting metabolic rate, 10% from the diet, and the remaining 20–30% from physical activity (Bouchard *et al.*, 2007). Physical activity is the only part of total energy expenditure that can vary substantially between individuals – from less than 10% among sedentary individuals to more than 80% among extremely active individuals (McArdle , Katch, & Katch, 1996).

Physical activity is determined by duration, frequency, intensity, and type (Howley, 2001), and is derived from different domains such as work, transportation, household and gardening, sport and exercise, and general leisure time. The most commonly assessed domains of physical activity are work and leisure-time (Pereira *et al.*, 1997) and commuting activities (Barengo, Kastarinen, Lakka, Nissinen, & Tuomilehto, 2006; Hu *et al.*, 2003, 2004, 2007).

Type refers to the mode of contraction (static vs. dynamic) and whether the activity comprises small or large muscle groups. *Duration* refers to the amount of time an activity is performed, while *frequency* describes the number of sessions or bouts of activity undertaken per day, week or month. *Intensity* of physical activity is expressed as energy expenditure per unit of time (e.g. $\text{kJ} \cdot \text{min}^{-1}$) and most studies assess the absolute intensity against, for example, development of chronic diseases. *The relative intensity* – that is, the percentage of an individual's maximal oxygen uptake that is needed to perform a specific task (Bouchard *et al.*, 1994; Howley, 2001) – is rarely assessed, but it has been shown that the relative intensity may be a stronger predictor of chronic disease and premature death than absolute intensity (Lee, Sesso, Oguma, & Paffenbarger, 2003). Accordingly, the American College of Sports Medicine (1990) recommends that intensity of physical activity should be relative to one's maximum oxygen uptake. A common means of quantifying the intensity of physical activity is to use metabolic equivalents (MET) (Ainsworth *et al.*, 1993, 2000). One MET corresponds to an energy expenditure of one kilocalorie per kilogram of body mass (BM) per hour or 3.5 ml O_2 kg BM^{-1} min^{-1}. Any specific physical activity can be assigned a MET value, expressing what number of multiples of the resting energy expenditure it requires. Sleep has a MET value of 0.9, whereas cycling at moderate speed is assigned 4.0 MET. The MET values for a large number of specific physical activities are listed in a compendium of physical activities (Ainsworth *et al.*, 1993, 2000).

Instruments for measuring physical activity should be valid and reliable, but also able to be used in large populations. Any proposed instrument should preferably be tested against a "gold standard" (criterion validity) or, less optimally, assessed for agreement with other instruments that are supposed to be associated with physical activity (e.g. fitness test, metabolic parameters) (concurrent validity). Commonly cited gold standards include direct behavioural observation, direct or indirect calorimetry, and the doubly labelled water method. Direct behavioural observation can theoretically validate all domains and dimensions by following and observing free-living individuals. Direct calorimetry quantifies total energy expenditure by measuring heat production or heat loss, and is performed in a closed chamber. The principle in the doubly labelled water method (Ekelund *et al.*, 2001) is that a certain amount of water with an enrichment of ^{2}H and ^{18}O atoms is ingested and energy expenditure is calculated by estimating carbon dioxide production using isotope dilution. The doubly labelled water method can only measure total energy expenditure, but combined with indirect calorimetry, an estimate of physical activity energy expenditure can be made. The method does not provide information on the domain or type of the physical activity performed. These methods are rarely used in large-scale epidemiological studies, as they are expensive and cumbersome. They are primarily useful for validation of simple and more feasible instruments, but will not solve the problem of validating physical activity over longer periods.

With regard to the statistical methods used to compare instruments, the Bland-Altman method is recommended over correlation coefficients for validation purposes (Schmidt & Steindorf, 2006). Although commonly used, correlation coefficients do not provide information on systematic over- or under-estimation and can therefore yield misleading conclusions in validation studies (Schmidt & Steindorf, 2006).

Objective measures of physical activity

The technology associated with objective measures of physical activity among larger populations has developed quickly during the last couple of decades. In the following, we look at pedometers, heart rate monitors, accelerometers, and various combinations of the three.

Pedometers

Pedometers are easy to mount and wear and they are an inexpensive means of counting steps (Saris, 1985). There are a large number of pedometers on the market, of which the NL-2000 (New-Lifestlyes, Inc., Lee's Summit, MO, USA) seems to be the most reliable (Crouter, Schneider, & Bassett, 2005) for counting steps under different conditions. Pedometers grossly underestimate physical activity expressed as energy expenditure (Crouter, Schneider, Karabulut, & Bassett, 2003), even in studies where cycling is not a part of the physical activity, but they are suitable for monitoring campaigns (e.g. 10,000 steps a day; Crouter *et al.*, 2003) and other intervention-based changes over time.

Heart rate monitoring

Heart rate monitoring is based on the assumption of a linear relationship between heart rate and oxygen consumption in moderate to vigorous activities. At rest and during low-intensity activities, the relationship is not linear and is confounded by mood, temperature, and diet. Heart rate depends on the physical fitness of the individual and a valid estimation depends on individual calibration based on knowledge of maximal and resting heart rate, and even better against direct measurement of oxygen uptake. Some studies have used cut points $>50\%$ of the heart rate reserve to estimate time spent in moderate- and high-intensity exercise (Fairclough & Stratton, 2005). Estimation of total energy expenditure from heart rate was validated against the doubly labelled water method (Racette, Schoeller, & Kushner, 1995; Rafamantanantsoa *et al.*, 2002) and showed acceptable correlations. Modern heart rate monitors are easy to use and can store data for long periods of time. Their use is best in smaller studies due to the need for individual calibration. If only the relative load is of interest, estimates can be made using resting and maximal heart rate based on age and sex.

Accelerometers

Accelerometers measure movements in one, two or three planes (Plasqui, Joosen, Kester, Goris, & Westerterp, 2005) by piezoelectric transducers and microprocessors. The devices are small and easy to carry, and the units of measurement (counts per minute) quantify the magnitude and the direction of accelerations. The new models have a memory, where data can be stored for each minute or each 10 s, thus allowing analysis of short bursts of activity. Furthermore, data can be stored for up to 200 days. The method has developed rapidly during the last years and several accelerometers have been tested under laboratory conditions during standardized activities with a strong correlation being reported with energy expenditure (Plasqui *et al.*, 2005). For epidemiological purposes, it is relevant to evaluate the ability of different accelerometers to accurately assess physical activity under free-living conditions. Validation against the doubly labelled water method has shown in general that accelerometers underestimate total energy expenditure. The Actigraph uniaxial accelerometer (formerly MTI and CSA) and the Tracmor triaxial accelerometer show a reasonable correlation with energy expenditure calculated using the doubly labelled water method (Plasqui & Westerterp, 2007). Only the Actigraph is commercially available. The Actiwatch, Caltrac, Tritac, and Lifecorder devices did not show substantial correlation with the doubly labelled water method (Plasqui & Westerterp, 2007).

There are other problems associated with the use of accelerometers. The output is frequency dependent because of an electronic filter, which is used to filter noise. This has some importance in children where the step frequency of the movement depends on the size of the child, which creates some difficulties when the activity of different age groups is compared (Brage, Wedderkopp, Andersen, & Froberg, 2003). Another drawback is that accelerometer output levels off when speed increases to more than 10 km h^{-1} (Brage *et al.*, 2003). However, the time spent running above 10 km h^{-1} is limited during habitual physical activities, and this problem is not that great in epidemiological studies. Furthermore, accelerometers cannot register physical activity with no acceleration, such as rowing, cycling, skating, and hill climbing. Nor can accelerometers register isometric muscle contraction during muscular work against an external force such as weight lifting, carrying, and pushing. Cycling is quantitatively a problem in some countries (Holland and Denmark). The challenge of translating counts per minute into energy expenditure has not been solved yet, because different types of activity with the same energy demand reveal different output from the monitor. However, several studies have validated walking and running, and there is good agreement that a walking speed of 4 km h^{-1} on a treadmill or overground corresponds to about 2000 counts per minute (Brage *et al.*, 2003; Ekelund, Aman, & Westerterp, 2003; Puyau, Adolph, Vohra, & Butte, 2002; Trost *et al.*, 1998).

Other instruments

Recently, a device that combines heart rate monitoring and accelerometry has been developed. This device (Actiheart®) combines the best features of heart rate monitoring and accelerometers by using heart rate in the high-intensity range, where heart rate best reflects the workload, and the counts from the accelerometer in the low-intensity range (Brage *et al.*, 2004). In combination with accelerometers, individual calibration in relation to heart rate may not be necessary. However, validation under free-living conditions is required.

The Actireg unit can register acceleration and change in body position (bending down or changing from lying to standing). Wires from the accelerometer are attached to the arm and leg. It is a reliable instrument to assess energy expenditure, but mounting of the instrument should be done by the same experienced person, which makes it unsuitable for large-scale studies (Arvidsson, Slinde, Nordenson, Larsson, & Hulthen, 2006; Hustvedt *et al.*, 2004).

Finally, devices such as portable armbands, which combine two-axis accelerometers with skin temperature, are potentially suitable for calculation of energy expenditure (St-Onge *et al.*, 2007).

Self-report measurement of physical activity

Self-report tools for measuring physical activity include physical activity records, logs, and questionnaires (Ainsworth, Montoye, & Leon, 1994; LaMonte & Ainsworth, 2001; Sallis & Saelens, 2000). Physical activity records and logs are self-administered, whereas questionnaires may be interviewer- or self-administered.

Physical activity records and logs

Physical activity records are diaries kept by study participants (Ainsworth *et al.*, 1994; LaMonte & Ainsworth, 2001). They provide a detailed account of all or selected types of physical activity performed within a given time. They are demanding for respondents to administer and time-consuming for researchers to quantify and process. Physical activity logs are similar to records, except that they are structured as checklists of specified activities usually developed from population-specific physical activity focus groups (Ainsworth *et al.*, 1994; LaMonte & Ainsworth, 2001). An evident drawback of the activity log occurs if the relevant activities are not included in the log. Physical activity records show a reasonable correlation with the doubly labelled water method (Rafamantanantsoa *et al.*, 2002).

Self-report questionnaires

Questionnaires have for many years been the method of choice in epidemiological studies exploring the relationship between physical activity and different health outcomes (LaMonte & Ainsworth, 2001; Sallis & Saelens, 2000). Physical activity questionnaires include simple single-item global questionnaires that assess general levels of physical activity to classify individuals as active or inactive (Saltin & Grimby, 1968), recall questionnaires aimed at a fairly specific assessment, and more extensive questionnaires that assess the frequency, duration, and intensity of specific activities during a specified period (from days to life-time) in different domains (Pereira *et al.*, 1997). Many of the questionnaires that exist were gathered together and published in 1997 together with information on validation and reliability testing (Pereira *et al.*, 1997). Some of the most frequently used questionnaires in large adult study populations are the Minnesota Leisure-time Physical Activity Questionnaire (Taylor *et al.*, 1978), the Paffenbarger Physical Activity Questionnaire (Paffenbarger, Wing, & Hyde, 1995), and the Aerobics Center Longitudinal Study Physical Activity Questionnaire (Kohl, Blair, Paffenbarger, Macera, & Kronenfeld, 1988).

Other questionnaires, also developed for adult populations, have been developed since 1997. These include the European Prospective Investigation into Cancer and Nutrition Study (EPIC) questionnaire, which assesses previous year activity at home, at work, and during recreation (Khaw *et al.*, 2006; Wareham *et al.*, 2002, 2003); the "Vital" questionnaire, which measures usual recreational physical activity during the preceding 10 years (Littmann *et al.*, 2004); the "SQUASH" questionnaire, which assesses physical activity during an average week in the past months performed in different domains (Wendel-Vos, Schuit, Saris, & Kromhout, 2003); the "Star" questionnaire, which is used in telephone interviews to assess the overall moderate and vigorous activity performed in a usual week (Matthews *et al.*, 2005); a screening instrument for family doctors that identifies inactive patients in a primary care setting (Marshall, Smith, Bauman, & Kaur, 2005); the Brunel Lifestyle Physical Activity questionnaire, which is an Internet-based questionnaire intended for use in conjunction with a 12-week personalized fitness programme delivered through the Internet (Karageorghis, Vencato, Chatzisarantis, & Carron, 2005); and questionnaires intended to quantify and estimate energy expenditure 24 h a day within all domains of physical activity (Aadahl & Jørgensen, 2003; Aadahl, Kjaer, & Jørgensen, 2007; Trolle-Lagerros *et al.*, 2005).

An attempt to reach consensus on questionnaires on physical activity is the International Physical Activity Questionnaire (IPAQ). The IPAQ has been translated into different languages (www.ipag.ki.se) and consists of four long and four short versions using two different reference periods ("usual week" or "last 7 days") (Craig *et al.*, 2003). In the short version, time spent in moderate and vigorous activities and walking is estimated, but not reported separately for the different domains. In the long version, time spent sitting and time spent in occupational, transport, household, and leisure-time physical activity is estimated independently and the intensity is assessed in each domain. The IPAQ questionnaire shows a low (Rutten *et al.*, 2003) to good (Craig *et al.*, 2003) repeatability in test–retest analysis and a low correlation with other national physical activity questionnaires (Rutten *et al.*, 2003). Generally, the IPAQ instrument leads to higher estimates of total physical activity than other questionnaires (Rutten *et al.*, 2003). Validation against accelerometers shows relatively low correlation coefficients (Craig *et al.*, 2003; Ekelund *et al.*, 2006).

A general drawback is the lack of consensus on how to validate questionnaires with no obvious criterion measure. The many different questionnaires indicate that no one questionnaire is superior. A set of minimum requirements is needed. These include:

- some kind of qualitative testing [e.g. cognitive interviewing (Beatty & Willis, 2007; Conrad & Blair, 2004)] to ensure the respondent has the same conception of the questionnaire as the researcher;
- validation against physical activity records or detailed interview ranging from 24-h to one-month recall, which is routinely used in nutritional research (Willett, 1998);
- a test against the doubly labelled water method or direct observation; and
- comparison with biomarkers.

Strengths and limitations of self-report questionnaires and objective measures

From an epidemiological point of view, ideally instruments should measure all dimensions of physical activity in specific domains. In addition, information on individual physical capacity should be obtained, either as "perceived exertion" (Borg, 1998) or as maximum oxygen consumption ($\dot{V}O_{2max}$). As this is seldom realistic in large study populations, the choice of instrument often depends on the health outcome of interest. Total amount of physical activity may be the relevant exposure in relation to some health outcomes, whereas information on a specific domain (e.g. leisure time) or dimension (e.g. intensity) of activities may be of interest in others.

In spite of the large number of methods for measuring physical activity, no "perfect" method has emerged to date. No one instrument can measure all dimensions of physical activity in all domains, over a long period of time at low cost and in large study populations. In general, self-report questionnaires for measuring physical activity are easy to administer. Many appear to correlate with biological markers (von Huth Smith, Borch-Johnsen, & Jørgensen, 2007) and to predict development of chronic diseases and premature deaths. However, the self-report nature of questionnaires means there will always be a degree of misclassification (Sallis & Saelens, 2000; Shephard, 2003). Remembering the duration, frequency, intensity, and type of physical activity performed in the past can be difficult for respondents, especially if the recall time-frame is extensive (e.g. a year or a life time). This is a particular problem among children due to cognitive limitations (Baranowski *et al.*, 1984; Kohl, Fulton, & Caspersen, 2000; Sallis, 1991; Sallis, Buono, & Freedson, 1991; Saris, 1985), and in unstructured physical activities such as work, sports, and exercise (Levin, Jacobs, Ainsworth, Richardson, & Leon, 1999). Social desirability bias, whereby respondents distort self-report in a favourable direction, may also reduce the validity of self-reported physical activity measures (Motl, McAuley, & DiStefano, 2005).

Objective measurement of physical activity has the potential to produce better estimates of the true association between physical activity and health risk factors (Wareham & Rennie, 1998; Wong, Day, & Wareham, 1999) than self-report. Accelerometry seems to be the most attractive technology, and it is sufficiently well developed for general use in large populations. Among the disadvantages is the fact that accelerometry appears to grossly underestimate energy expenditure, due to the lack of registration of certain activities. This may be overcome to an extent by combining accelerometry with heart rate monitoring – an emerging technology. Accelerometry is subject to the risk of reactivity (van Sluijs, van Poppel, Twisk, & van Mechelen, 2006) – that is, the fact that wearing the accelerometer may cause changes in physical activity patterns. This also applies to other methods, including heart rate monitoring and direct observation. Research is required of the dose–response relationship between counts per minute in accelerometry and various physiological (e.g. blood pressure) and biochemical (e.g. cholesterol) measures. Finally, studies are required that compare counts per minute with hard end-points such as development of chronic diseases.

At present, there is no one superior method that should be recommended above all other methods for measuring physical activity in large study populations. In future research, both objective and self-report measures should be used simultaneously to assess various aspects of measuring physical activity. Using accelerometry as criterion validity for questionnaires may not be suitable.

Physical fitness

Physical fitness comprises several components, of which cardiorespiratory endurance is the most important, because of its strong relationship with the development of chronic diseases, including cardiovascular diseases, diabetes, cancer, and premature death (Blair *et al.*, 1989; Myers *et al.*, 2002). Two further components, muscle strength and endurance, show an inconsistent relationship with musculoskeletal disorders (Hamberg-van Reenen, Ariëns, Blatter, van Mechelen, & Bongers, 2007). Further components of physical fitness are musculoskeletal flexibility and body composition. Here, we restrict ourselves to simple methods for assessing cardiorespiratory endurance, muscle strength, and muscle endurance in large populations.

Cardiorespiratory fitness

The most reliable and valid measure of aerobic capacity is the direct measurement of maximal oxygen consumption ($\dot{V}O_{2max}$) (Safrit, Hooper, Ehlert, Costa, & Patterson, 1988), although it is not immune against inaccuracy (Shephard, 1984). The method is not suitable in larger population groups, since expensive and sophisticated equipment is required. A variety of less complex procedures have been developed to estimate $\dot{V}O_{2max}$, and their validity has been determined by comparing the estimates with the criterion measure, the direct measurement of $\dot{V}O_2$. Both maximal and submaximal exercise tests have been developed, of which the maximal tests provide the most accurate estimations. However, the decision to select a maximal or submaximal exercise test to estimate $\dot{V}O_{2max}$ depends on the participants studied and the availability of appropriate equipment. During the tests, some degree of risk management is required (American College of Sports Medicine, 2006).

Maximal exercise tests. The Cooper test (maximal 12-min run test; 12-MRT) is strongly related to the criterion measure of $\dot{V}O_{2max}$ in adults ($r = 0.84$–0.92) (Cooper, 1968; Grant, Corbett, Amjad, Wilson, & Aitchison, 1995; McCutcheon, Sticha, Giese, & Nagle, 1990) and in children ($r = 0.9$) (Jackson & Coleman, 1976). Although the estimation equation used yielded a systematic underestimation of $\dot{V}O_{2max}$ by 4 $\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ (McCutcheon *et al.*, 1990) in one study, no statistically significant difference was observed in two further studies (Cooper, 1968; Grant *et al.*, 1995; McCutcheon *et al.*, 1990). However, inexperienced runners have difficulty finding the optimal speed, and are therefore underestimated. The multi-stage 20-m shuttle run test (MST) was shown to be an accurate method to estimate $\dot{V}O_{2max}$ in adults in one study ($r = 0.90$) (Léger & Gadoury, 1989), whereas others failed to provide such strong correlations ($r = 0.79$–0.86) and in fact reported a statistically significant underestimation of $\dot{V}O_{2max}$ (3.0–7.5%) (Cooper, Baker, Tong, Roberts, & Hanford, 2005; Grant *et al.*, 1995; McNaughton, Hall, & Cooley, 1998; Ramsbottom, Brewer, & Williams, 1988). In children and adolescents, the correlations between MST estimates and criterion measures of $\dot{V}O_{2max}$ ranged between 0.71 and 0.87 (American College of Sports Medicine, 1990; Boreham, Paliczka, & Nichols, 1990; Léger, Mercier, Gadoury, & Lambert, 1988; Liu, Plowman, & Looney, 1992). A comparison between the Cooper test and multi-stage 20-m shuttle run test showed that the former is a better predictor of $\dot{V}O_{2max}$ (Grant *et al.*, 1995) in experienced runners. Both running tests are most appropriate for individuals with sufficient fitness and motor skills and who are considerably motivated. This problem has partly been solved in a recent intermittent running test (Andersen, Andersen, Andersen, & Anderssen, 2008). This test does not require any equipment for the test leader or experience on the part of the participant. Besides these running tests, a simple maximal exercise test on a cycle ergometer was developed for children (Hansen, Froberg, Nielsen, & Hyldebrandt, 1989) and for adults (Andersen, 1995; Andersen, Henckel, & Saltin, 1987). The correlation between the estimation of $\dot{V}O_{2\max}$ based on maximal workload and the criterion measure was very strong ($r = 0.90$ and 0.95 for boys and girls, respectively). The use of a calibrated cycle ergometer might be a limitation when using this test in larger population groups. Furthermore, the participants should be familiar with cycling to achieve maximal performance of the ergometer.

Submaximal exercise test. Estimates of $\dot{V}O_{2\max}$ derived from submaximal tests are often based on the linear relationship of workload, heart rate, and $\dot{V}O_2$. They are difficult to compare due to different methods, activities, and study samples used to verify the validity of these assessments. A classic method is the cycle ergometer test of Åstrand & Rhyming (1954). For the simple 2-km walk test, estimates of $\dot{V}O_{2\max}$ were compared with the criterion measure and correlation coefficients of 0.55, 0.79, and 0.60 were obtained for moderately active middle-aged women, moderately active middle-aged men, and highly active men, respectively (Laukkanen, Oja, Pasanen, & Vuori, 1993). In a recent study, 6-min walk test estimates of $\dot{V}O_{2\max}$ for female seniors were related to the criterion measure ($r = 0.44$; Rance *et al.*, 2005). A better correlation between the walking test distance and the criterion measure was found for the 6-min walk test in children ($r = 0.94$) (Li *et al.*, 2005). Step tests have been used for many years, but no validation study has been published (Howley, Colacino, & Swensen, 1992).

In large epidemiological studies, cardiorespiratory fitness may be estimated in adults from a non-exercise test model that includes gender, age, body mass index, resting heart rate, and self-reported physical activity. These estimates of fitness were strongly related to the criterion measure in large groups ($r = 0.76$–0.81) (Jurca *et al.*, 2005). However, most of the variances in these tests are explained by variables that cannot be changed such as age and sex or variables such as body weight that do not change over a short period with increased physical activity. Recent literature suggests that self-rated physical fitness assessed by a simple question correlates well with the criterion measure, maximal oxygen uptake (Aadahl, Kjaer, Kristensen, Mollerup, & Jørgensen, 2007).

Muscle strength and endurance

Different field test batteries include different tests for muscle strength and muscle endurance. Both strength and endurance tests can be static or dynamic, and most batteries include a mixture of tests to assess strength and endurance– trunk, legs or arms; static or dynamic – to describe participants' physical abilities in general. A major problem is that there is no standardization. For example, there are many versions of sit-and-reach, and many versions of the Sargent jump, with each research group making its own modifications. This makes comparison between populations and description of secular trends in muscle strength or endurance very difficult. In the 1980s, an attempt was made to standardize tests and test populations across Europe to compare health-related physical fitness between populations. This test battery was called the Eurofit Test Battery (Eurofit, 1984). It included tests of functional strength, muscle endurance, balance, agility, flexibility, and coordination. The original test battery was adapted to children, but a test battery was later constructed for adults. In the 1990s, the European Union supported an international group working with "Health Enhancing Physical Activity" (HEPA), and this group has now been revived and is supported by WHO-Europe and the European Commission. This work led to the construction of a test battery to assess health-related fitness (Suni *et al.*, 1996, 1998a, 1998b).

Examples of content of test batteries

The explosive power of the legs as measured by force plates may be accurately predicted by the conventional jump-and-reach test, if the result is corrected for body weight ($r = 0.83$) (Shetty, 2002). Otherwise, muscle strength and endurance tests are often evaluated by determining test–retest reliability, as there is no "gold standard". Maximal muscle strength is commonly determined with the one-repetition maximum strength procedure, where the resistance is progressively increased until the participant can no longer perform the exercise. As the procedure requires stationary equipment, its use might be limited among larger groups, but if the participants are familiarized with the procedure it is highly reliable (Philips, Batterham, Valenzuela, & Burkett, 2004). Portable hand-held dynamometers greatly facilitate field testing for maximum isometric muscle strength assessment. Hand grip measurements are easily accomplished, and the devices are able to determine the muscle strength of several muscle groups with high reliability ($r = 0.73$–0.91) (van den Beld, van den Sanden, Sengers, Verbeek, & Gabreels, 2006).

Measurements of muscular endurance are often made with simple tests, such as push-ups to determine upper-body muscle endurance and sit-ups to measure abdominal muscle groups. If the results in a standardized push-up test are corrected for weight, it is a valid tool to determine muscular endurance of the upper body ($r = 0.70$–0.73) (Pate, Burgess, Woods, Ross, & Baumgartner, 1993). Sit-ups may involve varying accessory muscles besides the abdominal muscles, such as the hip flexors. Therefore, the curl-up test that consists of a small head and upper-body lift was developed to minimize the use of the hip flexors. The curl-up test has shown acceptable reliability ($r = 0.92$) (Sparling, Millard-Stafford, & Snow, 1997), whereas the reliability of dynamic or isometric sit-up tests is limited ($r < 0.50$). Isometric muscle endurance of the back extensor muscles can be assessed by the Biering-Sorensen test. This type of fitness has been shown to be related to low back pain (Andersen, Wedderkopp, & Lebouef-Yde, 2006; Biering-Sørensen, 1984).

Summary

Different test protocols have been developed and evaluated for the measurement of cardiorespiratory fitness. While maximal exercise tests do provide the most accurate results, the validity of submaximal procedures is still acceptable in populations where maximal testing is perceived unsafe. We would recommend the use of either a cycle ergometer test or one of the run tests (e.g. Andersen test). Some examples, such as the jump-and-reach test, show that field testing for muscle strength can be undertaken. But for many muscle tests only test–retest data are available. This reliability can be good, but standardization remains an important issue.

General conclusions

To determine the physical condition of large populations, we recommend that several components be measured: physical activity, cardiorespiratory fitness, muscle strength, and muscle endurance. They represent fundamentally different aspects of physical health, but to study the development and prevention of chronic disease, we need to measure all aspects in detail. Although physical activity and physical fitness are strongly associated, genetics play an important role.

Self-report measurements of physical activity are subject to bias and misclassification, but we can only gain information on domains of physical activity by using questionnaires or interviews. Furthermore, only self-report is pertinent when the physical condition over a long period (years) is needed, as is the case in evaluating the effect on development of chronic disease. There is, therefore, a need for questionnaires that have been tested and well validated and allow for international comparisons among different study populations. Objective measurements provide accurate, precise, and valid estimates, but tend to be less useful in large study populations and to underestimate total energy expenditure. Hence, there is still a need for objective methods to be developed that are more reliable and easy to administer in large study populations. The combined heart rate and accelerometer method appears promising.

References

Aadahl, M., & Jørgensen, T. (2003). Validation of a new self-report instrument for measuring physical activity. *Medicine and Science in Sports and Exercise*, *35*, 1196–1202.

Aadahl, M., Kjaer, M., & Jorgensen, T. (2007). Perceived exertion of physical activity: Negative association with self-rated fitness. *Scandinavian Journal of Public Health*, *35*, 403–409.

Aadahl, M., Kjær, M., Kristensen, J. H., Mollerup, B., & Jørgensen, T. (2007). Self-reported physical activity compared with maximal oxygen uptake in adults. *European Journal of Cardiovascular Prevention and Rehabilitation*, *14*, 422–428.

Ainsworth, B. E., Haskell, W. L., Leon, A. S., Jacobs, D. R., Montoye, H. J., Sallis, J. F., et al. (1993). Compendium of physical activities: Classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise*, *25*, 71–80.

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., et al. (2000). Compendium of physical activities: An update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, *32*, S504.

Ainsworth, B. E., Montoye, H. J., & Leon, A. S. (1994). Methods of assessing physical activity during leisure and work. In C. Bouchard, R. J. Shephard, & T. Stephens (Eds.), *Physical activity, fitness, and health: International Proceedings and Consensus Statement* (pp. 146–159). Champaign, IL: Human Kinetics Publishers.

American College of Sports Medicine (1990). American College of Sports Medicine position stand. *Medicine and Science in Sports and Exercise*, *22*, 265–274.

American College of Sports Medicine (2006). Clinical exercise testing. In *ACSM's guidelines for exercise testing and prescription* (pp. 91–114). Philadelphia, PA: Lippincott Williams & Wilkins.

Andersen, L. B. (1995). A maximal cycle exercise protocol to predict maximal oxygen uptake. *Scandinavian Journal of Medicine and Science in Sports*, *5*, 143–146.

Andersen, L. B., Andersen, T.-E., Andersen, E., & Anderssen, S. A. (2008). An intermittent running test to estimate maximal oxygen uptake: The Andersen test. *Journal of Sports Medicine and Physical Fitness*, *48*, 434–437.

Andersen, L. B., Henckel, P., & Saltin, B. (1987). Maximal oxygen uptake in Danish adolescents 16–19 years of age. *European Journal of Applied Physiology*, *56*, 74–82.

Andersen, L. B., Wedderkopp, N., & Leboeuf-Yde, C. (2006). Association between back pain and physical fitness in adolescents. *Spine*, *31*, 1740–1744.

Arvidsson, D., Slinde, F., Nordenson, A., Larsson, S., & Hulthen, L. (2006). Validity of the ActiReg system in assessing energy requirement in chronic obstructive pulmonary disease patients. *Clinical Nutrition*, *25*, 68–74.

Åstrand, P.-O., & Ryhming, I. (1954). A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *Journal of Applied Physiology*, *7*, 218–221.

Baranowski, T., Dworkin, R. J., Cieslik, C. J., Hooks, P., Clearman, D. R., & Ray, L. (1984). Reliability and validity of self report of aerobic activity: Family Health Project. *Research Quarterly for Exercise and Sport*, *55*, 309–317.

Barengo, N. C., Kastarinen, M., Lakka, T., Nissinen, A., & Tuomilehto, J. (2006). Different forms of physical activity and cardiovascular risk factors among 24–64-year-old men and women. *European Journal of Cardiovascular Prevention and Rehabilitation*, *13*, 51–59.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviews. *Public Opinion Quarterly*, *71*, 287–311.

Berlin, J. A., & Colditz, G. A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *American Journal of Epidemiology*, *132*, 612–628.

Biering-Sørensen, F. (1984). A one-year prospective study of low back trouble in a general population: The prognostic value of low back history and physical measurements. *Danish Medical Bulletin*, *31*, 362–375.

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Medicine and Science in Sports and Exercise*, *33*, S379–S399.

Blair, S. N., Kampert, J. B., Kohl, H. W., III, Barlow, C. E., Macera, C. A., Paffenbarger, R. S., et al. (1996). Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *Journal of the American Medical Association*, *276*, 205–210.

Blair, S. N., Kohl, W., Paffenbarger, R. S., Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Journal of the American Medical Association*, *262*, 2395–2401.

Boreham, C. A. G., Paliczka, V. J., & Nichols, A. K. (1990). A comparison of the PWC170 and 20-MST test of aerobic fitness in adolescent schoolchildren. *Journal of Sports Medicine and Physical Fitness*, *30*, 19–23.

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Stockholm: Human Kinetics.

Bouchard, C., Shepard, R. J., & Stephens, T. (1994). Methods of assessing physical activity during leisure and work. In C. Bouchard, R. J. Shepard, & T. Stephens (Eds.), *Physical activity, fitness, and health. International proceedings and consensus statement* (pp. 146–159). Champaign, IL: Human Kinetics Publishers.

Bouchard, C., An, P., Rice, T., Skinner, J. S., Wilmore, J. H., Gagnon, J., et al. (1999). Familial aggregation of $VO_{2\max}$ response to exercise training: Results from the HERITAGE Family Study. *Journal of Applied Physiology*, *87*, 1003–1008.

Bouchard, C., Blair, S. N., & Haskell, W. L. (2007). *Physical activity and health*. Toronto: Human Kinetics.

Brage, S., Brage, N., Franks, P. W., Ekelund, U., Wong, M. Y., Andersen, L. B., et al. (2004). Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *Journal of Applied Physiology*, *96*, 343–351.

Brage, S., Wedderkopp, N., Andersen, L. B., & Froberg, K. (2003). Influence of step frequency on movement intensity predictions with the CSA accelerometer: A field validation study in children. *Pediatric Exercise Science*, *15*, 277–287.

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*, 126–131.

Conrad, F., & Blair, J. (2004). Aspects of data quality in cognitive interviews: The case of verbal reports. In S. Presser, J. Rothgeb, M. Couper, J. Lessler, E. Martin, J. Martin, et al. (Eds.), *Questionnaire development, evaluation and testing methods* (pp. 67–88). New York: Wiley.

Cooper, K. H. (1968). A means of assessing maximal oxygen intake: Correlation between field and treadmill testing. *Journal of the American Medical Association*, *203*, 201–204.

Cooper, S. M., Baker, J. S., Tong, R. J., Roberts, E., & Hanford, M. (2005). The repeatability and criterion related validity of the 20 m multistage fitness test as a predictor of maximal oxygen uptake in active young men. *British Journal of Sports Medicine*, *39*, e19.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A., Booth, M. L., Ainsworth, B. E., et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, *35*, 1381–1395.

Crouter, S. E., Schneider, P. L., & Bassett, D. R., Jr. (2005). Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. *Medicine and Science in Sports and Exercise*, *37*, 1673–1679.

Crouter, S. E., Schneider, P. L., Karabulut, M., & Bassett, D. R., Jr. (2003). Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Medicine and Science in Sports and Exercise*, *35*, 1455–1460.

De Backer, G., Ambrosioni, E., Borch-Johnsen, K., Brotons, C., Cifkova, R., Dallongeville, J., et al. (2003). European guidelines on cardiovascular disease prevention in clinical practice: Third joint task force of European and other societies on cardiovascular disease prevention in clinical practice. *European Journal of Cardiovascular Prevention and Rehabilitation*, *10*, S2–S78.

Ekelund, U., Aman, J., & Westerterp, K. R. (2003). Is the ArteACC index a valid indicator of freeliving physical activity in adolescents? *Obesity Research*, *11*, 793–801.

Ekelund, U., Sepp, H., Brage, S., Becker, W., Jakes, R., Hennings, M., et al. (2006). Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. *Public Health Nutrition*, *9*, 258–265.

Ekelund, U., Sjöström, M., Yngve, A., Poortvliet, E., Nilsson, A., Froberg, K., et al. (2001). Physical activity assessed by activity monitor and doubly labeled water in children. *Medicine and Science in Sports and Exercise*, *33*, 275–281.

Eurofit (1984). Experimental battery. *International Journal of Physical Education*, *21*, 1–16.

Fairclough, S., & Stratton, G. (2005). Improving health-enhancing physical activity in girls' physical education. *Health Education Research*, *20*, 448–457.

Grant, S., Corbett, K., Amjad, A. M., Wilson, J., & Aitchison, T. (1995). A comparison of methods of predicting maximum oxygen uptake. *British Journal of Sports Medicine*, *29*, 147–152.

Hamberg-van Reenen, H. H., Ariëns, G. A. M., Blatter, B. M., van Mechelen, W., & Bongers, P. M. (2007). A systematic review of the relation between physical capacity and future low back and neck/shoulder pain. *Pain*, *130*, 93–107.

Hansen, H. S., Froberg, K., Nielsen, J. R., & Hyldebrandt, N. (1989). A new approach to assessing maximal aerobic power in children: The Odense School Study. *European Journal of Applied Physiology and Occupational Physiology*, *58*, 618–624.

Howley, E. T. (2001). Type of activity: Resistance, aerobic and leisure versus occupational physical activity. *Medicine and Science in Sports and Exercise*, *33*, S364–S369.

Howley, E. T., Colacino, D. L., & Swensen, T. C. (1992). Factors affecting the oxygen cost of stepping on an electronic stepping ergometer. *Medicine and Science in Sports and Exercise*, *24*, 1055–1058.

Hu, G., Eriksson, J., Barengo, N. C., Lakka, T. A., Valle, T. T., Nissinen, A., et al. (2004). Occupational, commuting, and leisure-time physical activity in relation to total and cardiovascular mortality among Finnish subjects with type 2 diabetes. *Circulation*, *110*, 666–673.

Hu, G., Jousilahti, P., Borodulin, K., Barengo, N. C., Lakka, T. A., Nissinen, A., et al. (2007). Occupational, commuting and leisure-time physical activity in relation to coronary heart disease among middle-aged Finnish men and women. *Atherosclerosis*, *194*, 490–497.

Hu, G., Qiao, Q., Silventoinen, K., Eriksson, J. G., Jousilahti, P., Lindstrom, J., et al. (2003). Occupational, commuting, and leisure-time physical activity in relation to risk for type 2 diabetes in middle-aged Finnish men and women. *Diabetologia*, *46*, 322–329.

Hustvedt, B. E., Christophersen, A., Johnsen, L. R., Tomten, H., McNeill, G., Haggarty, P., et al. (2004). Description and validation of the ActiReg: A novel instrument to measure physical activity and energy expenditure. *British Journal of Nutrition*, *92*, 1001–1008.

Jackson, A. S., & Coleman, A. E. (1976). Validation of distance run tests for elementary school children. *Research Quarterly for Exercise and Sport*, *47*, 86–94.

Jurca, R., Jackson, A. S., LaMonte, M. J., Morrow, J. R., Jr., Blair, S. N., Wareham, N. J., et al. (2005). Assessing cardiorespiratory fitness without performing exercise testing. *American Journal of Preventive Medicine*, *29*, 185–193.

Karageorghis, C. I., Vencato, M. M., Chatzisarantis, N. L., & Carron, A. V. (2005). Development and initial validation of the Brunel lifestyle physical activity questionnaire. *British Journal of Sports Medicine*, *39*, 23e.

Khaw, K. T., Jakes, R., Bingham, S., Welch, A., Luben, R., Day, N., et al. (2006). Work and leisure time physical activity assessed using a simple, pragmatic, validated questionnaire and incident cardiovascular disease and all-cause mortality in men and women: The EPIC–Norfolk prospective population study. *International Journal of Epidemiology*, *35*, 1034–1043.

Kohl, H., Blair, S. N., Paffenbarger, R. S., Macera, C. A., & Kronenfeld, J. J. (1988). A mail survey of physical activity habits as related to measured physical fitness. *American Journal of Epidemiology*, *127*, 1228–1239.

Kohl, H. W., Fulton, J. E., & Caspersen, C. J. (2000). Assessment of physical activity among children and adolescents: A review and synthesis. *Preventive Medicine*, *31*, S54–S76.

LaMonte, M. J., & Ainsworth, B. E. (2001). Quantifying energy expenditure and physical activity in the context of dose response. *Medicine and Science in Sports and Exercise*, *33*, S370–S378.

Laukkanen, R. M. T., Oja, P., Pasanen, M. E., & Vuori, I. M. (1993). Criterion validity of a two-kilometer walking test for predicting the maximal oxygen uptake of moderate to highly active middle-aged adults. *Scandinavian Journal of Medicine and Science in Sports*, *3*, 267–272.

Lee, I. M., Sesso, H. D., Oguma, Y., & Paffenbarger, R. S., Jr. (2003). Relative intensity of physical activity and risk of coronary heart disease. *Circulation*, *107*, 1110–1116.

Léger, L., Mercier, D., Gadoury, C., & Lambert, J. (1988). The multistage 20 meter shuttle run test for aerobic fitness. *Journal of Sports Sciences*, *6*, 93–101.

Léger, L., & Gadoury, C. (1989). Validity of the 20 m shuttle run test for aerobic fitness. *Canadian Journal of Sports Sciences, 14*, 21–26.

Levin, S., Jacobs, D. R., Ainsworth, B. E., Richardson, M. T., & Leon, A. S. (1999). Intraindividual variation and estimates of usual physical activity. *Annals of Epidemiology*, *9*, 481–488.

Li, A. M., Yin, J., Yu, C. C. W., Tsang, T., So, H. K., Wong, E., et al. (2005). The six-minute walk test in healthy children: Reliability and validity. *European Respiratory Journal*, *25*, 1057–1060.

Littmann, A. J., White, E., Kristal, A. R., Patterson, R. E., Satia-Abouta, J., & Potter, J. D. (2004). Assessment of a one-page questionnaire on long-term recreational physical activity. *Epidemiology*, *15*, 105–113.

Liu, N. Y. S., Plowman, S. A., & Looney, M. A. (1992). The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Research Quarterly for Exercise and Sport*, *63*, 360–365.

Marshall, A. L., Smith, B. J., Bauman, A. E., & Kaur, S. (2005). Reliability and validity of a brief physical activity assessment for use by family doctors. *British Journal of Sports Medicine*, *39*, 294–297.

Matthews, C. E., Ainsworth, B. E., Hanby, C., Pate, R. R., Addy, C., Freedson, P. S., et al. (2005). Development and testing of a short physical activity recall questionnaire. *Medicine and Science in Sports and Exercise*, *37*, 986–994.

McArdle, W. D., Katch, F. L., & Katch, V. L. (1996). *Exercise physiology: Energy, nutrition, and human performance*. Baltimore, MD: Williams & Wilkins.

McCutcheon, M. C., Sticha, S. A., Giese, M. D., & Nagle, F. J. (1990). A further analysis of the 12-minute run prediction of maximal aerobic power. *Research Quarterly for Exercise and Sport*, *61*, 280–283.

McNaughton, L., Hall, P., & Cooley, P. (1998). Validation of several methods of estimating maximal oxygen uptake in young men. *Perceptual and Motor Skills*, *87*, 575–584.

Motl, R. W., McAuley, E., & DiStefano, C. (2005). Is social desirability associated with self-reported physical activity? *Preventive Medicine*, *40*, 735–739.

Myers, J. N., Prakash, M., Froelicher, V., Do, D., Partington, S., & Atwood, J. E. (2002). Exercise capacity and mortality among men referred for exercise testing. *New England Journal of Medicine*, *346*, 793–801.

Paffenbarger, R. S., Wing, A. L., & Hyde, R. T. (1995). Physical activity as an index of heart attack risk in college alumni. *American Journal of Epidemiology*, *142*, 889–903.

Pate, R. R., Burgess, M. L., Woods, J. A., Ross, J. G., & Baumgartner, T. (1993). Validity of field tests of upper body muscular strength. *Research Quarterly for Exercise and Sport*, *64*, 17–24.

Pereira, M. A., FitzerGerald, S. J., Gregg, E. W., Joswiak, M. L., Ryan, W. J., Suminski, R. R., et al. (1997). A collection of physical activity questionnaires for health-related research. *Medicine and Science in Sports and Exercise*, *29*, S1–S205.

Philips, W. T., Batterham, A. M., Valenzuela, J. E., & Burkett, L. N. (2004). Reliability of maximal strength testing in older adults. *Archives of Physical Medicine and Rehabilitation*, *85*, 329–334.

Plasqui, G., Joosen, A. M., Kester, A. D., Goris, A. H., & Westerterp, K. R. (2005). Measuring free-living energy expenditure and physical activity with triaxial accelerometry. *Obesity Research*, *13*, 1363–1369.

Plasqui, G., & Westerterp, K. R. (2007). Physical activity assessment with accelerometers: A evaluation against doubly labeled water. *Obesity*, *15*, 2371–2379.

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, *10*, 150–157.

Racette, S. B., Schoeller, D. A., & Kushner, R. F. (1995). Comparison of heart rate and physical activity recall with doubly labeled water in obese women. *Medicine and Science in Sports and Exercise*, *27*, 126–133.

Rafamantanantsoa, H. H., Ebine, N., Yoshioka, M., Higuchi, H., Yoshitake, Y., Tanaka, H., et al. (2002). Validation of three alternative methods to measure total energy expenditure against the double labeled water method for older Japanese men. *Journal of Nutritional Science and Vitaminology*, *48*, 517–523.

Ramsbottom, R., Brewer, J., & Williams, C. A. (1988). A progressive shuttle run test to estimate maximal oxygen intake. *British Journal of Sports Medicine*, *22*, 141–144.

Rance, M., Boussuge, P. Y., Lazaar, N., Bedu, M., Van Praagh, E., Dabonneville, M., et al. (2005). Validity of a VO2max prediction equation of 2-km walk test in female seniors. *International Journal of Sports Science*, *26*, 453–456.

Rutten, A., Vuillemin, A., Ooijendijk, W. T., Schena, F., Sjöström, M., Stahl, T., et al. (2003). Physical activity monitoring in Europe: The European Physical Activity Surveillance System (EUPASS) approach and indicator testing. *Public Health Nutrition*, *6*, 371–376.

Safrit, M. J., Hooper, L. M., Ehlert, S. A., Costa, M. G., & Patterson, P. (1988). The validity generalization of distance run tests. *Canadian Journal of Sport Sciences*, *13*, 188–196.

Sallis, J. F. (1991). Self-report measures of children's physical activity. *Journal of School Health*, *61*, 215–219.

Sallis, J. F., Buono, M. J., & Freedson, P. S. (1991). Bias in estimating caloric expenditure from physical activity in children: Implications for epidemiological studies. *Sports Medicine*, *11*, 203–209.

Sallis, J. F., & Saelens, B. E. (2000). Assessment of physical activity by self-report: Status, limitations, and future directions. *Research Quarterly for Exercise and Sport*, *71*, S1–S14.

Saltin, B., & Grimby, G. (1968). Physiological analysis of middle-aged and old former athletes: Comparison with still active athletes of the same ages. *Circulation*, *38*, 1104–1115.

Saris, W. H. (1985). The assessment and evaluation of daily physical activity in children: A review. *Acta Paediatrica Scandinavica, Supplement*, *318*, 37–48.

Schmidt, M. E., & Steindorf, K. (2006). Statistical methods for the validation of questionnaires – discrepancy between theory and practice. *Methods of Information in Medicine*, *45*, 409–413.

Shephard, R. J. (1984). Tests of maximum oxygen intake: A critical review. *Sports Medicine*, *1*, 147–155.

Shephard, R. J. (2003). Limits of the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine*, *37*, 197–206.

Shetty, A. B. (2002). Estimation of leg power: A two-variable model. *Sports Biomechanics*, *1*, 147–155.

Sparling, P. B., Millard-Stafford, M., & Snow, T. K. (1997). Development of a cadence curl-up test for college students. *Research Quarterly for Exercise and Sport*, *68*, 309–316.

St-Onge, M., Mignault, D., Allison, D. B., & Rabasa-Lhoret, R. (2007). Evaluation of a portable device to measure daily energy expenditure in free-living adults. *American Journal of Clinical Nutrition*, *85*, 742–749.

Suni, J. H., Miilunpalo, S., Asikainen, T.-M., Laukkannen, R., Oja, P., Pasanen, M., et al. (1998a). Safety and feasibility of a health-related fitness test battery for adults. *Physical Therapy*, *78*, 134–148.

Suni, J. H., Oja, P., Laukkanen, R., Miilunpalo, S., Pasanen, M., Vuori, I., et al. (1996). Health-related fitness test battery for adults: Aspects of reliability. *Archives of Physical Medicine and Rehabilitation*, *77*, 399–405.

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

Taylor, H. L., Jacobs, D. R., Schucker, B., Knudsen, J., Leon, A. S., & De Backer, G. (1978). A questionnaire for the assessment of leisure time physical activities. *Journal of Chronic Diseases*, *31*, 741–755.

Trolle-Lagerros, Y., Mucci, L. A., Kumle, M., Braaten, T., Weiderpass, E., Hsieh, C. C., et al. (2005). Physical activity as a determinant of mortality in women. *Epidemiology*, *16*, 780–785.

Trost, S. G., Ward, D. S., Moorehead, S. M., Watson, P. D., Riner, W., & Burke, J. R. (1998). Validity of computer science and applications (CSA) activity monitor in children. *Medicine and Science in Sports and Exercise*, *30*, 629–633.

van den Beld, W. A., van den Sanden, G. A., Sengers, R. C., Verbeek, A. L., & Gabreels, R. J. (2006). Validity and reproducibility of hand-held dynamometry in children aged 4–11 years. *Journal of Rehabilitation Medicine*, *38*, 57–64.

van Sluijs, E. M., van Poppel, M. N., Twisk, J. W., & van Mechelen, W. (2006). Physical activity measurements affected participants' behaviour in a randomized controlled trial. *Journal of Clinical Epidemiology*, *59*, 404–411.

von Huth Smith, L., Borch-Johnsen, K., & Jørgensen, T. (2007). Commuting physical activity is favourably associated with biological risk factors for cardiovascular disease. *European Journal of Epidemiology*, *22*, 771–779.

Wareham, N. J., Jakes, R. W., Rennie, K. L., Mitchell, J., Hennings, S., & Day, N. E. (2002). Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. *International Journal of Epidemiology*, *31*, 168–174.

Wareham, N. J., Jakes, R., Rennie, K. L., Schuit, J., Mitchell, J., Hennings, S., et al. (2003). Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutrition*, *6*, 407–413.

Wareham, N. J., & Rennie, K. L. (1998). The assessment of physical activity in individuals and populations: Why try to be more precise about how physical activity is assessed? *International Journal of Obesity and Related Metabolic Disorders*, *22*, S30–S38.

Wendel-Vos, G., Schuit, J., Saris, W. H., & Kromhout, D. (2003). Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *Journal of Clinical Epidemiology*, *56*, 1163–1169.

Willett, W. (1998). *Nutritional epidemiology.* New York: Oxford University Press.

Wong, M. Y., Day, N. E., & Wareham, N. J. (1999). Measurement error in epidemiology: The design of validation studies II: Bivariate situation. *Statistics in Medicine*, *18*, 2831–2845.