Jørgensen T., Andersen, L. B., Froberg, K., Maeder, U., von Huth Smith, L., Aadahl, M. (2009). Position statement: Testing physical condition in a population: how good are the methods?. *European Journal of Sport Science, 9*, 257-267.

Abstract

A poor physical condition – expressed as physical inactivity and poor physical fitness – is associated with development of chronic diseases and premature death. The aim of this position statement is to evaluate the currently available methods 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 longer time periods and no feasible general criterion measure exist. Both objective and subjective methods are available. Of the objective methods accelerometry seems to be the most attractive technology, and is well enough developed for general use in large populations. The advantage of the method is that it is not dependent on memory of the individual, but the disadvantage is that it grossly underestimates energy expenditure, due to the lack of registration of certain activities. This may be solved to a certain extent by combining it with heart rate measurement, but it still does not measure activity in different domains. Of the subjective methods self-report questionnaires are feasible and easy to administer. Many questionnaires have been developed, but we are in need of 1) consensus on which measures to use for validation and 2) further development of internationally standardised questionnaires to be used in different settings and to different scientific questions. Many questionnaires correlate well with biological markers and development of chronic diseases, but subjective measurement will always face a certain degree of misclassification. Furthermore, unstructured physical activity like e.g. housework and gardening may be subject to recall bias. So far no measurement seem superior to the other, and the choice of instrument will depend on the research question asked. Future research should combine information from both objective and subjective methods.

Physical fitness comprises several components with cardiorespiratory endurance and muscle strength and endurance as the most important. Direct measurement of oxygen consumption is the criterion measure for cardiorespiratory endurance. As regards muscle strengths and endurance only test-retest reliability is available. The hand held dynamometers greatly facilitate field testing for maximal isometric muscle strength assessment, whereas force plate measurements can be used for the lower extremities. For endurance simple tests like push-ups and sit-ups seems reliable.

Introduction

A sedentary lifestyle has become prevalent in modern society and physical inactivity(Berlin & Colditz, 1990; Blair et al., 2001) as well as poor physical fitness (Blair et al., 1996; Blair et al., 2001) are associated with development of chronic diseases and premature death. Our knowledge on 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 facilitate the surveillance and assess the effect of preventive undertakings. Therefore sensitive, valid and reliable instruments for measuring the physical condition in large populations are imperative.

Physical condition comprises physical activity and physical fitness. Physical fitness and physical activity should be regarded as different, but complementary aspects of physical condition. Although physical activity is an important determinant of physical fitness (Bouchard et al., 2007; De Backer G 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, but may be less important to determine the fitness level in a sedentary subject, i.e. overweight and bed rest may decrease fitness level 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 level through physical activity seems genetically determined. When physical activity and fitness are analysed 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.

The aim of the present paper is to evaluate the currently available methods for measuring the physical condition of the population and to recommend further steps to improve measurements to increase our understanding of the implication of the physical condition on a population level.

Physical activity

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

Physical activity is determined by duration, frequency, intensity and type (Howley ET, 2001), and it derives from different domains such as occupation, 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 MA et al., 1997) and commuting activities (Barengo et al., 2006; Hu et al., 2003; Hu et al., 2004; Hu et al., 2007).
Type refers to mode of contraction (static versus dynamic) and whether the activity comprises small or large muscle groups. *Duration* refers to number of minutes or hours an activity is performed, whereas *frequency* describes the number of sessions or bouts of activity that is undertaken per day, week or month. *Intensity* of physical activity is expressed as EE per unit of time (e.g. kJ/min) and most studies assess the absolute intensity against e.g. development of chronic diseases. *The relative intensity*, i.e. the percentage of an individual's maximal oxygen uptake that is needed to perform a specific task (Bouchard et al., 1994; Howley ET, 2001) is rarely assessed, but it has been shown that the relative intensity may be a stronger predictor for chronic disease and premature death than absolute intensity (Lee et al., 2003). Accordingly, the American College of Sports Medicine recommends that intensity of physical activity should be relative to one's maximum oxygen uptake (American College of Sports Medicine position stand., 1990). A frequently used method for quantifying the intensity of physical activity is the use of metabolic equivalents (MET) (Ainsworth et al., 1993; Ainsworth et al., 2000). One MET corresponds to an EE of one kcal per kg body mass per hour or 3.5 ml O_2 per kg body mass per minute. 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; Ainsworth et al., 2000).

Instruments for measuring physical activity should be valid and reliable, but also feasible for use in large populations. Preferably it should be tested against a gold standard (criterion validity) or – less optimal – tested 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 are direct behavioural observation, direct or indirect calorimetry and the double labelled water method (DLW). Direct behavioural observation can theoretically validate all domains and dimensions by following and observing free living individuals. Direct calorimetry measures the total EE by measuring heat production or heat loss, and is performed in a closed chamber. The principle in the double labelled water method (DLW) (Ekelund U 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. DLW can only measure total EE, but combined with indirect calorimetry, an estimate of physical activity EE 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 covering longer time periods.

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

Objective measures of physical activity

The technology of objective measures for measuring physical activity among larger population groups has developed fast during the last couple of decades. In the following the knowledge on pedometers, heart rate monitoring, accelerometers and various combinations will be described.

Pedometers

Pedometers are easy to mount and wear and they are an inexpensive method for counting steps (Saris WH, 1985). A large number of pedometers are on the market, where NL-2000 seems to be mose reliable (Crouter et al., 2005) in counting steps under different conditions. Pedometers

grossly underestimate physical activity expressed as EE (Crouter et al., 2003) even in studies where cycling was not a part of the physical activity, but they are suitable for monitoring campaigns as e.g. 10,000 steps a day (Crouter et al., 2003), and other intervention based changes over time.

Heart rate monitoring

Heart rate (HR) monitoring is based on the assumption of a linear relationship between HR and oxygen consumption in moderate to vigorous activities. During rest and low-intensity activities the relationship is not linear and is confounded by mood, temperature, and diet. HR depends on the physical fitness level of the subject and a valid estimation depends on individual calibration based on knowledge of maximal and resting HR, and even better against direct measurement of oxygen uptake. Some studies have used cut points >50 % of the HR reserve to estimate time spent in moderate and high intensity exercise (Fairclough S & Stratton G, 2005). Estimation of total EE from HR has been validated against DLW (Racette et al., 1995; Rafamantanantsoa HH et al., 2002) and showed acceptable correlations. Modern HR monitors are easy to use and can store data for longer periods of time. The method is mainly feasible in smaller studies due to the necessity of individual calibration. If only the relative load is of interest, estimates can be made using assumption on resting and maximal heart rate based on age and sex.

Accelerometers

Accelerometers measure movements in one, two or three planes (Plasqui et al., 2005) by piezoelectric transducers and microprocessors. The devices are small and easy to carry, and measurement units, counts per minute (cpm), 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 seconds allowing analyses of short bursts of activity. Furthermore, data can be stored up till 200 days. The method has developed rapidly during the last years and several accelerometers have been tested under laboratory conditions during standardised activities showing very good correlation with EE (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 double labelled water shows in general that accelerometers underestimate the total EE. The uniaxial accelerometer Actigraph (former MTI, former CSA) and the triaxial accelerometer Tracmor shows a reasonable correlation with EE calculated from double labelled water method (Plasqui G & Westerterp, 2007). Only Actigraph is commercially available, and

devices Actiwatch, Caltrac, Tritac and Lifecorder did not show substantial correlation with the double labelled water method (Plasqui G & Westerterp, 2007).

There are other problems associated with the use of accelerometer. The output is frequency dependent because of an electronic filter, which is constructed 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 S et al., 2003). Another drawback is that accelerometer output is levelling off when speed increased to more than 10 km h^{-1}(Brage S et al., 2003). However, the time spent in running above 10 km h^{-1} is limited during habitual physical activities, and the problem is small in epidemiological studies. Furthermore, accelerometers cannot register physical activity with no acceleration such as rowing, cycling, skating, and hill climbing. Neither can accelerometers register isometric muscle contraction, muscular work against and external force such as weight lifting, carrying and pushing. Bicycling is quantitatively a problem in some countries (Holland and Denmark). The challenge of translation of cpm into energy expenditure has not been solved yet, because different types of activity with the same energy demand reveals different output from the monitor, However, a number of studies have validated walking and running, and there is good agreement that a walking speed of 4 km h^{-1} on a treadmill or over-ground corresponds to about 2000 cpm (Brage S et al., 2003; Ekelund U et al., 2003; Puyau MR et al., 2002; Trost SG et al., 1998).

Other instruments

Recently a device, which combines heart rate monitoring and accelerometers, have been developed. This device (Actiheart®) combines the best features of heart rate monitoring and accelerometers by using HR in the high intensity range where HR best reflects the work load and the counts from the accelerometer in the low intensity range (Brage S et al., 2004). In combination with accelerometers the individual calibration as regard heart rate may not be necessary. Still, validation under free living conditions is needed.

The Actireg is a unit which can register acceleration and change in body position (bending down or changing from lying to standing). Wires from the accelerometer are attached to 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 less suitable in large scale studies (Arvidsson D et al., 2006; Hustvedt BE et al., 2004).
Finally, devices such as portable armbands, which combine 2-axis accelerometers with skin temperature are potentially suitable for calculation of EE (St-Onge M et al, 2007).

Self-report measurement of physical activity

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

Record and logs

Physical activity records are diaries kept by study participants (Ainsworth BE et al., 1994; LaMonte & Ainsworth BE, 2001). They provide 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 BE et al., 1994; LaMonte & Ainsworth BE, 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 double labelled water method (Rafamantanantsoa HH 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 BE, 2001; Sallis JF & Saelens BE, 2000). Physical activity questionnaires range from simple single-item global questionnaires assessing general level of physical activity in order to classify individuals as active or inactive (Saltin B & Grimby G, 1968), over recall questionnaires with fairly specific assessments to more extensive questionnaires assessing frequency, duration and intensity of specific activities during a specified time frame (from days to life-time) in different domains (Pereira MA et al., 1997). There exists a large number of questionnaires and a number of these were gathered and published in 1997 along with information on validation and reliability testing

(Pereira MA et al., 1997). Some of the most frequently used questionnaires in large adult study populations are the Minnesota Leisure-time Physical Activity Questionnaire (Taylor HL et al., 1978), the Paffenbarger Physical Activity Questionnaire (Paffenbarger RS et al., 1995) and the Aerobics Center Longitudinal Study Physical Activity Questionnaire (Kohl H et al., 1988).

Other questionnaires, also developed for adult populations, have been developed since 1997. The European Prospective Investigation into Cancer and Nutrition Study (EPIC) questionnaire, which assessed past year activity at home, at work and in recreation (Khaw KT et al., 2006; Wareham NJ et al., 2003; Wareham NJ et al., 2002); the "Vital" questionnaire, which measure usual recreational physical activity during the preceding 10 years (Littmann AJ et al., 2004); the "SQUASH" questionnaire, which assesses physical activity on an average week in the past months performed in different domains (Wendel-Vos G et al., 2003); the "Star" questionnaire, which is used in telephone interviews assessing the overall moderate and vigorous activity in a usual week (Matthews CE et al., 2005); a screening instrument for family doctors identifying inactive patients in a primary care setting (Marshall AL et al., 2005); the "Brunel Lifestyle physical activity questionnaire", which is an Internet based questionnaire intended for use in conjunction with a 12-week personalised fitness programme delivered through the Internet (Karageorghis CI et al., 2005), and questionnaires intending to quantify and estimate energy expenditure 24 hours a day within all domains of physical activity (Aadahl & Jørgensen, 2003; Aadahl et al., 2007; Trolle-Lagerros et al., 2005).

An attempt to reach consensus on questionnaires on physical activity is the development of 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 CL 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 A et al., 2003) to good (Craig CL et al., 2003) repeatability in test-retest analysis and a low correlation with other national physical activity questionnaires (Rutten A et al., 2003). Generally, the IPAQ instrument leads to higher estimates of total physical activity than other questionnaires (Rutten A et al., 2003).

Validation against accelerometer show relatively low correlation coefficients (Craig CL et al., 2003; Ekelund U et al., 2006).

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

- some kind of qualitative testing (e.g. cognitive interviewing (Beatty PC & Willis GB, 2007; Conrad F & Blair J, 2004) to make sure that respondent have the same conception of the questionnaire as the researcher.
- validation against physical activity records or detailed interview covering from 24 hours to one month recall, which is routinely used in nutritional research (Willett W, 1998).
- if possible test against DLW or direct observation.
- comparison with biomarkers.

Strengths and limitations of self-report questionnaires and objective measures

From an epidemiological point of view ideal 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 VO2max. As this is seldom realistic in large study populations, the choice of instrument often depends on the specific health outcome of interest. Total amount of physical activity may be the relevant exposure in relation to some health outcomes, whereas information on specific domains (e.g. leisure time) and 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 so far. 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 feasible and easy to administer. Many seem to correlate with biological markers (von Huth Smith L et al., 2007) and to predict development of chronic diseases and premature deaths (1-3). Still, it should be remembered that the self-report nature of questionnaires, will always face a certain degree of misclassification (Sallis JF & Saelens BE, 2000; Shephard RJ, 2003). Remembering the duration, frequency, intensity and type of physical activity performed in the past can be difficult for respondents, especially if the recall

timeframe is extensive, e.g. a year of a life time. This is a special problem among children due to cognitive limitations (Baranowski T et al., 1984; Kohl HW et al., 2000; Sallis JF et al., 1991; Sallis JF, 1991; Saris WH, 1985). Unstructured physical activities such as e.g. work, sports and exercise (Levin S et al., 1999). Social desirability bias, where respondents distort self-report in a favourable direction may also reduce the validity of self-reported physical activity measures (Motl RW et al., 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 NJ & Rennie KL, 1998; Wong MY et al., 1999) compared to 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 seems to grossly underestimate the energy expenditure, due to the lack of registration of certain activities. This may be solved to a certain extent by combining accelerometry with heart rate measurement – an emerging technology. Accelerometry is subject to the risk of reactivity (van Sluijs EM et al., 2006), i.e. to the fact that wearing the accelerometer may cause changes in physical activity pattern. This applies to other types of measurement as well, e.g. heart rate monitoring and direct observation. Also we need studies showing a dose-response relationship between cpm in accelerometry and various physiological (e.g. blood pressure) and biochemical (e.g. cholesterol) measures. Finally, we are in need of studies comparing cpm with hard end-points such as development of chronic diseases.

So far, 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 instruments should be used simultaneously to assess various aspects of measuring physical activity. Using accelerometer as criterion validity for questionnaires may not be relevant.

Physical fitness

Physical fitness comprises several components with cardio respiratory endurance as the most important, because of its strong relations to development of chronic diseases like cardiovascular diseases, diabetes, cancer and premature death (Blair SN et al., 1989; Myers JN et al., 2002), whereas muscle strength and endurance show inconsistent relation to musculoskeletal disorders (Hamberg-van Reenen HH et al., 2007) Additional components are musculoskeletal flexibility and

body composition. This report is restricted to simple methods for assessing cardio respiratory endurance, muscle strength and muscle endurance in larger populations.

Cardiorespiratory fitness

The most reliable and valid measure of aerobic capacity is the direct measurement of maximal oxygen consumption (VO_{2max}) (Safrit et al., 1988), although it is not immune against inaccuracy (Shephard RJ, 1984). The method is not feasible in larger population groups, since expensive and sophisticated equipment is required. A variety of less complex procedures have been developed to estimate VO_{2max} and their validity has been determined by comparing the estimates with the criterion measures, the direct measurement of oxygen consumption. 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 VO_{2max} depends on the type of subjects and the availability of appropriate equipment. During the tests, some degree of risk management is required (Clinical Exercise Testing, 2006).

Maximal exercise tests

The Cooper test (12-MRT; maximal 12 min run test) is strongly related to the criterion measure VO_{2max} in adults (r = 0.84 – 0.92) (Cooper, 1968; Grant S et al., 1995; McCutcheon et al., 1990) and in children (r = 0.9) (Jackson AS & Coleman AE, 1976). While estimation equation used yielded a systematic underestimation of VO_{2max} by 4 ml/kg/min (McCutcheon et al., 1990) in one study, no statistically significant difference was found in two other studies (Cooper, 1968; Grant S et al., 1995; McCutcheon et al., 1990). However, inexperienced runners have difficulties to find the optimal speed, and are therefore underestimated. The multistage 20-meter shuttle run test (MST) was shown to be an accurate method to estimate VO_{2max} in adults in one study (r = 0.90) (Léger L & Gadoury C, 1989), whereas other studies failed to provide such strong correlations (r = 0.79 – 0.86) and reported a statistically significant underestimation of VO_{2max} by MST (3.0 – 7.5 %) (Cooper et al., 2005; Grant S et al., 1995; McNaughton L et al., 1998; Ramsbottom R et al., 1988). In children and adolescents, the correlations between MST estimates and criterion measures of VO_{2max} ranged between 0.71 and 0.87 (American College of Sports Medicine position stand., 1990; Boreham CAG et al., 1990; Léger L et al., 1988; Liu NYS et al., 1992). A comparison between Cooper test and MST showed that the former is a better predictor of VO_{2max} (Grant S et al., 1995) in experienced runners. Both running tests are most appropriate for individuals with sufficient

fitness level and motor skills and require a considerable motivation. This problem has partly been solved in a recent intermittent running test (Andersen et al., 2008). This test does not require any equipment for the test leader or experience for the subject. Besides these running tests, a simple maximal exercise test on a cycle ergometer was developed for children (Hansen HS et al., 1989) and for adults (Andersen, 1995; Andersen et al., 1987). The correlation between the estimation of VO_{2max} 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 for using this test in larger population groups. Furthermore, the participants should be familiar with cycling in order to achieve maximal performance of the ergometer.

Sub maximal exercise test

Estimates of VO_{2max} derived from submaximal tests are often based on the linear relationship of workload, heart rate, and VO_2. They are difficult to compare due to different methods, activity and study-samples that were used to verify the validity of these assessments. A classic method is the cycle ergometer-test by Åstrand & Rhyming (Åstrand PO & Ryhming I, 1954). For the simple 2-km walk test (2-KWT) the estimates of VO_{2max} were compared with the criterion measure and correlation coefficients were 0.55, 0.79 and 0.60 for moderately active middle-aged women, men, and highly active men, respectively (Laukkanen RMT et al., 1993). In a recent study 6-min walk test estimates of VO_{2max} for female seniors were related to the criterion measure with r = 0.44 (Rance M et al., 2005). A better correlation between the walking test distance and the criterion measure was found from the 6-min walk test in children (r = 0.94) (Li AM et al., 2005). Step tests have been used for many years, but no validation studies have been published (Howley et al., 1992).

In large epidemiologic studies cardio respiratory fitness may be estimated in adults from a non-exercise test model, including 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 R et al., 2005). However, most of the variances in these tests are explained by variables which cannot be changed such as age and sex or variables such as body weight that do not change over a shorter period of time with increased physical activity. Recent literature suggests that self-rated physical fitness assessed by a simple question is well correlated to the criterion measure, maximal oxygen uptake (Aadahl M et al., 2007).

Muscle strength and endurance

A great number of field test batteries have been used in different studies. They include different tests for muscle strength and muscle endurance. Both strength and endurance tests can be static or dynamic, and batteries include in general a mixture of tests assessing a combination of strength or endurance; trunk, legs or arms; static or dynamic tests – in order to describe the subjects' physical abilities in general. A major problem is related to the fact that no standardization has occurred. There are many versions of sit-and-reach, and many versions of Sargent jump, where each research group has modified tests, just to mention some of the problems. This makes comparison between populations and description of secular trends in muscle strength or endurance very difficult. In the 1980's an attempt was made to standardize tests and test populations across Europe in order 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 EU-Commission. This work led to the construction of a test battery especially to assess health related fitness (Suni JH et al., 1996; Suni JH et al., 1998b; Suni JH et al., 1998a).

Examples of content in 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 AB, 2002). Otherwise, muscle strength and endurance tests are often evaluated by determining test-retests-reliability, as a gold standard is missing. 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, the application might be limited among larger groups, but if the participants are familiarized with the procedure, it is highly reliable (Philips WT et al., 2004). The portable or so called hand-held dynamometers greatly facilitate field testing for maximum isometric muscle strength assessment. The hand grip measurement is commonly accomplished, and the devices are able to determine the muscle strength of several muscle groups with a high reliability (r = 0.73 – 0.91) (van den Beld WA et al., 2006).

Measurements of muscular endurance are often made with simple tests, such as the push-ups to determine upper-body muscle endurance and sit-ups to measure abdominal muscle groups. If the results in a standardised push-up test are corrected for weight, it is a valid measurement tool to determine muscular endurance of the upper-body ($r = 0.70 – 0.73$) (Pate RR et al., 1993). Sit-ups may involve varying accessory muscles besides 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 minimise the use of the hip flexors. The curl-up test achieved an acceptable reliability ($r = 0.92$) (Sparling PB et al., 1997), whereas the reliability of dynamic or isometric sit-up test seems to be limited ($r < 0.50$). Isometric muscle endurance in 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 LB et al., 2006; Biering-Sørensen, 1984).

Conclusion

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

General conclusions

In order to determine the physical condition of large populations, we recommend that several components should be measured: physical activity, cardio respiratory fitness, muscle strength and muscle endurance. They represent fundamentally different aspects of physical health, but in order 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 feasible when the physical condition over a longer time period (years) is needed, as is the case in evaluating the effect on development of chronic disease. There is, therefore, still a need for development of questionnaires that are tested and well validated and allow for international comparisons among different study populations. Objective measurements give accurate, precise and valid estimates, but tend to be less feasible in large study populations and to underestimate total EE. Hence, there is still a need for development of objective methods that are more reliable and easy to administer in large study populations. The combined heart rate and accelerometer method seems promising.

References

American College of Sports Medicine position stand. (1990). *Med Sci Sports Exerc, 22,* 265-274.

(2006). Clinical Exercise Testing. In Whaley M.H., Brubraker P.H., and Otto R.M (Eds.), *ACSM's Guidelines for Exercise Testing and Prescription* (pp. 93-114). Philiadelphia: Lippincott Williams & Wilkins.

Aadahl M, Kjær M, Kristensen JH, Mollerup B, & Jørgensen T (2007). Self-reported physical activity compared with maximal oxygen uptake in adults. *Eur J Cardiovasc Prev Rehabil, 14,* 422-428.

Aadahl, M., Jørgensen, T. (2003). Validation of a new self-report instrument for measuring physical activity. *Med Sci Sports Exerc, 35,* 1196-1202.

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

Ainsworth BE, Montoye HJ, & Leon AS (1994). Methods of Assessing Physical Activity During Leisure and Work. In Bouchard, C, Shephard RJ, and Stephens T (Eds.), *Physical Activity, Fitness, and Health. International Proceddings and Consensus Statement* (pp. 146-159). Champaign, IL: Human Kinetics Publishers.

Ainsworth, B. E., Haskell, W. L., Leon, A. S., Jacobs DR, Montoye HJ, Sallis JF, & Paffenbarger RS (1993). Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc, 25,* 71-80.

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., O'Brien, W. L., Bassett, D. R., Schmitz, K. H., Emplaincourt, P. O., Jacobs, D. R., & Leon, A. S. (2000). Compendium of physical activites: an update of activity codes and MET intensities. *Med Sci Sports Exerc, 32,* S504

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

Andersen, L. B. (1995). A maximal cycle exercise protocol to predict maximal oxygen uptake. *Scand J Med Sci 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. *J Sports Med Phys Fitness, 48,* 434-437.,

Andersen, L. B., Henckel, P., & Saltin, B. (1987). Maximal oxygen uptake in Danish adolescents 16-19 years of age. *Eur J Appl Physiol, 56,* 74-82.

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. *Clin Nutr, 25,* 68-74.

Åstrand PO, Ryhming I (1954). A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *J Appl Physiol, 7*, 218-221.

Baranowski T, Dworkin RJ, Cieslik CJ, Hooks P, Clearman DR, & Ray L (1984). Reliability and validity of self report of aerobic activity: Family Health Project. *Res Q, 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. *Eur J Cardiovasc Prev Rehabil, 13*, 51-59.

Beatty PC, Willis GB (2007). Research Synthesis: The practice of cognitive interviews. *Public Opin Q, 71*, 287-311.

Berlin, J. A., Colditz, G. A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol, 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. *Dan Med Bull, 31*, 362-375.

Blair SN, Kohl W, Phaffenbarger RS, Clark DG, Cooper KH, & Gibbons LW (1989). Physical fitness and all-cause mortality: a prospective study of healthy men and women. *Journal of the American Medical Association, 262*, 2395-2401.

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc, 33*, S379-S399.

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

Boreham CAG, Paliczka VJ, & Nichols AK (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, Blair SN, and Haskell, W. L. (2007). Physical Activity and Health. Toronto: Human Kinetics.

Bouchard, C, Shepard RJ, & Stephens T (1994). Methods of assessing physical activity during leisure and work. In Bouchard, C, Shehard RJ, and Stephens T (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., Perusse, L., Leon, A. S., & Rao, D. C. (1999). Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol, 87*, 1003-1008.

Brage S, Brage N, Franks PW, Ekelund U, Wong MY, Andersen LB, Froberg K, & Wareham NJ (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 LB, & 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 Rep, 100*, 126-131.

Conrad F & Blair J (2004). Aspects of data quality in cognitive interviews: The case of verbal reports. In Presser S, Rothgeb J, Couper M, Lessler J, Martin E, Martin J, and et al (Eds.), *Questionnaire Development, Evaluation and Testing Methods* (pp. 67-88). New York: John Wiley and Sons.

Cooper, K. H. (1968). A means of assessing maximal oxygen intake. Correlation between field and treadmill testing. *JAMA, 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. *Br J Sports Med, 39*, e19-e19.

Craig CL, Marshall AL, Sjöström M, Bauman A, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, & Oja P (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc, 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. *Med Sci Sports Exerc, 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. *Med Sci Sports Exerc, 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. *Eur J Cardiovasc Prev Rehabil, 10*, S2-S78.

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

Ekelund U, Sepp H, Brage S, Becker W, Jakes R, Hennings M, & Wareham NJ (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, Wedderkopp N, & Westerterp, K. R. (2001). Physical activity assessed by activity monitor and doubly labeled wated in children. *Med Sci Sports Exerc, 33*, 275-281.

Eurofit (1984). Experimental battery. *Int J Phys Educ, 21*, 1-16

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

Grant S, Crobett K, Amjad AM, Wilson J, & Aitchison T (1995). A comparison of methods of predicting maimum oxygen uptake. *British Journal of Sports Medicine, 29*, 147-152.

Hamberg-van Reenen HH, Ariëns GAM, Blatter BM, van Mechelen W, & Bongers PM (2007). A systematic review of the relation between physical capacity and future low back and neck/shoulder pain. *Pain, 130*, 107-

Hansen HS, Froberg K, Nielsen JR, & 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 ET (2001). Type of activity: resistance, aerobic and leisure versus occupational physical activity. *Med Sci Sports Exerc, 33*, S364-S369.

Howley ET, Colacino DL, & Swensen TC (1992). Factors affecting the oxygen cost of stepping on an electronic stepping ergometer. *Med Sci Sports Exerc, 24*, 1055-1058.

Hu, G., Eriksson, J., Barengo, N. C., Lakka, T. A., Valle, T. T., Nissinen, A., Jousilahti, P., & Tuomilehto, J. (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., & Tuomilehto, J. (2007). Occupational, commuting and leisure-time physical activity in relation to coronary heart diseae among middle-aged Finish men and women. *Atherosclerosis, 194*, 490-497.

Hu, G., Qiao Q, Silventoinen K, Eriksson, J. G., Jousilahti, P., Lindstrom, J, Valle, T. T., Nissinen, A., & Tuomilehto, J. (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 BE, Christophersen A, Johnsen LR, Tomten H, McNeill G, Haggarty P, & Lovo A (2004). Description and validation of the ActiReg: a novel instrument to measure physical activity and energy expenditure. *Br J Nutr, 92*, 1001-1008.

Jackson AS, Coleman AE (1976). Validation of distance run tests for elementary school children. *Research Quarterly for Exercise and Sport, 47*, 86-94.

Karageorghis CI, Vencato MM, Chatzisarantis NL, & Carron AV (2005). Development and initial validation of the Brunel lifestyle physical activity questionnaire. *Br J Sports Med, 39*, e23-e23.

Khaw KT, Jakes R, Bingham S, Welch A, Luben R, Day N, & Wareham N (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 European Prospective Investigation into Cancer in Norfolk prospective population study. *Int J Epidemiol, 35*, 1034-1043.

Kohl H, Blair SN, Paffenbarger RS, Macera, C. A., & Kronenfeld JJ (1988). A mail survey of physical activity habits as related to measured physical fitness. *Am J Epidemiol, 127*, 1228-1239.

Kohl HW, Fulton JE, Caspersen CJ, & ' (2000). Assessment of physical activity among children and adolescents: a review and synthesis. *Prev Med, 31*, S54-S76.

LaMonte, M. J., Ainsworth BE (2001). Quantifying energy expenditure and physical activity in the context of dose response. *Med Sci Sports Exerc, 33*, S370-S378.

Laukkanen RMT, Oja P, Pasanen ME, & Vuori IM (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 Medical Science and 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, Gadoury C (1989). Validity of the 20 m shuttle run test for aeroblic fitness. *Journal of Sports Sciences, 14*, 21-26.

Léger L, Mercier D, Gadoury C, & Lambert J (1988). The multistage 20 meter shuttle run test for aerobic fitness. *Journal of Sports Sciences, 14*, 21-26.

Levin S, Jacobs DR, Ainsworth BE, Richardson MT, & Leon AS (1999). Intraindividual variation and estimates of usual physical activity. *Ann Epidemiology, 9*, 481-488.

Li AM, Yin J, Yu CCW, Tsang T, So HK, Wong E, Chang D, Hon EKI, & Sung R (2005). The six-minute walk test in healthy children: reliability and validity. *European Respiratory Journal, 25*, 1057-1060.

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

Liu NYS, Plowman SA, & Looney MA (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 AL, Smith BJ, Bauman AE, & Kaur S (2005). Reliability and validity of a brief physical activity assessment for use by family doctors. *Br J Sports Med, 39*, 294-297.

Matthews CE, Ainsworth BE, Hanby C, Pate RR, Addy C, Freedson PS, Jones DA, & Macera CA (2005). Development and testing of a short physical activity recall questionnaire. *Med Sci Sports Exerc, 37,* 986-994.

McArdle WD, Katch FL, and Katch VL (1996). *Exercise physiology: energy, nutrition, and human performance*. Baltimore: 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. *Res Q Exerc 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 RW, McAuley E, & DiStefano C (2005). Is social desirability associated with self-reported physical activity? *Prev Med, 40,* 735-739.

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

Paffenbarger RS, Wing AL, & Hyde RT (1995). Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol, 142,* 889-903.

Pate RR, Burgess ML, Woods JA, Ross JG, & Baumgartner T (1993). Validity of field tests of upper body muscular strength. *Research Quarterly for Exercise and Sport, 64,* 17-24.

Pereira MA, FitzerGerald SJ, Gregg EW, Joswiak ML, Ryan WJ, Suminski RR, Utter AC, & Zmuda JM (1997). A collection of Physical Activity Questionnaires for health-related research. *Med Sci Sports Exerc, 29,* S1-205.

Philips WT, Batterham AM, Valenzuela JE, & Burkett LN (2004). Reliability of maximal strenght testing in older adults. *Archives of Physical Medicine and Rehabilitation, 85,* 329-334.

Plasqui G, Westerterp, K. R. (2007). Physical activity assessment with accelometers: a evaluation against doubly labeled water. *Obesity, 15,* 2371-2379.

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. *Obes Res, 13,* 1363-1369.

Puyau MR, Adolph AL, Vohra FA, & Butte NF (2002). Validation and calibration of physical activity monitors in children. *Obes Res, 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. *Med Sci Sports Exerc, 27,* 126-133.

Rafamantanantsoa HH, Ebine N, Yoshioka M, Higuchi H, Yoshitake Y, Tanaka H, Saitoh S, & Johnes PS (2002). Validation of three alternative methods to measure total energy

expenditure against the double labeled water method for older Japanese men. *J Nutri Sci Vitaminol, 48*, 517-523.

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

Rance M, Boussuge PY, Lazaar N, Bedu M, Van Praagh E, Dabonneville M, & Duche P (2005). Validity of a VO2 max prediction equation of 2-km walk test in female seniors. *International Journal of Sports Science, 26*, 453-456.

Rutten A, Vuillemin A, Ooijendijk WT, Schena F, Sjöström M, Stahl T, Vanden Auweele Y, Welshman J, & Ziemainz H (2003). Physical activity monitoring in Europe. The Euorpean Physical Activity Surveillance System (EUPASS) approach and indicator testing. *Public Health Nutrition, 6*, 371-376.

Safrit, M. J., Glaucia, Costa M., Hooper, L. M., Patterson, P., & Ehlert, S. A. (1988). The validity generalization of distance run tests. *Can J Sport Sci, 13*, 188-196.

Sallis JF (1991). Self-report measures of children's physical activity. *J Sch Health, 61*, 215-219.

Sallis JF, Buono MJ, & Freedson PS (1991). Bias in estimating caloric expenditure from physical activity in children. Implications for epidemiological studies. *Sports Medicine, 11*, 203-209.

Sallis JF, Saelens BE (2000). Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport, 71*, S1-14.

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 WH (1985). The assessment and evaluation of daily physical activity in children. A review. *Acta Paediatr Scand Suppl, 318*, 37-48.

Schmidt ME, Steindorf K (2006). Statistical methods for the validation of questionnaires - discrepancy between theory and practice. *Methods Inf Med, 45*, 409-413.

Shephard RJ (1984). Tests of maximum oxygen intake. A critical review. *Sports Medicine, 1*, 147-155.

Shephard RJ (2003). Limits of the measurement of habitual physical activity by questionnaires. *Br J Sports Med, 37*, 197-206.

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

Sparling PB, Millard-Stafford M, & Snow TK (1997). Development of a candence curl-up test for college students. *Research Quarterly for Exercise and Sport, 68*, 309-316.

Suni JH, Miilunpalo S, Asikainen T-M, Laukkannen R, Oja P, Pasanen M et al (1998a) Safely and feasibility of a health-related fitness test battery for adults. *Physical Therapy, 78,* 134-48,

Suni JH, Oja P, Laukkanen R, Miilunpalo S, Pasanen M, Vuori I et al (1996). Health-related fitness test battery for adults: aspects of reliability. *Arch Phys Med Rehabil, 77,* 399-405.

Suni JH, Oja P, Miilunpalo S, Pasanen M, Vuori I, Bös K (1998b). Health-related fitness test battery for adults: associations with percieved health, mobility, and back function and symptoms. *Arch Phys Med Rehabil, 79,* 000-000.

Taylor HL, Jacobs DR, Schucker B, Knudsen J, Leon AS, & De Backer G (1978). A questionnaire for the assessment of leisure time physical activities. *J Chronic Dis, 31,* 741-755.

Trolle-Lagerros, Y., Mucci, L. A., Kumle, M., Braaten, T., Weiderpass, E., Hsieh, C. C., Sandin, S., Lagiou, P., Trichopoulos, D., Lund, E., & Adami, H. O. (2005). Physical activity as a determinant of mortality in women. *Epidemiology, 16,* 780-785.

Trost SG, Ward DS, Moorehead SM, Watson PD, Riner W, & Burke JR (1998). Validity of computer science and applications (CSA) activity monitor in children. *Med Sci Sports Exerc, 30,* 629-633.

van den Beld WA, van den Sanden GA, Sengers RC, Verbeek AL, & Gabreels RJ (2006). Validity and reproducibility of hand-held dynamometry in childred aged 4-11 years. *Journal of Rehabilitation Medicine, 38,* 57-64.

van Sluijs EM, van Poppel MN, Twisk JW, & van Mechelen, W. (2006). Physical activity measurements affected participants' behaviour in a randomized controlled trial. *J Clin Epidemiol, 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. *Eur J Epidemiol, 22,* 771-779.

Wareham NJ, Jakes R, Rennie KL, Schuit J, Mitchell J, Hennings S, & Day NE (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 NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, & Day NE (2002). Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. *Int J Epidemiol, 31,* 168-174.

Wareham NJ, Rennie KL (1998). The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Int J Obes Relat Metab Disord, 22,* S30-S38.

Wendel-Vos G, Schuit J, Saris WH, & Kromhout D (2003). Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *J Clin Epidemiol, 56*, 1163-1169.

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

Wong MY, Day NE, & Wareham NJ (1999). Measurement error in epidemiology: the design of validation studies II: bivariate situation. *Stat Med, 18*, 2831-2845.