**Author(s):** Hyvärinen, Matti; Sipilä, Sarianna; Kulmala, Janne; Hakonen, Harto; Tammelin, Tuija H.; Kujala, Urho M.; Kovanen, Vuokko; Laakkonen, Eija K.

**Title:** Validity and Reliability of a Single Question for Leisure-Time Physical Activity Assessment in Middle-Aged Women

**Year:** 2020

**Version:** Accepted version (Final draft)

# Validity and reliability of a single question for leisure time physical activity assessment in middle-aged women

Matti Hyvärinen[1], Sarianna Sipilä[1], Janne Kulmala[2], Harto Hakonen[2], Tuija H Tammelin[2], Urho M Kujala[3], Vuokko Kovanen[3], Eija K Laakkonen[1]

[1]Gerontology Research Center, Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[2]LIKES Research Centre for Physical Activity and Health, Jyväskylä, Finland

[3]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

Corresponding author:
Eija K Laakkonen
PO Box 35/viv, FIN-40014 University of Jyväskylä, Finland
eija.k.laakkonen@jyu.fi
+358-40-5326410

# Abstract

***Purpose:*** The study investigates the validity and test-retest reliability of a single seven-level scale physical activity assessment question (SR-PA L7) and its three-level categorization (SR-PA C3). ***Methods:*** The associations of SR-PA L7 and C3 with accelerometer-measured leisure time physical activity (ACC-LTPA) and with results of four different physical performance tests (six-minute walk (n=733), knee extension (n=695), vertical jump (n=731) and grip force (n=780)) were investigated among women aged 47–55 years participating in the Estrogenic regulation of muscle apoptosis (ERMA) study (n=795). The reliability was studied using Spearman correlations with four-month test-retest period (n=152). ***Results:*** SR-PA L7 and C3 had low correlations with ACC-LTPA ($r_s$=0.105–0.337). SR-PA L7, SR-PA C3 and ACC-LTPA explained comparable but small amount of variance of the physical performance test results. The reliability analysis provided moderate agreement ($r_s$=0.707 and 0.622 for SR-PA L7 and C3, respectively). ***Conclusions:*** SR-PA L7 and C3 demonstrated limited validity and reasonable repeatability.


*Keywords:* physical activity measurement, self-reported physical activity, leisure time physical activity, accelerometry, test-retest reliability

# Introduction

It is widely known that aging is associated with increased risk of cardiovascular and other diseases as well as reduced physical performance. In women, these aging decrements may accelerate at middle age coinciding menopause (Colpani et al., 2018; Tseng et al., 2012). Physical activity is well known to provide health benefits although occupational physical activity may be less beneficial than leisure time and higher occupational physical activity may predispose to lower leisure time physical activity (LTPA) (Holtermann, Krause, van der Beek, & Straker, 2018; Prince, Elliott, Scott, Visintini, & Reed, 2019). Furthermore, higher LTPA is associated with a decreased health risks (Colpani et al., 2018) and especially the moderate and vigorous physical activity with increments in physical performance (Nelson et al., 2004; Pahor et al., 2006) thus being associated with better cardiorespiratory and muscular fitness (Corcoran et al., 2016; Kulinski et al., 2014; Wanderley et al., 2011). Therefore, the measurements that describe the physical performance, including walking speed (Cesari et al., 2009; Newman, Simonsick et al., 2006; Studenski et al., 2011), grip strength (Gale, Martyn, Cooper, & Sayer, 2007; Rantanen et al., 1999; Sasaki, H., Kasagi, Yamada, & Fujita, 2007) and lower-extremity muscle strength and power (Newman, Kupelian et al., 2006; Reid & Fielding, 2012), are of importance for assessing health-related physical functioning. However, the implementation of performance tests for assessing the physical performance level in large epidemiological studies can be very expensive and time-consuming. On the other hand, estimating LTPA by validated tools will also provide insight regarding physical performance.

Although doubly labeled water method is considered the gold standard for measuring energy expenditure caused by physical activity it is not often used due to methodological impracticalities (Sylvia, Bernstein, Hubbard, Keating, & Anderson, 2014). Other more commonly used objective assessment methods for measuring overall LTPA (e.g. accelerometer, pedometer and heart rate monitor) are generally also considered to provide more precise estimates compared to self-reported methods (Prince et al., 2008). However, each have their own limitations and sources of error (Ndahimana & Kim, 2017). For instance, it is widely known that hip-attached accelerometers register poorly activities involving only a minimal movement of the body's center of gravity such as bicycling and rowing (Dishman, Washburn, & Schoeller, 2001; Prince et al., 2008).

The self-report questionnaires and single question scales are often preferred in epidemiologic studies due to their relatively low cost and ease of implementation regardless of their limited reliability and validity due to the potential response bias and issues related to recalling the physical activity (Kowalski, Rhodes, Naylor, Tuokko, & MacDonald, 2012; Shephard, 2003). However, compared to accelerometer-based methods, clearly less knowledge exists about the relationship between single self-report questions for physical activity assessment and physical performance including muscular fitness. The few studies that have been conducted have focused on the elderly (Frändin & Grimby, 1994; Rantanen, Era, & Heikkinen, 1997). Furthermore, scanty knowledge exists about the associations between the single self-report questions and objective methods for physical activity assessment and, again, the studies that have been carried out focus mostly on the elderly (Emaus et al., 2010; Portegijs, Sipilä, Viljanen, Rantakokko, & Rantanen, 2017).

Perhaps the most widely used single self-report question for assessing habitual physical activity in older Nordic populations is the single-question scale developed by Saltin and Grimby (Saltin & Grimby, 1968) and its different formats. The original four-level scale has been modified over the years to its current six- and seven-level forms to capture also the activities performed at low intensity levels. These activities include e.g. household and daily living activities, which for some elderly may be the major form of habitual physical activity. The development of the different forms of this scale has been recently reviewed (Grimby & Frändin, 2018). Another fairly similar seven-level single-question scale for self-reported LTPA (SR-PA L7) assessment was first introduced by Hirvensalo et al. (Hirvensalo, Lampinen, & Rantanen, 1998). The main differences between SR-PA L7 and different formats of Saltin and Grimby question are in the formulation of physical activity frequency and focusing solely on the outdoor activities. The SR-PA L7 uses the format "how many times per week" while Saltin and Grimby uses the format "how many hours a week". In addition, the SR-PA L7 uses verbalizations such as "casual walks" and "light outdoor recreation", while Saltin and Grimby include also terminology referring to indoor domestic work such as cooking, dusting, straightening up and making beds. The difference in verbalization may cause SR-PA L7 to be a useful tool for middle-aged populations, who in general are considered to be somewhat more active than elderly.

SR-PA L7 and its categorization to three-level or two-level item has been used in research on pre- to postmenopausal women (Bondarev et al., 2018; Laakkonen, Soliymani et al., 2017; Ronkainen et al., 2008; Ronkainen et al., 2009; Sillanpää et al., 2017) and on older people (Hirvensalo et al., 1998). However, to our knowledge, no validation studies have been performed. Therefore, the aim of this study is to scrutinize the concurrent validity of the SR-PA L7 – a single seven-level scale physical activity question – and its three-level categorization (SR-PA C3) by studying its associations with accelerometer-measured LTPA (ACC-LTPA) and physical performance test results in middle-aged women. The physical performance was measured with comprehensive test battery including the tests for six-minute walk distance, knee extension force, vertical jump height and grip strength. In addition, a test-retest reliability analysis was performed.

## Methods

### Cohort description

Cross-sectional baseline data of the study Estrogenic Regulation of Muscle Apoptosis (ERMA) was analyzed. The Ethics Committee of the XXX Health Care District (XXXXX Dnro 8U/2014) approved the ERMA study and all study participants gave written informed consent. The more detailed characteristics of the study design, subject recruitment and exclusion criteria have been described previously (Kovanen et al., 2018). Briefly, the study focused on relatively healthy women aged 47 to 55 years undergoing their menopausal transition. Only mild health concerns were allowed to be present among the participants included in the current study. Any of the concerns were not such that it would hinder participant's ability for being physically active. The most common reported types of health concerns were musculoskeletal disorder, mental or neural disorder, elevated blood pressure, allergies or respiratory disorder and arrhythmia having prevalence of 36%, 15%, 14%, 10% and 8%, respectively. The number of participants that answered the SR-PA L7 and had valid accelerometer data in the baseline was 795. Within this group, the amount of participants that attended to each performance test varied from 695 to 788. Of the 795 participants answering the SR-PA L7 during the baseline measurement, 152 participants administered the question again during the next 120 days. This period was considered to be sufficiently short enough to minimize the potential true change that may occur in the physical activity but long enough to bring out the incidental variation in the responses to the SR-PA L7.

As the study participants represent the age group undergoing menopausal transition and menopause has been shown to affect their physical performance (Bondarev et al., 2018), participants' menopausal status was controlled in the current study. Participants were divided into four menopausal groups – premenopausal (PRE), early perimenopausal (PERI1), late perimenopausal (PERI2) and postmenopausal (POST). The categorization was implemented based on follicle-stimulating hormone (FSH) serum concentrations that were measured using IMMULITE 2000 XPi (Siemens Healthcare Diagnostics, UK) and, if available, the menstrual diary that the participants were advised to keep at least 12 weeks before the onset of the study (Kovanen et al., 2018).

**Self-reported physical activity**

For the laboratory visit, the participants were instructed to fill the baseline questionnaire that included the SR-PA L7. The SR-PA L7 is a single-question scale for assessment of current physical activity ranging from necessary daily activities and routines to participation in competitive sports (Hirvensalo et al., 1998). Although the question does not specify the type of the physical activity, it is distinctly more focused on the leisure time outdoor physical activities due to the design of the level descriptions (Table 1).

In some earlier studies that have used the SR-PA L7, the participants have been categorized to three activity groups (SR-PA C3) using the SR-PA L7 by merging the groups of the original question (Bondarev et al., 2018; Laakkonen et al., 2017; Ronkainen et al., 2008; Ronkainen et al., 2009; Sillanpää et al., 2017). Thus, in order to scrutinize the validity and reliability of this three-level categorization, the participants were also categorized into three physical activity groups based on the indicative volume of their physical activity; low (levels 1 and 2), medium (levels 3 and 4), and high (levels 5 to 7). The design of SR-PA L7 question and SR-PA C3 categories with frequencies of the responders are illustrated in Table 1.

**Physical activity assessment using accelerometer**

Physical activity was also assessed by using an accelerometer accompanied by physical activity diary. After the laboratory visit, the participants were personally instructed to wear the monitors for seven consecutive days on their right hip during the waking hours, except while doing aquatic activities. Furthermore, the participants were advised to record their wake-up time, working hours, and periods when the accelerometer was removed for over 30 minutes. GT3X and wGT3X ActiGraph accelerometers (Pensacola, FL) were used and the raw data was collected at the frequency of 60 Hz. The collected data was filtered, converted into 60-second epoch counts and divided into mean times spend at different physical activity intensities. The non-wear time was defined as periods of longer than 60 minutes of continuous zero counts (Migueles et al., 2017). Valid accelerometer measurement included three or more total valid measurement days in which the accelerometer wear time was at least 10 hours (Matthews, Hagströmer, Pober, & Bowles, 2012), however, 86 % (n = 681) of the participants had complete seven valid measurement days. Workdays and weekends were not addressed separately since many participants were working shifts and on weekends as well. The intensity levels were sedentary time in addition to light, moderate and vigorous physical activity, and the corresponding tri-axial vector magnitude cut-off points were 450, 2690 and 6166 counts per minute (Laakkonen, Kulmala et al., 2017; Sasaki, J. E., Dinesh, & Freedson, 2011). Furthermore, 25 000 counts per minute was used as the upper limit for tri-axial data (Laakkonen et al., 2017). Moderate-to-vigorous physical activity was defined by computing the sum of moderate and vigorous physical activity.

Working time and leisure time data was separated by utilizing the accompanied diary. The leisure time physical activity (ACC-LTPA) was normalized to 10-hour wearing time and whole day to 16-hour wearing time per day (Brakenridge et al., 2016). The leisure time data was utilized in the main analysis in order to produce more reliable comparisons with the SR-PA L7 that focused on the LTPA.

**Physical performance tests**

A set of laboratory measurements including a six-minute walk test (6MWT), maximal isometric knee extension test (KE), vertical jump test (VJ) and maximal isometric hand grip test (GF) to assess cardiorespiratory and muscular fitness were performed during the laboratory visit. However, some participants were unable to perform every performance test. The number of participants within each of the performance test are presented along with the test description in the following text. Cardiorespiratory fitness was tested using a 6MWT (n = 733). It was performed on a 20-meter indoor track, and the participants were instructed to complete as many laps as possible during the six-minute time (Enright, 2003). The measured attributes were distance walked, heart rate and perceived exertion measured on the Borg scale (Borg, 1982). Muscular fitness was evaluated with three different tests: KE (n = 695) was used to assess quadriceps muscle force, VJ (n = 739) was used to assess lower body muscle power and GF (n = 788) to assess hand and forearm force. The KE was measured at knee angle of 60° from full extension. Participants were instructed to extend the knee to produce the maximal force, and the peak force value was recorded. The lower body muscle power was measured by performing a vertical jump on a contact mat. Flight time *(t)* was measured, and the height of the jump *(h)* in meters was calculated as $h = gt^2/8$, where $g$ denotes the acceleration of gravity (9.81 $m/s^2$) (Bosco, Luhtanen, & Komi, 1983; Sipilä et al., 2001). In the GF measurement, participant's dominant arm was fixed to the armrest of the custom-made dynamometer chair (Good Strength; Metitur Oy, Palokka, Finland) with the elbow flexed at 90º angle (Ronkainen et al., 2009). Participants were instructed to squeeze the handle as forcefully as possible and maintain the contraction for two to three seconds. The peak force value was recorded for the analysis. In each test, the effort was repeated three to five times, and the effort with highest value was recorded.

**Background variables**

The baseline questionnaire was also used to determine the characteristics of the study population related to socioeconomic status and life habits. The participants were categorized by their smoking status into non-smokers, former smokers and current smokers, and their weekly alcohol consumption in alcohol units was calculated. Furthermore, based on the questions related to socioeconomic status, the participants were categorized by their marital status (single, in relationship), level of education (primary, secondary, tertiary), current work status (employed, not regularly employed, retired) and the physical load of the current work (sedentary, light, moderate, heavy). The anthropometrics were measured during the laboratory visit between 7:00 and 10:00 am after overnight fasting. Height was measured with a stadiometer and total body composition including fat mass in kilograms was assessed with a multifrequency bioelectrical impedance analyzer (InBody 720; Biospace, Seoul, Korea).

**Statistical analysis**

The population characteristic differences within the SR-PA C3 groups were studied using the Kruskal-Wallis test with continuous variables, and cross tabulation and chi-squared test with categorical and ordinal variables. The chi square values and p-values are presented in Table 2. Furthermore, the concurrent validity of the SR-PA L7 and C3 was scrutinized by studying

Spearman correlations between the self-reported and accelerometer-measured leisure time light physical activity, moderate-to-vigorous physical activity and total counts. Additional Pearson and Spearman correlation analysis between SR-PA L7 and C3 with whole day accelerometer-measured physical activity variables and with self-reported physical activity obtained by two independent questionnaires (Kujala, Kaprio, Sarna, & Koskenvuo, 1998; Lakka & Salonen, 1997; Rottensteiner et al., 2015) were carried out and the results are reported in the supplementary file.

Linear regression was utilized for studying the associations of SR-PA scales and ACC-LTPA variables with physical performance measures. The associations were studied by using univariate models separately for each physical activity variable as independent variable and with performance test results as the dependent variable. Additionally, hierarchical linear regression was used to study the additional changes in the coefficient of determination by entering the confounding factors into step 1 and adding the physical activity measures into step 2. The confounding factors were determined based on their significance of the standardized regression coefficient and change in the coefficient of determination. The examined variables were all the background variables shown in Table 2. The ones used in the multivariate models showed significant effect on at least one of the physical performance measures.

Preliminary analyses for multiple linear regression were performed to ensure there was no violation of the assumption of linearity, homoscedasticity, multicollinearity and normality. Initially, correlations were studied, and scatterplots were plotted to examine the linearity. The Durbin-Watson statistic was used to test the independence of observations and the Variance Inflation Factors were studied to test the multicollinearity. Furthermore, P-P plots and histograms were plotted to check that the residuals were approximately normally distributed, and scatterplots were used to test the homoscedasticity.

Level of agreement percentages and Spearman correlations were used to assess the test-retest reliability of the SR-PA L7 and C3. Data analysis was carried out using IBM SPSS Statistics software version 24 (Chicago, IL). The level of significance was set at $p \leq 0.05$.

# Results

## Characteristics of study population

Based on the self-reported physical activity, the participants were considerably active. Majority of the participants (62.5%) reported engaging in brisk or vigorous physical activities several times a week (Table 1). However, 10.8% of the study population reported engaging only in light activities or not moving more than is necessary in daily routines and chores and only 0.3% (2 participants) reported engaging competitive sports. In table 2, characteristics of the study population are presented within the SR-PA C3 groups, because of uneven distribution of participants within SR-PA L7 categories (Table 1). Of the background variables, fat mass, level of education and smoking status were significantly different between the SR-PA C3 groups in the study population. Especially, the fat mass was significantly lower in the groups that reported greater amount of physical activity.

On average, participants recorded over three hours of light physical activity and almost 40 minutes of moderate-to-vigorous physical activity, the rest of the daily leisure time recorded by accelerometer being sedentary time. Participants that reported greater amount of physical activity (high PA group) also recorded more physical activity with accelerometer compared to low and medium PA groups. Medium PA group differed from the low PA group only in accelerometer-measured moderate-to-vigorous physical activity. Furthermore, the table shows that the participants that reported greater amount of physical activity (high PA group) performed significantly better in 6MWT, KE and VJ than participants reporting to be at low and medium PA level. However, GF measurements results showed no significant differences between the SR-PA C3 groups.

**Concurrent validity**

Both SR-PA L7 and C3 correlated with all ACC-LTPA variables (Table 3). SR-PA L7 and C3 had very low correlations with light physical activity but the correlations were relatively higher with moderate to vigorous physical activity ($r_s$ = 0.318 and 0.337, respectively) and total counts ($r_s$ = 0.333 for both). Generally, lower self-reported physical activity was significantly associated with lower total counts and lower amount of ACC-LTPA at all physical activity levels. In addition, there was no systematic difference between SR-PA L7 and C3 associations with ACC-LTPA. Furthermore, two independent self-reported questionnaires demonstrated similar associations with the ACC-LTPA compared to SR-PA L7 and SR-PA C3 (supplementary table 1). SR-PA L7 and C3 were also associated with these independent self-reported questionnaires for LTPA assessment, whole day accelerometer-measured moderate-to-vigorous physical activity and total counts, but not with light physical activity (supplementary table 2).

**Associations between physical activity and physical performance**

Table 4 shows that both SR-PA L7 and C3 were rather weakly associated with 6MWT, KE and VJ results in univariate models (β = 0.134 – 0.265) and in multivariate models with the confounding factors (β = 0.099 – 0.156). Thus, the increase of one standard deviation in all physical activity measures corresponds for only a less than increment of 0.265 standard deviations in these physical performance measures. Additionally, moderate-to-vigorous ACC-LTPA and total counts were associated with 6MWT, KE and VJ results in univariate models (β = 0.125 – 0.285) and in multivariate models with the confounding factors (β = 0.072 – 0.177). However, the light ACC-LTPA was associated only with the 6MWT results in univariate model (β = 0.090). Generally, all physical activity measures had weaker associations with the GF results compared to other physical performance test results. The SR-PA scales and light ACC-LTPA had no associations with the GF results and the associations between the GF results with moderate-to-vigorous ACC-LTPA and total counts were fairly modest both in univariate and multivariate models. The portions of the variance explained by all physical activity measures ($R^2 \leq 0.081$) and the changes in the variance explained after adding them to the model with confounders ($\Delta R^2 \leq 0.030$) were relatively small with all physical performance measures.

Overall. The SR-PA L7 and C3 were slightly associated with 6MWT, KE and VJ results and the associations were little bit stronger compared to light ACC-LTPA. Furthermore, the SR-PA L7 and C3 associations with physical performance measurements had approximately similar magnitude compared to accelerometer-measured moderate-to-vigorous physical activity and total counts. With 6MWT results, moderate-to-vigorous ACC-LTPA and total counts had slightly stronger associations compared to SR-PA scales. On the other hand, SR-PA had stronger associations with the KE and VJ results compared to ACC-LTPA variables. The associations between the GF results and all physical activity variables were relatively low but SR-PA L7 and C3 had weaker associations (non-significantly different from zero) with GF results compared to ACC-LTPA variables.

**Test-retest reliability**

The mean time between the first and the second delivery of the SR-PA L7 was 85.9 ± 11.3 days while maximum duration of the allowed period was 120 days. The correlation coefficients in table 5 indicate acceptable test-retest reliability for SR-PA L7 ($r_s$ = 0.707) and questionable reliability for SR-PA C3 ($r_s$ = 0.622). The values for the level of agreement indicate that 59.9 % (SR-PA L7) and 73.7 % (SR-PA C3) of the participants reported exactly the same physical activity level when administering the question second time.

## Discussion

This paper focused on studying the concurrent validity of a single self-report seven-level scale question (SR-PA L7) and its three-level categorization (SR-PA C3) by examining their association with accelerometer-measured leisure time physical activity (ACC-LTPA). Associations between physical activity assessments and four different physical performance test results were also investigated. Additionally, the test-retest reliability of the SR-PA L7 and C3 was studied by using data from second administration of the questionnaire within four months of the first administration. Notably, the study utilizes the unique approach of focusing on the leisure time activity with both accelerometer and self-reported methods. The results show that the SR-PA L7 and C3 scales are slightly associated with ACC-LTPA and physical performance test results, especially, with measure of cardiorespiratory fitness (6MWT) along with lower extremity strength (KE) and power (VJ). With these physical performance measures, the associations are analogous to the ones between the ACC-LTPA and physical performance test results.

Although the associations between the self-reported and accelerometer-measured LTPA are indisputable within the study population, the strength of the correlations are only weak or moderate ($r_s \leq 0.337$). Furthermore, the correlations between the SR-PA items and ACC-LTPA were weaker with light ACC-LTPA compared to moderate-to-vigorous ACC-LTPA and total counts. In line with our results, analogous results with similar single self-report question for physical activity assessment have been reported with community-dwelling older people (Portegijs et al., 2017) and adults aged 30-69 years (Emaus et al., 2010). The limited overlap between SR-PA scales and ACC-LTPA is potentially due to not fully comparable ability to capture the same parameters of physical activity. It has been shown that the self-reported physical activity questionnaires often fail to capture the light and moderate activities, such as household chores, occupational activities and otherwise spontaneous or incidental movement, but they are able to capture how the individual perceives the physical activity (Prince et al., 2008; Tudor-Locke & Myers, 2001). However, accelerometer objectively measures the movement of the body, and, therefore, is able to capture even very light physical activities. In addition, weaker correlation with lighter physical activity might be due to the SR-PA L7 design in which the categories 0-2 focus on the light activities and categories 3-6 are distinctly more focused on the moderate to vigorous activities, although the subjects might often engage to diverse intensities. Nevertheless, it is likely that the subjects report the highest activity category that corresponds to their general activity level.

In the current study, according to the multiple linear regression analysis, the SR-PA L7 and C3 were associated with 6MWT, KE and VJ. The associations were similar or even stronger than with accelerometer-measured moderate-to-vigorous physical activity and total counts and, therefore, the results show that the self-reported physical activity assessed via SR-PA L7 is associated with performance test results related to cardiorespiratory fitness and lower extremity strength and power results as much as ACC-LTPA. Most of the ACC-LTPA variables were also associated with performance test results in univariate and multivariate models. Especially, total counts and moderate-to-vigorous physical activity were associated with physical performance. The results indicate that more intensive physical activity in addition to overall physical activity provide the most benefits for improving the physical performance level. Generally, these results are consistent with previous studies (Corcoran et al., 2016; Kulinski et al., 2014; Wanderley et al., 2011).

GF was distinctly more weakly associated with the SR-PA scales and ACC-LTPA compared to other performance test results. The SR-PA scales did not have significant associations with the GF results, and of the ACC-LTPA variables, only the total counts were associated with the GF results in the adjusted model. These low associations are related to the fact that, in general, most of the physical activity including walking, running and bicycling, is emphasized for lower extremities (Kuh, Bassey, Butterworth, Hardy, & Wadsworth, 2005). Furthermore, the SR-PA L7 question did not separate strong hand movements from the general physical activity and the hip-attached accelerometers register poorly gym exercises and upper-extremity movement that are associated with grip strength (Prince et al., 2008; Sievänen & Kujala, 2017).

Although self-report questionnaires and accelerometer-measured physical activity are known to be associated with energy expenditure and thus the physical activity level (Conway, Seale, Jacobs, Irwin, & Ainsworth, 2002; Plasqui & Westerterp, 2007; Schuit, Schouten, Westerterp, & Saris, 1997), the associations of SR-PA scales and ACC-LTPA variables with physical performance test results in the current study are somewhat low in both univariate ($\beta$ = 0.015 – 0.285) and multivariate models ($\beta$ = 0.024 – 0.177). This might be due to the fact that there are several other factors that affect the physical performance level in addition to physical activity. These factors include e.g. age, sex and anthropometrics (Jackson, Sui, Hébert, Church, & Blair, 2009; Samson et al., 2000; Tseng et al., 2014) in addition to genetics (Bouchard, Dionne, Simoneau, & Boulay, 1992), psychological factors (Lord, Murray, Chapman, Munro, & Tiedemann, 2002) and socioeconomic status (Shishehbor, Litaker, Pothier, & Lauer, 2006). Additionally, some covariates, such as fat mass, might be moderating variables thus affecting associations between the physical activity variables and physical performance measures. In this study, all participants were females, but due to the practicalities, only the effect of the age, anthropometrics and some socioeconomic factors could be controlled for.

The reliability analysis indicated acceptable repeatability for SR-PA L7 ($r_s$ = 0.707) and questionable repeatability C3 ($r_s$ = 0.622) within the four-month test-retest period. However, the level of agreement is somewhat poor varying only from 59.9 to 73.7 %. The observed correlations are similar or even slightly stronger compared other studies that have investigated the repeatability of a single-item physical activity assessment tools using correlation coefficients (Rantanen et al., 1997; Sihvonen, Rantanen, & Heikkinen, 1998). Relatively long period between administrating the questionnaires (85.9 ± 11.3 days) in the current study increases the chance that true changes have occurred in the physical activity between the administrations, which may have caused a decrease in the observed repeatability. However, the longer period between the administrations brings out the incidental variation in the responses by decreasing the change of choosing the same alternative by default when re-administrating the questionnaire.

The data used in the research were derived from the large study that focused on the middle-aged women undergoing their menopausal transition. Since the study population consisted of healthy women in a narrow age range, the study groups were quite homogenous. Therefore, one can assume that the results can be generalized for populations comprising of Nordic and other mainly Caucasian middle-aged females. However, because of the homogenous sample, the results might not be generalizable to populations with males, different age groups and generally to populations that are more heterogeneous. In the regression models used in the study, the effect of variables related to life habits, socioeconomic factors and anthropometrics was controlled. However, some of these variables were defined using a straightforward two- or three-level categorization from a single question in a baseline questionnaire to simplify the answers and to ensure the linearity. Therefore, the used variables were very rough categorizations.

## Conclusions

With the limitations considered in the discussion in mind, the results show that SR-PA L7 and SR-PA C3 have limited concurrent validity for assessment of physical activity compared to ACC-LTPA in Caucasian middle-aged women. However, SR-PA L7 and C3 demonstrated analogous associations with accelerometer-measured physical activity compared to other widely used self-report tools for physical activity assessment. The self-report items were associated with the physical performance results but they only explained a very modest, although similar to ACC-LTPA, amount of the variance in the physical performance test results. Thus, they are not applicable for assessing the physical performance. Furthermore, the SR-PA items demonstrated reasonable repeatability especially with SR-PA L7, the repeatability being slightly weaker with SR-PA C3. Yet, more research is needed for validation of the used seven-level scale question for more heterogeneous populations comprising of different age or ethnic groups along with both males and females.

## Acknowledgements

This work was supported by the Academy of Finland (X.X.: Grant nro 275323, X.X.: grant nro 309504), PANINI (X.X: Horizon 2020 – the framework Programme for Research and

Innovation, Marie Sklodowska-Curie Actions, ITN, ref 15-0667), and the Juho Vainio Foundation (X.X.X.). We are grateful to all the ERMA participants who contributed their time and information. In addition, we thank the staff members at the XXX and at the XXX who helped with the study.

## Competing interests

The authors declare that they have no financial or other competing interests.

## References

Bondarev, D., Laakkonen, E. K., Finni, T., Kokko, K., Kujala, U. M., Aukee, P., . . . Sipilä, S. (2018). Physical performance in relation to menopause status and physical activity. *Menopause (New York, N.Y.), 25*(12), 1432-1441. doi:10.1097/GME.0000000000001137

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise, 14*(5), 377-381.

Bosco, C., Luhtanen, P., & Komi, P. V. (1983). A simple method for measurement of mechanical power in jumping. *European Journal of Applied Physiology and Occupational Physiology, 50*(2), 273-282.

Bouchard, C., Dionne, F. T., Simoneau, J. A., & Boulay, M. R. (1992). Genetics of aerobic and anaerobic performances. *Exercise and Sport Sciences Reviews, 20*, 27-58.

Brakenridge, C. L., Fjeldsoe, B. S., Young, D. C., Winkler, E. A., Dunstan, D. W., Straker, L. M., & Healy, G. N. (2016). Evaluating the effectiveness of organisational-level strategies with or without an activity tracker to reduce office workers' sitting time: A cluster-randomised trial. *The International Journal of Behavioral Nutrition and Physical Activity, 13*(1), 3. doi:10.1186/s12966-016-0441-3 [doi]

Cesari, M., Kritchevsky, S. B., Newman, A. B., Simonsick, E. M., Harris, T. B., Penninx, B. W., . . . Health, Aging and Body Composition Study. (2009). Added value of physical performance measures in predicting adverse health-related events: Results from the health, aging and body composition study. *Journal of the American Geriatrics Society, 57*(2), 251-259. doi:10.1111/j.1532-5415.2008.02126.x [doi]

Colpani, V., Baena, C., Jaspers, L., van Dijk, G., Farajzadegan, Z., Dhana, K., . . . Franco, O. (2018). Lifestyle factors, cardiovascular disease and all-cause mortality in middle-aged and elderly women: A systematic review and meta-analysis. *European Journal of Epidemiology, 33*(9), 831-845. doi:10.1007/s10654-018-0374-z

Conway, J. M., Seale, J. L., Jacobs, D. R., Irwin, M. L., & Ainsworth, B. E. (2002). Comparison of energy expenditure estimates from doubly labeled water, a physical activity questionnaire, and physical activity records. *The American Journal of Clinical Nutrition, 75*(3), 519-525. doi:10.1093/ajcn/75.3.519 [doi]

Corcoran, M. P., Chui, K. K., White, D. K., Reid, K. F., Kirn, D., Nelson, M. E., . . . Fielding, R. A. (2016). Accelerometer assessment of physical activity and its association with physical function in older adults residing at assisted care facilities. *The Journal of Nutrition, Health & Aging, 20*(7), 752-758. doi:10.1007/s12603-015-0640-7 [doi]

Dishman, R. K., Washburn, R. A., & Schoeller, D. A. (2001). Measurement of physical activity. *Quest, 53*(3), 295-309. doi:10.1080/00336297.2001.10491746

Emaus, A., Degerstrøm, J., Wilsgaard, T., Hansen, B. H., Dieli-Conwright, C. M., Furberg, A., . . . Thune, I. (2010). Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? results from the tromsø study. *Scandinavian Journal of Public Health, 38*(5_suppl), 105-118. doi:10.1177/1403494810378919

Enright, P. L. (2003). The six-minute walk test. *Respiratory Care, 48*(8), 783.

Frändin, K., & Grimby, G. (1994). Assessment of physical activity, fitness and performance in 76-year-olds. *Scandinavian Journal of Medicine & Science in Sports, 4*(1), 41-46. doi:10.1111/j.1600-0838.1994.tb00404.x

Gale, C. R., Martyn, C. N., Cooper, C., & Sayer, A. A. (2007). Grip strength, body composition, and mortality. *International Journal of Epidemiology, 36*(1), 228-235. doi:10.1093/ije/dyl224

Gill, D. P., Jones, G. R., Zou, G., & Speechley, M. (2012). Using a single question to assess physical activity in older adults: A reliability and validity study. *BMC Medical Research Methodology, 12*(1), 20. doi:10.1186/1471-2288-12-20

Grimby, G., & Frändin, K. (2018). On the use of a six-level scale for physical activity. *Scandinavian Journal of Medicine & Science in Sports, 28*(3), 819-825. doi:10.1111/sms.12991

Hirvensalo, M., Lampinen, P., & Rantanen, T. (1998). Physical exercise in old age: An eight-year follow-up study on involvement, motives, and obstacles among persons age 65-84. *Journal of Aging and Physical Activity, 6*(2), 157-168.

Holtermann, A., Krause, N., van der Beek, Allard J, & Straker, L. (2018). The physical activity paradox: Six reasons why occupational physical activity (OPA) does not confer the cardiovascular health benefits that leisure time physical activity does. *British Journal of Sports Medicine, 52*(3), 149-150. doi:10.1136/bjsports-2017-097965

Jackson, A. S., Sui, X., Hébert, J. R., Church, T. S., & Blair, S. N. (2009). Role of lifestyle and aging on the longitudinal change in cardiorespiratory fitness. *Archives of Internal Medicine, 169*(19), 1781-1787. doi:10.1001/archinternmed.2009.312

Kovanen, V., Aukee, P., Kokko, K., Finni, T., Tarkka, I. M., Tammelin, T., . . . Laakkonen, E. K. (2018). Design and protocol of estrogenic regulation of muscle apoptosis (ERMA) study with 47 to 55-year-old women's cohort: Novel results show menopause-related differences in blood count. *Menopause (New York, N.Y.), 25*(9), 1020-1032. doi:10.1097/GME.0000000000001117

Kowalski, K., Rhodes, R., Naylor, P., Tuokko, H., & MacDonald, S. (2012). Direct and indirect measurement of physical activity in older adults: A systematic review of the literature. *The International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 148. doi:10.1186/1479-5868-9-148

Kuh, D., Bassey, E. J., Butterworth, S., Hardy, R., & Wadsworth, M. E. J. (2005). Grip strength, postural control, and functional leg power in a representative cohort of british men and women: Associations with physical activity, health status, and socioeconomic conditions. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 60*(2), 224-231. doi:10.1093/gerona/60.2.224

Kujala, U. M., Kaprio, J., Sarna, S., & Koskenvuo, M. (1998). Relationship of leisure-time physical activity and mortality: The finnish twin cohort. *Jama, 279*(6), 440-444. doi:10.1001/jama.279.6.440

Kulinski, J. P., Khera, A., Ayers, C. R., Das, S. R., de Lemos, J. A., Blair, S. N., & Berry, J. D. (2014). Association between cardiorespiratory fitness and accelerometer-derived physical activity and sedentary time in the general population. *Mayo Clinic Proceedings, 89*(8), 1063-1071. doi:10.1016/j.mayocp.2014.04.019

Laakkonen, E. K., Kulmala, J., Aukee, P., Hakonen, H., Kujala, U. M., Lowe, D. A., . . . Sipilä, S. (2017). Female reproductive factors are associated with objectively measured physical activity in middle-aged women. *PLoS One, 12*(2) doi:10.1371/journal.pone.0172054

Laakkonen, E. K., Soliymani, R., Karvinen, S., Kaprio, J., Kujala, U. M., Baumann, M., . . . Lalowski, M. (2017). Estrogenic regulation of skeletal muscle proteome: A study of premenopausal women and postmenopausal MZ cotwins discordant for hormonal therapy. *Aging Cell, 16*(6), 1276-1287. doi:10.1111/acel.12661

Lakka, T. A., & Salonen, J. T. (1997). The physical activity questionnaires of the kuopio ischemic heart disease study (KIHD): A collection of physical activity questionnaires for health related research. *Medicine & Science in Sports & Exercise, 29*, S58.

Lord, S. R., Murray, S. M., Chapman, K., Munro, B., & Tiedemann, A. (2002). Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 57*(8), M543. doi:10.1093/gerona/57.8.M539

Matthews, C. E., Hagströmer, M., Pober, D. M., & Bowles, H. R. (2012). Best practices for using physical activity monitors in population-based research. *Medicine & Science in Sports & Exercise, 44*(1 Suppl 1), S76. doi:10.1249/MSS.0b013e3182399e5b

Migueles, J., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., . . . Ortega, F. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine, 47*(9), 1821-1845. doi:10.1007/s40279-017-0716-0

Ndahimana, D., & Kim, E. (2017). Measurement methods for physical activity and energy expenditure: A review. *Clinical Nutrition Research, 6*(2), 68-80. doi:10.7762/cnr.2017.6.2.68

Nelson, M. E., Layne, J. E., Bernstein, M. J., Nuernberger, A., Castaneda, C., Kaliton, D., . . . Fiatarone Singh, M. A. (2004). The effects of multidimensional home-based exercise on functional performance in elderly people. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 59*(2), M160. doi:10.1093/gerona/59.2.M154

Newman, A. B., Kupelian, V., Visser, M., Simonsick, E. M., Goodpaster, B. H., Kritchevsky, S. B., . . . Harris, T. B. (2006). Strength, but not muscle mass, is associated with mortality in the health, aging and body composition study cohort. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 61*(1), 72-77. doi:10.1093/gerona/61.1.72

Newman, A. B., Simonsick, E. M., Naydeck, B. L., Boudreau, R. M., Kritchevsky, S. B., Nevitt, M. C., . . . Harris, T. B. (2006). Association of long-distance corridor walk performance with mortality, cardiovascular disease, mobility limitation, and disability. *Jama, 295*(17), 2018-2026. doi:10.1001/jama.295.17.2018

Pahor, M., Blair, S. N., Espeland, M., Fielding, R., Gill, T. M., Guralnik, J. M., . . . Studenski, S. (2006). Effects of a physical activity intervention on measures of physical performance: Results of the lifestyle interventions and independence for elders pilot (LIFE-P) study. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 61*(11), 1157-1165. doi:10.1093/gerona/61.11.1157

Plasqui, G., & Westerterp, K. R. (2007). Physical activity assessment with accelerometers: An evaluation against doubly labeled water. *Obesity, 15*(10), 2371-2379. doi:10.1038/oby.2007.281

Portegijs, E., Sipilä, S., Viljanen, A., Rantakokko, M., & Rantanen, T. (2017). Validity of a single question to assess habitual physical activity of community-dwelling older people. *Scandinavian Journal of Medicine & Science in Sports, 27*(11), 1423-1430. doi:10.1111/sms.12782

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *The International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 56. doi:10.1186/1479-5868-5-56

Prince, S. A., Elliott, C. G., Scott, K., Visintini, S., & Reed, J. L. (2019). Device-measured physical activity, sedentary behaviour and cardiometabolic health and fitness across occupational groups: A systematic review and meta-analysis. *The International Journal of Behavioral Nutrition and Physical Activity, 16*(1), 30.

Rantanen, T., Era, P., & Heikkinen, E. (1997). Physical activity and the changes in maximal isometric strength in men and women from the age of 75 to 80 years. *Journal of the American Geriatrics Society, 45*(12), 1439-1445. doi:10.1111/j.1532-5415.1997.tb03193.x

Rantanen, T., Guralnik, J. M., Foley, D., Masaki, K., Leveille, S., Curb, J. D., & White, L. (1999). Midlife hand grip strength as a predictor of old age disability. *Jama, 281*(6), 558-560. doi:10.1001/jama.281.6.558

Reid, K. F., & Fielding, R. A. (2012). Skeletal muscle power: A critical determinant of physical functioning in older adults. *Exercise and Sport Sciences Reviews, 40*(1), 4-12. doi:10.1097/JES.0b013e31823b5f13

Ronkainen, P. H. A., Pöllänen, E., Törmäkangas, T., Tiainen, K., Koskenvuo, M., Kaprio, J., . . . Kovanen, V. (2008). Catechol-O-methyltransferase gene polymorphism is associated with skeletal muscle properties in older women alone and together with physical activity. *PLoS One, 3*(3), e1819. doi:10.1371/journal.pone.0001819

Ronkainen, P. H. A., Kovanen, V., Alén, M., Pöllänen, E., Palonen, E., Ankarberg-Lindgren, C., . . . Sipilä, S. (2009). Postmenopausal hormone replacement therapy modifies skeletal muscle composition and function: A study with monozygotic twin pairs. *Journal of Applied Physiology, 107*(1), 25-33. doi:10.1152/japplphysiol.91518.2008

Rottensteiner, M., Leskinen, T., Niskanen, E., Aaltonen, S., Mutikainen, S., Wikgren, J., . . . Kujala, U. (2015). Physical activity, fitness, glucose homeostasis, and brain morphology in twins. *Medicine & Science in Sports & Exercise, 47*(3), 509-518. doi:10.1249/MSS.0000000000000437

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*(6), 1104-1115. doi:10.1161/01.CIR.38.6.1104

Samson, M. M., Meeuwsen, I. B., Crowe, A., Dessens, J. A., Duursma, S.A., Verhaar, H.J. (2000). Relationships between physical performance measures, age, height and body weight in healthy adults. *Age and Ageing, 29*(3), 235-242. doi:10.1093/ageing/29.3.235

Sasaki, H., Kasagi, F., Yamada, M., & Fujita, S. (2007). Grip strength predicts cause-specific mortality in middle-aged and elderly persons. *The American Journal of Medicine, 120*(4), 337-342. doi:S0002-9343(06)00447-5 [pii]

Sasaki, J. E., Dinesh, J., & Freedson, P. S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport, 14*(5), 411-416. doi:10.1016/j.jsams.2011.04.003

Schuit, A. J., Schouten, E. G., Westerterp, K. R., & Saris, W. H. M. (1997). Validity of the physical activity scale for the elderly (PASE): According to energy expenditure assessed by the doubly labeled water method. *Journal of Clinical Epidemiology, 50*(5), 541-546. doi:10.1016/S0895-4356(97)00010-3

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires commentary. *British Journal of Sports Medicine, 37*(3), 197-206. doi:10.1136/bjsm.37.3.197

Shishehbor, M. H., Litaker, D., Pothier, C. E., & Lauer, M. S. (2006). Association of socioeconomic status with functional capacity, heart rate recovery, and all-cause mortality. *Jama, 295*(7), 784-792. doi:10.1001/jama.295.7.784

Sievänen, H., & Kujala, U. M. (2017). Accelerometry—Simple, but challenging. *Scandinavian Journal of Medicine & Science in Sports, 27*(6), 574-578. doi:10.1111/sms.12887

Sihvonen, S., Rantanen, T., & Heikkinen, E. (1998). Physical activity and survival in elderly people: A five-year follow-up study. *Journal of Aging and Physical Activity, 6*(2), 133-140. doi:10.1123/japa.6.2.133

Sillanpää, E., Niskala, P., Laakkonen, E. K., Ponsot, E., Alén, M., Kaprio, J., . . . Sipilä, S. (2017). Leukocyte and skeletal muscle telomere length and body composition in monozygotic twin pairs discordant for long-term hormone replacement therapy. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies, 20*(2), 119-131. doi:10.1017/thg.2017.1

Sipilä, S., Taaffe, D. R., Cheng, S., Puolakka, J., Toivanen, J., & Suominen, H. (2001). Effects of hormone replacement therapy and high-impact physical exercise on skeletal muscle in post-menopausal women: A randomized placebo-controlled study. *Clinical Science, 101*(2), 147-157. doi:10.1042/CS20000271

Smith, B. J., Marshall, A. L., & Huang, N. (2005). Screening for physical activity in family practice: Evaluation of two brief assessment tools. *American Journal of Preventive Medicine, 29*(4), 256.

Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., . . . Guralnik, J. (2011). Gait speed and survival in older adults. *Jama, 305*(1), 50-58. doi:10.1001/jama.2010.1923

Sylvia, L. G., Bernstein, E. E., Hubbard, J. L., Keating, L., & Anderson, E. J. (2014). Practical guide to measuring physical activity. *Journal of the Academy of Nutrition and Dietetics, 114*(2), 199.

Tseng, L. A., Delmonico, M. J., Visser, M., Boueau, R. M., Goodpaster, B. H., Schwartz, A. V., . . . Newman, A. B. (2014). Body composition explains sex differential in physical performance among older adults. *Journals of Gerontology. Series A: Biological Sciences & Medical Sciences, 69*(1), 93-100. doi:10.1093/gerona/glt027

Tseng, L. A., El Khoudary, S. R., Young, E. A., Farhat, G. N., Sowers, M., Sutton-Tyrrell, K., & Newman, A. B. (2012). The association of menopausal status with physical function: The study of women's health across the nation (SWAN): Menopausal status and physical function. *Menopause (New York, NY), 19*(11), 1186.

Tudor-Locke, C. E., & Myers, A. M. (2001). *Challenges and opportunities for measuring physical activity in sedentary adults*. Cham: Adis International. doi:10.2165/00007256-200131020-00002

Wanderley, F. A. C., Silva, G., Marques, E., Oliveira, J., Mota, J., & Carvalho, J. (2011). Associations between objectively assessed physical activity levels and fitness and self-reported health-related quality of life in community-dwelling older adults. *Quality of Life Research, 20*(9), 1371-1378. doi:10.1007/s11136-011-9875-x

TABLE 1: The seven-level scale question for assessing the physical activity (SR-PA L7), its three-level categorization (SR-PA C3) and the frequencies of responders in the full sample with self-report and accelerometer-measured physical activity (n = 795).

| **Which of the following descriptions best corresponds to your physical activity at the moment?** | | **Frequencies % (n)** | |
|---|---|---|---|
| | | SR-PA L7 | SR-PA C3 |
| 1 | I do not move more than is necessary in my daily routines/chores. | 2.9 (23) | Low PA 10.8 (86) |
| 2 | I go for casual walks and engage in light outdoor recreation 1-2 times a week. | 7.9 (63) | |
| 3 | I go for casual walks and engage in light outdoor recreation several times a week | 6.9 (55) | Medium PA 26.7 (212) |
| 4 | I engage, 1–2 times a week, in brisk physical activity (e.g. yard work, walking, cycling) to the point of perspiring and some degree of breathlessness. | 19.7 (157) | |
| 5 | Several times a week (3–5), I engage in brisk physical activity (e.g. yard work, walking, cycling) to the point of perspiring and some degree of breathlessness. | 43.0 (342) | High PA 62.5 (497) |
| 6 | I do keep-fit exercises several times a week in a way that causes rather strong shortness of breath and sweating during the activity. | 19.2 (153) | |
| 7 | I participate in competitive sports and maintain my fitness through regular training | 0.3 (2) | |

Horizontal lines represent the cut-off points used for SR-PA C3 variable categorizing study participants into groups indicating low, medium and high levels of physical activity.

TABLE 2: Background variables, accelerometer-measured physical activity and physical performance of the study population within the three groups categorized based on the self-reported three-level scale physical activity question (SR-PA C3).

| | **Full Sample** (n = 795) | **Low PA** (n = 86) | **Medium PA** (n = 212) | **High PA** (n = 497) | **Test statistics** |
|---|---|---|---|---|---|
| **Background variables** | | | | | |
| **Age** [years]§ | 51.71 (1.91) | 51.73 (1.85) | 51.77 (1.90) | 51.68 (1.92) | $\chi^2$=0.261 p=0.878 |
| **Height** [cm]§ (missing n = 4) | 165.58 (5.67) | 165.08 (6.70) | 165.81 (5.65) | 165.57 (5.50) | $\chi^2$=0.512 p=0.774 |
| **Fat mass** [kg]§ | 22.00 (8.17) | 25.18 (7.78) | 23.72 (7.98) | 20.74 (8.04) | $\chi^2$=37.712 **p<0.001** |
| **Menopausal status** [%(n)]£ | | | | | $\chi^2$=2.923 p=0.818 |
| Pre | 25.9 (206) | 22.1 (19) | 26.4 (56) | 26.4 (131) | |
| Peri1 | 19.4 (154) | 24.4 (21) | 18.9 (40) | 18.7 (93) | |
| Peri2 | 20.5 (163) | 16.3 (14) | 21.1 (45) | 20.9 (104) | |
| Post | 34.2 (272) | 37.2 (32) | 33.5 (71) | 34.0 (169) | |
| **Alcohol consumption** [AU/week] § | 3.88 (3.83) | 3.73 (4.17) | 4.03 (3.90) | 3.84 (3.74) | $\chi^2$=2.056 p=0.358 |
| **Level of education** [%(n)]£ | | | | | $\chi^2$=15.084 **p=0.005** |
| Primary | 2.3 (18) | 7.0 (6) | 1.9 (4) | 1.6 (8) | |
| Secondary | 57.7 (459) | 66.3 (57) | 57.5 (122) | 56.3 (280) | |
| Tertiary | 40.0 (318) | 26.7 (23) | 40.6 (86) | 42.1 (209) | |
| **Marital status** [%(n)]£ (missing n = 1) | | | | | $\chi^2$=5.491 p=0.064 |
| Single | 24.9 (198) | 34.9 (30) | 22.2 (47) | 24.4 (121) | |
| In relationship | 75.1 (596) | 65.1 (56) | 77.8 (165) | 75.6 (375) | |
| **Smoking status** [%(n)]£ (missing n = 3) | | | | | $\chi^2$=10.843 **p=0.028** |
| Non-smoker | 66.9 (530) | 65.1 (56) | 68.4 (145) | 66.6 (329) | |
| Former smoker | 25.8 (204) | 23.3 (20) | 21.2 (45) | 28.1 (139) | |
| Current smoker | 7.3 (58) | 11.6 (10) | 10.4 (22) | 5.3 (26) | |
| **Work status** [%(n)]£ (missing n = 29) | | | | | $\chi^2$=8.573 p=0.073 |
| Employed | 91.5 (701) | 88.9 (72) | 95.1 (193) | 90.5 (436) | |
| Not regularly employed | 8.0 (61) | 9.9 (8) | 3.9 (8) | 9.3 (45) | |
| Retired | 0.5 (4) | 1.2 (1) | 1.9 (2) | 0.2 (1) | |
| **Work physical load**£ (missing n = 55) | | | | | $\chi^2$=1.672 p=0.947 |
| Sedentary | 52.0 (385) | 55.1 (43) | 51.5 (104) | 51.7 (238) | |
| Light | 20.5 (152) | 15.4 (12) | 21.3 (43) | 21.1 (97) | |
| Moderate | 25.7 (190) | 28.2 (22) | 25.2 (51) | 25.5 (117) | |
| Heavy | 1.8 (13) | 1.3 (1) | 2.0 (4) | 1.7 (8) | |
| **Accelerometer-measured leisure time physical activity** | | | | | |
| **Light PA** [min]§ | 196.44 (42.71) | 195.19 (49.74) | 187.73 (44.31) | 200.37 (41.16) | $\chi^2$=13.423 **p=0.001** |

| | | | | | |
|---|---|---|---|---|---|
| **MVPA** [min][§] | 38.55 (20.22) | 26.07 (15.45) | 32.01 (17.00) | 43.50 (20.52) | $\chi^2$=90.855 **p<0.001** |
| **Total counts** [counts*$10^5$][§] | 4.34 (1.29) | 3.64 (1.08) | 3.85 (1.11) | 4.67 (1.28) | $\chi^2$=88.304 **p<0.001** |
| **Physical performance** | | | | | |
| **6MWT** [m][§] (missing n = 62) | 668.26 (60.92) | 639.30 (59.23) | 653.09 (55.06) | 679.09 (60.92) | $\chi^2$=44.085 **p<0.001** |
| **KE** [N][§] (missing n = 100) | 461.88 (95.32) | 443.02 (98.27) | 443.08 (96.73) | 472.26 (92.32) | $\chi^2$=17.189 **p<0.001** |
| **VJ** [cm][§] (missing n = 56) | 19.16 (4.25) | 17.46 (4.15) | 18.16 (3.74) | 19.84 (4.31) | $\chi^2$=37.728 **p<0.001** |
| **GF** [F][§] (missing n = 7) | 312.93 (58.96) | 304.57 (68.31) | 311.45 (52.76) | 315.00 (59.74) | $\chi^2$=1.254 p=0.534 |

[§] Characteristics are illustrated as mean (standard deviation) and the differences between the groups are tested with Kruskal-Wallis test, [£] Characteristics are illustrated as % (n) and the differences between the groups are tested with cross tabulation and chi-squared test, statistically significant differences between the groups are highlighted by bolding, MVPA = moderate-to-vigorous physical activity, 6MWT = six-minute walk test, KE = maximal isometric knee extension force measurement, VJ = vertical jump measurement, GF = grip force measurement
Missing values respectively within SR-PA3 groups are Height: 1, 1, 2; Marital status: 0, 0, 1; Smoking status: 0, 0, 3; Work status: 5, 9, 15; Work physical load: 8, 10, 37; 6MWT: 10, 23, 29; KE: 15, 36, 49; VJ: 10, 21, 25;
GF: 2, 1, 4.

Table 3.

Spearman correlations for the seven-level scale question (SR-PA L7) and its three-level categorization (SR-PA C3) with accelerometer-measured physical activity (ACC-LTPA) (N = 795).

| | **SR-PA L7** | | | **SR-PA C3** | | |
|---|---|---|---|---|---|---|
| | **$r_s$** | **95% CI [lower bound, upper bound]** | **p-value** | **$r_s$** | **95% CI [lower bound, upper bound]** | **p-value** |
| **Light PA** | 0.105 | [0.035, 0.174] | 0.003 | 0.109 | [0.040, 0.179] | 0.001 |
| **MVPA** | 0.318 | [0.252, 0.384] | <0.001 | 0.337 | [0.271, 0.403] | <0.001 |
| **Total counts** | 0.333 | [0.267, 0.399] | <0.001 | 0.333 | [0.267, 0.399] | <0.001 |

MVPA = moderate-to-vigorous physical activity, 95% CI = 95% confidence interval, p-values show if the correlation is different from zero.

Table 4.

Single self-reported (SR-PA) and accelerometer-measured leisure time physical activity (ACC-LTPA) variable regression model and hierarchical regression models with physical performance test results as the dependent variable.

| | **6MWT [m]** | | **KE [F]** | | **VJ [m]** | | **GF [F]** | |
|---|---|---|---|---|---|---|---|---|
| | **Univariate model**<br>(n = 733) | **Multivariate model**[†]<br>(n = 725) | **Univariate model**<br>(n = 695) | **Multivariate model**[†]<br>(n = 687) | **Univariate model**<br>(n = 739) | **Multivariate model**[†]<br>(n = 731) | **Univariate model**<br>(n = 788) | **Multivariate model**[†]<br>(n = 780) |
| **Model with confounders** | | $R^2 = 0.301$ | | $R^2 = 0.058$ | | $R^2 = 0.341$ | | $R^2 = 0.080$ |
| **SR-PA L7** | $R^2 = 0.064$<br>β = 0.252<br>**p < 0.001** | $\Delta R^2 = 0.018$<br>β = 0.140<br>**p < 0.001** | $R^2 = 0.023$<br>β = 0.153<br>**p = 0.001** | $\Delta R^2 = 0.023$<br>β = 0.156<br>**p < 0.001** | $R^2 = 0.070$<br>β = 0.265<br>**p < 0.001** | $\Delta R^2 = 0.016$<br>β = 0.133<br>**p < 0.001** | $R^2 = 0.002$<br>β = 0.044<br>p = 0.216 | $\Delta R^2 = 0.001$<br>β = 0.031<br>p = 0.382 |
| **SR-PA C3** | $R^2 = 0.058$<br>β = 0.241<br>**p < 0.001** | $\Delta R^2 = 0.019$<br>β = 0.141<br>**p < 0.001** | $R^2 = 0.018$<br>β = 0.134<br>**p = 0.001** | $\Delta R^2 = 0.018$<br>β = 0.137<br>**p < 0.001** | $R^2 = 0.045$<br>β = 0.213<br>**p = 0.001** | $\Delta R^2 = 0.009$<br>β = 0.099<br>**p = 0.002** | $R^2 = 0.003$<br>β = 0.054<br>p = 0.127 | $\Delta R^2 = 0.002$<br>β = 0.048<br>p = 0.176 |
| **ACC-LTPA, Light** | $R^2 = 0.008$<br>β = 0.090<br>**p = 0.015** | $\Delta R^2 = 0.002$<br>β = 0.046<br>p = 0.150 | $R^2 = 0.000$<br>β = 0.015<br>p = 0.688 | $\Delta R^2 = 0.001$<br>β = 0.033<br>p = 0.382 | $R^2 = 0.005$<br>β = 0.070<br>p = 0.057 | $\Delta R^2 = 0.001$<br>β = 0.024<br>p = 0.439 | $R^2 = 0.003$<br>β = 0.051<br>p = 0.156 | $\Delta R^2 = 0.003$<br>β = 0.055<br>p = 0.116 |
| **ACC-LTPA, MVPA** | $R^2 = 0.069$<br>β = 0.263<br>**p < 0.001** | $\Delta R^2 = 0.025$<br>β = 0.163<br>**p < 0.001** | $R^2 = 0.016$<br>β = 0.125<br>**p = 0.001** | $\Delta R^2 = 0.013$<br>β = 0.117<br>**p = 0.002** | $R^2 = 0.028$<br>β = 0.168<br>**p < 0.001** | $\Delta R^2 = 0.005$<br>β = 0.072<br>**p = 0.021** | $R^2 = 0.006$<br>β = 0.078<br>**p = 0.029** | $\Delta R^2 = 0.004$<br>β = 0.063<br>p = 0.072 |
| **ACC-LTPA, Total counts** | $R^2 = 0.081$<br>β = 0.285<br>**p < 0.001** | $\Delta R^2 = 0.030$<br>β = 0.177<br>**p < 0.001** | $R^2 = 0.022$<br>β = 0.148<br>**p < 0.001** | $\Delta R^2 = 0.021$<br>β = 0.149<br>**p < 0.001** | $R^2 = 0.041$<br>β = 0.201<br>**p < 0.001** | $\Delta R^2 = 0.008$<br>β = 0.094<br>**p = 0.002** | $R^2 = 0.010$<br>β = 0.101<br>**p = 0.005** | $\Delta R^2 = 0.008$<br>β = 0.090<br>**p = 0.011** |

6MWT = six-minute walk test, KE = maximal isometric knee extension force measurement, VJ = vertical jump measurement, GF = grip force measurement, [†] Model is adjusted with age, height, fat mass, menopausal status, education level and smoking status, $R^2$ = model's coefficient of determination, $\Delta R^2$ = change in the coefficient of determination after including the physical activity variable in question, β = standardized regression coefficient related to physical activity variable in question, MVPA = moderate-to-vigorous physical activity. Statistically significant associations, which are different from zero, are highlighted by bolding.

Table 5.

Test-retest correlation and level of agreement for the single seven-level scale question (SR-PA L7) and its three-level categorization (SR-PA C3) (N = 152).

| | Level of agreement [%] | $r_s$ | 95% CI [lower bound, upper bound] | p-value |
|---|---|---|---|---|
| **SR-PA L7** | 59.9 | 0.707 | [0.593, 0.821] | <0.001 |
| **SR-PA C3** | 73.7 | 0.622 | [0.496, 0.749] | <0.001 |

95% CI = 95% confidence interval. Significant p-values means different from zero.

# Supplementary files

SUPPLEMENTARY TABLE 1: Pearson correlations for the physical activity questionnaire (Kujala et al. 1998) (N = 792) and 12-month physical activity questionnaire modified from Kuopio Ischemic Heart Disease Risk Factor Study (Lakka & Salonen, 1997; Rottensteiner et al., 2015) (N = 793) with accelerometer-measured leisure time and whole day physical activity.

| | **physical activity questionnaire† [n=792]** | | | **12-month physical activity questionnaire† [n=793]** | | |
|---|---|---|---|---|---|---|
| | **r** | **95% CI [lower bound, upper bound]** | **p-value** | **r** | **95% CI [lower bound, upper bound]** | **p-value** |
| **Accelerometer-measured leisure time physical activity** | | | | | | |
| **Light PA** | 0.076 | [0.007, 0.146] | **0.032** | 0.083 | [0.014, 0.153] | **0.019** |
| **MVPA** | 0.354 | [0.288, 0.419] | **<0.001** | 0.360 | [0.295, 0.426] | **<0.001** |
| **Total counts** | 0.349 | [0.284, 0.415] | **<0.001** | 0.339 | [0.274, 0.405] | **<0.001** |
| **Accelerometer-measured whole day physical activity** | | | | | | |
| **Light PA** | 0.029 | [-0.041, 0.099] | 0.419 | 0.085 | [0.016, 0.155] | **0.016** |
| **MVPA** | 0.325 | [0.259, 0.391] | **<0.001** | 0.356 | [0.292, 0.422] | **<0.001** |
| **Total counts** | 0.287 | [0.220, 0.354] | **<0.001** | 0.334 | [0.268, 0.400] | **<0.001** |

MVPA = moderate-to-vigorous physical activity, 95% CI = 95% confidence interval, † MET hours per day
Statistically significant correlations are highlighted

SUPPLEMENTARY TABLE 2: Spearman correlations for the seven-level scale question (SR-PA L7) and its three-level categorization (SR-PA C3) with accelerometer-measured whole day physical activity (N = 795), physical activity questionnaire (Kujala et al. 1998) (N = 792) and 12-month physical activity questionnaire modified from Kuopio Ischemic Heart Disease Risk Factor Study (Lakka & Salonen, 1997; Rottensteiner et al., 2015) (N = 793).

| | **SR-PA L7** | | | **SR-PA C3** | | |
|---|---|---|---|---|---|---|
| | **$r_s$** | **95% CI [lower bound, upper bound]** | **p-value** | **$r_s$** | **95% CI [lower bound, upper bound]** | **p-value** |
| **Accelerometer-measured whole day physical activity** | | | | | | |
| **Light PA** | 0.042 | [-0.028, 0.112] | 0.235 | 0.067 | [-0.003, 0.137] | 0.059 |
| **MVPA** | 0.280 | [0.214, 0.347] | **<0.001** | 0.296 | [0.229, 0.362] | **<0.001** |
| **Total counts** | 0.248 | [0.180, 0.315] | **<0.001** | 0.259 | [0.192, 0.327] | **<0.001** |
| **Physical activity questionnaires** | | | | | | |
| **physical activity questionnaire**† | 0.696 | [0.646, 0.746] | **<0.001** | 0.643 | [0.590, 0.697] | **<0.001** |
| **12-month physical activity questionnaire**† | 0.484 | [0.424, 0.546] | **<0.001** | 0.504 | [0.444, 0.564] | **<0.001** |

MVPA = moderate-to-vigorous physical activity, 95% CI = 95% confidence interval, † MET hours per day
Statistically significant correlations are highlighted