Association between self-reported and objectively measured physical fitness level in a middle-aged population in primary care

Kirstine H. Obling [a,*], Anne-Louise S. Hansen [b], Kristian Overgaard [c], Kasper Normann [b], Annelli Sandbaek [b], Helle T. Maindal [a]

[a] Aarhus University, Department of Public Health, Section for Health Promotion and Health Services, Aarhus, Denmark
[b] Aarhus University, Department of Public Health, Section for General Practice, Aarhus, Denmark
[c] Aarhus University, Department of Public Health, Section for Sport Science, Aarhus, Denmark

ARTICLE INFO

ABSTRACT

Aim. To investigate the association between self-reported physical fitness level obtained by a single-item question and objectively measured fitness level in 30- to 49-year-old men and women.

Methods. From the Danish 'Check Your Health Preventive Program' 2013–2014 fitness level was assessed in 2316 participants using the Aastrand test. Additionally, participants rated their physical fitness as high, good, average, fair or low. The association of self-reported- with objectively measured fitness level was analyzed by linear regression. Categories of self-reported- and objectively measured fitness level were cross-tabulated and agreement was quantified by Kappa statistics. Gender differences within categories were investigated by Poisson regression.

Results. Data from 996 men and 1017 women were analyzed (excluded, n = 303). In both men and women a higher self-reported fitness level was associated with a higher objectively measured fitness level ($R_{all} = 0.42$). Kappa agreement was 0.25. Poisson regression revealed that women rated their fitness level significantly lower than men ($p < 0.001$).

Conclusion. A single-item question is a cost-effective way of measuring physical fitness level, but the method has low association and fair agreement when compared to the Aastrand test. Men tend to overestimate physical fitness more than women, which should be accounted for if using the question in primary care settings.

Introduction

Low physical fitness level is strongly associated with major non-communicable diseases such as coronary heart disease, type 2 diabetes and breast- and colon cancers (Lee et al., 2012). Identifying at-risk individuals with low physical fitness level is therefore of considerable importance for public health.

Optimally, physical fitness is identified by measuring maximum oxygen uptake (VO_2max) directly using gas exchange respirometry at increasing workloads to exhaustion (gold standard) (Pescatello et al., 2014), but the extensive nature of this procedure makes it inappropriate for use in non-athletic at-risk individuals and in larger population-based samples (Pescatello et al., 2014). Consequently, indirect and less exhausting methods such as the Aastrand test can be used (Pescatello et al., 2014; Cink and Thomas, 1981). Both the direct and the indirect measurements, however, are often not feasible for use in daily clinical

settings and other preventive services due to the need of trained personnel, special equipment and time.

To address this issue, methods such as prediction models that estimate physical fitness based on e.g. sex, Body Mass Index and self-reported physical activity have been developed but also, multiple- and single-item questionnaires on physical fitness have been used (Cardinal, 1996; Jackson et al., 1990; Jurca et al., 2005; Knapik et al., 1992; Ortega et al., 2013; Mikkelsson et al., 2005).

In daily clinical settings, where time is limited, single-item questionnaires may be the most efficient and cost-effective method to estimate physical fitness. Few studies have investigated the association between single-item questions and objectively measured physical fitness. Findings from these studies are heterogeneous and in general disagree on whether answers of the questions may be influenced by e.g. gender and age (Jurca et al., 2005). Young et al. found a negative association between a single-item question and fitness level estimated from a 600-yard run in 193 seventh to tenth grade girls ($r = -0.52; -0.56$) (Young, 1985) whereas, Lamb et al. found a positive association between a single-item question and fitness level estimated from the Aastrand test ($r_{all} = 0.52$) in 118 university employees and students

* Corresponding author at: Aarhus University, Department of Public Health, Section for Health Promotion and Health Services Bartholins Allé 2, 8000 Aarhus C, Denmark.
E-mail address: Kirstine.obling@ph.au.dk (K.H. Obling).

(Lamb, 1992). Finally, Aadahl et al found that a single-item question reflected objectively measured fitness level using a graded bicycle test with increasing workload until exhaustion in a voluntary population of 102 people aged 35 to 65 years (p for trend <0.0001) (Aadahl et al., 2007).

However, to identify individuals with low fitness level in daily clinical settings, a single-item question must reflect physical fitness not only in people who are able to perform the direct tests but also in a general population. Therefore we aimed to establish whether the single-item question used by Aadahl et al. would reflect objectively measured fitness level in a large general population of 30-to-49-year-old men and women using the Aastrand test as objective measure. Additionally, we aimed to investigate whether men and women rate their physical fitness differently.

Materials and methods

Design

The association between self-reported fitness level obtained by a single-item-question and objectively measured physical fitness level was investigated in a cross-sectional design.

Study population

A total of 2316 men and women, aged 30–49 years, were selected consecutively from the ongoing Danish 'Check Your Health Preventive Program' (Maindal et al., 2014) in the period January 31th 2013 to January 31th 2014. The program is the result of a collaboration between a municipality, regional authorities and general practices and offers all citizens in the age group a health check including clinical- and behavioral measurements. In agreement with the Danish Health Law each participant provided written informed consent for data to be used for research purposes. The program was approved by The Danish Data Protection Agency (j.no: 2012-41-0183) and presented to the regional Scientific Ethics Committee which did not find approval of the program necessary since the study was deemed to not fall under the category biomedical research.

Anthropometric- and questionnaire data

Information on Body weight; Height; Body Mass Index (BMI); Waist circumference and VO_2max were obtained from the health examination at the Check Your Health Preventive Program. Body weight (Seca 769 measuring station) and height (Seca 222 Mechanical telescopic measuring rod) were measured without shoes to the nearest 0.1 kg and 0.5 cm, respectively. BMI was calculated dividing weight by height^2. Waist circumference (seca 203 Ergonomic circumference measuring tape) was measured to the nearest 0.5 cm between anterior superior iliac spine and the lower rib.

VO_2max was estimated using the Aastrand sub-maximal cycle test (Aastrand and Rodahl, 1986) carried out by using a Monark 939 E Pendulum Ergometer cycle with an initial constant workload of 75 watt for women and 100 watt for men. Pedaling frequency was set between 60–70 rounds per minute. Heart rate was measured continuously (polar T31 coded transmitter, Polar, Denmark) and recorded after six minutes, if participants achieved a steady state pulse within a target interval of 120–170 beats/min. If the target interval was not reached after two minutes of cycling, the workload was increased with 35 and 50 watt for women and men, respectively. The test proceeded until a steady-state was reached. Fitness level was estimated from the recorded heart rate and workload standardized to age and sex (Aastrand and Rodahl, 1986) (Monark 939 E Analysis Software, Version 3.0.12, Monark Exercise AB, Sweden) and in absolute (L/min) terms (Pescatello et al., 2014). The clinical examinations were standardized, administered by health professionals and took place in a local health care centre.

Prior to the health examination participants were asked to answer a web-based questionnaire regarding e.g. self-reported fitness level, smoking habits and alcohol use. Data concerning self-reported physical fitness was obtained by the question: 'How do you rate your own physical fitness?' Participants were asked to rate their own physical fitness as high, good, average, fair or low. Similar questions have previously been used in national and international studies (Mikkelsson et al., 2005; Jorgensen et al., 2003). Questions on smoking habits and alcohol use were based on items from the Danish National Health Profile (Christensen et al., 2010).

Statistical analyses

Individuals with missing data on self-reported physical fitness level and missing or non-valid data regarding VO_2max were excluded from analysis. Characteristics of the study population are presented as means and standard deviations for continuous normally distributed data, and numbers (n) and proportions (%) for categorical variables.

The association between categories of self-reported fitness level and VO_2max (ml/kg/min) was explored separately for men and women using linear regression analysis with smoking status as a covariate, self-reported physical fitness as the independent variable and VO_2max (ml/kg/min) as the dependent variable. Results are presented as means with 95% confidence intervals and correlation coefficients (r) derived from the linear regression analyses. To display the frequency distribution of self-reported fitness level in relation to objectively measured fitness level, VO_2max (ml/kg/min) was divided into cut-offs specified by Aastrand et al. (Table 1): Low; Fair; Average; Good and high(Aastrand and Rodahl, 1986). Self-reported fitness level and objectively measured physical fitness level was then cross-tabulated separately for men and women. Cohen's quadratic weighted kappa was used to assess the extent of agreement between self-reported physical fitness and VO_2max (ml/kg/min) in the cross-tabulation. The correlation coefficients and the Kappa values were interpreted based on Mukaka and Viera et al. respectively (Mukaka, 2012; Viera and Garrett, 2005). P-values less than 0.05 were considered statistically significant.

Finally, whether men and women rated their fitness level differently within categories was explored in a Poisson regression model based on count data for all combinations of sex, self-reported fitness level and VO_2max. The counts were regressed on all main effects and all pairwise interactions of the three variables. The interaction term for sex and self-reported fitness was the basis for the test.

Data was analyzed using STATA version 12 (Stata Statistical Software).

Results

A total of 2316 participants were included in the study. Data from 201 participants was excluded from the analysis due to missing VO_2max values (n = 159), missing self-reported fitness level (n = 31) or invalid test results (n = 11). Another 102 tests were declared invalid due to pulse values below 120 or above 170 beats per minute

Table 1
Standard cut-offs in VO_2max (ml/kg/min) relative to gender and 30-to-49-year-old Scandinavian non-athletic men and women specified by Aastrand et al.(Aastrand and Rodahl, 1986).

		VO_2max (ml/kg/min)			
	Low	Fair	Average	Good	High
Men, years					
30–39	<35	≥35 – <40	≥40 – <48	≥48 – <52	≥52
40–49	<31	≥31 – <36	≥36 – <44	≥44 – <48	≥48
Women, years					
30–39	<28	≥28 – <34	≥34 – <42	≥42 – <48	≥48
40–49	<26	≥26 – <32	≥32 – <41	≥41 – <46	≥46

during the Aastrand test. This left data from 2013 participants for analysis. The mean age was 41.5 (sd =6.0) years and 1017 (50.5%) were women. Men had a slightly higher VO_2max than women in absolute terms (Table 2). However, men had a relatively lower VO_2max than women when compared to age related norms (Table 1). Anthropometric- and self-reported-measures of physical fitness level of the study population are presented in Table 2.

Self-reported fitness level in relation to VO_2max (ml/kg/min) is presented in Table 3 distributed with regard to gender. In both men and women a higher self-reported fitness level was associated with a higher objectively measured fitness level ($r_{men} = 0.39$; $r_{women} = 0.44$; $r_{all} = 0.42$). Including smoking status as a covariate did not change the association.

In Table 4 the cross-tabulation of self-reported fitness level and VO_2max (ml/kg/min) relative to age and gender (Table 1) shows that both men and women tend to overestimate physical fitness although more pronounced among men ($\kappa = 0.25$). A total of 489 out of 2013 people were able to rate their physical fitness in the same category as measured objectively, 324 (66.3%) of these were women. Approximately four times more men (n = 121) than women (n = 30) with self-reported 'good' fitness had a 'low' fitness when measured objectively and twice as many men (n = 256) than women (n = 125) with self-reported 'average' fitness had a 'low' objectively measured fitness. Only 6% of men with an objectively measured 'low' fitness rated themselves as having so. Among women it was 19%. In addition to this finding, the Poisson regression revealed that the number of women rating themselves as fair, average, good and high was significantly lower than the number of men (p < 0.001), indicating that men tend to overestimate physical fitness more than women.

Discussion

We found a low association (r = 0.42) and a fair agreement ($\kappa = 0.25$) between self-reported physical fitness level, obtained by a single-item question, and objectively measured fitness level in men and women. This is in line with findings from other studies investigating the association between single-item questions on self-reported physical fitness and objectively measured physical fitness level (Young, 1985; Lamb, 1992 Jun; Aadahl et al., 2007). However, only one study that we know of has used the same measure of objective physical fitness level as in the present study (Lamb, 1992). Comparing our results to the results by Lamb et al. (Lamb, 1992) the overall association in our study was somewhat lower (r = 0.42 vs r = 0.52). This difference may be due to the fact that we included a more general population whereas Lamb et al. included students and teachers from a university.

Additionally, we found that both men and women overestimated self-reported fitness level. This overestimation was markedly more

Table 2
Characteristics of 2013 participants enrolled in the Danish 'Check Your Health Preventive Program' 2013–2014*.

	Number with available data	Men (n = 996)	Women (n = 1017)
Age	2013	41.6 (6.0)	41.5 (6.0)
Body weight, kg	2013	88.6 (14.5)	73.3 (15.1)
Body Mass Index, kg/m^2	2012	27.1 (4.1)	26.3 (5.2)
Waist circumference, cm	2010	96.4 (11.8)	87.3 (13.1)
VO$_2$max, ml/kg/min	2013	32.7 (9.2)	31.6 (8.9)
VO$_2$max, l/min	2013	2.8 (0.8)	2.2 (0.6)
Smoking daily, n (%)	1997	148 (14.9)	170 (16.9)
Alcohol, unit/week)	1720	6.0 (7.6)	3.0 (4.4)
Self-reported fitness level, n (%)	2013		
High		40 (4.0)	24 (2.4)
Good		312 (31.3)	233 (22.9)
Average		453 (45.5)	468 (46.0)
Fair		157 (15.8)	202 (19.9)
Low		34 (3.4)	90 (8.8)

* Mean (SD) if not otherwise stated.

Table 3
Self-reported fitness level in relation to VO_2max (ml/kg/min) in 2013 participants enrolled in the Danish 'Check Your Health Preventive Program' 2013-2014.

Self-reported physical fitness	VO$_2$max (ml/kg/min)	
	Men Mean (95% CI)	Women Mean (95 % CI)
High	41.6 (39.0–44.3)	41.7 (38.5–44.9)
Good	36.0 (35.1–37.0)	36.1 (35.1–37.1)
Average	32.0 (31.2–32.8)	32.0 (31.3–32.7)
Fair	27.8 (26.4–29.1)	27.3 (26.2–28.4)
Low	24.7 (21.8–27.5)	24.3 (22.6–25.9)

pronounced among men with objectively measured average, fair and low physical fitness. Several studies find similar results regarding perceived fitness or fitness related capacities among men. Delignieres et al. found that middle-aged men have higher perceived fitness than middle aged women (Delignieres et al., 1994). Godin et al. found that men have significantly higher physical self-efficacy and perceived physical ability than women (Godin and Shephard, 1985) and finally, Lintunen et al. found that boys have significantly higher perceptions of fitness than girls (Lintunen et al., 1995). A contributing factor to the gender differences could be that self-esteem is highly correlated with masculinity and self-efficacy but not with femininity (Delignieres et al., 1994). Furthermore, it has been suggested that the masculine-role endorsement could have a major influence of physical self-worth (Delignieres et al., 1994). In contrast, a meta-analysis by Germain et al. concludes that gender does not modify the association between self-reported and objectively measured fitness level (Germain and Hausenblas, 2006).

The strength of our study is the inclusion of the large general population of 30–49 year-old men and women. The use of the Aastrand test made this possible since it estimates maximum oxygen uptake without a great deal of effort for the participant. Consequently, selection bias due to health problems was minimized.

Although the Aastrand test has a good reliability compared to the direct test (r = 0.83–0.93) (Cink and Thomas, 1981), using the test is also a limitation of our study since the standard error for predicting the maximal oxygen uptake from the test is estimated to be ± 15 % (Cink and Thomas, 1981). Thus, some participants may have had a lower or higher true VO_2max (ml/kg/min) than measured, which may have affected the association between self-reported and objectively measured fitness level (Whaley et al., 1992; Greiwe et al., 1995; Siconolfi et al., 1982; Gore et al., 1999). Additionally, some studies have found the Aastrand test to consistently underestimate VO_2max in both genders (Davies, 1968; Metz and Alexander, 1971), whereas, other studies report that the Aastrand test underestimates VO_2max by 5–25 % in men only. Therefore, the apparent overestimation of self-reported fitness among men in our study might in part be due to a gender bias in the Aastrand test. Finally, some studies find the Aastrand test to overestimate VO_2max by 8.5–18.5% in women (Hartung et al., 1995) which may explain the apparently higher VO_2max relative to age and gender seen among women in our population. On the other hand, the possibility exists that the women in this population have a higher true VO_2max relative to gender norms than men, but we have no indications that the women were more physically active than men. Another limitation of our study (and other similar studies) is the fact that there is no clear conceptual definition of physical fitness since it is a multidimensional construct which besides VO_2max includes aspects such as muscular strength, flexibility and balance (Pescatello et al., 2014). The use of a single-item question may therefore fail to address the multidimensionality of physical fitness and participants in our study may have defined fitness individually in a way he or she found appropriate. Thus the

Table 4

Cross tabulation of self-reported fitness level and VO₂max (ml/kg/min) relative to age and gender in 2013 participants enrolled in The Danish 'Check Your Health Preventive Program' 2013–2014. Values are numbers (%).

Vo2max men/women	Self-reported fitness level category men/women*					
	High	Good	Average	Fair	Low	Total
High	11/8	36/27	17/17	0/2	0/0	64/54
	(17.2/14.8)	(9.7/50.0)	(26.6/31.5)	(0.0/3.7)	(0.0/0.0)	(6.4/5.3)
Good	2/3	17/38	12/36	3/3	0/0	34/80
	(5.9/3.8)	(50.0/47.5)	(35.3/45.0)	(8.8/3.8)	(0.0/0.0)	(3.4/7.8)
Average	13/8	78/81	75/151	8/31	1/11	175/282
	(7.4/2.8)	(44.8/28.7)	(42.9/53.5)	(4.6/11.0)	(0.57/3.9)	(17.6/27.7)
Fair	7/2	60/57	93/139	29/66	0/18	189/282
	(3.7/0.0)	(31.7/20.2)	(49.2/49.2)	(15.8/26.2)	(0.0/6.6)	(19.0/27.7)
Low	7/3	121/30	256/125	117/100	33/61	534/319
	(1.3/0.9)	(22.7/9.4)	(47.9/39.9)	(21.9/31.3)	(6.2/19.1)	(53.7/31.4)
Total	40/24	312/233	453/468	157/203	34/90	996/1017
	(4.0/2.3)	(31.3/22.9)	(45.5/46.0)	(15.8/20.0)	(3.4/8.8)	(100.0/100.0)

* κ = 0.25.

objective measure used in our study may not have corresponded entirely with the individual interpretation of the single-item question on physical fitness.

To better reflect the multidimensionality of physical fitness, multiple-item questionnaires have been developed and compared to different measures of objectively measured fitness level. Strøjer et al. investigated five components of self-assessed physical fitness by five visual analogue scales with illustration of aerobic fitness, muscle strength, endurance, flexibility and balance in 159 men and women. They found, that the interclass correlation coefficients between the self-assessed parameters and the objective measures (Aastrand test, maximal isometric voluntary contraction of the back extension and flexion muscles, the finger to floor method, isometric back extension and sitting on a wobble board) ranged from 0.62 to 0.80 (Stroyer et al., 2007). Mikkelsson et al. found that correlations between self-reported endurance, strength, speed and flexibility and physical fitness tests (bicycle ergometer test with three submaximal loads, ergojump test, 30-s sit-up test, hand-grip tests and a sit-and-reach test) ranged from 0.13 to 0.57 (Spearman) among 40-year-old men (n = 29) and women (n = 35), who voluntarily returned a questionnaire regarding physical fitness (Mikkelsson et al., 2005). Furthermore, Ortega et al. found Kappa coefficients between 0.54 and 0.65 between self-reported physical fitness (the 'International Fitness Scale') and measured parameters (20-m shuttle run test, hand-grip test, the standing long jump test and the sit and reach test) in a Spanish population (n = 276) aged 10 to 30 years (Ortega et al., 2013). Comparison of studies within this field should, however, be made with caution due to the large variation in methodology and ages of the populations.

Since only low to modest associations between single-and multiple-item self-reported fitness and objective measures of physical fitness have been reported, a single-item question may not be ideal to measure the absolute fitness level. However, the conception that measurement itself influences behavior is well-established (French and Sutton, 2010). As such, it could be hypothesized that although the question may not identify individuals with low fitness level in daily clinical settings it may motivate people to change physical activity behavior.

Conclusion

Although a single-item question is a cost-effective way of measuring physical fitness level the method only has a moderate association and a fair agreement when compared to the Aastrand test. If the question is used for estimation of physical fitness in primary care settings, health personnel should be aware that middle-aged women, and in particular men, tend to overestimate their physical fitness level.

Conflict of interest statement

The study was initiated and analyzed by the authors who declare that there is no conflict of interest.

References

Aadahl, M., Kjaer, M., Kristensen, J.H., Mollerup, B., Jorgensen, T., 2007. Self-reported physical activity compared with maximal oxygen uptake in adults. Eur. J. Cardiovasc. Prev. Rehabil. 14 (3), 422–428 (Jun).

Textbook of work physiology. In: Aastrand, P.O., Rodahl, K. (Eds.), 3rd ed. McGraw Hil.

Cardinal, B.J., 1996. Predicting cardiorespiratory fitness without exercise testing in epidemiologic studies: a concurrent validity study. J. Epidemiol. 6 (1), 31–35 (Mar).

Christensen, A., Davidsen, M., Ekholm, O., Hansen, S., Holst, M., Juel, K., 2010. Den nationale sundhedsprofil 2010: - Hvordan har du det? version: 1.0; versionsdato: 10. marts 2011 ed. Sundhedsstyrelsen, København.

Cink, R.E., Thomas, T.R., 1981. Validity of the Astrand-Ryhming nomogram for predicting maximal oxygen intake. Br. J. Sports Med. 15 (3), 182–185 (Sept).

Davies, C.T., 1968. Limitations to the prediction of maximum oxygen intake from cardiac frequency measurements. J. Appl. Physiol. 24 (5), 700–706 (May).

Delignieres, D., Marcellini, A., Brisswalter, J., Legros, P., 1994. Self-perception of fitness and personality traits. Percept. Mot. Skills 78 (3 Pt 1), 843–851 (Jun).

French, D.P., Sutton, S., 2010. Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? Br. J. Health Psychol. 15 (Pt 3), 453–468 (Sep).

Germain, J.L., Hausenblas, H.A., 2006. The Relationship Between Perceived and Actual Physical Fitness: A Meta-Analysis. J. of Appl. Sport Psychol 18 (4), 283–296 (12/01 2014/07).

Godin, G., Shephard, R.J., 1985. Gender differences in perceived physical self-efficacy among older individuals. Percept. Mot. Skills – LA English 60 (2), 599.

Gore, C.J., Booth, M.L., Bauman, A., Owen, N., 1999. Utility of pwc75% as an estimate of aerobic power in epidemiological and population-based studies. Med. Sci. Sports Exerc. 31 (2), 348–351 (Feb).

Greiwe, J.S., Kaminsky, L.A., Whaley, M.H., Dwyer, G.B., 1995. Evaluation of the ACSM submaximal ergometer test for estimating VO2max. Med. Sci. Sports Exerc. 27 (9), 1315–1320 (Sep).

Hartung, G.H., Blancq, R.J., Lally, D.A., Krock, L.P., 1995. Estimation of aerobic capacity from submaximal cycle ergometry in women. Med. Sci. Sports Exerc. 27 (3), 452–457 (Mar).

Jackson, A.S., Blair, S.N., Mahar, M.T., Wier, L.T., Ross, R.M., Stuteville, J.E., 1990. Prediction of functional aerobic capacity without exercise testing. Med. Sci. Sports Exerc. 22 (6), 863–870 (Dec).

Jorgensen, T., Borch-Johnsen, K., Thomsen, T.F., Ibsen, H., Glumer, C., Pisinger, C., 2003. A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. Eur. J. Cardiovasc. Prev. Rehabil. 10 (5), 377–386 (Oct).

Jurca, R., Jackson, A.S., LaMonte, M.J., et al., 2005. Assessing cardiorespiratory fitness without performing exercise testing. Am. J. Prev. Med. 29 (3), 185–193 (Oct).

Knapik, J.J., Jones, B.H., Reynolds, K.L., Staab, J.S., 1992. Validity of self-assessed physical fitness. Am. J. Prev. Med. 8 (6), 367–372 (Nov-Dec).

Lamb, K.L., 1992. Correlates of self-perceived fitness. Percept. Mot. Skills 74 (3 Pt 1), 907–914 (Jun).

Lee, I.M., Shiroma, E.J., Lobelo, F., et al., 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 380 (9838), 219–229 (Jul 21).

Lintunen, T., Heikinaro-Johansson, P., Sherrill, C., 1995. Use of the Perceived Physical Competence Scale with adolescents with disabilities. Percept. Mot. Skills 80, 571 (H.-W.Wilson - SSA).

Maindal, H.T., Stovring, H., Sandbaek, A., 2014. Effectiveness of the population-based Check your health preventive programme conducted in primary care with 4 years

follow-up [the CORE trial]: study protocol for a randomised controlled trial. Trials 15 (Aug 29, 341-6215-15-341).

Metz, K.F., Alexander, J.F., 1971. Estimation of maximal oxygen intake from submaximal work parameters. Res. Q. 42 (2), 187–193 (May).

Mikkelsson, L., Kaprio, J., Kautiainen, H., Kujala, U.M., Nupponen, H., 2005. Associations between self-estimated and measured physical fitness among 40-year-old men and women. Scand. J. Med. Sci. Sports 15 (5), 329–335 (Oct).

Mukaka, M.M., 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24 (3), 69–71 (Sep).

Ortega, F.B., Sanchez-Lopez, M., Solera-Martinez, M., Fernandez-Sanchez, A., Sjostrom, M., Martinez-Vizcaino, V., 2013. Self-reported and measured cardiorespiratory fitness similarly predict cardiovascular disease risk in young adults. Scand. J. Med. Sci. Sports 23 (6), 749–757 (Dec).

Pescatello, L.S., Arena, R., Riebe, D., Thompson, P.D., 2014. ACSM's guidelines for exercise testing and prescription. 9 ed. ACSM's Guidelines for Exercise TEsting and Prescription. Wolters Kluwer/Lippincott Williams & Wilkins Health, Philadelphia, pp. 60–113.

Siconolfi, S.F., Cullinane, E.M., Carleton, R.A., Thompson, P.D., 1982. Assessing VO2max in epidemiologic studies: modification of the Astrand-Rhyming test. Med. Sci. Sports Exerc. 14 (5), 335–338.

Stata Statistical Software: Release 12, 2015. College Station, TX: StataCorp LP.

Stroyer, J., Essendrop, M., Jensen, L.D., Warming, S., Avlund, K., Schibye, B., 2007. Validity and reliability of self-assessed physical fitness using visual analogue scales. Percept. Mot. Skills 104 (2), 519–533 (Apr).

Viera, A.J., Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. Fam. Med. 37 (5), 360–363 (May).

Whaley, M.H., Kaminsky, L.A., Dwyer, G.B., Getchell, L.H., Norton, J.A., 1992. Predictors of over- and underachievement of age-predicted maximal heart rate. Med. Sci. Sports Exerc. 24 (10), 1173–1179 (Oct).

Young, M.L., 1985. Estimation of fitness and physical ability, physical performance, and self-concept among adolescent females. J. Sports Med. Phys. Fitness 25 (3), 144–150 (Sep).