CROSS-VALIDATION OF METHODS ASSESSING MAXIMAL OXYGEN CONSUMPTION IN HUNGARIAN SCHOOL AGED CHILDREN

THESIS OF DOCTORAL (Ph.D) DISSERTATION

Mónika Kaj

University of Pécs

Department of Health Sciences

Doctoral School of Health Sciences

Head of Doctoral School: Prof. Dr. Bódis József professor, doctor of the Hungarian Academy of Sciences, president

Head of Program: Prof. Dr. Bódis József professor, doctor of the Hungarian Academy of Sciences, president

Supervisor: Dr. Wilhelm Márta professor, PhD., head of the institution, PTE Faculty of Sciences, Institution of Sport Sciences and Physical Education

Pécs, 2017

INTRODUCTION

One of the most determining components of physical fitness is the cardiovascular, also known as aerobic fitness, which is characterized as the maximal oxygen consumption per minute relative to bodyweight (relative VO_{2max} – ml/kg/minute[1]). Aerobic fitness level is closely related to the current health status and the risk of developing various cardiovascular diseases (Caspersen et al., 1985; Janssen and Cramp, 2007; Buchan et al., 2015). Determining the value of $relVO_{2max}$ not only provides useful information to sports practitioners in the course of training planning and performance assessment, but also for healthcare professionals with great significance in health diagnostics and epidemiological studies. There are several methods for direct determination of maximal oxygen consumption among which the most accurate and reliable results are obtained by direct measurement in laboratory measuring maximal load intensity (vita maxima testing). However, considerable time, expense, and specific personal and material conditions are required to perform accurate measurements. As an alternative to laboratory tests requiring special conditions, practitioners working in sports and health sciences have developed simple procedures (field tests) that allow VO_{2max} to be well estimated. The main criteria for using the developed techniques are that they meet scientifically proven requirements of the validity and reliability criteria.

The first topic elaborated in my doctoral dissertation is the validation of a new prediction formula[2] for the PACER test developed by the experts in the COOPER Institute (Dallas, USA), and is used by Physical Educational and sports professionals. This calculation method is applied by both the FITNESSGRAM and the Hungarian National Student Fitness Test (NETFIT®) from 2014 (Burns et al., 2015). In the course of fitness assessment during Physical Education class, pedagogical aspects have to play a significant role in addition to the importance of epidemiological considerations, when it comes to processing the dataset. In the PACER test, due to the previously used complicated conversion formula, the results became apparent only later, so the experience of running and the evaluation of the result could not be related, reducing the pedagogically desired effect of the test. The number of laps performed during the test is the absolute result, and the relation between $relVO_{2max}$ is not linear, and many factors, in particular the BMI and age greatly affect it. Also, the reliability of the new calculation method appropriate for pedagogical purposes has not yet been verified on a large sample.

In the first part of my doctoral thesis, I was studying the question whether the estimation method used by NETFIT® in the PACER test is valid for the determination of the aerobic fitness level. I elaborated a detailed statistical analysis of this topic on a statistically significant sample: N = 482 (10 to 18 year old girls and boys divided into age groups).

The second topic of my doctoral dissertation was the validation[3] of the quasi-laboratory endurance sub-maximal walking test developed by the National Health and Nutrition Examination Survey (NHANES), which is primarily used for healthcare purposes in school-age children. The protocol has regularly been used in the United States for medical screening since 1999 (Wang et al., 2010) and also the FITNESSGRAM® Scientific Committee has used the results of the NHANES study to develop the criterion-referenced FITNESSGRAM® standards in the 2011 review (Welk et al., 2011a).

---

[1] To simplify, maximal oxygen uptake (VO_{2max}) is defined as relative oxygen uptake throughout my dissertation

[2] $VO_{2max}=45.619 + (0.353 \text{ x laps}) – (1.121 \text{ x age})$

[3] Validation: a procedure that is used together for proving the validity of the chosen method by systematic investigation

The aim of the sub-maximal walking test on the treadmill is to achieve 75% of the estimated maximal (220-year) heart rate, and at this point extrapolation of the heart rate values recorded in the test can be used to estimate maximal oxygen consumption (NHANES, 2012). This prediction method has assessed a large number of screening tests involving school-age populations, however, to date, no evidence of the accuracy of the estimation has been made for school-age children (Pate et al., 2006). The responses given by adults and children may differ for different physiological parameters. In the period of growth and maturity, the organs develop at different pace, temporary differences may significantly affect the functional performance of the body, so the prediction methods used in adults, including principally regression equations, can not be applied in the same way for all ages. I wanted to focus on this problem with presenting the studies underlying.

AIM OF THE STUDY, ARISING QUESTIONS

The Hungarian School Sport Federation has aimed to set up a new, diagnostic pedagogical evaluation and measurement fitness system within the framework of the priority project TÁMOP-3.1.13, which was implemented with the collaboration of the American Cooper Institute in a nationally representative research. In this study, we investigated the fitness status of 10-19 year old pupils using anthropometric measurements, field tests of NETFIT® and laboratory tests (Karsai et al., 2013). As part of the systematic research, we investigated the probability of occurrence of the metabolic syndrome and the maximal oxygen consumption by complex spiroergometric assessments (Csányi et al., 2015). The aim of the research was to determine, among other things, how the criterion referenced standard values used in a similarly health-oriented FITNESSGRAM® can be applied to the Hungarian population. In this research, for the students' physiological measurement vita maxima walking test used by NHANES was extended monitoring continuous heart rates and respiratory condition hence allowing us to examine the validation and accuracy of the maximal oxygen consumption methods (PACER test and NHANES sub-maximal walking test).

In my dissertation, I examine:
- the extent to which the VO_{2max} data of different groups (divided by age and gender) differ from each other and how these data differ from the results of previous measurements
- whether the predictive model based on the performance of PACER test of the NETFIT® aerobic fitness profile can be reliably applied to the Hungarian school-age population, and how these predicted VO_{2max} estimates reflect the actual VO_{2peak} values
- In addition to assessing the congruence of estimated and measured individual absolute VO_{2max} values, we were looking at the extent of differences in student ratios after qualifying into health-oriented fitness categories used by NETFIT®, when the evaluation was based on the estimated and measured VO_{2max} values
- As part of the research, the vita maxima protocol was an extended version of the NHANES sub-maximal walking test protocol. Our goal was to determine the equivalence of the VO_{2max} values estimated by extrapolation from HR values measured during the NHANES protocol with the actual VO_{2max} measured at the end of the assessment
- our goal was also to determine whether gender and age have an effect and, if so, what effect do they have on the differences of the VO_{2max} values between the two prediction methods.

BIBLIOGRAPHY

Cardiovascular / cardiorespiratory fitness is one of the most important factors of health-related fitness, as its direct relationship with the health status is proven by numerous scientific studies. Its value characterizes the upper limit of the aerobic functioning of the body, and is the most commonly used indicator of the effective functioning of the central nervous system, the cardiopulmonary and metabolic system (Coyle et al., 1984, Gledhill, 1982, Hickson et al, 1977, Noakes et al. 2001, Saltin, 1969, 1973, Wagner, 1988, West et al., 1983). A number of studies have shown that aerobic fitness levels are associated with several health factors in adulthood, as the occurrence of cardiovascular disorders (Arraiz et al., 1992; Blair et al., 1989; Sandvik et al., 1993), type-2 diabetes (Colberg et al., 2010), hypertension (Blair et al., 1984) and certain types of cancer (Oliveria et al., 1996). The most reliable and accurate method for determining aerobic fitness is to measure the maximal oxygen consumption. The most accurate value is assured with direct measurement in laboratory conditions during a complete exhaustion test (Safrit et al., 1988), the so-called vita maxima test. The laboratory vita maxima test a less applicable method for measuring aerobic fitness in places where the necessary conditions (expensive devices, skilled staff, time requirements, etc.) are not available such as sports clubs, schools or large sample researches. For this reason, in practice, determining VO_{2max}, various field tests and quasi-laboratory measurements are used as an alternative instead of direct measurement. The PACER test is most widely used to estimate maximal oxygen consumption, because of its ease of use, low tool demand and reliability. In Hungary, the NETFIT[®] test battery introduced in the academic year of 2014/2015, also uses this field test to measure the aerobic fitness of school-age children (Kaj et al., 2014). The field test and the method itself (Leger, Mercier, Gadoury, & Lambert, 1988) are the same as it used in FITNESSGRAM[®] (The Cooper Institute, 2010), and is also part of numerous internationally accepted fitness test batteries (Eurofit 1998, The Brockport Physical Fitness Test, 1999, Assessment Aims of Physical Activity (ALPHA), and its role in international comparisons of aerobic fitness measurements is important (Tomkinson and Olds, 2007; Tomkinson et al., 2016). A total of four studies were published about the reliability of the test stating moderate to high reliability coefficient greater than 0.64 and significantly non-different mean values (Léger et al., 1988, Liu et al., 1992, Mahar et al. 1997, Beets and Pitetti, 2006). In a systematic analysis of Mayorga-Vega et al. (2015) 57 studies were reviewed, stating that criterion validity values estimating the maximal oxygen uptake of the PACER test was moderate to strong ($r^2 = 0,66\text{-}0,84$) and when the performances achieved in the test were combined with age, gender, body weight, the value was even higher ($r^2 = 0,78\text{-}0,95$). According to the same meta-analysis, the validity of the Léger protocol for adults was statistically higher ($r^2 = 0.94$, 0.87-1.00) than in children ($r^2 = 0.78$, 0.72-0.85), while the aerobic fitness level and the gender of the individual does not affect the validity value. The equation published by Léger et al. for the estimation of VO_{2max} was developed on a sample of 188 students (8-19 years old). Subsequently, with the widespread use of the test, more and more people looked at the accuracy of the Léger's formula in other populations and developed even more accurate prediction models. The prediction models of Barnett et al. (1993) were developed on a sample of young people aged 12 to 17 (N=55) and the formulas included gender and body weight in addition to age and maximal test speed. Matsuzaka et al. (2004) developed a regression model for the estimation of VO_{2max} by using data from 132 Japanese 8-17 year-olds based on the test performance (number of laps or maximum speed achieved), gender, age and BMI. Mahar et al. (2006, 2011) improved their previous models, and in their cross-validation study it was found that the regression model provides the most accurate feedback on the actual VO_{2max} value, which includes the number of laps squared, the BMI, age and gender interaction. The FITNESSGRAM[®] uses the regression model by Burns et al. (2015) since 2013 for assessing aerobic fitness and since the cross-validation on the Hungarian population (Saint-Maurice et al., 2015) the National Hungarian Student Fitness Test (NETFIT[®]) applies it from the

academic year of 2014/2015 as well. The accuracy of prediction models can be improved by adding newer variables, but it makes testing difficult and reduces the applicability of the field test. The purpose of the FITNESSGRAM Scientific Committee was to create a new formula simplifying its application in school settings and the evaluation of the results, giving pupils immediate feedback on the performance achieved in the test. In order to apply the previously used methods (Mahar et al., 2011), it was necessary to know the BMI value to calculate the VO_{2max} values from the performance achieved in the test, causing some practical problems with its applicability (Cureton et al., 2014):

- it resulted in an inappropriate relationship between the one-mile walking test (alternative test for PACER in FITNESSGRAM) and the VO_{2max} values estimated from the PACER test,
- it emphasized the role of body composition in the ability of students whether they reach the healthy fitness zone or not,
- it excluded the possibility of evaluating the results of students who were not measured for body height and body weight.

The estimation model currently in use is based only on the number of laps achieved and the age of the person.

Out of field test alternatives of the vita maxima laboratory measurements are the so-called quasi-laboratory tests, which are usually performed on a bicycle or a treadmill ergometer with increasing loads up to sub-maximal intensity while monitoring the heart rate. It is well known that heart rate elevates linearly during the increasing intensity until the maximal performance is reached (Astrand, 1955). Maximal oxygen consumption and maximal heart rate appear at the same intensity, so the performance at a certain heart rate correlates closely with the oxygen uptake and performance output measured at the maximal intensity. Sub-maximal tests with heart rate monitoring like the most well-known and widespread Physical Working Capacity-170, PWC-170 Test (Wahlund, 1948), and the National Health and Nutrition Examination Survey (NHANES) walking test are based on this connection as well. The NHANES is a national program aimed at examining the health and nutritional status of adults and children in the US and collecting data for the National Center for Health Statistics (NCHS) to assess the relationship between unfavorable cardiovascular fitness and other health risk factors (12-49 ages). The FITNESSGRAM health-oriented fitness battery used the NHANES data to set its criterion referenced standards (Cooper Institute, 1992) and to refine the standards in 2011 (Welk et al., 2011a), and so obtaining new standards that determine a criterion-referenced assessment of the PACER test results. The NHANES program estimated young people's cardiovascular fitness during 1999-2002 in a representative sample of 12-18 year-olds (N=1966) based on the heart rate measured in a sub-maximal treadmill walking test. Based on these values and the probability of occurance of the symptoms of metabolic syndrome, FITNESSGRAM's Scientific Committee set the VO_{2max} standards, which distinguish three fitness categories for the health-related classification: Healthy Fitness Zone, Needs Improvement Zone, and the Needs Improvement - Health at Risk Zone (Welk et al., 2011b).

In studying the adaptability of FITNESSGRAM's aerobic fitness standards in Hungary (Csányi et al., 2015), we used the VO_{2max} results measured during the vita maxima test to set up the NETFIT® aerobic fitness profile. The protocol meant an extended version of the two levels of the NHANES walking test until complete exhaustion, so we were able to validate the sub-maximal test protocol on the Hungarian sample taking the "gold standard" into account.

HYPOTHESES

In the light of the literature references and results referenced in the literature review, the following hypotheses have been formed:

1. Aerobic fitness values show a different pattern in my sample for boys and girls by age.

2. Based on the performance in the PACER test, there is **no** significant **difference** between the estimated VO_{2max} and the currently measured VO_{2peak} value according to the regression model used by NETFIT® (Burns et al., 2015).

3. Based on the estimated and the measured VO_{2peak} values calculated by the new formula of the PACER test, the **difference** in the ratio of students achieving the NETFIT® fitness zones is **not** statistically significant.

4. There is no significant difference between the VO_{2max} estimated by the NHANES walking test protocol and the currently measured VO_{2peak} values, therefore the test can reliably be used in Hungarian school-age children

5. The difference in the VO_2 values for the two prediction methods is affected by gender and age effects at the individual level, but at group level this effect can not be perceived.

RESEARCH SAMPLE AND METHOD

Method of sample selection

The sample selected in the research protocol for the foundation of the NETFIT® was implemented in joint collaboration of the Hungarian School Sport Federation and the Cooper Institute in cooperation with an economic research institute. The sample was selected with a stratified random sampling procedure. First, a total of 53 schools were selected from 7 different regions of the country based on the following stratification criteria: size of the settlement (village, city), school size / number of pupils (<200,200-400,400-800,800<), type of school (state-run, foundation, ecclesiastical, other). The designated 53 institutions were informed in an official letter about the research and invited to participate. On the basis of the feedback, 3 refused to attend the study, thus the participation rate was 94.4%.

Next simple random selection was made to designate participants in the selected schools. Participants were sequenced by grade and put into alphabetical order and identified by their previously determined assessment identifier numbers by the school administrators with the help of trained staff. If a student was not fit to perform the tests (injury, release, etc.), the next student was involved. Pupils involved in the laboratory measurements were marked with different colored codes, therefore provided sample for this thesis.

Sample

The sample of the NETFIT® research included a total of 2686 students aged 11-19 (grades 5-12). Of this sample, 578 students were selected to participate in laboratory measurements beyond the assessments of the field tests. The legal representatives of the pupils involved in the study were informed about the research and made a written statement about the child's assessment. 96 students rejected the participation, so a total of 482 students attended both the field tests and the laboratory assessment (55.6% boys and 44.4% girls).

Methods and tools for data collection

PACER test

The ability of maximal oxygen uptake of the subjects was first estimated based on PACER test. The field tests were carried out uniformly in January/April 2013 in the sports hall/gym of the designated schools by means of trained staff in the morning hours, after a standard 10-min warm-up

following the other field tests of the NETFIT® battery. In all cases, before the field tests were completed, students were given detailed information on the procedure and purpose of the research.

The 20-meter endurance test was performed uniformly. The participants had to run continuously on a designated 20-meter long rack following the pace set by the beeps that accelerated per minute. The test continued until the subject did not reach the end line by the time of the signal. The number of completed laps (including the first mistake) was recorded by the staff. After the test, the following formula was used to estimate the maximal oxygen consumption (Burns et al., 2015): $VO_{2peak} = 0.353$ x (number of laps) - 1,121 x (age) + 45,619

Laboratory assessment

Anthropometric and resting, as well as training physiological parameters were measured in five different exercise-physiological laboratories throughout the country between May-June 2013 involving the heads of the laboratories.

The pupils were measured by trained staff members, who were up in exercise physiological measurements, at different locations using a given standard protocol, always with a pediatric presence. 6-9 participants arrived at the venues per day, with empty stomach.
The test was conducted as follows:

- After arrival following a 3-minute resting (seated) session, the physician measured the student's resting blood pressure (systole/diastole, mmHg) by manual auscultation as well as by an automatic blood pressure monitor (OMRON M2) depending on the laboratory's location.
- Students were guided to a designated room where their fasting blood parameters (glucose, cholesterol, triglyceride levels) were measured. Measurement of fasting glucose, cholesterol and triglyceride levels were taken after disinfection from the fingertips, using a drop of blood, assessed by Multicare IN instrument calibrated for the measurement, and equipped with the suitable test strips.
- After that, students changed to light sportswear to be measured for anthropometric parameters. The body height was determined to millimeter precision with standard SECA 206 wall-mounted stadiometer. NHANES protocol (2012) was used to determine the hip and waist circumference and skinfolds (biceps, triceps, subscapular, suprailiaca, pectoral, abdomen, mid-axilla, thigh, calf). Standard Lange calipers were used to measure skinfolds. InBody 720 bioimpedance analyzer was used to analyze the body mass and body composition. Body weight and body height values were used to calculate body mass index (BMI) according to Cole et al. (2012).
- Students in the study were interviewed and examined to identify possible factors that exclude the initiation of the measurement (medication, physical injury, hypertension, irregular heartbeat, etc.) before the start of the progressive spiroergometric exercise.

Students were tested with one of the 7 automatically configured protocols. The protocols differed in the initial intensity, the degree of inclination, and the rate of speed rise between the levels. The protocol was selected based on the method of Jackson et al. (1990) according to the estimated maximal oxygen consumption using gender, age, BMI and physical activity level.

> estimated VO_{2max} =56.363 + [1.921 * (PAR)] – [0.38 * (decimal age)] – [0.754 * (BMI)] + [10.987 * (girl=0, boy=1)]

In order to establish PAR (Physical Activity Recode) values used in the equation, a short questionnaire was applied to determine a physical activity levels. On the basis of the questions, the PAR value ranged from 0 to 7.

For the testing, SCHILLER AT-104 complete spiroergometer was used at 3 venues and Sensormedic Vmax Encore spirometer at 2 locations.

The testing was implemented according to the National Health and Nutrition Examination Survey's (NHANES) National Youth Fitness Survey treadmill protocol up until the second phase (2-minute warm-up, two times 3-minute exercise). The protocol was developed so that by the end of the second phase, the subject would reach 75-80% of the estimated maximal HR predicted by age. For our study, the test was continued with an additional increase in intensity per minute until the subjects were completely exhausted. The measured VO_{2peak} value was used as a criterion for aerobic fitness for these validation studies. During the exercise, HR was monitored with a 12-lead ECG (heart rate, HR), and ventilation (VE) registered in every 10 seconds, oxygen uptake (VO_2) and carbon dioxide release (VCO_2), oxygen pulse (O_2 pulse), respiratory quotient (RQ), exercise time (s), etc.

Based on the NHANES protocol, the selected exercise protocol is appropriate only if the measured HR at the end of the two-minute warm-up phase is between 50-60% of the expected age-related maximal HR (220-age). If the HR was higher than 60% of the predicted HR_{max}, the participant started with a lower protocol level, and if the measured HR was less than 50% of the expected maximal HR, a higher exercise level was initially obtained. The 2-min warm-up phase was followed by twice the 3-minute phases recommended by the NHANES protocol, during which the inclination increased while the speed remained constant. At the end of each exercise section, blood pressure was measured, then velocity was increased every minute until complete fatigue. The velocity increase per minute was maintained until the participant indicated his/her intention to end, since he/she was no longer able to keep up with the pace, showed signs of fatigue and demonstrated metabolic responses near expected values for adolescents (oxygen uptake was not increasing any longer (plateau phase)), RQ greater than 1.00 (Rowland, 1996). After reaching maximal oxygen uptake, a 2-minute cool-down phase followed, during which we continuously examined the process of calm-down.

Statistical analysis

During data clarification, we first extracted and encoded the raw database and filtered out cases with no data recorded (demographic data such as sex and age, measured VO_{2max} and/or the number of laps completed in the PACER test).Student less than 10 years old and over 18,99 years old were excluded, others were categorized into age groups (10 year olds: 10,00 to 10,99 years, 11 year olds: 11,00 to 11,99 years, etc.). As a criterion for data processing only those students remained in the dataset who achieved complete exhaustion (heavy breathing, ruddy face, sweating) and the measured VO_{2peak} value was considered valid if the RQ was 1.0 or more at the end of the exercise (Welsman et al., 2005). In the third step, we excluded the possible extreme values from the analysis, where the difference between the estimated and measured VO_{2peak} value was greater than 20 ml/kg/min.

Testing for normality of the selected variables (measured and estimated VO_{2max} values, end speed of the treadmill test, number of laps in the PACER) was performed with the Kolmogorov-Smirnov test. The gender differences between the variables were analyzed by independent two-sample t-test. The comparison of the measured maximal oxygen consumption by gender and age groups was performed using the univariate ANOVA test. In the bivariate correlation analysis, the relationship between the measured maximal oxygen consumption, some anthropometric data and the performance achieved in the PACER test (number of laps) was characterized by correlation coefficients.

Validation of the new formula of the PACER test

In the first step, the validity of the PACER test was evaluated by linear regression coefficients calculated from the maximal speed measured at the end of the laboratory treadmill test, VO_{2peak} value and the number of laps in the test using the following formula:

Estimated VO_{2peak} = 45.619 + (0.353 * number of laps) - (1,121 * age)

The validity of the PACER test was first examined through linear regression coefficients resultant from the relation between laboratory measures - namely, final treadmill
speed (in km/hr) and measured peak VO_2 and PACER laps. Standardized and unstandardized beta weights for each bivariate relation (measured vs. PACER) were provided. All analyses were conducted separately for boys and girls. A statistically significant unstandardized beta weight ($\beta > 0$; p < .05) was indicative of an association between measured and predicted peak VO_2. The magnitude of the R^2 and the limits of agreement (LOA) proposed by Bland and Altman (mean difference ± 1.96 *standard deviation; Bland & Altman, 1986) were indicative of the overall accuracy of the current equation. Standard error of estimates (SEE) was calculated as well, which is also informative for examining similarity between methods.

A second set of analysis determined the equivalence of the two methods by using bioequivalence testing procedure (Schuirmann, 1987). Employing two one-sided tests, establishing predefined acceptance range makes possible to test if the estimated values are equivalent to the measured values and determine whether the difference of means are within a certain predefined range (Robinson&Robert, 2004; Walker&Nowacki, 2011). In the present study, based on an expert decision, 10% of the possible range of equivalence was indicated.

All in all, equivalence is achieved if the two one-sided 95% confidence intervals for the average difference are within ±10% of the measured peak VO_2.

Analyzes of agreement of the criterion-referenced NETFIT[®] fitness zone classification based on measured and the estimated PACER test VO_{2peak}

Using the gender- and age-specific, health-oriented, criterion-referenced cut-off points in NETFIT[®], the VO_{2peak} values obtained with two different methods were categorized into fitness zones: Healthy Fitness Zone (HFZ), Needs Improvement Zone (NI), Needs Improvement – Health at Risk Zone (NIHR). The impact of error on criterion-referenced evaluation of VO_2 scores was examined using kappa statistics. Kappa values were interpreed using the guidelines from Altman (1990) (poor: K≤0,20, fair:0,20-0,40, moderate:0.40-0,60; good:0,60-0,80; very good: K≥0,80).

Sensitivity and specificity values were computed to determine the type of error when evaluating aerobic fitness accordig to the estimated VO_{2max}. Sensitivity in the context of this study reflects the proportion (in percent) of the estimated value classifications of subjects in the HFZ, NI zone, or NIHR zone matching those obtained reference values, so the student actually belongs to the fitness zone reported by the NETFIT[®] system. After analyzing, the proportion of boys and girls in the different fitness zones results were plotted to provide a direct comparisons of the two assessments.

<u>Validating the estimation method of the NHANES treadmill protocol</u>

Estimation of the maximal oxygen consumption was calculated from the extrapolation of the HR measured at the end of the first two levels of the NHANES protocol according to the algorithm based on NHANES. The HR response and oxygen uptake showed parallel linear increase during the exercise, so the VO_{2max} value can be estimated by monitoring the HR in the knowledge of the sub-maximal intensity level (ACSM, 2009).

Estimated VO_{2max} values of those students who had an unrealistically high estimated VO_{2max} due to extrapolation were maximized at 80 mL/kg/min. Those subjects were excluded from the analysis who had a difference of more than 20 mL/kg/min between the measured and estimated values, and where the estimate was negative.

During the analysis, the similar statistical procedures used to cross-validate the PACER test were applied. The average difference between the estimated and measured VO_{2max} values was analyzed by agegroup and gender with univariate ANOVA test. Standardized and unstandardized beta weights for each bivariate relation were provided, whereas the relation between the estimated and measured VO_{2max} values were analyzed by linear regression analysis based on NHANES. The differences between the methods were analyzed and visualized by the Bland-Altman plot model. Based on the results of the previously described equivalence test, I concluded whether the difference of mean values is within the ± 10% range of measured VO_{2max} values, and therefore can be considered equivalent to the 95% confidence interval of the two one-sided tests.

RESULTS

A total of 482 students participated in both the laboratory and field tests, out of whom 45 reported anomalies that excluded participation in the study or caused abortion of the exercise (abnormal resting ECG, hypertension, ventricular septum defectus, tonsillitis follicularis, diarrhea, previous injury, chondromalatia patellae or medication). A further 78 students were excluded from the statistical analysis after failing to meet one of the expected inclusion criteria, so the total number of subjects included in the analysis was 354.

14% of our sample was overweight and 4,7% obese according to the Centers for Disease and Prevention (CDC) recommendation for children. The boys were taller, heavier, had lower body fat percentages than girls, and achieved better results in the PACER as well as in the treadmill test (p <0,001) with higher physiological parameters (max VE, max O_2pulse, max systolic BP). The symptoms of the metabolic syndrome occurred in 6,7% of the sample (boy: 5,7%, girl: 7,9%).

The reference for boys in the mean values of VO_{2peak} among agegroups vary between 52,54 ± 5,86 ml/kg/min (11 years) and 47,37±8,19 mL/kg/min (12 years), there is a relative stagnation for boys between 12 to 15 years, whereas the values are increasing from the age of 15. The values for girls vary between 47,21±8,24 mL/kg/min (12 years) and 36,28±9,14 mL/kg/min (16 years), and contrary to boys, from the age of 12 mean values are decreasing. The gender and age group together explain 31,6% of the variance of the evolution of VO_{2max} values (p <0.001, $F_{(15)}$=10,395, R^2=0,316). Gender (p<0,001, $F_{(1)}$ =110,801, partial η^2=0.247) and age group (p=0,43, $F_{(7)}$=2,096, partial η^2 = 0,042) separately and its interaction (p<0,001; $F_{(7)}$ = 3,230; partial η^2 = 0,063) also have a significant effect on the evolution of values.

Validation of the new formula of PACER test

The difference of means between PACER test and the VO_{2max} results measured in the laboratory treadmill test range between 0,15 and 4,64 mL/kg/min, and the measured and estimated values between the genders significantly differed ($t_{(352)} = 2,792$; CI: 0,66-3,91; $p<0,01$). The measured values for boys were underestimated by an average of 2,12 mL/kg/min by the prediction model, while the measured values for girls was overestimated by an average of 0,12 mL/kg/min. The differences for 17 ($t_{(31)}=2,428$, CI: 0,58-6,69 mL/kg/min; $p=0,02$) and 18+ year old boys ($t_{(33)}= 2,396$, CI: 0,55-6,69 mL/kg/min; $p=0.,02$) were significant. The number of PACER laps showed moderately strong correlation with the final speed at treadmill ($r_{(354)} = 0,503, p<0,001$), as well as with the measured VO_{2peak} values ($r_{(354)} = 0.563, p <0.001$).

The predicted and measured VO_{2peak} values shared 14,3% and 20,9% of the total variance for boys and girls, respectively, and were low to moderately correlated (for boys: $r_{(197)}=0,378, p<0,001$; for girls: $r_{(168)}=0,457, p<0,001$). A 1 mL/kg/min increase in predicted peak VO_2 was associated with a 0,41 mL/kg/min increase in measured peak VO_2 ($p<0,001$) in boys and 0,64 mL/kg/min increase in girls. The lower LOA for boys was -14,16 mL/kg/min, for girls -15,28 mL/kg/min, whereas the upper LOA for boys +18,66 mL/kg/min and for girls +14,66 mL/kg/min. The LOA were indicative of an error ranging from 28,4% to 37,5% and 33,2% to 34,7% of measured peak VO_2 in boys and girls, respectively. The mean absolute percent error values (taking into account both overestimation and underestimation) were 13,6% for boys and 15,46% for girls. In our sample, the standard error of estimation (SEE) was 6,47 mL/kg/min for boys and 4.91 mL/kg/min for girls.

The average peak VO_2 obtained from the maximal laboratory test for boys was 49,73±7,81 mL/kg/min and the 10% equivalence region ranged from 44,73 mL/kg/min to 54,67 mL/kg/min. Averaged estimated peak VO_2 for boys was 47,51±7,18 mL/kg/min, and the 95% confidence interval for the mean difference (95% CI [44,1; 48,61 mL/kg/min]) was partially equivalent because the lower predicted limit was not within the lower bound for the 10% range. The same comparisons for girls indicated that the average estimated peak VO_2 of 41,28±5,86 mL/kg/min and the 95% confidence interval for the mean difference (95% CI [41,34; 42,76 mL/kg/min]) were within the 10% region of equivalence for this group (10% equivalence range 36,83 mL/kg/min to 45,01 mL/kg/min) and therefore were considered equivalent.

Agreement of the criterion-referenced NETFIT® fitness zone classification based on measured and the estimated PACER test VO_{2peak}

The agreement between classifications of aerobic fitness obtained from measured and predicted peak VO_2 was fair for both boys (Kappa=0,23) and girls (Kappa=0,24). The agreement was 72,0% and 57,1% for boys and girls respectively, on average 65,3%. Overall, the highest balance of sensitivity and specificity was achieved for the HFZ (boys: sensitivity=83,8%, specificity=51,3%; girls: sensitivity=71,1%, specificity=59,4%). The agreement on classifications indicates that the proportion of boys and girls in the HFZ would be equal to 79,4% and 59,2%, respectively, if peak VO_2 was obtained from a maximal treadmill test and would be equal to 76,3% and 59,8%, for boys and girls respectively, if obtained from the PACER test.

Validation of the prediction method of the NHANES treadmill test protocol

The average difference between the measured and estimated VO_{2peak} calculated from HR values varied between 0,24 to 6,96, where the average difference was -2,96 and -1,77 for boys and girls respectively. The differences between the age group ($F_{(7)}=2,402$, partial $\eta^2=0,053$) and the age

and gender interaction ($F_{(7)}$= 2.223, partial η^2=0.049) were fair but significant (p <005) in our sample, altogether defining deviations of 10.5%. Estimated VO_{2peak} values showed a moderate to strong correlation with laboratory values for both sexes (boys $R_{(171)}$ =0,568, p <0,001; girls $R_{(145)}$=0,658, p<0,001). The linear regression line summarizing the relationship between estimated and measured VO_{2peak} values explains 32,2% of total standard deviation ($F_{(1)}$= 80,354, p<0,001) and 43,3% in girls ($F_{(1)}$=109,35; p<0,001). A 1 mL/kg/min increase in predicted peak VO_2 was associated with a 0,42 mL/kg/min increase in measured peak VO_2 (p<0,001) in boys and 0,55 mL/kg/min increase in girls (Figure 11). The standard error of estimation (SEE) was 8,03 mL/kg/min for boys and 7,11 ml/kg/min for girls.

Following the Bland-Altman method, the LOA were indicative of an error ranging from -19,01mL/kg/min to +13,09 mL/kg/min and -16,00 mL/kg/min to +12,46 mL/kg/min for boys and girls respectively, which represents 26% to 38%, and 30,2% to 38,8% range of error for measured peak VO_2 in boys and girls, respectively. The mean absolute percent error values (taking both overestimation and underestimation into account) were 14,44% for boys and 15,12% for girls.

The average peak VO_2 obtained from the maximal laboratory test for boys was 50,1±7,3 mL/kg/min and the 10% equivalence region ranged from 45,1 mL/kg/min to 55,1 mL/kg/min. Mean value of VO_{2peak} obtained from the HR values at the end of NHANES levels, was 47,1±9,7 mL/kg/min. 95% confident interval of the means of error (95% CI: 43,14-49.07 ml/kg/min) was only partially equivalent to the 10% range of agreement of the measured VO_{2peak}, since the lower predicted limit was not within the lower bound of the 10% range. The same comparisons for girls indicated that the average measured peak VO_2 of 39,5±9,4 mL/kg/min of which the 10% equivalence region ranged from 35,6 mL/kg/min to 43,5 mL/kg/min). In the case of the girls, the estimated mean values (39.5 ± 9.4 mL/kg/min) can be considered to be equivalent to the measured values, namely the lower and upper confidence interval value (95% CI: 36,73 – 40,27 ml/kg/min) was within the 10% range of LOA.

KEY FINDINGS OF THE STUDY

- In Hungary standing alone, we measured the aerobic fitness status of 10-19 year-olds on a nationally representative sample in both field tests and laboratory measurements.

- Based on the measured values of complex spiroergometric testing, the maximal oxygen consumption values show a different pattern for boys and girls in respect of age, with significantly better results for boys than girls. The values of girls from the age 11-12 are significantly decreasing to young adults, while boys show relative stagnation in the examined age range.

- Based on the NETFIT® gender and age-specific criteria, the proportion of students achieving the HFZ is significantly lower for girls than for boys.

- It was the first study to cross-validate the new regression model of the PACER test developed in 2014 (Cooper Institute), which is based on the number of laps achieved and age.

- The mean absolute percent error for estimating aerobic capacity from PACER test using the new regression model was approximately 15%

- At group level, the average differences derived from the estimation are within the 10% error limit of the measured VO_{2max} values, and thus are considered statistically equivalent to the data measured in the laboratory.

- The health-oriented fitness zone categorization of the results obtained in the PACER test based on the estimation method is precise enough in the NETFIT®.

- The applicability of the simplified regression model was verified for the Hungarian school-aged children. Involving more variables in the formula is not recommended for easy institutional use.

- Beyond age, gender also has a significant impact on the accuracy of the estimates, but the health-oriented categorization of NETFIT® is based on gender- and age-specific standards, so the usage in the model is negligible.

- In interpreting results, the potential of increased degree of individual mistakes and being conscious of the error of estimates, teachers and practitioners should emphasize the educational value of the results and not use the test for estimating aerobic capacity in individual children

- The PACER test and its estimation model in NETFIT® are reliable and provide appropriate institutional use to categorize health-oriented aerobic fitness levels in Hungarian school-age children, however, no accurate testing/assessment of individual values is recommended.

- We performed a completion study for validation of the NHANES sub-maximal walking test protocol for children unexampled.

- The average error rate of the estimation extrapolating the $HR-VO_2$ linear correlation measured at the end of the NHANES levels is 15%.

- For girls at group level, the average differences derived from the estimation are within the 10% error limit for the measured VO_{2max} values, and thus are considered equivalent to the

data measured in the laboratory. For boys, the lower values have a difference of 1,96 ml/kg/min for which the validity of the method is statistically unacceptable.

- It is recommended to develop a multivariate regression model for the NHANES sub-maximal walking test protocol to improve the accuracy of the estimation of the heart rates at the end of the levels.

LIST OF SCIENTIFIC PUBLICATIONS

(Cumulated impact factors: 12,2; As first author: 2,2)

Kaj, M., Vass, Z., Király A., Hernádi, Á., & Csányi, T. (2017). *A magyar 10-18 éves tanulók egészségközpontú fizikai fittségi állapota: Kutatási jelentés a Nemzeti Egységes Tanulói Fittségi Teszt (NETFIT®) 2015/2016. tanévi országos eredményeiről.* Magyar Diáksport Szövetség, 2017.

Kaj, M., Király A., Hernádi, Á., & Csányi, T. (2017) A NETFIT 2015/2016. tanév országos eredményeinek bemutatása. *Magyar Sporttudományi Szemle*, 18(70), 48-49.

Váczi, M., Tékus, É., Atlasz, T., Cselkó, A., Pintér, G., Balatincz, D., **Kaj, M**., & Wilhelm, M. (2016) Ballroom dancing is more intensive for the female partners due to their unique hold technique. *Physiology International (Acta Physiologica Hungarica)*, 103(3), 392-401. **(IF: 0,9)**

Király, A., **Kaj, M.,** Vass, Z., Boronyai, Z., & Csányi, T. (2016) Köznevelési típusú sportiskolás tanulók egészségközpontú fittségi állapota: Összehasonlító elemzés a 2014/15 tanévi NETFIT® mérések alapján. *Magyar Sporttudományi Szemle*, 2, 1.

Csányi, T., **Kaj, M.,** Vass, Z., Boronyai, Z., & Saint-Maurice, P. F. (2016) Health-related Physical Fitness among 10-18 y Hungarian Students: Results of the 2015 Hungarian National Student Fitness Test (NETFIT®) survey. In: AIESEP International Conference, Laramie, USA, 2016. 1p.

Csányi, T., **Kaj, M.,** Vass, Z., Boronyai, Z., Király, A., & Saint-Maurice, P. F.: *A magyar 10-18 éves tanulók egészségközpontú fizikai fittségi állapota (2015): Kutatási jelentés a Nemzeti Egységes Tanulói Fittségi Teszt (NETFIT®) 2014/2015. tanévi országos eredményeiről.* Magyar Diáksport Szövetség, Budapest, 2016.

Ihász, F., Karsai, I., **Kaj, M.,** Marton, O., Finn, K. J., & Csányi, T. (2015) Characteristics of cardiorespiratory output determining factors among 11–19-year-old boys at rest and during maximal load: Its impact on systolic hypertension. *Acta Physiologica Hungarica*, 102(3), 263-273. **(IF: 0,98)**

Kaj, M., Tékus, É., Juhász, I., Stomp, K., & Wilhelm, M. (2015) Changes in physical fitness of Hungarian college students in the last fifteen years, *Acta Biologica Hungarica*, 66(3), 270-281. **(IF: 0,605)**

Laurson, K. R., Welk, J. G., Marton, O., **Kaj, M.,** & Csányi, T. (2015) Agreement and Diagnostic Performance of FITNESSGRAM®, International Obesity Task Force, and Hungarian National BMI Standards. *Research Quarterly For Exercise And Sport,* 86(S1), S21-S28. **(IF: 1,566)**

Kaj, M., Saint-Maurice, P. F., Karsai, I., Vass, Z., Csányi, T., Boronyai, Z. & Révész, L. (2015) Associations Between Attitudes Toward Physical Education and Aerobic Capacity in Hungarian High School Students. *Research Quarterly For Exercise And Sport,* 86(S1), S74-S81. **(IF: 1,566)**

Saint-Maurice, P. F., Welk, G.J., Finn K. J., & **Kaj, M.** (2015) Cross-Validation of a PACER Prediction Equation for Assessing Aerobic Capacity in Hungarian Youth. *Research Quarterly For Exercise And Sport,* 86(S1), S66-S73. **(IF: 1,566)**

Saint-Maurice, P. F., Laurson, K. R., **Kaj, M.,** & Csányi, T. (2015) Establishing Normative Reference Values for Standing Broad Jump Among Hungarian Youth. *Research Quarterly For Exercise And Sport,* 86(S1), S37-S44. **(IF: 1,566)**

Saint-Maurice, P. F., Laurson, K. R., Karsai, I., **Kaj, M.,** & Csányi, T. (2015) Establishing normative reference values for handgrip among Hungarian Youth. *Research Quarterly For Exercise And Sport,* 86(S1), S29-S36. **(IF: 1,566)**

Vowell, J., Welk, G. J., Saint-Maurice, P. F., Csányi, T., & **Kaj, M.** (2015) Distribution Of Health-related Physical Fitness In Hungarian Youth: An Examination With Fitnessgram Standards. *American College of Sports Medicine 62 nd Annual Meeting, 6 th World Congress on Exercise is Medicine and World Congress on the Basic Science of Exercise Fatigue: Medicine & Science In Sports & Exercise*, 46,5, S558.

Finn, K. J., **Kaj, M.,** Marton, O., Karsai, I., Saint-Maurice, P. F., & Csányi, T. (2015) Accuracy Of The NHANES Treadmill Test In Estimating Aerobic Fitness In Adolescents. *American College of Sports Medicine 62 nd Annual Meeting, 6 th World Congress on Exercise is Medicine and World Congress on the Basic Science of Exercise Fatigue: Medicine & Science In Sports & Exercise*, 46:5, S238.

Kaj, M., Csányi, T., Karsai, I., & Marton, O.: *Kézikönyv a Nemzeti Egységes Tanulói Fittségi Teszt /NETFIT®/ alkalmazásához.* Budapest: Magyar Diáksport Szövetség, 2014.

Kaj, M., Karsai, I., Marton, O., Ihász, F., Finn, K. J., Saint-Maurice, P. F, & Csányi, T. (2014) A NHANES futószalagos protokoll VO2max predikciós módszerének validitási vizsgálata magyar iskoláskorú gyermekeken. XI. Országos Sporttudományi Kongresszus, *Magyar Sporttudományi Szemle,* 18, 11.

Ihász, F., **Kaj, M.**, Csányi, T., Karsai, I., Marton, O., & Vass, Z. (2014) Vérnyomásváltozások jellemzôi (11-19 éves) fiúk körében nyugalomban és a terhelés maximumán. XI. Országos Sporttudományi Kongresszus, *Magyar Sporttudományi Szemle*, 18, 11.

Csányi, T., Welk, G., Saint-Maurice, P. F, **Kaj, M.**, Marton, O., Ihász, F., Laurson, K., Zhu, W., Finn, K. J., & Karsai, I. (2014) Health-related Fitness among 10-18 y Hungarian Students: Results of a nationally representative study with the Hungarian National Student Fitness Test (NETFIT®). In: 10th Annual Meeting and 5th Conference of HEPA Europe, Zurich, Svájc. p. 45.

Csányi, T., **Kaj, M.**, Karsai, I., & Marton, O.: Oktatófilm a NETFIT® alkalmazásához. Budapest, Magyar Diáksport Szövetség, 2014.

Karsai, I., **Kaj, M. ,** Marton, O., Saint-Maurice, P. F., Finn, K. J., Ihász, F., & Csányi, T. (2014). 20 m-es ingafutás teszt alapján VO2max értéket becslô regressziós modellek kritérium alapú validitásvizsgálata 11-18 éves magyar populáción. XI. Országos Sporttudományi Kongresszus, *Magyar Sporttudományi Szemle,* 18, 9.

Marton, O., **Kaj, M.**, Karsai, I., Ihász, F., Finn, K. J., & Csányi, T. (2014) A metabolikus szindróma tüneteinek megjelenési gyakorisága és az aerob kapacitás vizsgálata 11-19 éves magyar tanulókon. XI. Országos Sporttudományi Kongresszus, *Magyar Sporttudományi Szemle*, 18, 9.

Csányi, T., Karsai, I.**, Kaj, M.**, Marton, O., Ihász, F., Welk, G., Zhu, W., Saint-Maurice, P. F., Laurson, K., & Finn, K. J.: *Assessment of health-related fitness in Hungary: the NETFIT, as the Hungarian Fitnessgram initiative.* In: Dragan Milanović, Goran Sporiš (szerk.) 7th International Scientific Conference on Kinesiology: Fundamental and Applied Kinesiology, Opatija, 2014. p. 330.

Csányi, T., Welk, G., Saint-Maurice, P. F., Marton, O., **Kaj, M.**, Zhu, W., & Karsai, I.: *Health-Related Physical Fitness achievement rates among Hungarian High School Students – Results from the Hungarian National Youth Fitness Study.* In: 9th FIEP European Congress and 7th International Scientific Congress "Sport, Stress, Adaptation", Sofia, Bulgária, 2014. p. 10.

Csányi, T., Révész, L., **Kaj, M.**, & Karsai, I.: *Quality Physical Education in Hungary.* In: Hungarian International Conference on Physical Education: HIPE 2014. Budapest, Magyar Diáksport Szövetség, 2014. pp. 8-9.

Karsai, I., **Kaj, M.**, Csányi, T., Ihász, F., Marton, O., & Vass, Z. (2013). Magyar 11-19 éves iskolások egészségközpontú fittségi állapotának keresztmetszeti vizsgálata - Első jelentés az Országos Reprezentatív Iskolai Fittségmérési Program eredményeiről. *Magyar Sporttudományi Szemle*, 14(56), 9-18.

Kaj, M., Németh, J., Tékus, É., & Wilhelm, M. (2013). Physique, body composition and physical fitness of Finish, Hungarian and American adolescents. *Exercise and Quality of Life*, 5(1), 19-29.

Váczi M., Tékus, É., **Kaj, M.,** Kőszegi, T., Ambrus, M., Tollár, J., Atlasz, T., Szabadfi, K., & Karsai, I. (2013). Changes in metabolic and muscle damage indicators following a single bout of jump training on stair versus at level. *Acta Physiologica Hungarica,* 100(4), 445-456. (**IF:1,06)**

Tékus, É., **Kaj, M.**, Szabó, E., Szénási, L. N., Figler, M., Gábriel, R., & Wilhelm, M.: *The effect of the sympathetic nervous system and dehydration on salivary lactate concentration*. In: Szabó István (szerk.): 1st International Doctoral Workshop on Natural Sciences, Pécs, 2012, p. 52

Kaj, M.: *Changes in health-related fitness status of the Hungarian university students in the last decades*. In: 20th International Congress on Sports Sciences for Students, Budapest, 2012, p. 75.

Kaj, M., Tékus, É., Juhász, I., Stomp, K., & Wilhelm, M.: *Immobility Stress - How has the health-related fitness status changed in the Hungarian young adults in the last decades?* In: Szabó István (szerk.): 1st International Doctoral Workshop on Natural Sciences, Pécs, 2012, p. 22.

Tékus, É., **Kaj, M.**, Szabó, E., Szénási, N. L., Kerepesi, I., Figler, M., Gábriel, R., & Wilhelm, M. (2012). Comparison of blood and saliva lactate level after maximum intensity exercise. *Acta Biologica Hungarica*, 63(S1), 89-98. **(IF: 0,82).**