VALIDITY OF TIME TO EXHAUSTION ON A FIXED INCREMENTAL TREADMILL PROTOCOL AS A MEASURE OF AEROBIC FITNESS IN 10-YEAR OLD CHILDREN

Eivind Aadland[1], Turid Skrede[1], Asgeir Mamen[2], Geir Kåre Resaland[1,3]

[1]Sogn og Fjordane University College, Faculty of Teacher Education and Sports, Box 133, 6851 Sogndal, Norway

[2]Norwegian School of Health Sciences – Kristiania University College,Institute of Physical Activity, Nutrition and Public Health, Oslo, Norway, Box 1190 Sentrum, 0107 Oslo, Norway

[3]Sogn og Fjordane Center for Health Research, Fřrde Central Hospital. Box 1000. 6807 Fřrde, Norway

Original scientific paper

Abstract

Purpose: High aerobic fitness is important for children's metabolic health. Direct measurement of peak oxygen consumption (VO_{2peak}) is often not feasible, but time to exhaustion (TTE) on a fixed incremental treadmill protocol has been used as a measure of aerobic fitness. The aim of the present study was to determine the validity of TTE as a measure of VO_{2peak} in 10-year old children. Methods: A total of 118 10-year old children (67 boys and 51 girls) were recruited from one school and performed an incremental VO_{2peak}-test where oxygen consumption and TTE were assessed. Of these, 113 children provided valid data and served as subjects for the analysis of validity, which was performed using linear regression analysis. Results The equation to estimate VO_{2peak} was: VO_{2peak} = (coefficient (95% bootstrapped confidence interval)) 26.909 (12.910 to 40.879) + 0.063(0.048 to 0.079) · TTE – 3.225 (-5.391 to -1.000)· gender - 0.330 (-0.460 to -0.196)· body mass (R^2 = 0.72, SEE = 4.82, $p < .001$). Conclusions Time to exhaustion yielded valid estimates of VO_{2peak} for 10-year old children. However, researchers should be aware of the amount of noise in indirect tests to estimate aerobic fitness. Future studies should assess reliability and external validity to determine if the current findings are reproducible and valid across different contexts.

Keywords: *cardiorespiratory fitness; maximal oxygen consumption; measurement; agreement*

INTRODUCTION

High aerobic fitness is consistently associated with a favorable metabolic risk profile in children (Andersen et al. 2008b; Anderssen et al. 2007). In adults, the relationship between aerobic fitness and health becomes evident through increased incidence of cardiovascular disease and mortality in those having a poor fitness level compared to their fitter peers (Fogelholm 2010; Myers et al. 2002). On this background, precise and simple determination of aerobic fitness in children is a prerequisite to inform the society regarding targets for and effectiveness of early public health action.

Various indirect field tests have been developed to estimate maximal oxygen consumption in certain populations (Jorgensen et al. 2009). In children, the 20 m multistage shuttle run test (MSRT) (Leger and Lambert 1982; Leger et al. 1988) has been applied extensively worldwide (Olds et al. 2006; Ortega et al. 2011). Also, several other tests, both intermittent tests, like the Andersen test (Ahler et al. 2012; Andersen et al. 2008a) and the Yo-Yo test (Ahler et al. 2012), along with various continuous running tests (5-minute run, 6-minute run, 12-minute run, 15-minute run, half-mile run etc.) (Castro-Pinero et al. 2009; Cooper 1968; Jackson and Coleman 1976; Macnaughton et al. 1990; van Mechelen et al. 1986) are used as measures of aerobic fitness. Such tests are often preferred over the direct measurement of peak oxygen consumption (VO_{2peak}) determined from an incremental treadmill or bicycle protocol to voluntary exhaustion, due to their simpler administration and thus superior feasibility. However, two recent external validation studies have shown that current equations to estimate VO_{2peak} from the MSRT in 8 – 13 year old children are inadequate (Batista et al. 2013; Melo et al. 2011) due to biased estimates and large individual errors. Thus, studies applying indirect tests such as the MSRT are prone to report erroneous results and may misinform the society regarding the level and importance of aerobic fitness. Large errors on an individual basis (noise in the x-variables) leads to attenuated associations with health due to regression dilution bias (Hutcheon et al. 2010), which increase the probability of performing type II errors. Moreover, limited indoor space or varying seasons may leave field test less convenient and reliable than laboratory tests.

Direct measurement of VO_{2peak} is time-consuming and requires expensive equipment and highly trained test personnel, and is therefore not feasible in many settings. However, performance (time to exhaustion (TTE), maximal speed or workload) on such tests can easily be obtained without any sophisticated equipment and may provide researchers with valid data regarding exercise capacity, as indicated in well-trained adults (Hawley and Noakes 1992; Noakes et al. 1990; Schabort et al.

2000). Moreover, a modest to good relationship between performance and VO_{2peak} have been shown for a treadmill (Pettersen et al. 2001) and cycle ergometer (Andersen 1995) protocolsin subjects aged 8 to 28 years old. However, we are not aware of any studies that have suggested a regression equation for estimation VO_{2peak} and thoroughly examined the validity of TTE obtained during a fixed incremental treadmill protocol as a measure of VO_{2peak} in children.

The aim of the present study was therefore to determine the validity of TTE as determined from an incremental treadmill protocol to exhaustion as a measure of aerobic fitness using VO_{2peak} as the criterion-measure in 10-year old children.

METHODS

Participants

All 121 children in fifth grade (10-year olds) at one school in the western part of Norway over two subsequent school years were invited to participate in the study. A total of 118 children (67 boys and 51 girls; 58 during 2012-2013 and 60 during 2013-2014) were included in the study. Three children were excluded from the study (one for each of the following: short of growth, heart problems, skeletal disease).
Children and their parents were given thorough oral and written information regarding the study protocol. Each child orally agreed to participate in the study and written informed consent was obtained from each child's parent(s)/guardian(s) prior to inclusion in the study. The study met the standards of the Declaration of Helsinki and was approved by the Regional Committee for Medical Research Ethics (RCMRE West) in Norway.

Instruments

Children performed one fixed incremental treadmill test to voluntary exhaustion to measure their time to exhaustion and peak oxygen consumption (VO_{2peak}). The inclination of the treadmill (PPS 55, Woodway, GmbH, Weil am Rhine, Germany) was constant of 5.3% during the test. Children started to walk at 5 km $h^{-1}$for 5 minutes. Thereafter the speed increased with 1 km h^{-1}each minute until the children were exhausted. Oxygen consumption was measured using the Moxus Modular Metabolic System (AEI Technologies Inc., Pittsburgh, PA, USA). A two-point gas calibration according to known concentrations and calibration according to atmospheric pressure were performed each test day. Volume calibration of the breathing valve (model 2700 placed on the face mask Vmask 7400; both Hans Rudolph Inc., Shawnee, KS, USA) was performed between each test using a 3 l syringe (Series 5530, Hans Rudolph, Shawnee, KS, USA).The oxygen analyzer has shown to be reliable and valid compared to the Douglas bag technique(Medbř et al. 2012; Rosdahl et al. 2013). To prevent injuries in case of falls during the test, children performed the test with a safety rope connected to a chest belt system from Cosmos (h/p/cosmos sports & medical GmbH, Nussdorf-Traunstein, Germany).Throughout the test, a test-assistant was in charge of the safety of the subject by tightly holding the safety rope. If the subject stumbled, the test assistant could pull the rope, thereby raising the subject, and prevent a fall.The child and parent(s)/guardian(s) were also informed of test procedures before testing, and the child's parent(s)/guardian(s) were allowed and encouraged to observe the testing.

After each test, test leader and associates discussed several subjective criteria to verify a near maximal performance; hyperpnoea, unsteady running pattern, and verbal and body language clearly indicating that the child wanted to stop the test, despite repeated strong verbal encouragement to continue. Additionally, the respiratory exchange ratio (RER) andmaximal heart rate (HR_{peak}) (Polar Electro OY, Kempele, Finland) were noted. The presence of a plateau were defined as an increase in oxygen consumption of less than 50% of the expected increase over one or more minute(s)(< 2.9 $mlkg^{-1}min^{-1}$). The reliability of VO_{2peak} tested directly in children is shown to be approximately 4%, which compares favorably with the reliability of adults' results (Welsman et al. 2005).

The VO_{2peak} is presented as absolute (l min^{-1}) and relative values ($mlkg^{-1}min^{-1}$), defined as the highest value of two successive 30 s. measurements. Time to exhaustion was recorded to the nearest second. Height and body mass were measured without shoes and socks before the children started the VO_{2peak} test. Height was measured to the nearest 0.1 cm using a wall-mounted stadiometer. Body mass was measured to the nearest 0.1 kg (subtracting 0.2 kg for light clothes) using an electronic scale (Seca 770, SECA GmbH, Hamburg, Germany).

Children were instructed not to eat the last two hours prior to testing and to engage in normal physical activity the day before the test and the day of testing.

Statistical analyses

The anthropometric characteristics and data on $VO_{2peak}$and TTE are presented as the mean values and standard deviation (SD). Differences between genders and classes were tested using an independent group's t-test (continuous variables) and a Chi-Square test (dichotomous variables).

Validity of TTE as a measure of VO_{2peak} was assessed using Pearson's r, linear regression and Bland Altman plots in two steps: 1) We split our sample in two (children tested during 2012-2013 vs. 2013-2014) and made an equation based on the first group (training dataset) to predict VO_{2peak} for the second group (testing dataset). Three variables were included in the model (VO_{2peak} = a + b· TTE + c · gender + d · body mass, [boys = 0; girls = 1]). The predicted and measured VO_{2peak} were then compared using linear regression and a Bland Altman plot, showing the difference between two subsequent tests as a function of the mean of the two tests (Bland and Altman 1986). Because the data were deemed to be homoscedastic, the standard error of the measurement (SEM) and limits of agreement (LoA) were calculated according to Hopkins (Hopkins 2000) (SEM = SD of the differences/$\sqrt{2}$; LoA = SD of the differences · 1.96).Means were compared using a one-sample t-test. 2) As the derived equation was found to perform adequately in the new sample, we finally calculated a new equation based on the whole sample (n = 113) using the following model (VO_{2peak} = a + b· TTE + c · gender + d · body mass, [boys = 0; girls = 1]). The final model is reported as regression coefficients with 95% bootstrapped CIs.

The between-subject variance in work economy were calculated as the variance for the random intercept term for subjects divided by the total variance over the whole incremental protocol as assessed on a minute-by-minute basis (using the mean of the two last minutes for 5 km h^{-1} and the last 30 seconds of other speeds until one minute prior to exhaustion) using a linear mixed-effect model. To test whether work economy and/or the presence of a plateau for VO_{2peak} could account for some of the unexplained variance in the relationship between TTE and VO_{2peak}, work economy and plateau (0 vs. 1) were included as independent variables in the final regression model. Work economy used in this model were calculated as the mean VO_2 between 5 and 8 km h^{-1}, and reported as ml $O_2 kg^{-1} min^{-1} km^{-1}$.

All analyses were performed using SPSS v. 20 (IBM SPSS Inc., Armonk, New York, USA). A p-value <.05 indicated statistically significant findings.

RESULTS

Subjects' characteristics

Except for a significantly higher VO_{2peak} and TTE in boys compared to girls (p≤ .002), t here were no significant differences between genders or the two subsamples (classes) included (p>.095) (table 1). Of the total sample included (n = 118), 113 children provided valid data for VO_{2peak} and TTE (two children did not perform the test, one child were excluded for not performing a maximal test and two children were excluded due to technical errors).

Table 1. Subject's characteristics (mean (SD)).

	Overall	**Boys**	**Girls**	**1. year**	**2. year**
Number (%)	118	67 (57)	51 (43)	58 (49)	60 (51)
Age (years)	10.3 (0.3)	10.3 (0.3)	10.3 (0.3)	10.3 (0.3)	10.2 (0.3)
Height (cm)	143.4 (5.8)	143.9 (6.2)	142.7 (5.1)	143.5 (5.5)	143.3 (6.0)
Body mass (kg)	38.0 (8.4)	38.3 (9.5)	37.6 (6.8)	38.2 (9.0)	37.7 (8.0)
Body mass index (kg m^{-2})	18.4 (3.2)	18.3 (3.6)	18.4 (2.7)	18.4 (3.3)	18.3 (3.2)
% overweight/obese	20.3	20.9	19.6	22.4	18.3
Waist circumference (cm)	64.3 (8.6)	64.9 (9.5)	63.6 (7.3)	63.9 (8.6)	64.7 (8.8)
VO_{2peak} (l min^{-1})*	2.00 (0.34)	2.10 (0.35)	1.88 (0.29)	2.06 (0.33)	1.96 (0.35)
VO_{2peak} (ml$kg^{-1}min^{-1}$)*	54.0 (9.0)	56.6 (9.1)	50.5 (7.5)	55.4 (9.5)	52.6 (8.2)
TTE (s)*	652 (87)	672 (97)	624 (62)	658 (95)	646 (79)

*overall n = 113; n = 66 for boys and 47 for girls; n = 57 for 1. year and 56 for 2. year

Maximal heart rate (HR_{peak}) and respiratory exchange ratio (RER) obtained during the VO_{2peak} test were (mean (SD)) 201 (8.9) beats minute^{-1}and 1.07 (0.07), respectively. A total of 22 (20%) children obtained a plateau in VO_{2peak}. Besides a significantly higher RER in girls (1.08 (0.06)) compared to boys (1.05 (0.08)) (p = .032), there were no differences between genders or classes (p > .418).

Validity and prediction equation

The bivariate relationships between TTE and VO_{2peak} were r = 0.79 (p < .001).

To develop an equation to predict VO_{2peak} from TTE, we initially split our sample in two to perform a validation of our equation in an independent dataset. The children included during 2012-2013 served as the training dataset (n = 57) from which the equation was developed, whereas the group included during 2013-2014 served as the testing dataset (n = 56). The equation developed was as follows (regression coefficients and 95% CI): VO_{2peak} = 31.879(17.243 to 46.514) + 0.059 (0.044 to 0.075)· TTE – 4.475 (-7.069 to -1.882)· gender - 0.355 (-0.517 to -0.194) · body mass (R^2 = 0.77, standard error of the estimate (SEE) = 4.69 $mlkg^{-1}min^{-1}$, p < .001, boys = 0, girls = 1). Predicted vs. measured VO_{2peak} yielded R^2= 0.66 and SEE = 4.84. Mean predicted values were higher than measured VO_{2peak} (54.7 (7.0) vs. 52.6 (8.2) $mlkg^{-1}min^{-1}$, p = .002). Typical error (SEM) was 3.40 $mlkg^{-1}min^{-1}$and limits of agreement were ± 9.43 $mlkg^{-1}min^{-1}$ (± 18% of mean VO_{2peak}) between the predicted and measured values (figure 1).

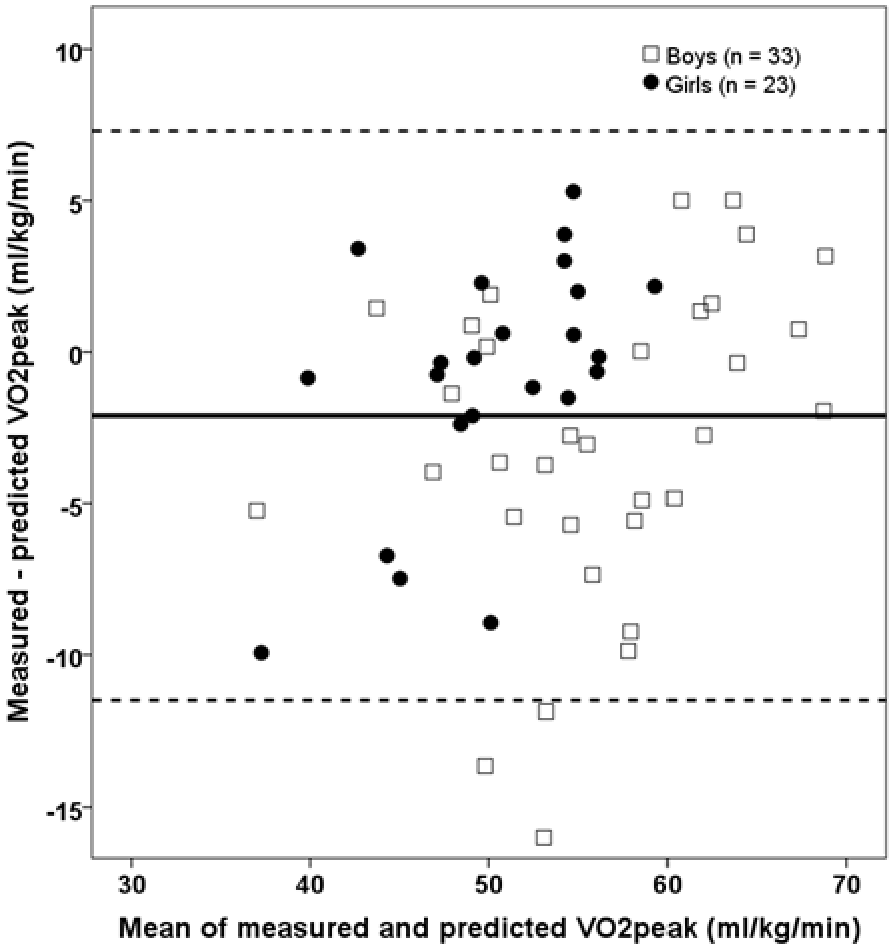
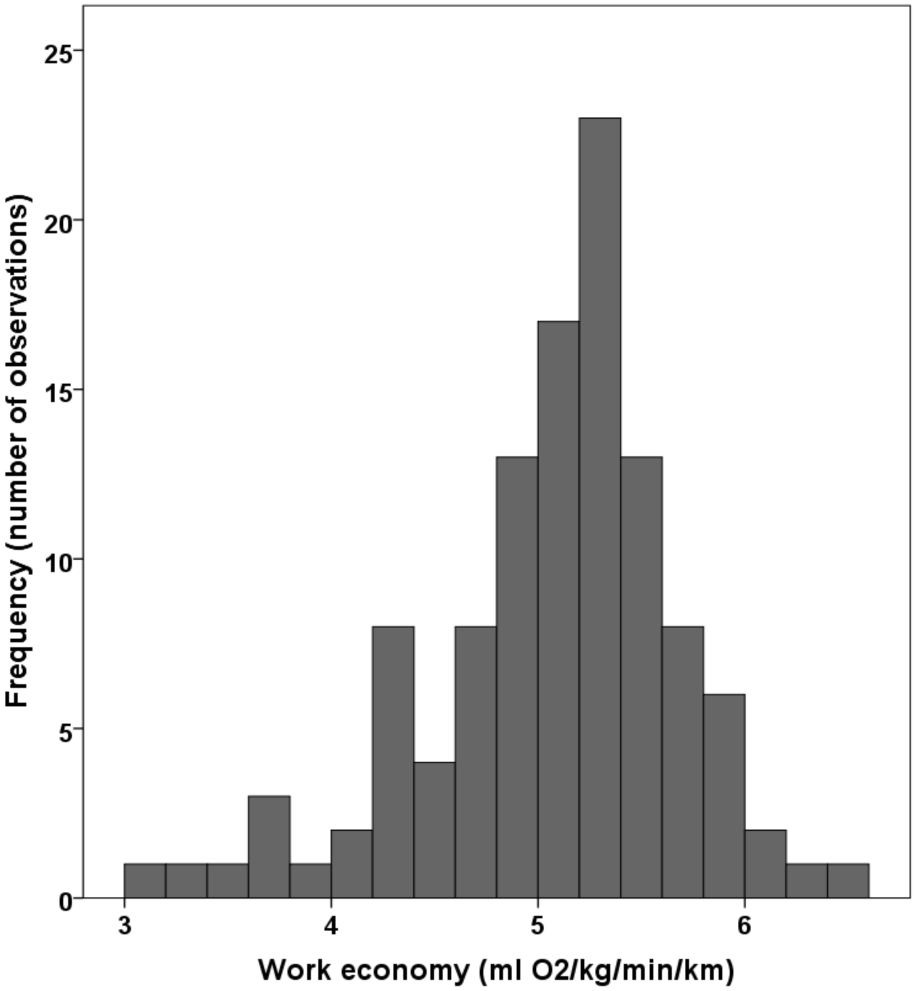


Figure 1. Bland Altman plot showing the differences between measured VO_{2peak} and predicted VO_{2peak} from TTE as a function of the mean of the two values in class 2 (testing dataset) based on the regression equation derived from class 1 (training dataset) (VO_{2peak} = 31.879 (17.243 to 46.514) + 0.059 (0.044 to 0.075) · TTE – 4.475 (-7.069 to -1.882)· gender - 0.355 (-0.517 to -0.194)· body mass (R^2 = 0.77, standard error of the estimate (SEE) = 4.69$mlkg^{-1}min^{-1}$, p < .001, boys = 0, girls = 1). The solid line is the mean difference; dotted line is limits of agreement (bias ± 1.96 SD of the difference).

As the derived equation performed sufficiently in the independent testing set, we established a new equation based on the whole sample (regression coefficients and 95% bootstrapped CI): VO_{2peak} = 26.909 (12.910 to 40.879) + 0.063(0.048 to 0.079)· TTE – 3.225 (-5.391 to -1.000) · gender - 0.330 (-0.460 to -0.196)· body mass (R^2 = 0.72, SEE = 4.82 $mlkg^{-1}min^{-1}$, p < .001, boys = 0, girls = 1, n = 113).

As the findings indicated a certain degree of errors in estimates of VO_{2peak} based on TEE, we explored how individual variability in work economy and the presence of a plateau influenced the results. Children exhibited great variation in work economy, as the variance of intercepts accounted for 77% of the total variation in O_2-consumption during the protocol. Mean energy expenditurefor 5 to 8 km h^{-1}varied from 3.19 (minimum) to 6.52 (maximum) ml $O_2kg^{-1}min^{-1}$ km^{-1}, having a group mean of 5.08 (0.61) ml $O_2kg^{-1}min^{-1}$ km^{-1}(figure 2). After controlling for body mass, girls exhibited superior work economy compared to boys (mean difference: -0.34 (95% CI -0.54 to -0.14) ml $O_2kg^{-1}min^{-1}$ km^{-1}). When included in the final regression model, both work economy ($p < .001$) and the presence of a plateau ($p = .003$) were significant predictors of VO_{2peak}. Inclusion of these variables (to a total of 5 predictors) increased the model fit from $R^2 = 0.72$ and SEE = 4.82 $mlkg^{-1}min^{-1}$ (based on TTE, gender and body mass), to $R^2 = 0.87$ and SEE = 3.27 $mlkg^{-1}min^{-1}$.

Figure 2.Histogram showing the inter-individual distribution of work economy.



DISCUSSION

The present study has proposed an equation to estimate VO_{2peak} from TTE derived from a fixed incremental treadmill protocol. We suggest using TTE from such a treadmill protocol to test aerobic fitness, without performing any sophisticated measures of oxygen consumption, provides researchers with a valid measure of aerobic fitness that is superior to other indirect field tests.

Despite the main finding that TTE may provide valid estimates of VO_{2peak}, our findings clearly showed that TTE was not equivalent to VO_{2peak}. Consistent with previous external validation studies of the MRST (Batista et al. 2013; Melo et al. 2011), we found relatively wide limits of agreement for estimates of VO_{2peak}. Our results showed that one must expect individual deviations in VO_{2peak} of ± 9.43 $mlkg^{-1}min^{-1}$ (± 18% of mean VO_{2peak}) based on estimation from TTE, gender and body mass, although the typical error was considerably less (SEM = 3.40 $mlkg^{-1}min^{-1}$). Clearly, VO_{2peak} and TTE are related, but they are measures of two distinct features: VO_{2peak} is a measure of one's capacity to consume oxygen undermaximal work, whereas TTE is a measureof performance. Amongst many candidates to explain the less than perfect relationship between VO_{2peak} and TTE, work economy may be one of the most important. Consistent with previous studies (Morgan et al. 2000; Morgan et al. 2002) we found great individual variation in work economy between children. Morgan et al (Morgan et al. 2002) reported a 56% difference in oxygen consumption between 10-year old children (highest VO_2– lowest VO_2/mean VO_2) at approximately 6.4 km h^{-1}. The corresponding finding in the present study (mean VO_2 between 5 and 8 km h^{-1}) was 69%, and figure 2 shows that the differences in work economy between the most extreme children were more than two-fold. Some of this variation was attributed to girls exhibiting better work economy than boys in the present study, as also shown previously (Ariens et al. 1997; Morgan et al. 1999). Moreover, those children who exhibited a plateau at VO_{2peak} were found to obtain a lower VO_{2peak} for a given TTE, as could be expected. Finally, neither VO_{2peak} nor TTE is measured without error. Typical coefficients of variation for repeated measurements of VO_{2peak} in children are 4-5% (Fredriksen et al. 1998; Welsman et al. 2005). Studies has shown that variation in TTE or performance on a maximal exhaustive protocol are similar to (Andersen 1995) or somewhat greater (Fredriksen et al. 1998) than for VO2peak. Thus, there are several sources of variability that could influence the relationship between TTE and VO_{2peak}.

The less than perfect relationship between VO_{2peak} and TTE may limit the usefulness of TTE to estimate VO_{2peak} on an individual basis, or at least, it should remind researchers to be aware of the amount of "noise" in indirect tests, because they may dilute any relationship between aerobic fitness and health or some other outcome of interest to a great degree (Hutcheon et al. 2010). However, the importance of such a dilution may depend on how we understand and interpret our results in relation to how we believe aerobic fitness is related to health. The dilution effect would be obvious if we believe VO_{2peak} is the "true" health-related measure, and we apply TTE as a proxy measure of VO_{2peak}. However, if TTE and VO_{2peak}, which are both measured with errors, provide a measure of the body's capacity to undertake maximal aerobic work, then both measures are representations of the underlying "latent" variable "aerobic fitness", and associations with health could be expected to be similar. As far as we are aware of, this hypothesis has not been subject to investigation.Thus, future studies should directly compare the use of direct and indirect measures of aerobic fitness regarding their ability to predict health outcomes.

Although a high VO_{2peak} is a prerequisite for good race or time trial performance, maximal work or speed performed on the protocol (which is equivalent to TTE on a fixed protocol) has proved to be as good as or superior to VO_{2peak}as a measure of performance in adults in many studies (Hawley and Noakes 1992; Noakes et al. 1990; Schabort et al. 2000). Thus, for training interventions where the primary outcome would be performance (not health), TTE would probably be the preferred measure of aerobic fitness.Moreover, TTE can probably provide a relatively precise measure of change in aerobic fitness over time in longitudinal studies including pre- and post-testing, given that work economy in children seem to be rather stable over days (Amorim et al. 2009) and years (Morgan et al. 2004), and probably are marginally influenced by physical training as indicated in adults (Morgan et al. 1995). However, we are left to render this a speculation, since we did not perform the test more than once in the current investigation. Therefore, further research should perform repeated measurements to investigate this question.

To create a new equation we performed an external validation within our sample (using a training dataset and an independent test dataset) prior to establishing the final equation (Steyerberg 2009). The procedure showed a significant bias, but no significantly different slope between the predicted and measured VO_{2peak} in the test dataset. We have no good explanation for the biased estimate, other than ascertaining that VO_{2peak} was 2.8 $mlkg^{-1}min^{-1}$ higher in class 1 (the training dataset) than in class 2 (the testing dataset) (p = .095), which is very similar to the bias found (2.1 $mlkg^{-1}min^{-1}$). There was no difference in RER, HR_{max} or the proportion that obtained a plateau in VO_{2peak} between the two classes. The merging of the two subgroup datasets and establishment of a final equation for the total group (n = 113), makes the equation more stable and provide more accurate estimates of VO_{2peak} in new samples. Nevertheless, the finding of a biased estimate in the testing dataset does indicate that the equation may be susceptible to render biased estimates in varying contexts (i.e. varying sample characteristics). This finding therefore clearly indicates that the proposed equation need to be further externally validated in samples having characteristics heterogeneous to the present sample from which the equation was created.

Considering that measurement of oxygen consumption requires sophisticated equipment and highly trained test personnel, we believe TTE on a fixed incremental treadmill protocol might be well-suited to measure aerobic fitness in some contexts. However, given the nature of the maximal treadmill protocol being quite laborious as it is limited to the testing of one child at a time, various indirect field-tests, like the MSRT (Leger and Lambert 1982; Leger et al. 1988), the Andersen test (Ahler et al. 2012; Andersen et al.

2008a) or other tests (Ahler et al. 2012; Castro-Pinero et al. 2009; Cooper 1968; Jackson and Coleman 1976; Macnaughton et al. 1990; van Mechelen et al. 1986) still have superior feasibility. The main benefit when choosing a field test, is that a relatively large number of subjects can be tested in parallel, thus, leaving the tests very efficient. However, a perfectly feasible test has no value, if the test is not valid. Two recent external validation studies found the MSRT to be unsuitable to measure aerobic fitness on an individual level (Batista et al. 2013; Melo et al. 2011), thus, better indirect measures of aerobic fitness is needed.

Practical aspects

Choosing the best test for a given purpose requires a thorough consideration of pros and cons in a given context, to reach an acceptable balance between several important features. We believe TTE could be applied as a valid laboratory-measure of aerobic fitness in studies where field tests are considered to provide too large limits of agreement, or where such tests are unpractical due to restricted indoor space (gym hall less than 20 m long or wide) or varying seasons (where standardization of outdoor running tests are severely compromised), but where an oxygen-analyzer or personnel to handle such equipment are unavailable. Moreover, we suggest TTE would perform well as a measure of aerobic fitness in intervention studies where a limited number of participants are to be tested. However, the reliability and thus the sensitivity of TTE on a fixed incremental treadmill protocol should be determined before such use is initialized.

Strengths and limitations

The main strength of the present study is the inclusion of a relatively large sample of heterogeneous children of both genders. The sample size made it possible to perform an external validation of our equation and arrive at relatively stable estimates for VO_{2peak}. Still, the present study has several limitations. First, the sample used for the external validation of our equation may be a limitation of the present study. One could argue that our test dataset were not fully independent, as the children composing both datasets came from the same school and performed the tests in the same laboratory, led by the same testers (Steyerberg 2009). Thus, the equation could be expected to perform worse in other contexts, as discussed. Second, our sample was restricted to 10-year old children. Although the homogeneity of our sample may be a strength in terms of internal validity, as running economy improves with age (Ariens et al. 1997; Rowland 2005), the equation suggested to estimate VO_{2peak} in the present study may not be valid in other age-groups.Third, we did not assess reliability of TTE in the present study. Therefore, some of the discrepancy between TTE and VO_{2peak}could be caused by variability in TTE over time. Furthermore, assessment of reliability is critical to determine if TTE could be applied as a reliable measure in longitudinal studies of aerobic fitness.

Conclusions

We conclude that TTE yielded valid estimates of VO_{2peak} for 10-year old children. However, a certain amount of individual variability was found for estimates of VO_{2peak} based on TTE, which to some extent could be explained by between-subject variability in work economy and the presence of a plateau for VO_{2peak}. Researchers should be aware of the amount of noise in indirect tests to estimate aerobic fitness, because "real" relationships between aerobic fitness and health may be diluted and increase the probability of performing type II errors. In the future, studies of reliability and external validation studies should be performed to determine if the current findings are reproducible and valid across different contexts.

Acknowledgements

We thank all children and teachers at the participating school for their excellent cooperation during the data collection. We also thank students at SognogFjordane University College for their assistance during the test sessions.

Conflict of interest statement

The authors declare that they have no competing interests.

REFERENCES

1. Ahler T, Bendiksen M, Krustrup P, Wedderkopp N (2012). Aerobic fitness testing in 6- to 9-year-old children: reliability and validity of a modified Yo-Yo IR1 test and the Andersen test. Eur J Appl Physiol 112:871-876
2. Amorim PRS, Byrne NM, Hills AP (2009). Within- and between-day repeatability and variability in children's physiological responses during submaximal treadmill exercise. Res Q Exerc Sport 80:575-582

3. Andersen LB (1995) A maximal cycle exercise protocol to predict maximal oxygen uptake. Scand J Med Sci Sports 5:143-146
4. Andersen LB, Andersen TE, Andersen E, Anderssen SA (2008a). An intermittent running test to estimate maximal oxygen uptake: the Andersen test. The Journal Of Sports Medicine And Physical Fitness 48:434-437
5. Andersen LB, Sardinha LB, Froberg K, Riddoch CJ, Page AS, Anderssen SA (2008b). Fitness, fatness and clustering of cardiovascular risk factors in children from Denmark, Estonia and Portugal: the European Youth Heart Study. Int J Pediatr Obes Suppl 1:58-66
6. Anderssen SA, Cooper AR, Riddoch C, Sardinha LB, Harro M, Brage S, Andersen LB (2007). Low cardiorespiratory fitness is a strong predictor for clustering of cardiovascular disease risk factors in children independent of country, age and sex. EurJCardiovasc Prev & Rehabil 14:526-531
7. Ariens GAM, vanMechelen W, Kemper HCG, Twisk JWR (1997). The longitudinal development of running economy in males and females aged between 13 and 27 years: The Amsterdam Growth and Health Study. Eur J Appl Physiol 76:214-220
8. Batista MB et al. (2013). Validity of equations for estimating VO2peak from the 20-m shuttle run test in adolescent aged 11-13 years. J Strength Cond Res 27:2774-2781
9. Bland JM, Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310
10. Castro-Pinero J, Ortega FB, Mora J, Sjostrom M, Ruiz JR (2009). Criterion related validity of 1/2 mile run-walk test for estimating VO2peak in children aged 6-17 years. Int J Sports Med 30:366-371
11. Cooper KH (1968). A means of assessing maximal oxygen intake. Correlation between field and treadmill testing. JAMA 203:201-204
12. Fogelholm M (2010). Physical activity, fitness and fatness: relations to mortality, morbidity and disease risk factors. A systematic review. Obes rev 11:202-221
13. Fredriksen PM, Ingjer F, Nystad W, Thaulow E (1998) Aerobic endurance testing of children and adolescents - a comparison of two treadmill-protocols. Scand J Med Sci Sports 8:203-207
14. Hawley JA, Noakes TD (1992). Peak power output predicts maximal oxygen uptake and performance time in trained cyclists. Eur J Appl Physiol 65:79-83
15. Hopkins WG (2000). Measures of reliability in sports medicine and science. Sports Med 30:1-15
16. Hutcheon JA, Chiolero A, Hanley JA (2010). Random measurement error and regression dilution bias. Br Med J 340:c2289
17. Jackson AS, Coleman AE (1976). Validation of distance run tests for elementary-school children. Res Q 47:86-94
18. Jorgensen T, Andersen LB, Froberg K, Maeder U, Smith LV, Aadahl M (2009). Position statement: Testing physical condition in a population - how good are the methods? Eur J Sport Sci 9:257-267
19. Leger LA, Lambert J (1982). A maximal multistage 20-m shuttle run test to predict VO2max. Eur J Appl Physiol 49:1-12
20. Leger LA, Mercier D, Gadoury C, Lambert J (1988) The multistage 20 metre shuttle run test for aerobic fitness. J Sports Sci 6:93-101
21. Macnaughton L, Croft R, Pennicott J, Long T (1990). The 5-minute and 15-minute runs as predictors of aerobic capacity in high-school-students. J Sports Med Physical Fitness 30:24-28
22. Medbř J, Mamen A, Beltrami F (2012). Examination of the Moxus Modular Metabolic System by the Douglas-bag technique. Applied Physiology, Nutrition & Metabolism 37:860-871
23. Melo X, Santa-Clara H, Almeida JP, Carnero EA, Sardinha LB, Bruno PM, Fernhall B (2011). Comparing several equations that predict peak VO2 using the 20-m multistage-shuttle run-test in 8-10-year-old children European Journal of Applied Physiology 111:839-849 doi:10.1007/s00421-010-1708-z
24. Morgan D, Tseh W, Caputo JL, Craig I, Keefer D, PE M (2000). Variability in the aerobic demand of walking and running in young children Med Sci Sports Exerc 32:S116
25. Morgan DW, Bransford DR, Costill DL, Daniels JT, Howley ET, Krahenbuhl GS (1995). Variation in the aerobic demand of running among trained and untrained subjects Med Sci Sports Exerc 27:404-409
26. Morgan DW, Tseh W, Caputo JL, Craig IS, Keefer DJ, Martin PE (1999). Sex Differences in Running Economy of Young Children. Pediatr Exerc Sci 11:122
27. Morgan DW et al. (2002). Longitudinal profiles of oxygen uptake during treadmill walking in able-bodied children: the locomotion energy and growth study. Gait Posture 15:230
28. Morgan DW et al. (2004). Longitudinal stratification of gait economy in young boys and girls: the locomotion energy and growth study. Eur J ApplPhysiol91:30-34
29. Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood JE (2002) Exercise capacity and mortality among men referred for exercise testing N Engl J Med 436:793-801
30. Noakes TD, Myburgh KH, Schall R (1990). Peak treadmill running velocity during the VO2 max test predicts running performance. J Sports Sci 8:35-45

31. Olds T, Tomkinson G, Leger L, Cazorla G (2006). Worldwide variation in the performance of children and adolescents: An analysis of 109 studies of the 20-m shuttle run test in 37 countries. J Sports Sci 24:1025-1038
32. Ortega FB, Artero EG, Ruiz JR, Espana-Romero V, Jimenez-Pavon D, Vicente-Rodriguez G et al. (2011). Physical fitness levels among European adolescents: the HELENA study. Br J Sports Med 45:20-29
33. Pettersen SA, Fredriksen PM, Ingjer E (2001). The correlation between peak oxygen uptake VO2peak and running performance in children and adolescents. Aspects of different units. Scand J Med Sci Sports 11:223
34. Rosdahl H, Lindberg T, Edin F, Nilsson J (2013). The Moxus Modular metabolic system evaluated with two sensors for ventilation against the Douglas bag method. Eur J Appl Physiol 113:1353-1367
35. Rowland T (2005). Children's exercise physiology. 2nd edn. Human Kinetics, Champaign, USA
36. Schabort EJ, Killian SC, Gibson AS, Hawley JA, Noakes TD (2000). Prediction of triathlon race time from laboratory testing in national triathletes. Med Sci Sports Exerc 32:844-849
37. Steyerberg EW (2009). Clinical prediction models: a practical approach to development, validation and updating. Springer Science+Business Media, New York, USA
38. van Mechelen W, Hlobil H, Kemper HCG (1986). Validation of two running tests as estimates of maximal aerobic power in children. Eur J Appl Physiol 55:503-506
39. Welsman J, Bywater K, Farr C, Welford D, Armstrong N (2005). Reliability of peak VO2 and maximal cardiac output assessed using thoracic bioimpedance in children. Eur J Appl Physiol 94:228-234

Corresponding author:

Eivind Aadland, PhD
SognogFjordane University College,
Faculty of Teacher Education and Sports,
Box 133, 6851 Sogndal, Norway
Phone: +47 57 67 60 86;
Fax: +47 57 67 60 51;
E-mail: eivind.aadland@hisf.no

Received: 27. October 2014
Accepted: 11. December 2014