Validity and Reproducibility of a New Treadmill Protocol: The Fitkids Treadmill Test

Elles MW Kotte¹, Janke F de Groot²,³,⁴, Bart C Bongers²,³,⁴,⁵, Alexander MF Winkler¹, and Tim Takken³,⁴

¹Fitkids Foundation, Amsterdam, The Netherlands; ²University of Applied Sciences, Utrecht, the Netherlands; ³Child Development & Exercise Center, Wilhelmina Children’s Hospital, University Medical Center Utrecht, the Netherlands; ⁴Partner of Shared Utrecht Pediatric Exercise Research (SUPER) Lab, Utrecht, the Netherlands; ⁵Department of Epidemiology, School for Public Health and Primary Care (CAPHRI), Maastricht University, Maastricht, the Netherlands

Accepted for Publication: 11 February 2015
Validity and Reproducibility of a New Treadmill Protocol: The Fitkids Treadmill Test

Elles MW Kotte¹, Janke F de Groot²,³,⁴, Bart C Bongers²,³,⁴,⁵, Alexander MF Winkler¹, and Tim Takken³,⁴

¹Fitkids Foundation, Amsterdam, The Netherlands; ²University of Applied Sciences, Utrecht, the Netherlands; ³Child Development & Exercise Center, Wilhelmina Children’s Hospital, University Medical Center Utrecht, the Netherlands; ⁴Partner of Shared Utrecht Pediatric Exercise Research (SUPER) Lab, Utrecht, the Netherlands; ⁵Department of Epidemiology, School for Public Health and Primary Care (CAPHRI), Maastricht University, Maastricht, the Netherlands

Correspondence to:
Tim Takken, M.Sc., Ph.D.
Child Development & Exercise Center,
Wilhelmina Children’s Hospital, University Medical Center Utrecht,
Postal address: KB.02.056.0, P.O. Box 85090, 3508 AB Utrecht, the Netherlands.
Email: t.takken@umcutrecht.nl, phone: +31 88 75 540 30.

Running title: Clinimetric properties Fitkids Treadmill Test

This study was financed by the Johan Cruyff Foundation, the Rabobank Foundation and SIA RAAK (PRO-4-03). The results of the current study do not constitute endorsement by the American College of Sports Medicine (ACSM). The authors declare no conflict of interest.
ABSTRACT

Purpose: To investigate the validity and reproducibility of a new treadmill protocol in healthy children and adolescents: the Fitkids Treadmill Test (FTT). Methods: Sixty-eight healthy children and adolescents (6-18 years) were randomly divided into a validity group (14 boys, 20 girls; mean ± SD age: 12.9 ± 3.6 years) that performed the FTT and Bruce protocol, both with respiratory gas analysis within two weeks and a reproducibility group (19 boys, 15 girls; mean ± SD age: 13.5 ± 3.5 years) that performed the FTT twice within two weeks. A subgroup of 21 participants within the reproducibility group performed both FTTs with respiratory gas analysis. Time to exhaustion (TTE) was the main outcome of the FTT. Results: VO₂peak measured during the FTT showed excellent correlation with VO₂peak measured during the Bruce protocol (r=0.90; P<0.01). Backward multiple regression analysis provided the following prediction equations for VO₂peak (L·min⁻¹) for boys and girls respectively:

\[ VO₂_{peak\ FTT} = -0.748 + (0.117 \times TTE\ FTT) + (0.032 \times body\ mass) + 0.263 \]

\[ VO₂_{peak\ FTT} = -0.748 + (0.117 \times TTE\ FTT) + (0.032 \times body\ mass) \]

(R²=0.935, SEE=0.256 L·min⁻¹).

Cross-validation of the regression model showed an R² of 0.76. Reliability statistics for the FTT showed an ICC of 0.985 (95% confidence interval: 0.971 to 0.993; P<0.001) for TTE. Bland-Altman analysis showed a mean bias of -0.07 minutes, with limits of agreement between +1.30 and -1.43 minutes. Conclusion: Results suggest that the FTT is a useful treadmill protocol with good validity and reproducibility in healthy children and adolescents. Exercise performance on the FTT and body mass can be used to adequately predict VO₂peak when respiratory gas analysis is not available.

Key words: Exercise Testing, Physical Fitness, Reliability, Child.
INTRODUCTION

Standardized exercise testing remains an important tool that provides valuable diagnostic and prognostic information in daily clinical practice. Exercise testing allows individualized assessment of exercise tolerance and can be used to monitor exercise training programs. For use in daily clinical practice, non-sophisticated, inexpensive, reliable and valid exercise tests are of increasing interest, as they might help to increase the utilization of exercise testing (6).

The Bruce treadmill protocol is the most frequently used protocol in children and adolescents using a treadmill for cardiopulmonary exercise testing (12) and pediatric norm values have been published (4, 30, 31). While the Bruce treadmill protocol has good validity and reproducibility, the use in outpatient physical therapy practices can be difficult as the test requires a treadmill ergometer that can operate at an incline of 22%. Many of these practices are embedded in health- and sports centers with only standard treadmills available that can operate at a maximum incline of 15%. Therefore, the angle of inclination of the Bruce treadmill protocol is a major concern in outpatient physical therapy practices.

While there are several other established maximal treadmill protocols, including the Balke protocol (3), the Cornell protocol (21) and the German Society of Pediatric Cardiology protocol (18), the maximum incline of these treadmill protocols exceed 15% as well. In addition to this practical issue, many of the established maximal treadmill protocols are too demanding for children with a disability or chronic disease because of the high incline in the first stage of the protocol, leading to a premature exhaustion of the muscles of the lower limbs before achieving cardiac or respiratory limits (23).
In summary, there is a need for a maximal treadmill protocol that can be used in outpatient physical therapy practices when limited to a treadmill machine with maximal incline of 15%. In this paper, a new maximal treadmill protocol, the so-called Fitkids Treadmill Test (FTT), is introduced. The protocol starts with 0° incline making this protocol useful in children and adolescents with a disability or chronic disease. Before application in clinical practice, insight in the clinimetric properties of the FTT is crucial. The current study is the first step in the identification of the clinimetric properties of the FTT and aims to investigate the validity and the reproducibility of the new developed FTT in healthy children and adolescents between 6 and 18 years of age.

METHODS

Participants

Healthy children and adolescents (6-18 years of age) were recruited from primary and secondary schools, as well as from different sports clubs located in the Netherlands (convenience sample). The inclusion of participants started after approval of the Central Committee on Research Involving Human Subjects in the Netherlands (CCMO). The participants and their parents individually received written information about the study and informed consent forms. Signed informed consent forms were obtained from the parents or legal guardian of each participant and separately from each participant who was 12 years of age or older. The modified physical activity readiness questionnaire (PAR-Q) was used to evaluate the health status of willing participants and to assess safety for performing maximal exercise. Exclusion criteria were: use of medication affecting exercise capacity, cardiovascular or respiratory disease, impaired motor
development, morbid obesity (body mass index [BMI] >2.5 SDS) or positive responses to one or more of the modified PAR-Q questions.

**Study design**

Participants were randomly divided in a validity or a reproducibility group. To assess the validity of the FTT, the validity group performed the FTT and the Bruce test with respiratory gas analysis, in a counterbalanced order within two weeks. To assess the reproducibility of the FTT, the reproducibility group performed the FTT twice within two weeks. A subgroup of 21 participants within the reproducibility group performed both FTTs with respiratory gas analysis.

**Anthropometry**

Body mass and height were determined to the nearest 0.5 kg and 0.1 cm using respectively an analogue scale (Medisana PSD; Medisana Benelux N.V.; Kerkrade, the Netherlands), and a stadiometer (Seca 213; Seca, Hamburg, Germany). For both measurements, participants were wearing light clothes and no shoes. BMI was derived from body mass and height. BMI for age SD scores were calculated using Dutch reference values (28). Subcutaneous fat of the biceps, triceps, subscapular, and supra-iliac regions were measured using a Harpenden skinfold caliper. The average of three measures of each area was used. Body density was estimated using the equations proposed by Deurenberg et al. (17) and used to estimate percent body fat based on the Siri equation (32). Body Surface Area (BSA) was calculated using the equation of Haycock et al. which has been validated in infants, children and adults (20).
Exercise testing

Exercise tests were performed on a motor driven treadmill ergometer (Lode Valiant; Lode BV, Groningen, the Netherlands) using the Lode Ergometry Manager Software (Lode BV, Groningen, the Netherlands). During all treadmill tests, heart rate (HR) was monitored using a soft strap with a heart rate sensor (Polar H1 transmitter; Polar, Kempele, Finland). The participants of the validity group and 21 participants of the reproducibility subgroup breathed through a face mask (Hans Rudolph Inc., Kansas City, MO) during both maximal treadmill tests, which was connected to a mobile gas analysis system (Cortex Metamax B³; Cortex Medical GmbH, Leipzig, Germany) with an in-built gas analyzer. The mobile respiratory gas analysis system was calibrated for respiratory gas analysis measurements (ambient air and a gas mixture of 17% oxygen and 5% carbon dioxide) and volume measurements (3-L syringe). Values for oxygen uptake (VO₂), carbon dioxide production (VCO₂), minute ventilation (VE), and respiratory exchange ratio (RER) were collected at ten-second intervals. The Cortex Metamax B³ is a valid and reliable system for measuring ventilatory parameters during exercise (9, 25, 26).

Peak VO₂ (VO₂peak), peak VE (VEpeak) and peak RER (RERpeak), measured as the average value over the last 30 seconds before peak exercise, and peak HR (HRpeak), defined as the highest value achieved during the last 30 seconds before test termination, were used for analysis. The test was deemed maximal when at least one of the following criteria was met: an HRpeak >180 beats·min⁻¹ or an RERpeak >1.0 (1). Before and directly after the exercise test, participants were asked to rate their perceived exhaustion on a ten-point visual analog scale (VAS). Perceived exertion of the FTT and the Bruce test were determined by subtracting the pretest VAS score from the posttest VAS score (∆VAS; posttest VAS score minus pretest VAS score).
**Fitkids Treadmill Test**

The FTT protocol consists of a 90-second warm-up period (3.5 km/h, 0% grade) followed by the initiation of the test at 3.5 km/h and a 1% gradient for 90 seconds followed by incremental increases in both speed (0.5 km/h) and incline (2%) every 90 seconds until an incline of 15% was attained (see table 1). After this last step, the incline is held at 15% and incremental increases of speed (0.5 km/h) are performed every 90 seconds until volitional exhaustion. After the test, participants are monitored for two minutes to ensure a normal recovery of HR (2.0 km/h with a flat treadmill). The main outcome measure of the FTT is time to exhaustion (TTE) and is defined as the point at which the participant chooses to stop despite strong verbal encouragement. TTE is calculated as the total duration of the test minus the duration of the warm up phase.

**Bruce test**

The Bruce treadmill protocol consisted of a 90-second warm-up period (2.74 km/h and a flat treadmill) followed by the initiation of the test at 2.74 km/h and a 10% gradient for three minutes followed by incremental increases in speed and incline every three minutes until volitional exhaustion as described elsewhere (10). After the test, participants were monitored for two minutes ensure a normal recovery of HR (2.0 km/h with a flat treadmill). The main outcome measure of the Bruce protocol is time to exhaustion (TTE) and is defined as the point at which the participant chooses to stop despite strong verbal encouragement. TTE is calculated as the total duration of the test minus the duration of the warm up phase.

During both tests, participants are not allowed to use the handrails, except for touching the handrail with one or two fingers to maintain body position near the center of the moving belt.
Statistical analysis

The IBM SPSS Statistics for Windows, version 20.0 (IBM Corp, Armonk, New York) was used for data analyses. The distribution of the variables was assessed using visual inspection (histogram, boxplot and normal Q-Q plot) and the Shapiro-Wilk test for normality. Except for ∆VAS scores in both groups and VE_peak in the validity group, all variables were normally distributed with skewness and kurtosis z-scores between -1.96 and +1.96.

For both the validity group and the reproducibility group, data were checked on significant differences for TTE, cardiopulmonary variables, and perceived exertion between both tests. Paired samples t-tests were completed for normally distributed data, whereas Wilcoxon Signed Rank Tests were completed for non-normally distributed data.

Validity

To examine the validity of the FTT, correlation coefficients were calculated between TTE achieved at the FTT and TTE achieved at the Bruce test, as well as between the cardiopulmonary variables achieved at both tests. Backward multiple linear regression analysis was used to generate an equation to predict VO_{2peak} achieved at the FTT. Pearson’s product correlations were calculated between anthropometric variables (age, height, body mass, percentage body fat, fat free mass (FFM), BMI, BSA) and VO_{2peak} attained at the FTT. The best predictor candidate for VO_{2peak} attained at the FTT was entered into the model along with sex (0=girls, 1=boys) and TTE of the FTT. The goodness of fit and precision of the regression equation were evaluated using multiple coefficient of determination (R^2) and the absolute standard error of estimate (SEE). The model was tested for random errors, homoscedasticity and multicollinearity. The
regression model was cross-validated using the data from the respiratory gas analysis obtained in the reproducibility subgroup (n=21). Prior to the cross-validation analysis, the validity group and the reproducibility group were checked on significant differences in participant characteristics. Independent sample t-tests were completed for normally distributed data and Mann-Whitney Tests for non-normally distributed data.

Reproducibility

For test-retest reliability analysis, the two-way mixed (type: absolute agreement) intraclass correlation coefficient (ICC) was calculated for TTE achieved at both FTTs and for cardiopulmonary variables achieved at both FTTs. ICC values higher than 0.75 were considered acceptable (27). To determine if improvements are meaningful for a single patient, limits of agreement (LOA) were calculated for the main outcome measure of the FTT, TTE, according to the procedure described by Bland and Altman (5). TTE was defined as the main outcome measure in the reproducibility analysis because many clinicians might not be able to collect expired gasses during exercise. A P-value <0.05 was considered statistically significant.

RESULTS

Of the 74 participants approached for this study none were excluded due to a positive response on the PAR-Q. The study population was randomly divided in a validity group and a reproducibility group. From both the validity group and the reproducibility group, three participants were excluded because of the following reasons: invalid respiratory gas analysis (n=1), illness during the second day of testing (n=2), dyspnea during the test (n=1), scheduling issues (n=1) and a painful leg during testing (n=1). Eventually, the validity group and
reproducibility group both consisted of 34 participants. Participant descriptive statistics are found in table 2.

**Validity**

Participants from the validity group performed both the FTT and the Bruce protocol without any adverse effects. They all met the subjective criteria of maximal effort (sweating, unsteady walking, facial flushing and clear unwillingness to continue despite strong verbal encouragement) as well as the objective criteria of maximal effort (an \( \text{HR}_{\text{peak}} > 180 \text{ beats} \cdot \text{min}^{-1} \) and/or an \( \text{RER}_{\text{peak}} > 1.0 \)) during both the FTT and the Bruce treadmill protocol. Seventeen of the 34 participants (50%) reached the maximal incline of 15%. Mean ± standard deviation (SD) between-visit time: 6.2 ± 1.4 days.

Results of the FTT and the Bruce protocol completed by the validity group are presented in table 3. Significant higher values of TTE were found for the FTT. The cardiopulmonary variables and rating of perceived exertion were not significantly different between the FTT and the Bruce protocol. Pearson correlation coefficients for TTE, \( \text{HR}_{\text{peak}} \), \( \text{VO}_2\text{peak} \) (L min\(^{-1}\)) and \( \text{RER}_{\text{peak}} \) achieved at the Bruce protocol and the FTT were respectively 0.98, 0.67, 0.90 and 0.64 (all \( P < 0.01 \)), which are categorized as moderate to strong correlations according to Dancey and Reidy’s categorization (14). For \( \text{VE}_{\text{peak}} \) a Spearman correlation coefficients of 0.96 was found (\( P < 0.01 \)).

In order to construct an equation to predict \( \text{VO}_2\text{peak} \) attained at the FTT, Pearson’s correlation coefficients were calculated between anthropometric variables and \( \text{VO}_2\text{peak} \) attained at the FTT. FFM appeared to be the best predictor candidate to include into the regression model (\( r = 0.934 \),
P=0.01). However, FFM is impractical in daily use and was therefore not included into the final model. Body mass, which was the second best predictor and easily available in daily practice, was included into the final model along with sex and TTE on the FTT. The model that incorporated FFM can be found in the Supplemental Digital Content [See document, Supplemental Digital Content 1, Equations to predict VO2peak achieved during the FTT from the attained TTE at the FTT and fat free mass (FFM), http://links.lww.com/MSS/A530].

The following equations for boys and girls were generated to predict VO2peak achieved at the FTT from the attained TTE at the FTT and body mass:

Boys: VO2peak FTT = -0.748 + (0.117 × TTE FTT) + (0.032 × body mass) + 0.263
Girls: VO2peak FTT = -0.748 + (0.117 × TTE FTT) + (0.032 × body mass)

In the prediction equations, ‘VO2peak FTT’ represents the predicted VO2peak in L·min⁻¹, ‘TTE FTT’ is the total duration of the FTT in minutes minus the duration of the warm up phase in minutes, and ‘body mass’ is expressed in kg (R² = 0.935, SEE= 0.256).

For cross-validation purposes, the VO2peak estimated by the equations was plotted against the measured VO2peak during the first FTT in the reproducibility subgroup who performed the FTTs with respiratory gas analysis (see Figure 1). Estimated VO2peak (3.19 ± 0.51 L·min⁻¹) did not differ significantly from the observed VO2peak (3.31 ± 0.88 L·min⁻¹; p=0.291; R² = 0.76).
Reproducibility

Participants from the reproducibility group performed the FTT twice without any adverse effects. They all met the subjective criteria of maximal effort (sweating, unsteady walking, facial flushing and clear unwillingness to continue despite strong verbal encouragement) as well as the objective criteria of a maximal effort (an $HR_{\text{peak}}>180$ beats min$^{-1}$ and/or an $RER_{\text{peak}}>1.0$) during both FTTs. One participant showed an $HR_{\text{peak}}$ of 174 and 169 respectively on the first and second FTT, however, $RER_{\text{peak}}$ measured during these tests were respectively 1.28 and 1.25. Twenty-two of the 34 participants (65%) reached the maximal incline of 15%. Mean ± SD between-visit time: 8.9 ± 3.8 days.

The results of both FTTs performed by the reproducibility group are shown in table 4. There were no significant differences in TTE and $HR_{\text{peak}}$ between the two FTTs. Also for perceived exertion ($\Delta$VAS), no significant difference was found between both FTTs. For the cardiopulmonary variables, a significant higher $RER_{\text{peak}}$ was found during the second FTT. To quantify the relation between TTE achieved at both tests and between the cardiopulmonary variables achieved at both tests, ICCs were calculated. The ICC for TTE, which is the main outcome for the test-retest reliability statistics, was 0.985 (95% CI: 0.971 to 0.993; $P<0.001$). The ICC values for $HR_{\text{peak}}$, $VO_2\text{peak}$, $RER_{\text{peak}}$ and $VE_{\text{peak}}$ were respectively 0.767 (95% CI: 0.584 to 0.876; $P<0.001$), 0.963 (95% CI: 0.912 to 0.985; $P<0.001$), 0.631 (95% CI: 0.269 to 0.834; $P<0.001$) and 0.948 (95% CI: 0.877 to 0.978; $P<0.001$). To analyze agreement between the two FTTs, a Bland Altman plot for TTE is depicted in Figure 2. The mean bias between the two FTTs was -0.07 minutes. The LOA for TTE were +1.30 minutes and -1.43 minutes.
DISCUSSION

The aim of the current study was to investigate the validity and reproducibility of a new maximal treadmill protocol for children and adolescents between 6 and 18 years of age: the FTT. The results of the present study suggest that the FTT is a useful treadmill protocol with good validity and reproducibility. Moderate to strong significant correlations between TTE and cardiopulmonary variables attained at the FTT and the Bruce protocol were found, indicating that the FTT is a valid test for the assessment of aerobic exercise capacity in healthy children and adolescents. At the same time, there were no significant differences between the $VO_{2peak}$ measured during the Bruce protocol and the $VO_{2peak}$ measured during the FTT. These results suggest that the tests produce quite similar estimates and that the tests may be interchangeable because of these similar results.

The current study showed that $VO_{2peak}$ can be adequately predicted from TTE and body mass in boys and girls, explaining 94% of the total variance in $VO_{2peak}$. When applying the multiple prediction equations developed in the validity group to the reproducibility subgroup, no significant difference was found between the observed $VO_{2peak}$ ($3.31 \pm 0.88 \text{ L} \cdot \text{min}^{-1}$) and the $VO_{2peak}$ estimated from the prediction equation ($3.19 \pm 0.51 \text{ L} \cdot \text{min}^{-1}$; $p=0.291$). The multiple correlation of the cross-validation regression equation was good ($R^2 = 0.76$). This means that, for clinicians who do not have the resources to directly measure $VO_{2peak}$, the TTE during the FTT is an adequate alternative that gives insight in the aerobic fitness of healthy children and adolescents. Nevertheless, measuring $VO_{2peak}$ using respiratory gas analysis during incremental exercise is still considered to be the gold standard for aerobic fitness by the World Health Organization (29).
Several other studies have predicted aerobic fitness from functional performance during exercise testing, both in adults and the pediatric population. Bruce et al. (10) developed the first predictive equations, which were population-specific for active and sedentary men with and without cardiac conditions as well as for healthy adults (Pearson correlation coefficients between predicted VO$_{2peak}$ and measured VO$_{2peak}$ ranged from r = 0.87 to r = 0.92). Foster et al. (19) later developed a more generalized equation, dependent only on Bruce treadmill test performance, to predict VO$_{2peak}$ (L·min$^{-1}$) ($R^2$=0.98, SEE= 3.35 mL·kg$^{-1}$·min$^{-1}$ or 8.5%). This generalized equation was developed for use in cardiac patients as well as healthy, sedentary and active individuals. Buono et al. (11) predicted VO$_{2peak}$ (L·min$^{-1}$) during a timed distance run on an oval dirt track in a healthy children and adolescents. The equation in their study (based on mile-run time, sex, skinfold thickness and body mass) explained 84% of the total variance in VO$_{2peak}$ ($R^2$= 0.84 and SEE 4.3 mL·kg$^{-1}$·min$^{-1}$ or 9%). Bongers et al. concluded that VO$_{2peak}$ (L·min$^{-1}$) could be validly predicted from the attained peak work rate at the Steep Ramp Test in healthy children and adolescents ($R^2$=0.917, SEE=0.24 L·min$^{-1}$ or 9%) (7, 8) and Dencker et al. (16) found that VO$_{2peak}$ (L·min$^{-1}$) could be predicted from the peak work rate reached during cardiopulmonary exercise testing (CPET) in healthy children ($R^2$=0.83, SEE=0.11 L·min$^{-1}$ or 8.4%). In the study of Dencker et al., CPET was performed on an electronically braked cycle ergometer using a protocol with an initial workload of 30 Watts (W) and an increase of 15 W·min$^{-1}$ (1 W·4 sec$^{-1}$). The equation developed in the current study is in agreement with previous studies reporting that the addition of body mass (11, 24) and sex (11) improves the prediction of VO$_{2peak}$ from TTE in children.
Test-retest reproducibility encompasses both reliability and agreement. Where the first refers to the consistency or stability of a test after repeated trials, the latter analyzes the variation within the individual scores during a test-retest situation (2, 15) and is used to determine the clinical value of a measurement. The test-retest reliability of the FTT can be considered excellent with an ICC for TTE of 0.985 (95% CI: 0.971 to 0.993; p<0.001). This means that the measurement error is small compared to the variability between the participants and that the discrimination of persons is hardly affected by measurement error. These results are comparable with those reported by Cumming et al. (13). They investigated the test-retest reliability of the Bruce protocol in 20 school children aged 7 to 13 years, and reported a correlation coefficient of 0.94 between TTE achieved on trial one and TTE achieved on trial two. The mean TTE in the study of Cumming et al. was 13.9 ± 2.1 minutes for trial one and 13.7 ± 1.9 minutes for trial two. The mean TTE of the FTT observed in the current study was 14.2 ± 4.0 minutes for trial one and 14.3 ± 4.1 minutes for trial two. The results of the current study are also comparable with those of Johnston et al. (22) who determined test-retest reproducibility of cardiopulmonary variables during CPET in children using a treadmill protocol. The reported ICCs for VO$_2$peak, HR$_{peak}$ and VE$_{peak}$ were respectively 0.96, 0.87 and 0.91 against 0.96, 0.77 and 0.95 for respectively VO$_2$peak, HR$_{peak}$ and VE$_{peak}$ found in the current study.

The agreement analysis revealed narrow LOA (+1.30 to -1.43 minutes), indicating that the agreement of the FTT is good. The average difference of TTE attained at the two FFTs was roughly 4 seconds. There was no evidence for a significant learning effect as reflected by a symmetrically distribution of the differences around the zero difference line. A change score in
TTE between two concessive measurements within an individual can only be considered to represent a real change if it is outside the limits of agreement.

A limitation of this study is that only healthy participants were tested so the equations currently developed are appropriate only for healthy children and adolescents. As the FTT is developed for use in outpatient physical therapy practices, predicting models for the clinical population should be developed and evaluated. An additional limitation of the current study is the small sample size used in the cross-validation analysis. The predictive accuracy of the developed regression equation seems proportionally biased, in which true values less than 2.5 L·min\(^{-1}\) are systematically overestimated and true values more than 4.0 L·min\(^{-1}\) are systematically underestimated. The predictive accuracy of the developed regression equation should be evaluated using a larger sample in future studies. Future studies should also look at reference values for different populations.

In conclusion, the results of the current study suggest that the FTT is a useful treadmill protocol with good validity and reproducibility in healthy children and adolescents aged 6 to 18 years. The use of the FTT can be favored instead of the Bruce protocol when limited to a standard treadmill machine with maximum incline of 15%. We have shown that exercise performance on the FTT and body mass can be used to simple predict VO\(_{2}\)\(_{\text{peak}}\) in healthy children and adolescents in situations where it is not possible to measure VO\(_{2}\)\(_{\text{peak}}\) with respiratory gas analysis. Further testing of the FTT in clinical populations is warranted.
ACKNOWLEDGEMENTS

This study was financed by the Johan Cruyff Foundation, the Rabobank Foundation and SIA RAAK (PRO-4-03). The results of the current study do not constitute endorsement by the American College of Sports Medicine (ACSM). The authors are very grateful to Lode BV, Groningen, the Netherlands, and ProCare BV, Groningen, The Netherlands, for the technical support during this study. We thank the participating schools, “basisschool de Wiekslag”, Tubbergen, the Netherlands and “Bonhoeffer college”, Castricum, the Netherlands. We also want to thank the Child Development & Exercise Center, University Medical Center Utrecht, Utrecht, the Netherlands and the participating sports clubs. Finally we are especially grateful to all our participants.

CONFLICT OF INTEREST

The authors declare no conflict of interest.
REFERENCES


FIGURE CAPTIONS

**Figure 1:** Cross-validation plot: linear relationship between the predicted VO₂peak values based on the prediction equation of the regression model and the VO₂peak measured during the first Fitkids Treadmill Test in the reproducibility subgroup.

**Figure 2:** Bland-Altman plot of time to exhaustion as attained at the first Fitkids Treadmill Test versus the second Fitkids Treadmill Test.

SUPPLEMENTAL DIGITAL CONTENT

**Supplemental Digital Content 1:** Equations to predict VO₂peak achieved during the FTT from the attained TTE at the FTT and fat free mass (FFM).
Figure 1

Predicted $VO_{2peak}$ (L.min$^{-1}$)

Observed $VO_{2peak}$ (L.min$^{-1}$)

$R^2=0.76$
Figure 2
Table 1. Protocol Fitkids Treadmill Test.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Stage duration (s)</th>
<th>Speed (km/h)</th>
<th>Speed (mph)</th>
<th>Gradient (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warm-up</td>
<td>90</td>
<td>3.5</td>
<td>2.175</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>3.5</td>
<td>2.175</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
<td>4.0</td>
<td>2.485</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>4.5</td>
<td>2.796</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>90</td>
<td>5.0</td>
<td>3.106</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>5.5</td>
<td>3.417</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>90</td>
<td>6.0</td>
<td>3.728</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>90</td>
<td>6.5</td>
<td>4.039</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>90</td>
<td>7.0</td>
<td>4.350</td>
<td>15</td>
</tr>
<tr>
<td>9</td>
<td>90</td>
<td>7.5</td>
<td>4.661</td>
<td>15</td>
</tr>
</tbody>
</table>

... Speed is not limited until voluntary exhaustion
Maximum gradient is limited to 15%

Abbreviations: km/h=kilometers per hour; mph=miles per hour.
Table 2. Participant characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Validity group</th>
<th>Reproducibility group</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (boys/girls)</td>
<td>14/20</td>
<td>19/15</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>12.9 ± 3.6</td>
<td>13.5 ± 3.5</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>[6.5-18.6]</td>
<td>[6.6-18.2]</td>
<td></td>
</tr>
<tr>
<td>Body mass (kg)</td>
<td>48.9 ± 16.8</td>
<td>49.2 ± 15.9</td>
<td>0.928</td>
</tr>
<tr>
<td></td>
<td>[20.0-80.8]</td>
<td>[21.0-80.5]</td>
<td></td>
</tr>
<tr>
<td>Body height (m)</td>
<td>1.57 ± 0.18</td>
<td>1.61 ± 0.19</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td>[1.15-1.85]</td>
<td>[1.20-1.89]</td>
<td></td>
</tr>
<tr>
<td>BMI (kg∙m⁻²)</td>
<td>19.1 ± 3.2</td>
<td>18.3 ± 2.3</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>[13.6-25.3]</td>
<td>[12.9-22.5]</td>
<td></td>
</tr>
<tr>
<td>BSA (m²)</td>
<td>1.44 ± 0.34</td>
<td>1.47 ± 0.33</td>
<td>0.775</td>
</tr>
<tr>
<td></td>
<td>[0.80-2.04]</td>
<td>[0.83-2.05]</td>
<td></td>
</tr>
<tr>
<td>FFM (kg)</td>
<td>39.1 ± 13.0</td>
<td>40.1 ± 13.2</td>
<td>0.750</td>
</tr>
<tr>
<td></td>
<td>[17.2-64.0]</td>
<td>[16.7-61.9]</td>
<td></td>
</tr>
<tr>
<td>Body fat (%)</td>
<td>19.4 ± 4.5</td>
<td>18.6 ± 4.2</td>
<td>0.486</td>
</tr>
<tr>
<td></td>
<td>[12.8-27.2]</td>
<td>[10.8-26.1]</td>
<td></td>
</tr>
<tr>
<td>BMI for age (SD score)</td>
<td>-0.02</td>
<td>0.30</td>
<td>0.048*</td>
</tr>
<tr>
<td></td>
<td>(-0.45-0.37)</td>
<td>(-0.20-1.24)</td>
<td></td>
</tr>
</tbody>
</table>

Date are presented as mean ± SD, [range].

Abbreviations: BMI=body mass index, BSA=body surface area; FFM=fat free mass; SD=standard deviation; *p<0.05. *: Non-parametric Mann-Whitney Test, data presented as Median (Inter Quartile Range).
Table 3. Fitkids Treadmill Test and Bruce protocol results of the validity group.

<table>
<thead>
<tr>
<th></th>
<th>Fitkids Treadmill Test (n=34)</th>
<th>Bruce Protocol (n=34)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to exhaustion (min)</td>
<td>13.5±3.7 [9.2-22.2]</td>
<td>12.1±3.0 [7.9-18.0]</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>HR&lt;sub&gt;peak&lt;/sub&gt; (beats∙min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>194.9±6.9 [181-208]</td>
<td>194.1±6.3 [183-207]</td>
<td>0.360</td>
</tr>
<tr>
<td>VO&lt;sub&gt;2peak&lt;/sub&gt; (L∙min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>2.5±1.0 [1.1-4.4]</td>
<td>2.5±1.0 [1.0-4.5]</td>
<td>0.779</td>
</tr>
<tr>
<td>VO&lt;sub&gt;2peak&lt;/sub&gt; (mL∙kg&lt;sup&gt;-1&lt;/sup&gt;∙min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>51.1±7.6 [28.0-71.0]</td>
<td>51.8±8.2 [34.0-72.0]</td>
<td>0.366</td>
</tr>
<tr>
<td>RER&lt;sub&gt;peak&lt;/sub&gt;</td>
<td>1.1±0.1 [1.0-1.3]</td>
<td>1.1±0.1 [1.0-1.3]</td>
<td>0.947</td>
</tr>
<tr>
<td>VE&lt;sub&gt;peak&lt;/sub&gt; (L∙min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>75.3 (56.3-96.6)</td>
<td>71.09 (55.0-103.3)</td>
<td>0.918&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>∆VAS</td>
<td>10 (9.0-10.0)</td>
<td>9.5 (8.75-10)</td>
<td>0.491&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Date are presented as mean ± SD, [range].

Abbreviations: HR<sub>peak</sub> = peak heart rate; VO<sub>2peak</sub> = peak oxygen uptake; RER<sub>peak</sub> = peak respiratory exchange ratio; VE<sub>peak</sub> = peak minute ventilation; ∆VAS = visual analog scale difference addressing the participants’ level of fatigue (posttest VAS score minus pretest VAS score). *p≤0.001. <sup>a</sup>: Non-parametric Wilcoxon Singed Ranks Test, data presented as Median (Inter Quartile Range).
Table 4. Fitkids Treadmill Test results of the reproducibility group.

<table>
<thead>
<tr>
<th></th>
<th>First FFT</th>
<th>Second FFT</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to exhaustion (min)</td>
<td>n=34 14.3±4.0</td>
<td>14.3±4.1</td>
<td>0.584</td>
</tr>
<tr>
<td></td>
<td>[7.7-24.5]</td>
<td>[8.0-24.3]</td>
<td></td>
</tr>
<tr>
<td>HR&lt;sub&gt;peak&lt;/sub&gt; (beats·min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>n=34 197±8</td>
<td>196±10</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>[174-215]</td>
<td>[169-222]</td>
<td></td>
</tr>
<tr>
<td>VO&lt;sub&gt;2peak&lt;/sub&gt; (L·min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>n=21 3.3±0.9</td>
<td>3.3±0.8</td>
<td>0.645</td>
</tr>
<tr>
<td></td>
<td>[2.0-5.1]</td>
<td>[2.2-5.0]</td>
<td></td>
</tr>
<tr>
<td>VO&lt;sub&gt;2peak&lt;/sub&gt; (mL·kg&lt;sup&gt;-1&lt;/sup&gt;·min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>n=21 54.5±12.7</td>
<td>55.2±10.1</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>[26.0-80.0]</td>
<td>[40.0-77.0]</td>
<td></td>
</tr>
<tr>
<td>RER&lt;sub&gt;peak&lt;/sub&gt;</td>
<td>n=21 1.1±0.1</td>
<td>1.2±0.1</td>
<td>0.029*</td>
</tr>
<tr>
<td></td>
<td>[1.0-1.31]</td>
<td>[1.0-1.4]</td>
<td></td>
</tr>
<tr>
<td>VE&lt;sub&gt;peak&lt;/sub&gt; (L·min&lt;sup&gt;-1&lt;/sup&gt;)</td>
<td>n=21 105.8±25.8</td>
<td>107.3±27.0</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td>[64.4-152.2]</td>
<td>[65.7-151.2]</td>
<td></td>
</tr>
<tr>
<td>Δ VAS</td>
<td>n=34 10</td>
<td>10</td>
<td>0.241&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>(8.0-10.0)</td>
<td>(8.38-10.0)</td>
<td></td>
</tr>
</tbody>
</table>

Date are presented as mean ± SD, [range].

Abbreviations: HR<sub>peak</sub> = peak heart rate; VO<sub>2peak</sub> = peak oxygen uptake; RER<sub>peak</sub> = peak respiratory exchange ratio; VE<sub>peak</sub> = peak minute ventilation; ΔVAS = visual analog scale difference addressing the participants’ level of fatigue (posttest VAS score minus pretest VAS score); *p<0.05. <sup>a</sup>: Non-parametric Wilcoxon Singed Ranks Test, data presented as Median (Inter Quartile Range).
SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content 1. Equations to predict VO$_{2peak}$ achieved during the FTT from the attained TTE at the FTT and fat free mass (FFM).

Boys: VO$_{2peak}$ FTT = -0.673 + (0.097 × TTE FTT) + (0.046 × FFM) + 0.160
Girls: VO$_{2peak}$ FTT = -0.673 + (0.097 × TTE FTT) + (0.046 × FFM)

In the prediction equations, ‘VO$_{2peak}$ FTT’ represents the predicted VO$_{2peak}$ in L·min$^{-1}$, ‘TTE FTT’ is the total duration of the FTT in minutes minus the duration of the warm up in minutes, and FFM is expressed in kg ($R^2 = 0.934$, SEE=0.258).