

Development of Youth Aerobic-Capacity Standards Using Receiver Operating Characteristic Curves

Gregory J. Welk, PhD, Kelly R. Laurson, PhD, Joey C. Eisenmann, PhD, Kirk J. Cureton, PhD

Background: Cardiovascular fitness has important implications for current and future health in children.

Purpose: In this paper, criterion-referenced standards are developed for aerobic capacity (an indicator of cardiovascular fitness) based on receiver operating characteristic (ROC) curves.

Methods: The sample was drawn from participants aged 12–18 years in the National Health and Nutrition Examination Survey (1999–2002, N=1966). Subjects completed a treadmill exercise test from which maximal oxygen uptake (VO_2max) was estimated from heart rate response. Metabolic syndrome was classified using previously published standards based on the National Cholesterol Education Program/Adult Treatment Panel III adult values at age 20 years. Using aerobic fitness z-scores as the test and metabolic syndrome as the criterion, ROC curve analysis was used to identify aerobic-capacity thresholds.

Results: The area under the curve (AUC) value for boys (83.1%) was high, indicating good utility for detecting risk of metabolic syndrome with aerobic fitness values. The AUC for girls (77.2%) was slightly below the recommended value of 80%. Although the ROC plots identified a defensible point for classifying levels of fitness, the approach in the present study was to establish two independent thresholds, one aimed at high specificity and one aimed at high sensitivity. The resulting z values for the low- and higher-risk threshold lines were then converted back to VO_2max estimates using published LMS (L = skewness, M = median, and S = coefficient of variation) parameters. Values at the low-risk threshold ranged from 40 to 44 mL/kg/min for boys and from 38 to 40 mL/kg/min for girls.

Conclusions: In summary, aerobic fitness can be used with moderate accuracy to differentiate between adolescents with and without metabolic syndrome. Age- and gender-specific aerobic-capacity thresholds for creating separate risk groups were identified using nationally representative growth percentiles.

(Am J Prev Med 2011;41(4S2):S111–S116) © 2011 American Journal of Preventive Medicine

Introduction

Aerobic capacity, also referred to as cardiorespiratory or cardiovascular fitness, is considered to be the most important dimension of health-related fitness. Numerous studies have documented the importance of an adequate aerobic capacity for good health in adults,^{1–3} and the evidence appears equally compelling in children and

adolescents. Aerobic capacity in youth is associated with cardiovascular disease risk factors,^{4–6} and prospective studies have demonstrated that aerobic capacity tracks reasonably well from childhood/adolescence into adulthood.⁷ Declines in aerobic capacity from childhood to adolescence are also associated with an increased risk of overweight and metabolic syndrome in adults.^{8–10} Recent reports have indicated that approximately one third of U.S. adolescents possess inadequate levels of aerobic capacity.^{11,12} Collectively, these studies provide strong evidence to support the importance of focused efforts to monitor and promote aerobic capacity in youth.

Youth fitness testing is a common part of most physical education programs, and aerobic capacity is perhaps the most commonly assessed component. Field tests of aerobic capacity (e.g., 1-mile run) are typically administered

From the Department of Kinesiology, Iowa State University (Welk), Ames, Iowa; the School of Kinesiology and Recreation, Illinois State University (Laurson), Normal, Illinois; The Healthy Weight Center at Helen DeVos Children's Hospital (Eisenmann), Grand Rapids, Michigan; and the Department of Kinesiology, University of Georgia (Cureton), Athens, Georgia

Address correspondence to: Gregory J. Welk, PhD, Department of Kinesiology, Iowa State University, 257 Forker Building, Ames IA 50011.
E-mail: gwelk@iastate.edu
0749-3797/\$17.00
doi: 10.1016/j.amepre.2011.07.007

by physical education teachers with the goal of providing personalized information to children and/or parents. A unique advantage of the FITNESSGRAM® program is that fitness levels are evaluated using criterion-referenced standards that indicate how much fitness is needed for good health.¹³ The standards were established initially using a metabolic model to account for developmental changes and then modified by expert opinion.¹⁴ Later, these standards were linked to health by extrapolating adult standards backwards to values appropriate for youth.¹⁵ The standards have been well supported, and recent studies have demonstrated that they have good utility for detecting health risks.¹⁶ However, it is important to evaluate new approaches to improve the accuracy and clinical utility of youth fitness standards.

A challenge in setting criterion-referenced standards in youth is the need to link fitness to clinical health outcomes. Although morbidity and mortality are commonly used in adults, these are not appropriate indicators for establishing risk in youth. The availability of nationally representative data on aerobic capacity and clinical risk in the National Health and Nutrition Examination Survey (NHANES) provides the opportunity to develop standards that reflect potential health risk in youth. In this paper, new criterion-referenced standards are developed for aerobic capacity based on receiver operating characteristic (ROC) curves, an established procedure for establishing clinical thresholds.¹⁷ Although recent studies have used a similar methodology,^{16,18} the present study also utilizes recently developed LMS (L=skewness, M=median, and S=coefficient of variation) growth curves¹⁹ to better characterize the developmental changes in aerobic capacity during adolescence.

Methods

Subjects

Data for the present study were obtained from the NHANES (1999–2000 and 2001–2002), conducted by the National Center for Health Statistics (NCHS), CDC. The NHANES is designed to assess the health and nutritional status of adults and children in the U.S. through interviews and direct physical examinations. For this project, a combination of data was used: aerobic-capacity data, anthropometric data, and metabolic examination data of children aged 12–18 years across two cross-sectional waves of NHANES testing (1999–2000 and 2001–2002). The unweighted sample sizes for each 2-year wave of NHANES (1999–2000 and 2001–2002) is approximately 2000 school-aged children and adolescents aged 8–18 years (note: aerobic-capacity data are not available for children aged <12 years). Thus, a total of 1240 children and adolescents who completed the treadmill test were included in the current analysis (591 boys and 649 girls).

The NHANES uses a nationally representative sample so results can be generalized to U.S. children and adolescents. IRB approval was not obtained since the data are publicly available and de-identified. Doc-

umentation on the NHANES (1999–2000 and 2001–2002) data set can be obtained online at www.cdc.gov/nchs/nhanes.htm.

Measures

Aerobic capacity. Aerobic capacity was operationally defined as the estimated maximal oxygen uptake (VO_2max). The assessments of aerobic capacity in NHANES (1999–2000 and 2001–2002) were conducted by trained health technicians using a submaximal treadmill exercise test. Based on gender, age, BMI, and self-reported level of physical activity, participants were assigned to one of eight treadmill test protocols.

The goal of each protocol was to elicit a heart rate that is approximately 75% of the age-predicted maximum ($220 - \text{age}$) by the end of the test. Each protocol included a 2-minute warm-up, two 3-minute exercise stages, and a 2-minute cool-down period. Heart rate was monitored continuously using an automated monitor with four electrodes connected to the thorax and abdomen of the participant and was recorded at the end of warm-up, at the end of each exercise stage, and after each minute of recovery. VO_2max was estimated by extrapolation of heart rate from the two 3-minute exercise stages to the age-specific maximal heart rate. Because adolescents with a true VO_2max greater than 75 mL/kg/min are uncommon, VO_2max values greater than this value were set equal to 75 mL/kg/min (<1% of subjects).

Cardiovascular disease risk factors/metabolic syndrome.

Metabolic syndrome was used as the outcome variable in the ROC analyses. The clinical examination included detailed anthropometric measures as well as a variety of clinical data. Stature and body mass were measured according to standard procedures, with stature measured to the nearest 0.1 cm and body mass measured to the nearest 0.1 kg. Waist circumference was measured to the nearest 0.1 cm using a steel measuring tape, just above the uppermost lateral border of the ileum at the end of a normal expiration. Systolic and diastolic blood pressures were recorded as the average of three or four consecutive measurements with a mercury sphygmomanometer after the participant sat and rested quietly for 5 minutes. High-density lipoprotein cholesterol (HDL-C) and fasting triglycerides (TG) were analyzed at the Johns Hopkins Lipoprotein Analytical Laboratory, and fasting glucose was analyzed at the University of Missouri–Columbia. All measurements described in the preceding text were taken by trained health technicians in the Mobile Examination Center, and detailed quality control checks are included to ensure high-quality data across the survey. The training, examination protocol, and quality control procedures are outlined in the NHANES procedures manuals available at www.cdc.gov/nchs/nhanes.htm.

Presence of individual metabolic syndrome risk factors (waist circumference, systolic blood pressure/diastolic blood pressure, HDL-C, TG, and fasting glucose) was evaluated using age- and gender-specific threshold values derived with LMS techniques linked to the National Cholesterol Education Program/Adult Treatment Panel III adult values at age 20 years.²⁰ These thresholds were used to identify subjects in the current study whose data were above each threshold. Subjects with three or more of the five metabolic components were identified as having metabolic syndrome. The standards were also based on the same NHANES (1999–2000 and 2001–2002) data set so the parameters can be considered to be based on nationally representative samples of the U.S. population.

Data Processing and Analyses

The aerobic-capacity standards in this study were developed using ROC analysis, which provide an empirical basis for selecting an optimal threshold value. To employ ROC analyses for establishing health standards, it is necessary to have a single binary variable denoting the presence or absence of disease or disease risk. In the context of this study, the diagnostic accuracy of the “test” then refers to the ability of aerobic capacity to discriminate adverse metabolic health, as assessed by the metabolic syndrome (yes/no).

Youth with the metabolic syndrome who are classified correctly represent the true-positive cases, whereas metabolic syndrome subjects classified as normal metabolic health represent false-negative cases. The sensitivity (Se) of the diagnosis is the probability that the aerobic-capacity value will classify a subject with metabolic syndrome when the subject is truly with metabolic syndrome; the specificity (Sp) is the probability that the aerobic-capacity value will classify a subject without metabolic syndrome when the subject is truly metabolically healthy (true negative).

A unique challenge in establishing aerobic-capacity standards in youth is to account for normal physical growth and maturation. In the present study, growth and maturation were partially accounted for by converting the aerobic-capacity values into z-scores using previously derived LMS parameters.¹⁹ Readers are encouraged to consult this article in this supplement to the *American Journal of Preventive Medicine* for technical details, but a brief summary is provided. The LMS parameters summarize the changing distribution from three curves representing the median (M), coefficient of variation (S), and skewness (L).²¹ The availability of LMS values allows the age-related distribution of aerobic capacity across this age range to be summarized with the three associated variables (L, M, and S).

The z-score was computed relative to the LMS distribution using the following formula:

$$z = ((Y/M)^L - 1)/(L \times S).$$

This z-score was then used in the ROC analyses to evaluate whether aerobic capacity can detect the metabolic syndrome with reasonable accuracy. The ROC analyses in the present study were conducted using customized routines in SAS that adjusted for weight, cluster, and strata variables provided with the NHANES (1999–2000 and 2001–2002) data sets (Dr. Mithat Gönen, Memorial Sloan-Kettering Cancer Center, personal communication, 2008). The macro made it possible to apply the NHANES (1999–2000 and 2001–2002) sample weights to the data and to calculate ROC statistics (such as Se and Sp) for each possible threshold value. Separate ROC curves were created for boys and girls since there are established gender differences in aerobic capacity and metabolic syndrome.

The ROC plots were generated within a customized SAS macro, and supplemental output yielded values reflecting the total AUC, the key diagnostic indicator in ROC curve analyses. Associated output files were examined to determine the relative changes in sensitivity and specificity for different z-score values. The point generally selected as the optimal threshold is the point that is closest to the upper-left part of the ROC plot. This point maximizes the sum of Se and Sp and can be considered to produce the best overall classification agreement.

The approach in the present study, however, was to determine two threshold values, one emphasizing Se and the other emphasizing

Table 1. Physical characteristics of the study sample

	Boys (n=649)	Girls (n=591)
Characteristic, M (SE)		
Age (years)	15.1 (0.1)	14.9 (0.1)
Height (cm)	171.0 (0.6)	161.3 (0.4)
Mass (kg)	66.97 (0.9)	58.9 (0.9)
BMI	22.6 (0.3)	22.5 (0.3)
VO ₂ max (mL/kg/min)	47.3 (0.6)	39.6 (0.4)
Prevalence of risk factors, % (SE)		
High WC	8.2 (1.1)	23.8 (3.2)
High BP	12.1 (2.8)	4.2 (1.2)
High TG	11.7 (1.8)	9.4 (2.1)
Low HDL-C	30.7 (2.4)	49.0 (2.9)
High fasting glucose	11.3 (1.8)	3.4 (1.2)
Metabolic syndrome	6.3 (1.2)	5.9 (1.6)

Note: Risks based on established standards are developed from the NHANES.

BP, blood pressure; HDL-C, high-density lipoprotein cholesterol; NHANES, National Health and Nutrition Examination Survey; SF, skinfold; TG, triglycerides; VO₂max, maximal oxygen consumption; WC, waist circumference

ing Sp. A variety of thresholds were evaluated to find an appropriate balance between Se and Sp for each threshold. The VO₂max values associated with these thresholds were determined by inserting the z values into Equation 1 and solving for Y (Equation 1: $Y = M(1 + LSz)^{1/L}$). The equation is solved for each age and gender combination, resulting in unique threshold values for each subgroup.

Results

The descriptive statistics for the sample population are summarized in Table 1. The average VO₂max values (mL/kg/min) for the sample ranged from 39.1 to 41.1 in girls and from 43.5 to 49.4 in boys. However, there were no appreciable age-related trends evident in the cross-sectional analyses. The prevalence of individual risk factors varied between boys and girls, but the prevalence of metabolic syndrome (based on the present calculations) was 6.3% (1.2%) in boys and 5.9% (1.6%) in girls.

The resulting ROC plots demonstrated good utility for detecting risk of metabolic syndrome based on aerobic capacity. The AUC value for boys was high (AUC=83.1) but slightly lower in girls (AUC=77.2). For boys, the optimal classification threshold (z= -0.60) resulted in a Se=92.3% (95% CI=79, 100) and Sp=64.0% (95% CI=59, 69). For girls, the optimal classification threshold

($z=0.13$) resulted in a $Se=92.8\%$ (95% CI=81, 108) and a $Sp=62.7\%$ (95% CI=56, 68).

The VO_2 max z-scores identified in the ROC curve analyses denote the value that maximizes classification agreement (i.e., point that yields the maximum value for the sum of Se and Sp). Although the optimized threshold is defensible from a statistical perspective, it is also important to consider the relative importance of Se and Sp for youth fitness reporting. Thus, an alternative threshold was derived by identifying the point that yielded more-equivalent values of Se and Sp . The resulting thresholds yielded equated values of Se and Sp (~ 0.70 for girls and for boys). This approach yields similar types of errors (e.g., similar rates of false positives and false negatives).

Although this approach also is defensible, Se and Sp guard against different types of errors and therefore may not need to be weighted equally. Rather than identifying a single value, two independent thresholds were developed (one characterized as a low-risk threshold, and the other characterized as a higher-risk threshold). The advantage of this approach is that it makes it possible to categorize levels of aerobic capacity into three zones rather than two (moderate-risk zone is defined as values between the low- and higher-risk threshold values). The low-risk (high-fit) threshold was established by emphasizing Se over Sp . The high sensitivity of this threshold would ensure that most children with metabolic syndrome would have aerobic-capacity levels below this threshold. A child above this threshold would have a low risk of metabolic syndrome and can be considered as possessing a good level of aerobic capacity.

The Se threshold was set at a higher value for boys ($Se \sim 0.85$) than girls ($Se \sim 0.75$) because of the stronger link between aerobic capacity and metabolic syndrome in boys. Achieving the same level of diagnostic classification accuracy in girls would have necessitated setting standards at an exceptionally high level (values higher than boys for most age groups). The higher-risk (low-fit) threshold was established by emphasizing Sp over Se . The high Sp of this threshold ($>95\%$) would ensure that youth with low levels of aerobic capacity get appropriate feedback about potential risk. The diagnostics suggest that 95% of children without metabolic syndrome would have aerobic-capacity levels above this threshold. It is possible that children with metabolic syndrome could fall above this threshold (due to lowered Se), but there is clear evidence of increased risk (high Sp) for youth scoring below this threshold.

The resulting z-scores for the two thresholds and the associated L, M, and S values were then used to create corresponding VO_2 max estimates. The resulting values are shown in Table 2. As is apparent, the values for boys

increase with age whereas the values for girls decrease with age.

Discussion

This study describes the diagnostic characteristics of newly developed aerobic-capacity standards that could be used in school and sport programs or clinical settings to evaluate adolescents' level of aerobic capacity. The generally high AUC values and high Se/Sp values demonstrate that the aerobic-capacity thresholds have good utility for discriminating youth who may have metabolic syndrome.

The newly developed thresholds follow the same age-related patterns as the previous FITNESSGRAM standards; however, there are some key differences. In boys, the ROC-derived higher-risk (low-fit) threshold is lower than the previous healthy fitness zone (HFZ) for young children but gradually increases with age to approximate the previous standard of 42 mL/kg/min. It is noteworthy that the low-risk zone approaches the accepted adult standard of 42 mL/kg/min by age 18 years—corroborating the utility of this adult standard. The previous standards were set at this constant value of 42 mL/kg/min across the ages, but the new standards developed in this study vary with age. The ROC-derived low-risk (healthy) zone starts below the value of 42 mL/kg/min for young boys but increases above this value for older boys. These standards would result in more boys achieving the HFZ at young ages but fewer achieving it at older ages.

In girls, the ROC-derived higher-risk (low-fit) threshold is lower than the previous standard at young ages but increases with age to approximate the previous HFZ. The ROC-derived low-risk (high-fit) threshold is slightly higher than the previous HFZ but declines with age—at a rate slightly slower than the previous standards. This change would result in fewer girls achieving the HFZ, but differences may be more apparent with older girls.

A unique aspect of the revised standards is that they are equivalent for boys and girls who are aged 10 and 11 years. From a developmental perspective, young boys and girls are more similar than different in physical skills and fitness.²² During adolescence, boys and girls follow different developmental trends,¹⁹ and the use of LMS parameters enables these changes to be taken into account. The proposed standards tend to start diverging at age 14 years, with values decreasing for girls and increasing for boys. It is important to emphasize that these differences do not imply higher expectations for boys and lower expectations for girls. The gender differences are merely reflective of normal growth and maturation. It is important therefore to recognize that the proposed threshold values reflect the same percentile score across the developmental transition. This is a substantial advance from

Table 2. Aerobic-capacity thresholds, percentiles, and corresponding sensitivity and specificity in youth

	Boys		Girls	
	At risk	HFZ	At risk	HFZ
Age (years)				
10.0–10.9	37.3	40.2	37.3	40.2
11.0–11.9	37.3	40.2	37.3	40.2
12.0–12.9	37.6	40.3	37.0	40.1
13.0–13.9	38.6	41.4	36.6	39.7
14.0–14.9	39.6	42.5	36.3	39.4
15.0–15.9	40.6	43.6	36.0	39.1
16.0–16.9	41.1	44.1	35.8	38.9
17.0–17.9	41.2	44.2	35.7	38.8
18.0–18.9	41.2	44.3	35.3	38.6
Sensitivity (95% CI)	59 (55.1, 64.7)	85 (80.8, 88.7)	50.9 (43.2, 58.6.0)	72.2 (49.3, 95.1)
Specificity (95% CI)	92.3 (79.1, 100)	59 (33.4, 84.7)	92.8 (81.3, 100)	71.7 (65.8, 77.6)

HFZ, healthy fitness zone

previous thresholds since it enables the standards to take normal growth and development into account.

The proposed ROC-derived standards provide some clear advantages over the established values that have been used in FITNESSGRAM. The standards also offer advantages over recently proposed standards developed using similar ROC methodology.^{16,18} Using the same NHANES (1999–2000 and 2001–2002) data, Lobelo et al.¹⁶ used raw VO_2 max values rather than LMS z-scores, and a standardized continuous metabolic risk factor score rather than dichotomous metabolic syndrome. For boys, they reported that the threshold values that best discriminated risk for metabolic risk were 44.1 mL/kg/min for those aged 12–15 years and 40.3 mL/kg/min for those aged 16–19 years. For girls, the parallel values were 36.0 and 35.5 mL/kg/min for those aged 12–15 years and 16–19 years, respectively. Since their ROC-derived thresholds were similar to the current FITNESSGRAM values, the authors concluded that the current standards have reasonable utility.

A limitation is that youth were grouped by age bands (12–15 years and 16–19 years). This was necessary to have sufficient samples for the statistical analyses, but it imposed artificial age groups that limited power for the analyses. In the present study, the LMS-derived z-scores were used as the predictor variable, which allowed all of the data to be pooled for ROC analyses. More importantly, the use of LMS allowed the derived thresholds to be redistributed along the designated centiles to establish gender- and age-specific standards.

Another recent study¹⁸ used data from boys and girls aged 9 years and 15 years of the European Youth Heart Study to establish aerobic-capacity standards. The study indicated that the optimal thresholds for detecting elevated cardiovascular disease risk were 43.6 mL/min/kg in boys aged 9 years and 46.0 mL/min/kg for boys aged 15 years. In girls, the thresholds for detecting risk were 37.4 mL/min/kg for those aged 9 years and 33.0 mL/min/kg in those aged 15 years. They reported that specificity (range: 79.3%–86.4%) was higher than sensitivity (range: 29.7%–55.6%) for all threshold values. The thresholds reported in the previous two studies are similar to those proposed here, but the advantage of the current values is that the results provide age-specific values rather than clustered standards for specific age ranges.

Another advantage of the proposed standards is that they have clear diagnostic utility for identifying youth who may be at risk of metabolic syndrome—an indicator that captures risks related to cardiovascular disease, diabetes, and other chronic conditions, including some types of cancer.^{23,24} Historically, it has been assumed that youth were not susceptible to these “adult” health conditions, but given the current pediatric obesity epidemic, the prevalence of the metabolic syndrome has increased in the past decade.²⁵ In addition, the metabolic syndrome tracks reasonably well from adolescence into adulthood.²⁶ The results herein demonstrate that aerobic capacity has reasonable diagnostic utility for detecting risk of metabolic syndrome.

A second advantage is that the standards were derived using nationally representative data (NHANES [1999–2000 and 2001–2002]) and can be considered to be representative of the population aged 12–18 years in the U.S. Some consideration was given to the possible differences in standards for different ethnic groups, but comparisons of LMS curves did not reveal differences for blacks or Hispanics.¹⁹ A final advantage of the proposed ROC-derived standards is that the sensitivity and specificity are

equivalent for boys and girls and across the full age range tested in the analyses (12–18 years). This was accomplished by using LMS-derived z-scores as the predictor variable in the ROC analyses. The use of LMS to control for growth and maturation was the primary distinguishing feature from the two previous studies.^{16,18} Although the previous studies reported similar findings, the proposed standards offer advantages for characterizing levels of aerobic capacity in youth.

There are some limitations of the NHANES (1999–2000 and 2001–2002) design and data used to establish these standards. One limitation is that only about 6% of the sample was diagnosed with metabolic syndrome. The relatively small size of this group reduces the stability of the cut-point determinations. A second limitation is that the aerobic-capacity assessment uses a submaximal protocol. Therefore, it is possible that the data do not equate with actual aerobic-capacity values. Another limitation is that the indicator of metabolic syndrome used here reflects health risk. There are clearly other, alternative health indicators that could be used, as well as different definitions or criteria for evaluating metabolic syndrome.

In summary, aerobic capacity can be used with moderate accuracy to differentiate between adolescents with and without metabolic syndrome. Age- and gender-specific aerobic-capacity thresholds for creating separate risk groups were identified using nationally representative growth percentiles. These values could be useful in school and sport programs or clinical settings. The new FITNESSGRAM standards were based on the values developed in this paper.²⁷

Publication of this article was supported by The Cooper Institute through a philanthropic gift from Lyda Hill.

No financial disclosures were reported by the authors of this paper.

References

- Blair SN, Goodyear NN, Gibbons LW, Cooper KH. Physical fitness and incidence of hypertension in healthy normotensive men and women. *JAMA* 1984;252:487–90.
- Blair SN, Kohl HW, Paffenbarger RS, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality: a prospective study of healthy men and women. *JAMA* 1989;17:2395–401.
- Blair SN, Kohl HW III, Barlow CE, Paffenbarger RS, Gibbons LW, Macera CA. Changes in physical fitness and all-cause mortality: a prospective study of healthy and unhealthy men. *JAMA* 1995;273:1093–8.
- Boreham C, Riddoch C. The physical activity, fitness and health of children. *J Sports Sciences* 2001;19:915–29.
- Twisk JWR. Physical activity, physical fitness and cardiovascular health. In: Armstrong N, van Mechelen W, eds. *Pediatric exercise science and medicine*. London U.K.: Oxford University Press, 2000.
- Ortega FB, Ruiz JR, Castillo MJ, Sjöström M. Physical fitness in childhood and adolescence: a powerful marker of health. *Int J Obes* 2008;32:1–11.
- Malina RM. Tracking of physical activity and physical fitness across the lifespan. *Res Q Exerc Sport* 1996;67:S48–57.
- Dwyer T, Magnusson CG, Schmidt MD, Ukoumunne OC, Ponsonby A-L, Raitakari OT, et al. Decline in physical fitness from childhood to adulthood associated with increased obesity and insulin resistance in adults. *Diabetes Care* 2009;34:683–7.
- Cleland VJ, Dwyer T, Venn AJ. Physical activity and healthy weight maintenance from childhood to adulthood. *Obesity* 2008;16:1427–33.
- Ferreira I, Twisk JWR, Stehouwer CDA, van Mechelen W, Kemper HCG. Longitudinal changes in Vo2max: associations with carotid IMT and arterial stiffness. *Med Sci Sports Exerc* 2003;35:1670–8.
- Carnethon MR, Gulati M, Greenland P. Prevalence and cardiovascular disease correlates of low cardiorespiratory fitness in adolescents and adults. *JAMA* 2005;294:2981–8.
- Pate RR, Wang CY, Dowda M, Farrell SW, O'Neill JR. Cardiorespiratory fitness levels among U.S. youth 12 to 19 years of age: findings from the 1999–2002 National Health and Nutrition Examination Survey. *Arch Pediatr Adolesc Med* 2006;160:1005–12.
- Zhu W, Mahar MT, Welk GJ, Going SB, Cureton KJ. Approaches for development of criterion-referenced standards in health-related youth fitness tests. *Am J Prev Med* 2011;41(4S2):S68–76.
- Cureton KJ, Warren BL. Criterion-referenced standards for youth health-related fitness tests: a tutorial. *Res Q Exerc Sport* 1990;61:7–19.
- Cureton KJ, Plowman SA. Aerobic capacity assessments. In Welk GJ, Meredith MD, eds. *FITNESSGRAM®/ACTIVITYGRAM® reference guide*. 3rd ed. Dallas TX: The Cooper Institute, 2008.
- Lobelo F, Pate RR, Dowda M, Liese AD, Ruiz JR. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Med Sci Sports Exerc* 2009;41:1222–9.
- Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–77.
- Adegboye AR, Anderssen SA, Froberg K, et al. Recommended aerobic fitness level for metabolic health in children and adolescents: a study of diagnostic accuracy. *Br J Sports Med* 2011;45(9):722–8.
- Eisenmann JC, Laurson KR, Welk GJ. Aerobic fitness percentiles for U.S. adolescents. *Am J Prev Med* 2011;41(4S2):S106–110.
- Jolliffe CJ, Janssen I. Development of age-specific adolescent metabolic syndrome criteria that are linked to the Adult Treatment Panel III and International Diabetes Federation criteria. *J Am Coll Cardiol* 2007;49:891–8.
- Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
- Malina RM, Bouchard C, Bar-Or O. *Growth, maturation, and physical activity*. 2nd ed. Champaign IL: Human Kinetics, 2004.
- Hsu IR, Kim SP, Kabir M, Bergman RN. Metabolic syndrome, hyperinsulinemia, and cancer. *Am J Clin Nutr* 2007;86:s867–71.
- Grundy SM. Metabolic syndrome: a multiplex cardiovascular risk factor. *J Clin Endocrinol Metab* 2007;92:399–404.
- Cook S, Auinger P, Li C, Ford ES. Metabolic syndrome rates in U.S. adolescents, from the National Health and Nutrition Examination Survey, 1999–2002. *J Pediatr* 2008;152:165–70.
- Camhi SM, Katzmarzyk PT. Tracking of cardiometabolic risk factor clustering from childhood to adulthood. *Int J Pediatr Obes* 2010;5:122–9.
- Meredith M, Cooper Institute (Dallas TX). *FITNESSGRAM ACTIVITYGRAM: Test administration manual*. Updated 4th edition. Champaign IL: Human Kinetics, 2010.