Receiver operating characteristic analysis of body mass index, triceps skinfold thickness, and arm girth for obesity screening in children and adolescents\(^1,2\)

**Luis B Sardinha, Scott B Going, Pedro J Teixeira, and Timothy G Lohman**

**ABSTRACT**

**Background:** Valid and practical methods based on health-related criteria for obesity screening in children and adolescents are not available. Arbitrarily defined body mass index (BMI) cut-offs have been proposed to select adolescents at high risk of developing obesity in adulthood.

**Objective:** We assessed the usefulness of BMI, triceps skinfold thickness, and upper arm girth for screening for obesity by using a health-related definition of obesity (≥25% body fat in boys and ≥30% body fat in girls) and a criterion method (dual-energy X-ray absorptiometry) that estimates percentage fat without the potential bias associated with other methods in adolescents.

**Design:** This was a cross-sectional study of Portuguese boys (\(n = 165\)) and girls (\(n = 163\)) aged 10–15 y. Nonparametric receiver operating characteristic (ROC) analysis was used to define the best tradeoff between true-positive and false-positive rates.

**Results:** True-positive rates ranged from 67% to 87% and from 50% to 100% in girls and boys, respectively, and false-positive rates ranged from 0% to 19% and from 5% to 26%, respectively. For children aged 10–11 y, the areas under the curves (AUCs) for ROCs, an index of diagnostic accuracy, were close to 1.0, suggesting very good accuracy. For older boys and girls, AUCs for triceps skinfold thickness were similar to or greater than AUCs for BMI and upper arm girth.

**Conclusions:** The results suggest that triceps skinfold thickness gives the best results for obesity screening in adolescents aged 10–15 y. BMI and upper arm girth were reasonable alternatives, except in 14–15-y-old boys, in whom both indexes were only marginally able to discriminate obesity. *Am J Clin Nutr* 1999;70:1090–5.

**KEY WORDS** Obesity, receiver operating characteristic curve, ROC curve, dual-energy X-ray absorptiometry, DXA, anthropometry, body mass index, BMI, triceps skinfold thickness, arm girth, screening, children, adolescents

**INTRODUCTION**

Excess body fat is the hallmark of obesity. Numerous reports have confirmed the strong association between obesity and elevated risk of chronic disease, and the estimated health care costs of the sequelae of obesity are enormous (1). Despite widespread concern about obesity, the development of standard definitions of obesity for screening and intervention remains problematic. Because of the lack of simple, accurate methods for assessing body fat directly, anthropometric indexes such as the body mass index (BMI; \(\text{kg/m}^2\)) are often used as surrogates for body composition (2–6). Interpretation of results is difficult, however, because BMI reflects relative leg length, body frame size, and fat-free mass in addition to fatness (7). Consequently, 2 persons with the same amount of body fat can have quite different BMI values. These problems are underscored by the discordant estimates of prevalence when obesity is estimated from BMI and other anthropometric measures such as skinfold thicknesses (8, 9).

Despite the potential problems with the use of BMI, an expert committee recommended that the 85th percentile of BMI be used as a cutoff for routine preventive obesity screening in adolescents (4). The usefulness of this cutoff is unclear, however. Arbitrary cutoffs derived from a reference distribution have ≥2 limitations. First, in growing adolescents, the BMI-fatness relation is confounded by significant changes in the ratio of muscle and bone to height and the percentage fat associated with the 85th percentile of BMI can vary considerably (10). Second, changes over time in the distribution of weight-for-height in the reference population may change risk with no concomitant change in the estimate of the prevalence of obesity.

Lazarus et al (11) used receiver operating characteristic (ROC) plots to examine the performance of BMI in screening children and adolescents for excess body fat as defined by body fatness greater than or equal to the 85th percentile of percentage total body fat. The use of a percentile-based definition of excess body fat is problematic, however, because the average percentage body fat associated with the 85th percentile varies considerably, as it did in the study by Lazarus et al: from 18% to 33% in boys and from 24% to 37% in girls. Thus, although boys and...
girls with higher percentages of body fat relative to the group may be identified, these children may not be obese or have elevated risk factors for disease.

There is a need to better define the relation between BMI and percentage body fat in children and adolescents. Thus, the purpose of this study was to assess this relation in children and adolescents by ROC analysis, with the goal of defining BMI cutoffs with the greatest sensitivity and specificity for detecting adolescent obesity. For comparison, similar analyses were done to assess the usefulness of triceps skinfold thickness and upper arm girth as the diagnostic test. In contrast with Lazarus et al (11), we used a health-related criterion to define obesity as percentage body fat ≥25% in males and ≥30% in females (12) as measured by dual-energy X-ray absorptiometry (DXA). DXA is an appropriate criterion method in children and adolescents because it is based on a 3-compartment model in which bone mineral, lean soft tissue, and body fat are all assessed (13–15).

SUBJECTS AND METHODS

Subjects

One hundred sixty-three girls and 165 boys aged 10–15 y volunteered for the study. All subjects were white. Subjects were recruited from 4 high schools and were free of any known disease. All subjects who participated in this study underwent anthropometric assessment and measurement of body composition by DXA. For data analysis, subjects were categorized into subgroups according to sex and age as follows: girls 10–11 (n = 54), 12–13 (n = 60), and 14–15 (n = 49) y, and boys 10–11 (n = 55), 12–13 (n = 52), and 14–15 (n = 58) y. Written, informed consent was obtained from all subjects’ parents before the subjects participated in the study. The experimental protocol was approved by the Ethical Committee of the Faculty of Human Movement, Technical University of Lisbon.

Anthropometric measurements

Weight (to the nearest 0.1 kg) was measured with an electronic scale (model 770; Seca, Hamburg, Germany) and height (to the nearest 0.1 cm) was measured with a wall-mounted stadiometer. For height measurements, subjects were standing without shoes on a horizontal surface with their bodies stretched upward to the fullest extension and their heads in the Frankfurt plane. BMI was derived from weight (kg) and height² (m). Arm girth and triceps skinfold thickness were measured according to the procedures described in the Anthropometric Standardization Reference Manual (16). Arm girth was measured with an anthropometric tape and triceps skinfold thickness was measured with Lange calipers (Cambridge Scientific Instruments, Cambridge, MD). Arm girth and triceps skinfold thickness were each measured twice by a trained technician. The means of the 2 trials were used in all statistical analyses.

Body composition

Total percentage body fat was estimated from total-body DXA scans made with a QDR-1500 total-body scanner (Hologic Inc, Waltman, MA) in the pencil beam mode with the enhanced whole-body analysis (software version 5.67). With this technique, composition is estimated from the attenuation of X-rays pulsed synchronously between 70 and 140 kV with the line frequency for each pixel of the scanned image. Comprehensive reviews of the theory and methodology of DXA measurements of body composition have been published (15, 17). A step phantom with 6 fields of acrylic and aluminum of various thicknesses and known absorptive properties was scanned alongside each subject to serve as an external standard for the analysis of tissue composition. The same technician positioned the subjects, performed the scans, and performed the analysis according to the operator’s manual following the standard analysis protocol. The within-subjects SD and percentage CV for duplicate measurements of percentage body fat for 14 subjects were 0.4% and 1.8%, respectively.

Criterion-based definition of obesity

A health-related, criterion-based definition considers the known relation between a screening variable and selected health outcomes. To our knowledge, there are only 2 published criterion-based definitions of obesity in children and adolescents (12, 18). The results of Williams et al (12) suggested a criterion of ≥25% body fat for boys and ≥30% body fat for girls. For 9- and 15-y-old girls and boys, Dwyer and Blizzard (18) suggested the same value for girls, but a cutoff of 20% body fat for boys. Considering that boys and girls ranging from 5 to 18 y of age were included in the Williams et al (12) sample, whereas Dwyer and Blizzard (18) studied only 9- and 15-y-olds, we defined obesity as ≥25% body fat for boys and ≥30% body fat for girls in the present study. These standards were defined on the basis of findings that higher percentages of body fat increased the prevalence of selected cardiovascular disease risk factors, such as high systolic and diastolic blood pressure and unfavorable lipoprotein ratios, in 3320 children and adolescents aged 5–18 y relative to their peers with lower percentages of body fat (12). Note that the percentage body fat standards used by Williams et al (12) were derived from body density as estimated from skinfold-thickness measurements. The limitations of anthropometry and densitometry in children and adolescents are well known (8, 19, 20). However, it is unlikely that any bias occurred given that the equation used to convert skinfold thicknesses to density was derived from pooled data from several well-described samples and that percentage body fat was calculated from density by using age- and sex-specific adjustments for variation in fat-free body composition and density. Thus, the typical errors due to heterogeneity in fat-free mass were addressed and minimized.

Statistical analysis

ROC analysis is a way of evaluating the accuracy of a diagnostic test by summarizing the potential of the test to discriminate between the absence and presence of a health condition (21, 22). In the context of the present study, this diagnostic accuracy refers to the ability of the anthropometric variables to discriminate obesity from nonobesity, as assessed by percentage body fat measured by DXA with cutoffs of 25% and 30% body fat for boys and girls, respectively. With use of these criterion values, obese boys and girls who were classified correctly as obese by the anthropometric variables represent the true-positive cases, whereas obese subjects classified as nonobese represent false-negative cases. Nonobese subjects classified correctly as nonobese represent the true-negative cases, whereas nonobese subjects classified as obese represent false-positive cases.

The sensitivity of the anthropometric diagnosis is the probability that the anthropometric variables will classify a subject as obese when the subject is truly obese; the specificity is the probability that the anthropometric variables will classify a subject as...
Inc, Chicago). Cance for all analyses was set at
assessed by one-way analysis of variance (ANOV A), followed by
continuous data to introduce ties, it provides unbiased estimates
girth were performed by using a nonparametric approach (21,
tributions, the AUC is 1.0 (22). In the present study, the AUCs
fect separation between the 2 conditions with no overlap of dis-
variable cannot distinguish between the conditions of interest
The physical characteristics of the subjects are given in Table 1.
increased with age. Triceps skinfold thickness and arm girth were
years than in the oldest girls, whereas in boys arm girth was greater in the oldest group than in the
other 2 age groups. Percentage body fat was not significantly dif-
percentage body fat to be lower in older boys than in younger boys, with a significant difference between
between the youngest group and the 2 other groups. The prevalence of obe-
by the tests (1-false-positive rate) was greater than the propor-
positive rates ranged from 0% to 19% and from 5% to 26%.
true-positive rates ranged from 67% to 87%
and

TABLE 1
Physical characteristics of the subjects by sex and age.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Weight (kg)</th>
<th>Height (cm)</th>
<th>BMI (kg/m²)</th>
<th>Triceps skinfold thickness (mm)</th>
<th>Arm girth (cm)</th>
<th>Body fat (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10–11 y (n = 54)</td>
<td>40.6 ± 8.9¹</td>
<td>143.9 ± 7.4²</td>
<td>19.5 ± 3.4³</td>
<td>19.7 ± 6.8⁴</td>
<td>23.7 ± 3.3⁵</td>
<td>30.1 ± 8.7⁶</td>
</tr>
<tr>
<td>12–13 y (n = 60)</td>
<td>50.0 ± 10.2</td>
<td>155.1 ± 7.1</td>
<td>20.7 ± 5.7</td>
<td>20.1 ± 5.7⁷</td>
<td>25.1 ± 3.1</td>
<td>28.3 ± 6.8⁸</td>
</tr>
<tr>
<td>14–15 y (n = 49)</td>
<td>55.6 ± 9.8</td>
<td>158.7 ± 5.8⁹</td>
<td>21.9 ± 3.3</td>
<td>22.7 ± 5.8⁹</td>
<td>26.4 ± 3.4</td>
<td>29.3 ± 7.8³</td>
</tr>
<tr>
<td>Boys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10–11 y (n = 55)</td>
<td>42.2 ± 9.8⁴</td>
<td>142.6 ± 6.7⁷</td>
<td>20.5 ± 5.0</td>
<td>19.4 ± 8.3</td>
<td>24.1 ± 4.4</td>
<td>26.7 ± 10.8⁶</td>
</tr>
<tr>
<td>12–13 y (n = 52)</td>
<td>46.9 ± 9.7⁷</td>
<td>153.3 ± 9.7</td>
<td>19.7 ± 4.2</td>
<td>16.4 ± 6.9</td>
<td>23.9 ± 3.9</td>
<td>20.9 ± 9.9</td>
</tr>
<tr>
<td>14–15 y (n = 58)</td>
<td>57.8 ± 9.8</td>
<td>164.3 ± 7.9</td>
<td>21.3 ± 2.9</td>
<td>16.5 ± 7.2</td>
<td>26.0 ± 3.2⁸</td>
<td>17.3 ± 8.2</td>
</tr>
</tbody>
</table>

¹X ± SD.
²Percentage body fat by dual-energy X-ray absorptiometry.
³All age groups within a sex significantly different, P < 0.05.
⁴Significantly different from the 14–15-y-old group, P < 0.05.
⁵Significantly different from boys of the same age, P < 0.05.
⁶Significantly different from the other 2 age groups, P < 0.05.

RESULTS
The physical characteristics of the subjects are given in Table 1. In both girls and boys, weight, height, and BMI generally
boys. False-positive rates were lower in girls than in boys at all ages for all indexes except for BMI in 14–15-y-olds.

Shown in Table 4 are the AUC estimates for BMI, triceps skinfold thickness, and arm girth for the 3 age groups of boys and girls. In children aged 10–11 y, the AUCs for BMI, triceps skinfold thickness, and arm girth were not significantly different from each other. All values in this age group were close to 1.0, suggesting very good diagnostic accuracy for all anthropometric indexes. In girls aged 12–13 y, the AUCs for triceps skinfold thickness and arm girth were not significantly different from each other and were both better than the AUC for BMI. In the 14–15-y-old girls, however, the AUC for triceps skinfold thickness was higher than the AUCs for both BMI and arm girth (which were not significantly different from each other). In both older groups of boys, the AUC for triceps skinfold thickness was higher than the respective AUCs for BMI and arm girth (which were not significantly different from each other within each age group). In 14–15-y-old boys, the 95% CIs were 0.47 and 0.72 for BMI and 0.48 and 0.74 for arm girth. The lower limits of these 95% CIs are very close to 0.5, suggesting poor diagnostic performance by BMI and arm girth in the oldest boys.

DISCUSSION

The results of this study suggest that BMI, triceps skinfold thickness, and upper arm circumference can be used with reasonable success to detect childhood and adolescent obesity. The cutoffs with the best tradeoff between true-positive and false-positive rates varied predictably with sex and age, as did the specific index that gave the best overall discrimination. For our knowledge, in only one other study were ROC curves used to assess the usefulness of BMI for obesity screening in children and adolescents (11). Although there were methodologic differences between that study and ours, the results of that study were qualitatively similar to ours and supported the use of BMI as an obesity index in children and adolescents. The absolute value of BMI giving the best tradeoff between true-positive and false-positive rates approximated the 70–75th percentile of the age- and sex-matched weighted average BMI from a large, multiethnic sample of US children and adolescents, except in 14–15-y-old boys, whose BMI cutoff was closer to the 85th percentile (6). Thus, the recommended BMI cutoffs for Portuguese boys and girls correspond to lower percentiles than those recommended for US boys and girls when US norms were applied (4). Similar population-based normative data are not presently available for Portuguese children and adolescents.

In the past, anthropometric indexes of obesity were recommended largely on the basis of their correlations with body fat (5, 19, 28). Correlational analysis is a useful technique for describing the associations between anthropometric variables and body fat and for making recommendations for prediction equations. However, correlational analysis cannot describe the nature and extent of misclassifications, and when the purpose is to discriminate obesity from nonobesity, it is inappropriate to make recommendations based on correlations alone. For this purpose, categorical analyses and a valid measure of true obesity are required. Unlike correlational analysis, which is blind to systematic bias, categorical classification requires referent values that are especially sensitive to bias. Thus, an appropriate criterion technique for estimating body fatness is a critical requirement for obtaining valid results.

The choice of a reference test for use in children and adolescents has been problematic because the validity of the assumptions underlying most accessible laboratory tests is questionable in prepubescent and pubescent boys and girls. In contrast with Himes and Bouchard (29) and Johnston (30), who used underwater weighing, we used DXA as the criterion method. Because DXA estimates soft tissue composition independent of bone mineral (15, 17), it does not require the assumption of an invariant composition of fat-free mass, which introduces systematic overestimation of body fat in children and adolescents by underwater weighing and other techniques (20). Although variation in the hydration of fat-free mass is another potential source of error, empirical data and theoretical calculations have shown that large fluctuations in fluid must occur before estimates of body fat by DXA are biased (15, 31). The results of recent validation studies showing accurate estimation of body fat in children and adolescents (13, 14) also support the use of DXA as a valid criterion method.

<table>
<thead>
<tr>
<th>Age group</th>
<th>BMI (kg/m²)</th>
<th>Triceps skinfold thickness (mm)</th>
<th>Arm girth (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10–11 y (n = 55)</td>
<td>19.0</td>
<td>17.0</td>
<td>22.8</td>
</tr>
<tr>
<td>Cutoff</td>
<td>96.0</td>
<td>100.0</td>
<td>96.0</td>
</tr>
<tr>
<td>True-positive rate (%)</td>
<td>14.0</td>
<td>10.0</td>
<td>14.0</td>
</tr>
<tr>
<td>False-positive rate (%)</td>
<td>12–13 y (n = 52)</td>
<td>19.4</td>
<td>19.0</td>
</tr>
<tr>
<td>Cutoff</td>
<td>86.0</td>
<td>86.0</td>
<td>71.0</td>
</tr>
<tr>
<td>True-positive rate (%)</td>
<td>24.0</td>
<td>5.0</td>
<td>18.0</td>
</tr>
<tr>
<td>False-positive rate (%)</td>
<td>14–15 y (n = 58)</td>
<td>24.0</td>
<td>19.0</td>
</tr>
<tr>
<td>Cutoff</td>
<td>50.0</td>
<td>63.0</td>
<td>50.0</td>
</tr>
<tr>
<td>True-positive rate (%)</td>
<td>8.0</td>
<td>26.0</td>
<td>24.0</td>
</tr>
</tbody>
</table>
In the absence of widely accepted definitions of excess adiposity, age- and sex-specific percentile cutoffs of percentage body fat have been used as referent definitions of obesity (11, 29, 32, 33). Although statistically valid, this approach is arbitrary and the percentage body fat corresponding with a given percentile may vary considerably among samples. Moreover, because body fatness increases with age, the percentage body fat corresponding to a given percentile will likely be different at different ages. This point is illustrated in the study by Lazarus et al (11), in which the percentage body fat corresponding with the 85th percentile ranged from 18% to 33% in males and from 24% to 37% in females aged 4–20 y. A similar situation occurs in adults (33).

Some authors have used linear regression to adjust for the effect of age (11, 33). In this approach, the residual value for each subject from the regression line is divided by the SE to obtain a “studentized” residual analogous to a z score, and these standard scores are then converted to percentile ranks. The results are misleading, however, because the interpretation of a standard score of 2.0, for example, above an estimated percentage body fat of 35% compared with 25% is quite different in its relation to obesity, although the percentile ranking would be equivalent.

Ideally, when designing an obesity screening program, the choice of a cutoff should be based on health-related criteria. Unfortunately, the long-term health outcomes for different amounts of adiposity at different ages have not been described. Nevertheless, we used a health-related definition of obesity (≥25% body fat in boys and ≥30% body fat in girls) based on the work of Williams et al (12), who showed no “excess risk” of elevated values of disease risk factors at body fat percentages <25% in boys and 30% in girls. The findings were similar for white and black boys and girls aged 7–18 y, an age range similar to that in our study sample. The results of Dwyer and Blizzard (18) also support this approach, although these authors derived a lower cutoff for boys (20% compared with 25%) than did Williams et al (12). The results from these 2 studies (12, 18) suggest that a single, sex-specific criterion can be applied throughout adolescence, and, in the absence of contrary data, we believe that defining obesity by health-related criteria is more useful than definitions based on population distributions.

Our aim was to examine a range of values for each diagnostic test to determine the cutoffs with the greatest accuracy, ie, the cutoff with maximum true-positive and minimum false-positive tests. This approach resulted in BMI cutoffs that were lower than the predetermined cutoffs tested in previous studies (eg, age- and sex-matched 85th or 95th percentile). It is not surprising then that the true-positive rates (50–96% in boys and 67–83% in girls) were higher than those reported by other investigators with referent estimates of percentage body fat from underwater weighing (29, 30) and DXA (11) because lowering the cutoff should raise the true-positive rate. The corresponding false-positive rates (8–24% in boys and 3–13% in girls) were comparable with or higher than previously reported values, as would be expected given that lowering cutoffs to raise true-positive rates is expected to also increase the number of false-positive tests.

The cutoffs for BMI, upper arm girth, and triceps skinfold thickness increased predictably with age in boys and girls given the typical changes that occur in body weight and composition across the ages studied. The BMI and arm girth cutoffs increased more in boys than in girls, whereas the triceps-skinfold-thickness cutoffs increased more in girls than in boys. These changes are consistent with the greater gains in muscle and bone experienced by boys during adolescence and the greater gains in total body fat and arm and leg subcutaneous adipose tissue experienced by girls (34).

We used the areas under the ROC curves as an indication of the overall performance of the indexes, which is analogous to the indicator efficiency described by Himes and Bouchard (29). The AUCs for triceps skinfold thickness, the most direct index of fatness, were equal to or higher than the AUCs for BMI and upper arm girth at all ages, although there were no significant differences among the indexes in the youngest boys and girls. Note, however, that the triceps skinfold-thickness measurement probably has a lower reliability than BMI or arm girth, which reinforces the need for adequate technician training. There was little difference between the performances of BMI and upper arm girth, except in 12–13-y-old girls, for whom the AUC was higher for upper arm girth than for BMI. In general, the overall performances of all the indexes were good, except for BMI and upper arm girth in 14–15-y-old boys, for whom the 95% CI included an AUC of 0.5.

In summary, our results suggest that measurements of triceps skinfold thickness give the best results for obesity screening in boys and girls aged 10–15 y. BMI and upper arm girth are reasonable second choices, except in 14–15-y-old boys, for whom both indexes were only marginally able to discriminate between obesity and nonobesity. The best cutoffs determined by ROC analysis had higher true-positive rates and somewhat higher false-positive rates than reported for predetermined cutoffs. Still, with these cutoffs, the misclassification of nonobese children and adolescents as obese (false-positive rate) was lower than the misclassification of obese children and adolescents as nonobese (false-negative rate). Past investigators generally have recommended a conservative approach, opting for lower false-positive rates because of their concern for the psychosocial implications.
REFERENCES