Sensitivity and specificity of classification systems for fatness in adolescents

Martin G Neovius, Yvonne M Linné, Britta S Barkeling, and Stephan O Rossner

ABSTRACT

Background: Various body mass index (BMI) standards have been proposed for defining overweight in adolescence, but few studies have evaluated their diagnostic accuracy.

Objective: We compared the sensitivity and specificity of BMI-based classification systems for detecting excess fatness in adolescents.

Design: A cross-sectional analysis of 474 adolescents aged 17 y was used. Body composition was measured by using densitometry. The international BMI-based systems recommended by the International Obesity Task Force and the World Health Organization were evaluated on the basis of their sensitivity and specificity for detecting excess body fat. Receiver operating characteristic analysis was performed to derive cutoffs to maximize the sum of sensitivity and specificity. True positives were defined by using the percentage body fat cutoffs proposed by Williams et al (Am J Public Health 1992;82:358–63).

Results: For both classification systems, the specificity for overweight was high for both sexes (0.95–1.00). The sensitivity was fairly high for the males (0.72–0.84) but was very low for the females (0.22–0.25). For the males, a BMI cutoff equal to the 85th percentile on a Swedish BMI reference chart maximized the sum of sensitivity and specificity while having both high sensitivity (0.92) and high specificity (0.92). For the females, larger tradeoffs in specificity were needed to improve sensitivity. The mean (±SE) areas under the receiver operating characteristic curves for the males and the females were 0.97 ± 0.02 and 0.85 ± 0.02, respectively.

Conclusions: Recommended international classification systems have very high specificity, which results in few cases of nonoverweight adolescents being mislabeled as overweight. However, the sensitivity is very low in female adolescents. Thus, many overweight female adolescents could be missed in intervention programs that use the proposed international BMI cutoffs as selection criteria. Am J Clin Nutr 2004;80:597–603.

KEY WORDS Adolescent overweight, body mass index, classification systems, percentage body fat, diagnostic accuracy, sensitivity, specificity, receiver operating characteristic analysis

INTRODUCTION

Despite the growing concern about adiposity-related problems among the young, no universally accepted classification system for adolescent obesity exists. Although body mass index (BMI; in kg/m²) is widely used for classification of adult overweight and obesity, its use in adolescents is controversial (1–3). The limitations of BMI as a measure of adiposity in the pediatric population are larger than those in the adult population because BMI varies with age, sex, and maturation (4, 5). An additional complication, for all age groups, is that relative risks associated with certain BMI values seem to be population dependent (6). Thus, universal classification systems are difficult to design. Currently, there are a number of proposed systems. For example, the International Obesity Task Force (IOTF) and the World Health Organization (WHO) have recommended different international classification systems for childhood and adolescent obesity (7, 8). In addition to these systems, national variants exist (9). The controversy around the classification systems makes it difficult to monitor global and national trends, make comparisons between studies, stratify for public health measures, and screen in clinical practice. Furthermore, messages to the media and hence the public might be confusing when population prevalence estimates fluctuate depending on the choice of classification system (10).

The classification system proposed by Cole et al (IOTF/Cole) (7), which is recommended by the IOTF, is gaining increasing acceptance. This system was derived mainly for global monitoring (7). However, a trend toward recommending the use of the IOTF/Cole reference for clinical practice and public health measures at the national level has developed (11). Such recommendations are made despite the fact that the reference has not been thoroughly evaluated in terms of screening ability and relation to morbidity and mortality (11). Two studies have hitherto focused on evaluating classification systems on their ability to detect excess body fat among children and adolescents (12, 13). A limiting factor in such attempts is the absence of reference values defining overweight or obesity in terms of percentage body fat (%BF). The studies found that the IOTF/Cole system is highly specific but is insensitive for finding obesity (12, 13). However,
both studies arbitrarily defined true positives for obesity as the top 5% in the study population. Hence, the standards used were not anchored to health outcomes, which can be done by using a health-related criterion to define overweight or obesity as above certain threshold values of %BF (14, 15). The purpose of the present study was threefold: 1) to evaluate the sensitivity and specificity of recommended, international, BMI-based classification systems for detecting fatness, 2) to compare these systems with a national reference, and 3) to examine the influence on the analyses of the choice of reference values for excess fatness.

SUBJECTS AND METHODS

The subjects in the Stockholm Weight Development Study were 481 adolescents (n = 279 females and 202 males). Body-composition data were available for 474 of the subjects. The adolescents were a subset of the offspring of 1423 women who participated in the Stockholm Pregnancy and Weight Development Study in 1984–1985 (16). The local Ethical Committee of Huddinge University Hospital granted ethical approval for the study. Written informed consent was obtained from each mother, and verbal consent was also obtained from each adolescent.

The BodPod Body Composition System (Life Measurement Instruments, Concord, CA) was used to measure the subjects’ weight to the nearest 0.1 kg while they were dressed only in underwear. The subjects’ standing height was measured to the nearest 0.5 cm while they stood against a wall-mounted stadiometer. BMI was determined as Quetelet’s index (kg/m²).

%BF was measured by using air-displacement plethysmography with the BodPod. The equipment was used in an enclosed room without windows, where a constant environment could be kept. A series of repeated measurements was performed on phantoms of known weights and volumes for the assessment of methodologic error. Two measurements were performed on each fast- ing subject according to the manufacturer’s instructions and recommendations, with the subject wearing tight-fitting underwear or a swimsuit and a swim cap (17, 18). A single air-displacement plethysmography procedure consisted of 2 measurements of body volume. If these differed by >150 mL, a third measurement was performed. By using preprogrammed equations, predicted lung volume was used to calculate body volume. Appropriate corrections for thoracic gas volume and skin surface area artifact were applied to this raw measurement to obtain actual body volume. The final result reported by the instrumentation was calculated from the average of the raw measurements or from the average of the closest 2 measurements when 3 measurements were required. Data on body density were converted to %BF by using the equation of Siri (19), as used by the software supplied by the manufacturer.

BMI-based classification systems

The IOTF/Cole system consists of sex-specific BMI percentile curves that at age 18 y pass through the BMI cutoffs for adult overweight and obesity of 25 and 30, respectively (7). The definitions of adolescent overweight and obesity are thereby linked to adult risk. The percentile curves were produced from large-survey data from the United Kingdom, the United States, Holland, Singapore, Hong Kong, and Brazil (n = 97,876 males and 94,851 females) (7). The reference is recommended by the IOTF and is widely used (4).

The WHO/MDD system was derived by Must, Dallal, and Dietz (MDD) from data collected in 1971–1974 as part of the first US National Health and Nutrition Examination Survey (NHANES I) (8). It is a sex- and age-specific percentile-based system in which overweight and obesity (or at risk of overweight and overweight) are defined as BMI values above the 85th and 95th percentiles, respectively. The reference has been recommended by several health organizations, including a WHO Expert Committee (20).

He et al (21) derived age- and sex-specific percentiles from a longitudinal study of 3650 full-term infants born in Sweden in the 1970s. In comparison with American BMI reference values, the Swedish values are much lower, especially at the higher percentiles (21). Cutoffs for classification of overweight have not been derived. In the present article, the age- and sex-specific BMI cutoff of the IOTF/Cole system that corresponds to a BMI of 25 at age 18 y, the WHO/MDD 85th percentile, and the He et al 85th percentile were used in defining subjects as normal-weight or overweight.

Definition of excess body fat

There are no generally accepted %BF cutoffs for excess fatness or for overweight or obesity in children and adolescents. Several previous studies have defined childhood or adolescent obesity as the fattest 5% in the sample as determined by various measures of %BF (2, 12, 13). Such a method sets the true prevalence to a fixed percentage, and although persons with higher %BF than other persons in the group may be identified, the relation to increased morbidity risk remains unclear and may vary. However, Williams et al (15) published %BF cutoffs derived from findings of a significant overrepresentation of selected cardiovascular risk factors, such as high blood pressure and unfavorable lipoprotein profiles. In a sample of 3320 subjects aged 5–18 y, Williams et al (15) found that %BF values of 25% and 30% for males and females, respectively, were suitable to define excess fatness. %BF estimates were derived from skinfold thickness measurements, a method that has limitations in adolescents (4, 22). However, through the methodology used to convert the measurements to %BF, the typical errors due to heterogeneity in fat-free mass were minimized, as described by Sardinha et al (14). Therefore, it is less likely that any bias occurred (14). In the present study, these criterion-based cutoffs were used as reference values for defining overweight to avoid setting the prevalence to a fixed percentage by using %BF cutoffs that are unrelated to metabolic risk.

Statistical analyses

Statistical analyses were performed by using SPSS for WINDOWS (version 11.5; SPSS Inc, Chicago). Sensitivity for fatness was defined as the probability of the respective systems to classify subjects with excess fatness as overweight (true positives). Specificity was defined as the probability of classifying subjects without excess fatness as nonoverweight (true negatives). Receiver operating characteristic (ROC) analysis was performed to determine cutoff values to minimize the total number of misclassifications and evaluate the general performance of BMI in reflecting body fatness. ROC analysis describes the clinical performance of screening tests in terms of diagnostic accuracy or the ability to correctly classify subjects into clinically relevant subgroups, as defined by a reference test (23). The diagnostic accuracy of the screening measure is evaluated by summarizing the
potential of the test to discriminate between the absence and presence of a health condition. In the present study, the diagnostic accuracy referred to the ability of BMI to discriminate overweight from nonoverweight as assessed by %BF measured with the use of air-displacement plethysmography and as defined by the %BF cutoffs proposed by Williams et al (15). In the ROC analysis, the true-positive rate (sensitivity) is plotted against the false-positive rate (1 − specificity) across a range of values from the diagnostic test. In the present study, sex-specific curves were constructed with %BF as the reference test and BMI as the diagnostic test. Thereafter, BMI cutoffs maximizing the sum of sensitivity and specificity were derived.

The area under the ROC curve was used as a measure of the overall performance of the ROC curve because it reflects the probability that the diagnostic test will classify correctly (24). The area under the ROC curve can take values between 0 and 1, where 1 is a perfect screening test and 0.5 is a test equal to chance. In the ROC curves below, a line was plotted at a 45° angle to represent an area under the ROC curve of 0.5. Positive [sensitivity/(1 − specificity)] and negative [(1 − sensitivity)/specificity] likelihood ratios were also calculated to express the odds that a given value of a screening test outcome would be expected in a person with or without the target disorder, respectively.

### TABLE 1
Subject characteristics

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 200)</th>
<th>Females (n = 274)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>16.9 ± 0.4</td>
<td>16.8 ± 0.4</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>68.7 ± 12.0</td>
<td>59.6 ± 9.1</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.80 ± 0.06</td>
<td>1.68 ± 0.06</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>21.3 ± 3.3</td>
<td>21.6 ± 3.1</td>
</tr>
<tr>
<td>%BF</td>
<td>16.2 ± 7.5</td>
<td>29.4 ± 6.5</td>
</tr>
</tbody>
</table>

1 All values are x ± SD. %BF, percentage body fat measured by using air-displacement plethysmography.

2 Significantly different from males, P < 0.001 (independent-sample t test).

### RESULTS
Subject characteristics are presented in Table 1. The mean BMI values did not differ significantly between the sexes, whereas the mean %BF was significantly higher in the females than in the males (P < 0.001). The mean %BF for the males was almost 9 percentage points lower than the recommended 25%BF cutoff for overweight in males, whereas the mean %BF for the females was nearly equivalent to the proposed cutoff of 30% (Table 2). This explains the high prevalence of true positives for overweight among the females, in comparison with the prevalence of overweight as defined by various BMI-based references (Table 2).

%BF and BMI were significantly correlated in both the males and the females (males, r = 0.74, P < 0.01; females, r = 0.72, P < 0.01). However, high correlational validity does not guarantee clinical validity of classification systems. Therefore, the nature and extent of misclassifications were evaluated by ROC analysis.

The IOTF/Cole and WHO/MDD classification systems were highly specific for both sexes, but their sensitivity was very low for the females (Table 3). Thus, almost all adolescents labeled as overweight were truly overweight, whereas ≈75% of the truly overweight females were mislabeled as normal-weight. The result was similar for the widely used adult BMI cutoff of 25, which will be applied when the adolescents become 18 y of age.

Cutoffs were derived through ROC analysis to maximize the sum of sensitivity and specificity (Table 3). For the males, this cutoff was equivalent to the 85th percentile on the Swedish BMI percentile charts. The 85th percentile for the females traded some specificity for a moderate increase in sensitivity, which, however, was still very low. For the females, the optimal cutoff derived by ROC analysis improved the sensitivity relative to that of the international references but required large tradeoffs in specificity.

A comparison of the performance of the optimal cutoffs between the sexes is further illustrated by the resulting positive and negative likelihood ratios (Table 3). With the use of the optimal BMI cutoff derived from ROC analysis, a truly overweight male

### TABLE 2
BMI (in kg/m²) and percentage body fat (%BF) cutoffs for 17-y-old males (n = 200) and females (n = 274) from 4 references

<table>
<thead>
<tr>
<th>Definition of overweight</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOTF/Cole¹</td>
<td>BMI of ≥25</td>
<td>BMI, 24.5 (14.9)</td>
</tr>
<tr>
<td>WHO/MDD⁴</td>
<td>BMI ≥ 85th percentile</td>
<td>BMI, 25.3 (10.4)</td>
</tr>
<tr>
<td>He et al⁵</td>
<td>BMI ≥ 85th percentile</td>
<td>BMI, 23.9 (18.3)</td>
</tr>
<tr>
<td>Williams et al⁶</td>
<td>%BF cutoffs determined by biological endpoint</td>
<td>%BF, 25% (12.7)</td>
</tr>
</tbody>
</table>

¹ IOTF, International Obesity Task Force; WHO, World Health Organization; MDD, Must, Dallal, and Dietz.

⁴ Resulting prevalence estimates (percentages) within parentheses.

³ International age- and sex-specific BMI percentile curves that at age 18 y pass through the BMI cutoffs for adult overweight and obesity of 25 and 30, respectively (7). The reference is recommended by the IOTF.

⁵ Age- and sex-specific BMI percentile curves derived by MDD from data collected in 1971–1974 as part of the first National Health and Nutrition Examination Survey (NHANES I) (8). Overweight and obesity (or at risk of overweight and overweight) are defined as BMI values above the 85th and 95th percentiles, respectively. The reference has been recommended by the WHO for international use.

⁶ Sex-specific %BF cutoffs derived from findings of a significant overrepresentation of selected metabolic risk factors and high blood pressure in a sample of 3230 American children and adolescents aged 5–18 y (15).
TABLE 3
Sensitivity and specificity for excess fatness for BMI-based references

<table>
<thead>
<tr>
<th></th>
<th>Males (n = 200)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Positive L pos</td>
<td>Negative L neg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IOTF/Cole</td>
<td>0.84 (±0.02)</td>
<td>0.95 (±0.09)</td>
<td>16 (±2.0)</td>
<td>0.17 (±0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHO/MDD</td>
<td>0.72 (±0.04)</td>
<td>0.98 (±0.06)</td>
<td>42 (±2.1)</td>
<td>0.28 (±0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et al 15</td>
<td>0.92 (±0.06)</td>
<td>0.92 (±0.04)</td>
<td>12 (±2.0)</td>
<td>0.09 (±0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult 7</td>
<td>0.76 (±0.08)</td>
<td>0.97 (±0.06)</td>
<td>27 (±2.0)</td>
<td>0.25 (±0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal (ROC)</td>
<td>0.92 (±0.06)</td>
<td>0.92 (±0.04)</td>
<td>12 (±2.0)</td>
<td>0.09 (±0.03)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Females (n = 274)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Positive L pos</td>
<td>Negative L neg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IOTF/Cole</td>
<td>0.25 (±0.01)</td>
<td>1.00 (±0.05)</td>
<td>Inf (±Inf)</td>
<td>0.75 (±0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHO/MDD</td>
<td>0.22 (±0.02)</td>
<td>1.00 (±0.08)</td>
<td>Inf (±Inf)</td>
<td>0.78 (±0.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et al 15</td>
<td>0.39 (±0.06)</td>
<td>0.97 (±0.06)</td>
<td>15 (±2.0)</td>
<td>0.63 (±0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult 7</td>
<td>0.02 (±0.01)</td>
<td>1.00 (±0.05)</td>
<td>Inf (±Inf)</td>
<td>0.98 (±0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal (ROC)</td>
<td>0.77 (±0.06)</td>
<td>0.76 (±0.04)</td>
<td>3 (±2.0)</td>
<td>0.30 (±0.03)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mean (±SE) area under the curve, which equals the probability of correct classification, was 0.97 (±0.02 (95% CI: 0.94, 1.00) in the males and 0.85 (±0.02 (95% CI: 0.80, 0.89) in the females. IOTF, International Obesity Task Force; WHO, World Health Organization; MDD, Must, Dallal, and Dietz; ROC, receiver operating characteristic; Inf, infinity.

1 True-positive rate. True positives were defined by using the percentage body fat cutoffs proposed by Williams et al (15).

2 Positive likelihood ratio [sensitivity/(1 − specificity)].

3 Negative likelihood ratio [(1 − sensitivity)/specificity].

4 Age- and sex-specific BMI percentile curves derived by MDD from data collected in 1971–1974 as part of the first National Health and Nutrition Examination Survey (NHANES I) (8). Overweight and obesity (or at risk of overweight and overweight) are defined as BMI values above the 85th and 95th percentiles, respectively (7). The reference is recommended by the IOTF.

5 International age- and sex-specific BMI percentile curves that at age 18 y pass through the BMI cutoffs for adult overweight and obesity of 25 and 30, respectively (7). The reference is recommended by the WHO for international use.

6 National age- and sex-specific percentile curves for BMI from a longitudinal study of 3650 full-term infants born in Sweden in the 1970s (21). Recommendations on cutoffs for categorization have not been published.

7 The widely used BMI cutoff of 25 for overweight in adult men and women.

8 The cutoff maximizing the sum of sensitivity and specificity, which in this study was when equal values for sensitivity and specificity were obtained. Derived by ROC analysis.

9 The mean (±SE) area under the curve (area under the curve) between true positives and false positives.

10 The 45° line represents chance as a diagnostic test (area under the curve = 0.5). Excess fatness was defined according to the percentage body fat cutoffs proposed by Williams et al (15).

Because there are no generally accepted reference values to define overweight or obesity by %BF in adolescents, the results in the presented analysis will be determined by the choice of %BF cutoff. Therefore, the influence of different %BF cutoffs on the definition of true positives was also analyzed (Table 4). For the males, %BF cutoffs from 17.5% to 30% were examined. To produce equal sensitivity and specificity with the use of the IOTF/Cole system, true overweight in 17-y-old males would have to be considered at a %BF > 30%, ie, 5 percentage points higher than the cutoff proposed by Williams et al (15). For the
Recommendations on cutoffs for categorization have not been published. Percentiles, respectively. The reference has been recommended by the WHO for international use. Examination Survey (NHANES I) (8). Overweight and obesity (or at risk of overweight and overweight) are defined as BMI values above the 85th and 95th respectively (7). The reference is recommended by the IOTF.

For identification of Swedish adolescents with excess fatness, the corresponding %BF cutoff would need to be near 40%, ie, 10 percentage points above the proposed cutoff.

**DISCUSSION**

Few studies have evaluated proposed BMI-based classification systems for adolescent obesity for their respective diagnostic accuracy in detecting fatness (12, 13). In several studies, correlational analyses between different measures of fatness were conducted, but such studies can show only the closeness of association, not the extent and type of misclassifications. Therefore, ROC analysis was used to evaluate the clinical validity of BMI as a diagnostic tool for detecting excess fatness. We evaluated the sensitivity, specificity, and positive and negative likelihood ratios of the BMI-based classification systems recommended for international use by the IOTF and the WHO (7, 20).

The results were compared with a national BMI reference, and BMI cutoffs maximizing the sum of sensitivity and specificity were also derived from the sample.

For identification of Swedish adolescents with excess fatness, the IOTF/Cole and WHO/MDD classification systems were shown to have very high specificity in both sexes, but the sensitivity was very low in the females. All the females classified as overweight were truly overweight, but ≈75% of the truly overweight females were misclassified as having normal weight. Thus, many overweight females would be missed in intervention programs using BMI as the selection criterion. Through ROC analysis, cutoffs were derived for the males to further improve the tradeoff between sensitivity and specificity. This cutoff was identical to the 85th BMI percentile for 17-y-old Swedish males from a national reference (21). For the females, BMI proved to be less valid in classifying persons with excess fat as overweight. The cutoffs for optimizing the tradeoff in females were much lower than the ones recommended by the IOTF and the WHO, but the positive likelihood ratio was low, and the negative likelihood ratio was fairly high.

The choice of %BF cutoff to define true positives or true overweight determines the results in this kind of analysis and could be used as an argument to reject the results. However, to produce equal sensitivity and specificity when using the IOTF/Cole system to classify overweight, the %BF cutoffs would need to be 30% for males and nearly 40% for females. For males and females, respectively, the cutoffs suggested in the literature are 25% and 30% (criterion-based for 5–18 y of age) (15), 20% and 30% (criterion-based for 9–15 y of age) (25), and 21% and 34% (%BF at 17 y of age corresponding to a BMI of 25 at 18 y of age) (26). From these suggested cutoffs and the results from the various %BF cutoffs provided in the present article, it seems fairly safe to conclude that the IOTF/Cole and WHO/MDD classification systems are highly specific in both sexes but are insensitive for overweight in 17-y-old Swedish females.

Evidence of screening ability and relation to morbidity is available for some national BMI-based classification systems (27, 28) but is scarce for the proposed international systems (29). In a comparison of screening ability for obesity in a British population between 1990 reference data for the United Kingdom and the IOTF/Cole system, Reilly et al (30) found that the IOTF/Cole system had low sensitivity in girls and very low sensitivity in boys. This result is in fairly good agreement with our results for
and specificity for fatness of BMI compared with those of measurements to estimate %BF when comparing the sensitivity both dual-energy X-ray absorptiometry and skinfold-thickness potential variation in results between studies. Mei et al (32) used study. In the present study, %BF cutoffs defined by a biological endpoint were used instead (15). Proposed %BF cutoffs derived from a biological endpoint approach can be criticized on the grounds of study population, sample size, and chosen endpoints. The cutoffs used in the present study can specifically be criticized for not being age specific (26). Thus, a systematic underestimation of the proportion of excess adiposity in younger subjects and an overestimation in older subjects are likely to result, especially in females (26). These effects may have contributed to the large discrepancy in prevalence estimates between the BMI-based and %BF-based classification systems in the females. There are no reported prevalence estimates of overweight for 17-y-old Swedes based on %BF or morbidity. In 1998 the prevalence of overweight (BMI > 25) in 16–84-y-old women was ≈38%, but the prevalence in 16–24-y-old women was only 12% (31).

Furthermore, the %BF cutoffs used were derived from a biracial American sample. However, reference values for healthy %BF ranges have not been published for Swedish, Scandinavian, or European populations. Using cutoffs anchored to metabolic risk appears to be the best available alternative and has been used for other European samples in similar evaluations (14). In addition, possible population differences are likely to be within the %BF cutoff ranges included in the sensitivity analysis in the present article, which supports the conclusions of the present study. The choice of method for estimating %BF is also a source of potential variation in results between studies. Mei et al (32) used both dual-energy X-ray absorptiometry and skinfold-thickness measurements to estimate %BF when comparing the sensitivity and specificity for fatness of BMI compared with those of weight/height. Sardinha et al (14) used dual-energy X-ray absorptiometry when evaluating BMI, triceps skinfold thickness, and upper arm girth; Reilly et al (13) and Fu et al (12) used bioelectrical impedance when evaluating the IOTF/Cole classification system; and in the present study, densitometry by air-displacement plethysmography was used. Air-displacement plethysmography has been proven to produce %BF estimates of comparable accuracy to those produced by dual-energy X-ray absorptiometry and hydrostatic weighing (33).

With the assumption that air-displacement plethysmography produces accurate and valid %BF measurements and that true positives for overweight are defined by the %BF cutoffs proposed by Williams et al (15), the results from the present study clearly show the tradeoffs between sensitivity and specificity when applying different classification systems for overweight. Which system to recommend for national use is not obvious, because such a recommendation is dependent on the purpose of the system. The optimal cutoffs derived in the present study maximize the sum of sensitivity and specificity, which may be considered optimal for selective public health interventions. For clinical practice, minimizing the number of false positives is often preferred to avoid the stigma associated with being mislabeled as obese in adolescence. However, many true positives will be missed as a consequence, unless ancillary measures are used in conjunction.

This study examined only the diagnostic accuracy of the classification systems for detecting fatness. Future studies need to evaluate the diagnostic accuracy for directly detecting cardiovascular risk factors. Such studies have been conducted to some extent in adults and prepubertal children (34, 35).

In conclusion, the tradeoff between sensitivity and specificity should be analyzed in detail before making general recommendations about classification systems for overweight. The diagnostic demands on a classification system intended for use in clinical practice are different from those on systems intended for public health use or monitoring. Therefore, recommendations should be explicit regarding the setting in which suggested systems should be used. A multipurpose system may be the easiest to implement but would not suit the varying demands of public health, clinical practice, and monitoring. An international reference is a compromise to obtain acceptable, comparable prevalence estimates at the global level. At the national level, given the probable population differences in relative risks at certain BMI values, the seriousness of the adolescent obesity problem, and its character as a major cost driver through obesity-related illnesses, customized systems derived from national data are likely to be more efficient. Such systems should therefore be developed.

We especially thank Catharina Grimming, Eva Hedlund, Maria Saxer, and Karin Vagstrand for providing help and support to the study. We also thank James Stubbs (Rowett Institute) and Paul Higgins (University of Alabama at Birmingham) for valuable comments and discussions and the unit for Preventive Nutrition, Karolinska Institutet, for providing BodPod equipment support. YML and BSB were the lead epidemiologists on the project and were primarily responsible for developing the study design for the Stockholm Pregnancy and Women’s Nutrition Study (1999) and the follow-up Stockholm Weight Development Study (2002). They also supervised the data collection and helped in editing the manuscript. MGN provided critical input for the conception of this particular article, was responsible for conducting the analyses, performed the statistical analyses, and drafted the manuscript. SOR was the principal investigator; conceived the idea of the 3 studies in 1984, 1999, and 2002, respectively; assisted with the study design; and provided help with manuscript revision. None of the authors had any conflicts of interest.

REFERENCES

11. Reilly JJ. Assessment of childhood obesity: national reference data or
12. Fu WP, Lee HC, Ng CJ, et al. Screening for childhood obesity: interna-
18. McCrory MA, Gomez TD, Bernauer EM, Mole PA. Evaluation of a new
7. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard
definition for child overweight and obesity worldwide: international
8. Must A, Dallal GE, Dietz WH. Reference data for obesity: 85th and 95th
percentiles of body mass index (wt/ht2) and triceps skinfold thickness.
Nutr 1999;70(suppl):126S–30S.
7–9-y-old French children: frequency of obesity, overweight and thin-
11. Reilly JJ. Assessment of childhood obesity: national reference data or
12. Fu WP, Lee HC, Ng CJ, et al. Screening for childhood obesity: interna-
tional vs population-specific definitions. Which is more appropriate? Int
13. Reilly JJ, Dorosty AR, Emmett PM. Identification of the obese child:
adequacy of the body mass index for clinical practice and epidemiology.
characteristic analysis of body mass index, triceps skinfold thickness,
and arm girth for obesity screening in children and adolescents. Am J
elevated blood pressure, total cholesterol, and serum lipoprotein ratios in
17. Dempster P, Aitkens S. A new air displacement method for the deter-
18. McCrory MA, Gomez TD, Bernauer EM, Mole PA. Evaluation of a new
air displacement plethysmograph for measuring human body composi-
19. Siri WE. Body composition from fluid spaces and density: analysis of
20. WHO. Physical status: the use and interpretation of anthropometry.
index reference values from Goteborg, Sweden: birth to 18 years of age.
22. Lohman TG. Applicability of body composition techniques and con-
23. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots:
a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:
561–77.
24. Hanley JA. The robustness of the “binormal” assumptions used in fitting
25. Dwyer T, Blizzard CL. Defining obesity in children by biological end-
point rather than population distribution. Int J Obes Relat Metab Disord
26. Taylor RW, Jones IE, Williams SM, Goulding A. Body fat percentages
measured by dual-energy X-ray absorptiometry corresponding to re-
cently recommended body mass index cutoffs for overweight and obe-
sity in children and adolescents aged 3–18 y. Am J Clin Nutr 2002;76:
1416–21.
27. Dietz WH. Health consequences of obesity in youth: childhood predic-
28. Morrison JA, Barton BA, Biro FM, Daniels SR, Sprecher DL. Over-
weight, fat patterning, and cardiovascular disease risk factors in black
prevention, and treatment; evidence based answers to common ques-
30. Reilly JJ, Savage SA, Ruxton CH, Kirk TR. Assessment of obesity in a
community sample of prepubertal children. Int J Obes Relat Metab Disord
31. The Swedish Council on Technology Assessment in Health Care. Obe-
sity—problems and solutions: a systematic literature review. 1st ed.
Dietz WH. Validity of body mass index compared with other body-
composition screening indexes for the assessment of body fatness in
33. Fields DA, Goran MI, McCrory MA. Body-composition assessment via
air-displacement plethysmography in adults and children: a review.
factors by indices of obesity obtained from anthropometry and dual-
energy X-ray absorptiometry in Japanese individuals. Int J Obes Relat
35. Higgins PB, Gower BA, Hunter GR, Goran MI. Defining health-related