OBJECTIVE: Although the need for accurate anthropometric measurement has been repeatedly stressed, reports on growth and physical measurements in human populations rarely include estimates of measurement error. We describe the standardization process and reliability of anthropometric measurements carried out in a pilot study.

METHODS: For the intraobserver assessment of anthropometric measurements, we studied 101 adolescents (58 boys and 43 girls) from five cities. For interobserver assessment, we studied 10 adolescents from the same class in Zaragoza and different from those in the intraobserver sample.

RESULTS: For skinfold thickness, intraobserver technical errors of measurement (TEMs) in general were smaller than 1 mm; for circumferences, TEMs in general were smaller than 1 cm. Intraobserver reliability for skinfold thickness was greater than 95% for almost all cases; for circumferences, intraobserver reliability generally was greater than 95%. Interobserver TEMs ranged from 1 to 2 mm for the six skinfold thicknesses measured; for circumferences, TEMs were smaller than 1 cm for the arm, biceps, and waist and between 1 and 2 cm for the hip and thigh. Interobserver reliabilities for skinfold thickness and circumference were always greater than 90%; except for biceps skinfold.

CONCLUSIONS: Our results are in agreement with those recommended in the literature. Therefore, these anthropometric measures seem to be adequate to assess body composition in a multicenter survey in adolescents. Nutrition 2003;19:481–486. ©Elsevier Inc. 2003

KEY WORDS: skinfold thickness, circumference, reliability, technical error of measurement, obesity

INTRODUCTION

Because of its importance to health, body composition is commonly investigated in epidemiologic, clinical, and population studies. Reliable methods for measurement of body fat and fat distribution therefore are important. During the past decade, investigators have emphasized the accuracy of newer techniques, such as dual-energy x-ray absorptiometry, magnetic resonance imaging, and computed tomography, for measuring body composition; nevertheless, anthropometry is the most widely used method, and it recently has been used to estimate fat distribution.1,2 The distinct advantages of anthropometry are that it is portable, non-invasive, inexpensive, and useful in field studies, and there is a substantial literature available on the subject.3

Although the need for accurate anthropometric measurement has been repeatedly stressed, reports on growth and physical measurements in human populations rarely include estimates of measurement error. Reliability is the degree to which within-subject variability is due to factors other than measurement error variance or physiologic variation. The lower the variability between repeated measurements of the same subject by one (intraobserver differences) or two or more (interobserver differences) observers, the greater is the precision. The most commonly used
In September 2000, we conducted a 2-d theoretical and practical workshop in Madrid with five researchers who planned to perform the anthropometric measurements. All five anthropometrists had experience in the anthropometric assessment of nutrition status. The aim of the workshop was to standardize the methodology and use it as a reference, as determined by an experienced anthropometrist (L.A.M.).

As with any quantitative biological measure, in anthropometric assessment it is important to minimize error. Poor precision in measurement of an anthropometric variable will lead to underestimation of correlations with other variables. The main sources of error of imprecision are random imperfections in the measuring instruments or in the measuring and recording techniques.

Adolescence is a decisive period during human life because of multiple physiologic and psychological changes that take place. We developed a research project to evaluate the nutrition status of Spanish adolescents from five cities, Granada, Madrid, Murcia, Santander, and Zaragoza, which is called the Alimentación y Valoración del Estado Nutricional en Adolescentes (AVENA Study). Before carrying out the field work, we conducted a pilot study. For anthropometric assessment, in the pilot study we standardized the methods of measurement and obtained the intra- and interobserver errors of measurement. We describe the standardization process and the reliability of the anthropometric measurements carried out in the pilot study.

### MATERIALS AND METHODS

#### Population and Design

In September 2000, we conducted a 2-d theoretical and practical workshop in Madrid with the five researchers who planned to perform the anthropometric measurements. The use of two error estimates, TEM and R, can provide most of the information needed to determine whether a series of anthropometric measurements can be considered accurate. As with any quantitative biological measure, in anthropometric assessment it is important to minimize error. Poor precision in measurement of an anthropometric variable will lead to underestimation of correlations with other variables. The main sources of error of imprecision are random imperfections in the measuring instruments or in the measuring and recording techniques.

Material and methods

To perform the anthropometric measurements, the subject was in a standing position. For measuring the relaxed arm circumference, the subject stood relaxed with the arm hanging freely at the side; the tape was passed around the arm at the level of the midpoint of the upper arm. For measuring flexed upper arm circumference (biceps circumference), the subject contracted the biceps as much as possible, and the tape was passed around the arm so that it touched the skin surrounding the maximum circumference. To measure the waist circumference, the tape was applied to the following sites: triceps, halfway between the acromion process and the olecranon process; biceps, at the same level as the triceps skinfold and directly above the center of the cubital fossa; subscapular, about 20 mm below the tip of the scapula and 45 degrees to the lateral side of the body; suprailiac, about 20 mm above the iliac crest and 20 mm toward the medial line; thigh, in the midline of the anterior aspect of the thigh, midway between the inguinal crease and the proximal border of the patella; and calf, at the level of maximum calf circumference, on the medial aspect of the calf. The five circumferences were measured in centimeters with non-elastic tape to the nearest millimeter. In general, for these measurements, the subject was in a standing position. For measuring the relaxed arm circumference, the subject stood relaxed with his or her side to the observer, and the arm hanging freely at the side; the tape was passed around the arm at the level of the midpoint of the upper arm. For measuring flexed upper arm circumference (biceps circumference), the subject contracted the biceps as much as possible, and the tape was passed around the arm so that it touched the skin surrounding the maximum circumference. To measure the waist circumference, the tape was applied to the following sites: triceps, halfway between the acromion process and the olecranon process; biceps, at the same level as the triceps skinfold and directly above the center of the cubital fossa; subscapular, about 20 mm below the tip of the scapula and 45 degrees to the lateral side of the body; suprailiac, about 20 mm above the iliac crest and 20 mm toward the medial line; thigh, in the midline of the anterior aspect of the thigh, midway between the inguinal crease and the proximal border of the patella; and calf, at the level of maximum calf circumference, on the medial aspect of the calf.

The five circumferences were measured in centimeters with non-elastic tape to the nearest millimeter. In general, for these measurements, the subject was in a standing position. For measuring the relaxed arm circumference, the subject stood relaxed with his or her side to the observer, and the arm hanging freely at the side; the tape was passed around the arm at the level of the midpoint of the upper arm. For measuring flexed upper arm circumference (biceps circumference), the subject contracted the biceps as much as possible, and the tape was passed around the arm so that it touched the skin surrounding the maximum circumference. To measure the waist circumference, the tape was applied...
horizontally midway between the lowest rib margin and the iliac crest, near the level of the umbilicus, at the end of gentle expiration. The hip circumference measurement was taken at the point yielding the maximum circumference over the buttocks, with the tape held in a horizontal plane. Proximal thigh circumference was measured just below the gluteal fold and perpendicular to its long axis; the subject stood erect with the feet slightly apart and the body mass evenly distributed between both legs.  

**Statistical Analysis**

The TEM is the most commonly used measure of precision, which is the square root of measurement error variance. The TEM was obtained by performing a number of repeated measurements on the same subject by the same observer (three measures by that observer) or two or more observers (one measure by five observers).

The units of TEM were the same as those of the anthropometric measurement (millimeters for skinfold thicknesses and centimeters for circumferences). TEM was calculated with the following formula, where \( n \) is the number of subjects, \( K \) is the number of determinations of the variable taken from each subject (intraobserver analysis) or the number of subjects assuming one determination per observer (interobserver assessment), and \( M \) is the measurement:

\[
\text{TEM} = \sqrt{\frac{1}{n} \left[ \sum (\sum M^2) - \left( \frac{\sum M^2}{K} \right) \right]} \]

\( R \) as a percentage \((\% R)\), was calculated with the following equation:

\[
\% R = 1 - \frac{\left( \text{total TEM} \right)^2}{\text{SD}^2}
\]

where SD\(^2\) is the total intersubject variance for the study, including measurement error. This coefficient show the proportion of the between-subject variance in a measured population that is free from measurement error.

To assess whether the variation was higher for the highest measurements than for the lowest ones, we also calculated correlations between mean values of each measurement and their corresponding standard deviations for the intra- and interobserver results.

**RESULTS**

Table II shows the intraobserver TEM and \(\% R\) for each anthropometric measurement in the five cities. For skinfold thickness, TEMs in general were smaller than 1 mm, except for the suprailiac skinfold in Madrid and the thigh skinfold in Madrid and Santander. For circumferences, TEMs were smaller than 1 cm, except for waist circumference in Murcia and hip circumference in Granada, Madrid, Murcia, and Santander. Reliability for skinfold thickness was greater than 95% for all cases, except for biceps skinfold in Granada and Murcia. Reliability for circumferences was always greater than 95%, except for hip circumference in Murcia.

Table III shows the interobserver TEM and \(\% R\) for each anthropometric measurement. TEMs ranged from 1 to 2 mm for the six skinfold thicknesses measured. For circumferences, TEMs were smaller than 1 cm for the arm, biceps, and waist and between 1 and 2 cm for the hip and thigh. Reliabilities for skinfold thick-
ness and circumference require more training and produce different degrees of error. Skinfold thickness is accepted as body fatness predictor for two reasons: approximately 40% to 60% of total body fat is in the subcutaneous region of the body, and skinfold thickness can be directly measured with well-calibrated caliper. Some circumference measurements also have been used in equations for predicting body fatness. Circumferences measured at the mid-arm, mid-thigh, waist, and hip are used more frequently because they indicate differences across people in major regions of the body. Many recent studies have used circumferences for estimating skeletal muscle mass and fat distribution.13,14

Equations that predict body composition values provide a way of obtaining such data from variables that can be measured easily and accurately in large-scale epidemiologic and population studies, where sophisticated laboratory settings are impractical. The predicted values of body composition are less precise than those from measured laboratory procedures, but they are less expensive and are practical and easy to apply. This greater accessibility comes with larger errors.15

Reliability is the degree to which within-subject variability is due to factors other than measurement error variance or physiologic variation. The lower the variability between repeated measurements of the same subject by one (intraobserver differences) or two or more (interobserver differences) observers, the greater is the precision.9 The most commonly used measures of precision are the TEM and R. R indicates the proportion of between-subject variance in a measured population that is free from measurement error. Measures of R can be used to compare the relative reliability of different anthropometric measurements and of the same measurements in different age groups and to estimate sample size requirements in anthropometric surveys. A generous allowance for measurement error might be up to 10% of the observed variance; this is equivalent to an R value of 90% or greater. Although this might be an acceptable lower limit, even at R values of approximately 95%, there is the occasional gross measurement error that is likely to have important consequences. Only when R is in the region of 99% is such an error unlikely.4 Acceptable levels of measurement error are difficult to ascertain because TEM is related to the anthropometric characteristics of the group or population under investigation. However, R greater than 95% should be sought when possible.

The intra- and interobserver TEMs for skinfold thickness in our survey were lower than the reference values proposed by Ulijaszek

### TABLE IV.

<table>
<thead>
<tr>
<th>Skinfold thickness</th>
<th>Granada</th>
<th>Madrid</th>
<th>Murcia</th>
<th>Santander</th>
<th>Zaragoza</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biceps</td>
<td>0.349</td>
<td>0.054</td>
<td>0.650</td>
<td>0.001</td>
<td>0.210</td>
</tr>
<tr>
<td>Triceps</td>
<td>0.675</td>
<td>0.000</td>
<td>0.466</td>
<td>0.033</td>
<td>0.055</td>
</tr>
<tr>
<td>Subscapular</td>
<td>0.330</td>
<td>0.070</td>
<td>0.593</td>
<td>0.005</td>
<td>0.660</td>
</tr>
<tr>
<td>Suprailiac</td>
<td>0.512</td>
<td>0.003</td>
<td>0.377</td>
<td>0.092</td>
<td>0.032</td>
</tr>
<tr>
<td>Thigh</td>
<td>0.577</td>
<td>0.001</td>
<td>0.155</td>
<td>0.506</td>
<td>0.217</td>
</tr>
<tr>
<td>Calf</td>
<td>0.293</td>
<td>0.110</td>
<td>0.507</td>
<td>0.019</td>
<td>0.164</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Circumference</th>
<th>Arm</th>
<th>Madrid</th>
<th>Murcia</th>
<th>Santander</th>
<th>Zaragoza</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biceps</td>
<td>0.114</td>
<td>0.545</td>
<td>0.359</td>
<td>0.109</td>
<td>0.055</td>
</tr>
<tr>
<td>Triceps</td>
<td>0.179</td>
<td>0.339</td>
<td>0.179</td>
<td>0.441</td>
<td>0.662</td>
</tr>
<tr>
<td>Waist</td>
<td>0.241</td>
<td>0.191</td>
<td>0.219</td>
<td>0.341</td>
<td>0.315</td>
</tr>
<tr>
<td>Hip</td>
<td>0.052</td>
<td>0.899</td>
<td>0.109</td>
<td>0.638</td>
<td>0.563</td>
</tr>
<tr>
<td>Thigh</td>
<td>0.468</td>
<td>0.008</td>
<td>0.100</td>
<td>0.662</td>
<td>0.255</td>
</tr>
</tbody>
</table>

* Boldface indicates P < 0.05
Reliability of Anthropometric Measures in Adolescents

and Lourie. In almost all cases, intraobserver reliability for skinfold thickness also was greater than 95%: these results are very similar to or even better than those observed by other investigators (see review by Ulijaszek and Kerr). Interobserver reliability for skinfold thickness was greater than 90%, except for biceps skinfold; interobserver reliabilities observed by other researchers have ranged from 49% to 98% for the biceps, 48% to 99% for the triceps, 60% to 99% for the subscapular, 56% to 97% for the suprailiac, and 81% to 99% for the calf skinfold thicknesses.

Interobserver error is a major issue in measuring skinfold thickness. Standardized methodology, including positioning of the instrument and the subject, a well-trained data collector, and practicing until results are consistent, can increase reproducibility. Special attention to locating the site, grasping the skin, and ensuring that the caliper is at a 90-degree angle relative to the grasped skinfold are essential for high reproducibility.

Another form of unreliability is independability, which is due to variation in some biological characteristic of the individual being measured, which results in variation in the measurement; even if the technique used is exactly replicated each time. Size of skinfold measurement in any individual can differ according to duration and level of compression during measurement, which can differ according to the level of tissue hydration. There may be two components to skinfold compressibility: dynamic and static. Dynamic compressibility likely is due to the expulsion of water from subcutaneous tissue, and static compressibility is a function of the tension and thickness of the skin and subcutaneous tissue and of the distribution of fibrous tissue and blood vessels. Skinfold thicknesses are affected by individual and regional differences in compressibility that change with age, sex, and recent weight loss. When a skinfold thickness is measured, the pressure exerted by the calipers displaces some extracellular fluid. This displacement and the corresponding compressibility are marked in preterm infants soon after birth and in malnutrition, when the extracellular fluid content of subcutaneous adipose tissue is increased. In addition, pressure from skinfold calipers may force some adipose tissue lobules to slide into areas of lesser pressure; this sliding may be more marked for thick skinfold thicknesses in which the adipose tissue contains little connective tissue. The conformist view is that intersite and intersubject differences in skinfold compressibility reduce the utility of skinfold thickness. However, if variations in compressibility reflect differences in the fluid content of uncompressed skinfold thicknesses, the reduction of these differences by compression might increase the validity of skinfold thicknesses as measures of regional fatness.

The intra- and interobserver TEMs for circumferences in our survey were similar to the reference values proposed by Ulijaszek and Lourie. In almost all cases, intraobserver reliability for circumferences was greater than 95%; these results are very similar to or even better than those observed by other investigators (see review by Ulijaszek and Kerr). Interobserver reliability for circumferences was greater than 90%; interobserver reliabilities observed by other researchers ranged from 94% to 100% for the arm, 86% to 99% for the waist, and 68% to 99% for the hip circumferences. For hip circumference, the intraobserver R was lower in Murcia, where only girls were included, than in the other cities, and the highest interobserver TEM was also the highest R. Circumferences are more reliable than skinfold thicknesses, and they can always be measured regardless of body size and fatness. Reproducibility of circumferences can be increased by paying special attention to positioning the subject, using anatomic landmarks to locate measuring sites, taking readings in millimeters with the tape measure directly in contact with the subject’s skin without compression, and keeping the tape at 90 degrees to the long axis of the region of the body under the measured circumference.

Nordhamn et al. observed in adults that, because of their greater reliability, sagittal abdominal diameter and the waist have a higher predictive capacity for cardiovascular risk than does the waist-to-hip ratio. Several anthropometric measurements (waist circumference and subcapular and suprailiac skinfold thicknesses) had less reliability in overweight than in lean subjects. We also observed that the variability of measurements is greater when the measures taken are also greater. In extremely fat individuals, skinfold thicknesses cannot be measured accurately; in these cases, generally corresponding to a sum of skinfold thicknesses exceeding 120 to 140 mm, skinfold thicknesses cannot be used to estimate body fat percentages.

Skinfold thicknesses include skin and subcutaneous adipose tissue, with the latter consisting of adipocytes that contain triacylglycerols and connective tissue that contain blood vessels and nerves. The thickness of a double layer of skin is about 1.8 mm, but this varies among individuals and systematically by site and with age. The paucity of subcutaneous adipose tissue in the lean can make it difficult to elevate a skinfold, and it is not easy to elevate skinfold thicknesses with parallel sides in those with large amounts of subcutaneous adipose tissue. Consequently, skinfold thicknesses are less precise than circumferences in overweight individuals than in general populations, but skinfold thicknesses are less affected by edema than circumferences because caliper pressure reduces the fluid content of the subcutaneous adipose tissue.

Overweight and obesity in children and adolescents is a major public health concern. It would be important to define the adequate tools for the assessment of this condition. The International Obesity Task Force recently proposed using a new reference standard for body mass index. We agree with this proposal in terms of screening and as a public health indicator. However, if we want to precisely measure the increase in body fat tissue and detect the metabolic complications of obesity, we must use another criteria, such as the percentage of total body fat and the waist circumference, respectively.

Anthropometric measurement error is unavoidable and should be minimized by paying close attention to every aspect of the data-collection process. Regardless of the measurement made and the size of the error, it is better to know the error size because this will determine the confidence one has in the different measurements made and will influence the interpretation of anthropometric data collected. We also recommend that replicate measurements of anthropometric variables be made. In the pilot survey described in this paper, we minimized the intra- and interobserver errors to acceptable ranges. The quality of the anthropometric measurement also will be monitored during the multicenter survey.

ACKNOWLEDGMENTS

The authors thank all the adolescents who participated in this pilot study, especially those from the Colegio Escuelas Pias (Zaragoza, Spain) who participated in the intra- and interobserver assessments.

REFERENCES


19. Lee MMC, Ng CK. Postmortem studies of skinfold caliper measurement and actual thickness of skin and subcutaneous tissue. Hum Biol 1967;37:91


