Review article

Anthropometric measurement error and the assessment of nutritional status

Stanley J. Ulijaszek[1]* and Deborah A. Kerr[2]

[1]*Institute of Biological Anthropology, University of Oxford, 58 Banbury Road, Oxford OX2 6QS, UK*
[2]*Department of Nutrition, Dietetics and Food Science, Curtin University School of Public Health, GPO Box U1987, Perth WA6001, Australia*

Anthropometry involves the external measurement of morphological traits of human beings. It has a widespread and important place in nutritional assessment, and while the literature on anthropometric measurement and its interpretation is enormous, the extent to which measurement error can influence both measurement and interpretation of nutritional status is little considered. In this article, different types of anthropometric measurement error are reviewed, ways of estimating measurement error are critically evaluated, guidelines for acceptable error presented, and ways in which measures of error can be used to improve the interpretation of anthropometric nutritional status discussed. Possible errors are of two sorts; those that are associated with: (1) repeated measures giving the same value (unreliability, imprecision, undependability); and (2) measurements departing from true values (inaccuracy, bias). Imprecision is due largely to observer error, and is the most commonly used measure of anthropometric measurement error. This can be estimated by carrying out repeated anthropometric measures on the same subjects and calculating one or more of the following: technical error of measurement (TEM); percentage TEM, coefficient of reliability (R), and intraclass correlation coefficient. The first three of these measures are mathematically interrelated. Targets for training in anthropometry are at present far from perfect, and further work is needed in developing appropriate protocols for nutritional anthropometry training. Acceptable levels of measurement error are difficult to ascertain because TEM is age dependent, and the value is also related to the anthropometric characteristics of the group or population under investigation. R > 0·95 should be sought where possible, and reference values of maximum acceptable TEM at set levels of R using published data from the combined National Health and Nutrition Examination Surveys I and II (Frisancho, 1990) are given. There is a clear hierarchy in the precision of different nutritional anthropometric measures, with weight and height being most precise. Waist and hip circumference show strong between-observer differences, and should, where possible, be carried out by one observer. Skinfolds can be associated with such large measurement error that interpretation is problematic. Ways are described in which measurement error can be used to assess the probability that differences in anthropometric measures across time within individuals are due to factors other than imprecision. Anthropometry is an important tool for nutritional assessment, and the techniques reported here should allow increased precision of measurement, and improved interpretation of anthropometric data.

Anthropometry: Nutritional status: Measurement error: Imprecision

Anthropometry is literally 'the measurement of man', which could encompass any physiological, psychological or anatomical trait. In practice, anthropometry refers specifically to morphological traits which can be externally measured.

Anthropometry has an important place in nutritional assessment (Jelliffe & Jelliffe, 1989; Gibson, 1990), and in addition to use in the clinical setting (Gibson, 1990) is used in nutritional screening, surveillance, and monitoring

(Tomkins, 1994). The literature on methods of anthropometric measurement and interpretation is large (e.g. Weiner & Lourie, 1981; Cameron, 1984, 1986; Heymsfield *et al.* 1984; Lohman *et al.* 1988; Jelliffe & Jelliffe, 1989; Gibson, 1990; Ulijaszek & Mascie-Taylor, 1994; World Health Organization, 1995; Norton & Olds, 1996; Ulijaszek, 1997). However, the extent to which measurement error can influence both measurement and interpretation of nutritional status is usually little considered, beyond the determination of measurement error for training. As with any use of quantitative biological measure, it is important to minimize error, and to know and understand the various ways in which it is estimated and assessed. While anthropometric measurement error has been described by various authors (Heymsfield *et al.* 1984; Mueller & Martorell, 1988) and guidelines for acceptable measurement in training (Zerfas, 1986; Norton & Olds, 1996) and practice (Frisancho, 1990; Ulijaszek & Lourie, 1994) given, there has been little evaluation of different methods of measurement error estimation. In this review article, different types of anthropometric measurement error are described, ways of estimating measurement error are critically evaluated, guidelines for acceptable error are presented, and ways in which measures of error can be used to improve the interpretation of anthropometric nutritional status are discussed.

Limitations of anthropometry

Anthropometry is a relatively quick, simple, and cheap means of nutritional assessment. Its limitations include the extent to which measurement error can influence interpretation, and the length of time needed to take measurements. For large studies, or for nutritional screening and surveillance, a number of anthropometrists may be needed, and this influences the degree of measurement error, especially if there is between-observer bias. In choosing the instrument to assess nutritional status, workers often elect to measure only height and weight. These measures are quick, simple and require only limited training. More comprehensive measurement sets which include skinfolds and circumferences require more training and carry different degrees of error with them.

Types of measurement error

Various terms are used to describe anthropometric measurement error. These include: unreliability (Habicht *et al.* 1979); imprecision, undependability and inaccuracy (Heymsfield *et al.* 1984); precision, accuracy, validity and reliability (Pederson & Gore, 1996); as well as reproducibility and bias (Mueller & Martorell, 1988). Cameron (1986) noted the lack of standardized terminology to describe anthropometric measurement error. Despite the varied terminology, measurement error has predominantly two types of effect on the quality of the data collected (Habicht *et al.* 1979). These effects limit the extent to which: (1) repeated measures give the same value; and (2) measurements depart from 'true' values. In the first category of measurement error are the terms reliability and unreliability, reproducibility, undependability, precision and imprecision, while the second category of measurement error includes the terms bias, validity, accuracy and inaccuracy.

Of the first class of measurement error, reliability is the degree to which within-subject variability is due to factors other than measurement error variance or physiological variation. Unreliability is the within-subject variability due to those two factors alone. Imprecision is the variability of repeated measurements, and is due to intra- and inter-observer measurement differences. The greater the variability between repeated measurements of the same subject by one (intra-observer differences) or two or more (inter-observer differences) observers, the greater the imprecision and the lower the precision (Norton & Olds, 1996). Undependability is due to physiological variation (Mueller & Martorell, 1988). This includes non-nutritional factors that influence the reproducibility of the measurement, such as differences in height of an individual across the day as a consequence of compression of the spinal column. Unreliability is the sum of imprecision and undependability.

Of the second class of measurement error, accuracy is the extent to which the 'true' value of a measurement is attained (Mueller & Martorell, 1988). Inaccuracy is systematic bias, and may be due to instrument error, or to errors of measurement technique. Both of these factors may give systematic bias to all measurements relative to well-calibrated equipment used by an experienced anthropometrist. Validity is the extent to which a measurement actually measures a characteristic (Norton & Olds, 1996) and is conceptually close to the term accuracy, given that 'true' values of measurements are impossible to determine.

Unreliability

Unreliability is composed of imprecision and undependability. Imprecision is the measurement error variance (Mueller & Martorell, 1988) and is a function of biological error. It might be expected that error between two or more observers should be greater than that obtained for within-observer error, since systematic between-observer bias would contribute to between-observer measurement differences. However, most studies which report both intra- and inter-observer error show this not to be the case (Ulijaszek & Lourie, 1994), although in studies involving two or more observers, imprecision is an additive function of all within- and between-observer error values. The extent of imprecision is likely to be increased if anthropometry is carried out by poorly trained individuals. Since anthropometry is often regarded as less complicated to carry out than many other measures of nutritional status, measurement is often delegated to lower-qualified staff. This is acceptable provided that the potential anthropometrists receive adequate training from an expert or criterion anthropometrist to reach a measurable level of expertise before survey, and maintain it across the period of work.

The most commonly used measure of imprecision is the technical error of measurement (TEM) (Mueller & Martorell, 1988), which is the square root of measurement error variance. The TEM is obtained by carrying out a number of repeat measurements on the same subject, either by the same observer, or by two or more observers, taking the differences and entering them into an appropriate equation.

Table 1. Example calculation of technical error of measurement (TEM) from repeat measurements of stature (m) carried out by four observers on ten subjects

	Height (m) as determined by measurer:				(1) ΣM^2	(2) $(\Sigma M)^2/K$	(2)–(1)
	1	2	3	4			
Subject no.							
1	0.865	0.863	0.863	0.864	298.4259	298.4256	0.0003
2	1.023	1.023	1.027	1.025	419.8412	419.8401	0.0011
3	0.982	0.980	0.989	0.985	387.3070	387.3024	0.0046
4	0.817	0.816	0.812	0.817	266.0178	266.0161	0.0017
5	0.901	0.894	0.900	0.903	323.6446	323.6401	0.0045
6	0.880	0.876	0.881	0.881	309.4098	309.4081	0.0017
7	0.948	0.947	0.947	0.946	358.7238	358.7236	0.0002
8	0.906	0.905	0.907	0.908	328.6974	328.6969	0.0005
9	0.924	0.924	0.926	0.924	341.8804	341.8801	0.0003
10	0.989	0.987	1.002	0.993	394.2343	394.2210	0.0133
							$\Sigma = 0.0282$

$$\text{TEM} = \sqrt{0.0282/(N(K-1))} = \sqrt{0.0282/(10(4-1))} = 0.00307$$

M, measurement; K, number of observers (assuming one determination per observer); N, number of subjects.

The calculations for intra- and inter-observer error are broadly the same. For intra-observer TEM for two measurements, and inter-observer TEM involving two measurers, the equation is:

$$\text{TEM} = \sqrt{(\Sigma D^2)/2N}, \qquad (1)$$

where D is the difference between measurements and N is the number of individuals measured. When more than two observers are involved, the equation for estimating inter-observer TEM is more complex:

$$\text{TEM} = \sqrt{((\Sigma_1^N((\Sigma_1^K M^2) - ((\Sigma_1^K M)^2/K)))/N(K-1))}, \quad (2)$$

where N is the number of subjects, K is the number of observers (assuming one determination per observer) for the variable taken on each subject, and M is the measurement. The units of TEM are the same as the units of the anthropometric measurement in question. An example calculation of TEM from measurements of stature (m) made by four observers is given in Table 1. In this example, the functions ΣM^2 and $(\Sigma M)^2/K$ are calculated for each subject measured by the four measurers, then the latter is subtracted from the former. These differences are then summed, and divided by $N(K-1)$. The TEM is then obtained by taking the square root of this value. In this case, TEM is 0.00307 m.

Table 2. Correlations of size of measurement against size of technical error of measurement (TEM), and relative TEM (%TEM) calculated from data on elite athletes (from Ross *et al.* 1994)

	No. of different measurements	r	
(a) Mean value against TEM:			
Lengths	12	0.32	
Circumferences	12	0.92	$P < 0.001$
Skinfolds	8	0.55	
(b) Mean value against %TEM:			
Lengths	12	−0.79	$P < 0.01$
Circumferences	12	−0.23	
Skinfolds	8	−0.59	

The size of the TEM may be positively associated with the size of measurement, where large mean values are associated with high TEM and small mean values with low TEM. Table 2(a) shows correlations of anthropometric measurement means against TEM, for girths, lengths, and skinfolds, respectively, from data collected by Ross *et al.* (1994) on elite athletes. This data set represents the largest single population study of measurement error thus far published. While the mean value of measurement is positively associated with TEM (r 0.92, $P < 0.001$) for circumferences, there is a smaller but non-significant association with lengths and skinfolds.

The positive association between TEM and size of measurement is problematic, since comparative imprecision of different measurements cannot be assessed. In order to compare TEM collected on different variables or different populations, Norton & Olds (1996) have recommended the conversion of the absolute TEM to relative TEM (%TEM) in the following way:

$$\%\text{TEM} = (\text{TEM/mean}) \times 100. \qquad (3)$$

This is a measure of CV, which has the usual form of a standard deviation measure divided by the mean. This is simple to calculate, has no units, and according to the authors, allows direct comparisons of all types of anthropometric measure.

While the imprecision of different anthropometric variables can be easily compared as long as they have the same units of measurement, such comparisons may be misleading if the size of the measurement influences the size of measurement error. The choice between alternative anthropometric measures cannot rest purely on the basis of measurement error, since different measures give different information. However, for a wide variety of body length, breadth and circumference measures, the relationship between the mean values and their CV is a slightly negative one (Roebuck *et al.* 1975), with the potential for introduction of error when comparing TEM of different measures. This negative relationship has been attributed to declining measurement error with increasing dimension of the physical variable being measured (Pheasant, 1988). To

illustrate this effect, consider the %TEM of a measurement if the TEM is 0·001 m. If the dimension being measured is 1·70 m (for example, adult height), then the TEM represents a %TEM of 0·06. If the dimension is 0·80 m (for example, adult sitting height), then this TEM value represents a %TEM of 0·12. If the dimension is 0·30 m (for example, adult arm circumference), then this TEM values represents a %TEM of 0·33. This negative relationship does not apply to skinfolds, where larger values can be more error-bound (Wong *et al.* 1988; Ulijaszek & Lourie, 1994).

Despite making direct comparison of different anthropometric measures possible, %TEM provides no information for comparison of studies in which more than one observer is used, and where both intra- and inter-observer TEM are reported. There are two ways to overcome this problem. The first is to square the TEM, turning them into variances, summing them, then taking the square root, giving the total TEM. In the case of two observers and two measurements per observer, this is:

total TEM
$$= \sqrt{(((\mathrm{TEM(intra_1)}^2 + \mathrm{TEM(intra_2)}^2)/2) + \mathrm{TEM(inter)}^2)}, \tag{4}$$

where TEM(intra$_1$) is the intra-observer TEM for the first observer, TEM(intra$_2$) is the intra-observer TEM for the second observer, and TEM(inter) is the inter-observer TEM between the two of them. In this case, TEM(intra$_1$), TEM(intra$_2$) and TEM(inter) are calculated using equation (1). Where more than two observers are involved, a value for TEM(intra) for each observer, calculated using equation (1), is incorporated in equation (4). All values for TEM(intra) are squared, summed, and divided by the number of observers. Furthermore, with more than two observers, TEM(inter) is calculated using equation (2). For example, in the case of three observers, equation (4) becomes:

total TEM
$$= \sqrt{(((\mathrm{TEM(intra_1)}^2 + \mathrm{TEM(intra_2)}^2)}$$
$$\overline{+\mathrm{TEM(intra_3)}^2)/3) + \mathrm{TEM(inter)}^2)}, \tag{5}$$

where TEM(intra$_1$) is the intra-observer TEM for the first observer, TEM(intra$_2$) is the intra-observer TEM for the second observer, TEM(intra$_3$) is the intra-observer TEM for the third observer and TEM(inter) is the inter-observer TEM between the three of them. In this case, TEM(intra$_1$), TEM(intra$_2$) and TEM(intra$_3$) are calculated using equation (1), and TEM(inter) is calculated using equation (2).

Relative total TEM (% total TEM) can then be obtained using the equation:

$$\% \text{ total TEM} = ((\text{total TEM})/\text{mean}) \times 100. \tag{6}$$

This value could then be used to compare measurement error across studies, regardless of number of observers used. In general, the relative total TEM is smallest in studies that use only one observer. However, a comparison of size of measurement and %TEM using the same data set of Ross *et al.* (1994) (Table 2(b)) shows correlations of anthropometric

measures against %TEM to have no relationship for circumferences or skinfolds, but a negative relationship for lengths ($r - 0.79$, $P < 0.01$). This suggests that %TEM removes the size of measurement/TEM relativity for measures of skinfolds and circumferences, but over-compensates for measurements of length.

The TEM is a standard deviation measure, and can be used to determine the proportion of the total standard deviation for the study population which can be attributed to measurement error. Although maximum acceptable TEM have been recommended as reference values for a variety of measures by Frisancho (1990), these ignore the age dependence of TEM (Lourie & Ulijaszek, 1992) and fail to give values for height and weight.

Another approach to obtain comparability of anthropometric measurement error is to use the coefficient of reliability (R), which ranges from 0 to 1, and can be calculated using the equation:

$$\mathrm{R} = 1 - \left(\frac{(\text{total TEM})^2}{\mathrm{SD}^2}\right), \tag{7}$$

where SD^2 is the total inter-subject variance for the study in question, including measurement error. This coefficient is the most widely used measure of anthropometric precision in population studies (Mueller & Martorell, 1988) and reveals the proportion of between-subject variance in a measured population which is free from measurement error. In the case of a measurement with an R of 0·95, 95 % of the variance is due to factors other than 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 (Mueller & Martorell, 1988).

The relationship between TEM, %TEM and R can be established, knowing that %TEM = (TEM/mean) × 100, and that CV = (SD/mean) × 100:

$$\mathrm{R} = 1 - \left(\frac{\mathrm{TEM}^2/\mathrm{mean}^2}{\mathrm{SD}^2/\mathrm{mean}^2}\right) = 1 - \left(\frac{\%\mathrm{TEM}^2}{\mathrm{CV}^2}\right). \tag{8}$$

Equation (8) shows that R and TEM are related through the CV. Comparison of size of measurement, TEM, %TEM and R for a number of nutritional anthropometric measures from the same data set of Ross *et al.* (1994) shows some interesting anomalies across different measures (Table 3). In this subset of anthropometric measures, %TEM shows no significant relationship with size of circumference measures, while TEM does (r 0·99, $P < 0.001$); for skinfolds, there is no association between size of measurement and both TEM and %TEM. However, the relationship of %TEM to R varies according to specific measurement, and measurement type. Calculation of %TEM at given values of R, using equation (8) for values for means, standard deviations and TEM for the data of Ross *et al.* (1994) given in Table 3, shows that higher %TEM is not consistently associated with lower R. For example, at R = 0·994, calf circumference %TEM is 0·56, while for waist circumference the value is 0·79. At a very similar R value, %TEM for calf skinfold is 4·80, while for triceps skinfold it is 3·68. There is considerable variability in %TEM–R relationship within and

Table 3. Means and measurement error values for repeat measures data reported in Ross *et al.* (1994)

	Sample size	Mean (m)	CV (%)	TEM (m)	%TEM	R	%TEM at R:	
							0.990	0.950
Lengths:								
Arm span	55	1·796	7·16	0·0048	0·27	0·999	0·72	1·60
Stature	59	1·745	5·65	0·0022	0·12	0·999	0·56	1·26
Circumferences:								
Hip	874	0·942	5·56	0·0066	0·70	0·984	0·56	1·24
Waist	875	0·746	10·64	0·0059	0·79	0·994	1·06	2·38
Calf	369	0·358	7·32	0·0020	0·56	0·994	0·73	1·64
Upper arm	375	0·303	10·04	0·0035	1·16	0·987	1·00	2·24

Correlation coefficient, size of circumference measurement against: TEM (r 0·99, $P < 0.001$); %TEM (r 0·69)

Skinfolds:								
Suprailiac	898	0·107	42·70	0·00077	7·16	0·972	4·27	9·55
Triceps	898	0·106	41·79	0·00039	3·68	0·992	4·18	9·34
Calf	898	0·096	45·05	0·00046	4·80	0·989	4·50	10·07
Subscapular	898	0·091	31·33	0·00035	3·83	0·985	3·13	7·01
Biceps	898	0·051	41·62	0·00039	7·69	0·966	4·16	9·31

Correlation coefficient, size of skinfold measurement against: TEM (r 0·43); %TEM (r −0·53)

TEM, technical error of measurement; %TEM, relative technical error of measurement; R, coefficient of reliability.

between measurement types, and R and %TEM show different measures of anthropometric imprecision. This is a function of the CV associated with different types of measurement; in general, samples of circumference and length measurements have much smaller CV than do samples of skinfold measurements. At a given R value, %TEM, calculated using equation (8) varies approximately twofold across circumferences, and by approximately half across skinfolds and lengths respectively, for the Ross *et al.* (1994) dataset (Table 3). Furthermore, variation in %TEM across all nutritional anthropometric measures given in Table 3 shows an eightfold range between the highest and lowest calculated values.

Reliability can also be assessed using intraclass correlation coefficients (ICC). Values can range from 0 to 1, and this measure is an estimate of the proportion of the combined variance for the true biological value for any anthropometric measure, and for the measurement error associated with it (Norton & Olds, 1996). The ICC is close to 1 if there is low variability between repeated measures of the same subject; that is if measurement error is low. Another form of unreliability is undependability. This 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. The most common sources of undependability in nutritional anthropometry are with respect to measurements of height and skinfolds. Differences in height measurement of any individual may arise according to the time of day the measurement is made, due to increased compression of the spine later in the day. Size of skinfold measurement in any individual can vary according to duration and level of compression during measurement, which can vary according to level of tissue hydration (Ward & Anderson, 1993). It has been suggested that there are two components to skinfold compressibility: dynamic and static (Becque *et al.* 1986). Dynamic compressibility is probably due to the expulsion of water from subcutaneous tissue (Becque *et al.*

1986), while static compressibility is a function of the tension and thickness of the skin and subcutaneous tissue (Lee & Ng, 1967), and the distribution of fibrous tissue and blood vessels (Himes *et al.* 1979). Skinfold compressibility varies by site of measurement and between individuals, but not by sex. Caution has been urged when making comparisons of skinfolds between subjects and even between sites within the same subject (Martin *et al.* 1992), although it is possible that the degree of compressibility of cadaver skinfolds may differ from that of living subjects to some unknown degree.

Inaccuracy

Accuracy is the extent to which the 'true' value of a measurement is attained, while inaccuracy is systematic bias which reduces the likelihood of attaining the true value. The most common form of inaccuracy is that due to equipment bias, and the risk of inaccuracy is greater with a complex instrument than with a simple one. Thus, inaccuracy due to measurement by a simple tape measure is likely to be less than that due to measurement involving sliding scales, such as anthropometers and stadiometers, or spring-loaded calipers which are used to measure skinfolds. Different skinfold calipers are likely to give different degrees of compression at different sizes of actual skinfold, with corresponding differences in skinfold measurement. Compression differences between different makes of caliper have been identified (Schmidt & Carter, 1990) due to lack of standards for caliper-jaw surface area or spring tension (Gore *et al.* 1995); the standard jaw pressure of 10 g/mm gives greater compression by calipers with large surface area and heavier spring pressure, than by calipers with small surface area and lighter spring pressure (Schmidt & Carter, 1990). Thus, both Harpenden and Holtain skinfold calipers consistently give smaller values than Lange calipers (Gruber *et al.* 1990; Zillikens & Conway, 1990).

Differences in degree of compression and size of measurement have also been shown for calipers from the

same manufacturer. When comparing the dynamic compression of four different sets of Harpenden skinfold calipers relative to a fifth set which gave the greatest overall compression, Gore *et al.* (1995) found three sets to be within 5 % of the lowest measured value, with the fourth set giving measurements that were between 8 and 21 % greater than the measurements obtained with the calipers giving the greatest amount of compression. The fourth set had old, rather than new, springs, while the other three sets had new ones; this highlights the importance of regular inspection of equipment. Methods for dealing with such error of instrumentation include the regular calibration of skinfold calipers, and measurements involving the use of three sets of calipers for the mathematical estimation of uncompressed skinfolds (Ward & Anderson, 1993).

The timing of measurements can influence their accuracy in different ways, and has implications for the analysis and interpretation of data on short-term growth. If, for example, skinfolds are measured several times across a 5 min period in an attempt to increase accuracy, accuracy may actually decline, as later measurements are more compressed due to the expulsion of water from the adipose tissue at the site of measurement. If measurements are carried out hourly, the risk of inaccuracy due to observer fatigue and reduced motivation across the day rises. If measurements are made daily, then the timing of the measurement might bear upon accuracy, as with the increased compression of the spine across the course of the day. If measurements are carried out weekly for a year, then bias due to small changes in measurement technique across this period is possible.

Acceptable levels of measurement error under study conditions

Acceptable levels of measurement error are difficult to ascertain because TEM is age dependent, and related to the anthropometric characteristics of the group or population under investigation. However, $R > 0.95$ should be sought where possible (Ulijaszek & Lourie, 1994). While it has been largely unfashionable to report levels of measurement error in anthropometric studies in the past (possibly because these values were difficult to interpret meaningfully in the context of the data collected), this has changed, with many studies since 1990 reporting measurement error values. Tables 4 and 5 give intra- and inter-observer values for TEM and R from a number of studies for a range of nutritional anthropometric measures, including weight, length, height, and arm, waist, hip and calf circumferences, and biceps, triceps, subscapular, suprailiac (also known as supraspinale), and medial calf skinfolds.

The range of values for TEM and R varies enormously across measurement type, for both intra- and inter-observer error. The interpretation of this variation is difficult, because discussion of methodological sources of measurement error in any of the studies reported in Tables 4 and 5 is at best limited. The extremely high values for TEM in intra-observer measures of arm and calf circumferences, and suprailiac and medial calf skinfolds come from one study, the Hispanic Health and Nutrition Examination Survey (Chumlea *et al.* 1990). High inter-observer TEM for arm circumference, subscapular, suprailiac and medial calf

skinfolds are also found in this study. Extremely high values for intra-observer TEM in triceps and subscapular skinfolds come from another large study, that of Ferrario *et al.* (1995). Extremely high values for inter-observer TEM in weight, waist and hip circumferences come from self-reported data (Rimm *et al.* 1990), while high inter-observer TEM values for waist and hip circumferences, and for suprailiac and medial calf skinfolds are found among a group of recently trained nurse anthropometrists (Williamson *et al.* 1993). Thus, circumstances in which extremely high TEM are reported include: (1) large epidemiological studies in which anthropometry is but one part of a complex study design involving many methods; and (2) when data are either self-reported by the subjects, or anthropometrists are recently trained with limited experience of anthropometry.

Although the number of studies giving measurement error is rather limited with respect to length, demispan, calf circumference, biceps, suprailiac and calf skinfolds, some provisional generalizations can be made. The expectation that inter-observer error should be greater than intra-observer error is not met, from empirical observation. Intra-observer R is greater than inter-observer R for measures of hip circumference and biceps skinfolds. Intra- and inter-observer R values are similar to each other for measures of weight, height, length, demispan, arm and waist circumferences, and triceps, subscapular, suprailiac and calf skinfolds. Inter-observer R is greater than intra-observer R for calf circumference.

With respect to intra-observer measurement error, weight, length, height, demispan, arm, waist and hip circumferences and biceps skinfolds all give acceptable levels of R (>0.95). For inter-observer measurement error, acceptable levels of R are generally achieved for weight, length, height, demispan, arm and calf circumferences. Lower levels of R are achieved for calf circumference, triceps, subscapular, suprailiac and medial calf skinfolds (intra-observer), and for waist and hip circumferences, biceps, triceps, subscapular, suprailiac and medial calf skinfolds (inter-observer). Caution is needed in carrying out skinfold measures, regardless of whether one observer or several are involved in any particular study, it being more important to obtain correct training and maintenance of standardized techniques (Weiner & Lourie, 1981; Lohman *et al.* 1988; Norton & Olds, 1996), than to reduce between-observer variation. With respect to hip circumference measurement, the best way to minimize measurement error is to ensure adequate training and quality control across time, as well as to minimize the number of observers within any study. While it might be expected that the use of composite measures in nutritional assessment, such as BMI, waist : hip ratio, or the sum of four skinfolds for the prediction of percentage body fat might improve the overall precision of measurements, this appears to be true only where the R values of the individual measurements are high (Mueller & Kaplowitz, 1994). Where the R values of individual measurements are lower than 0·99, the composite R values are also low (Mueller & Kaplowitz, 1994).

Based on calculations of TEM at set levels of R using published data from the combined National Health and Nutrition Examination Surveys I and II (Frisancho, 1990), Ulijaszek & Lourie (1994) have put forward references for

Table 4. Reported values for intra-observer technical error of measurement (TEM) and coefficient of reliability (R)

Measurement	TEM				R			
	No. of studies	Sources*	Mean	Range	No· of studies	Sources*	Mean	Range
Weight (kg)	6	1, 2, 3, 4, 5	0·17	0·1–0·3	7	1, 2, 3, 4, 5, 6, 7	0·98	0·95–1·00
Length (m)	4	1, 8, 9, 10	0·0035	0·001–0·008	3	1, 9, 10	0·96	0·92–0·99
Height (m)	19	1, 2, 3, 4, 6, 7, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22	0·0038	0·001–0·013	10	1, 2, 3, 5, 6, 7, 11 12, 16, 22	0·98	0·93–0·99
Demispan (m)	1	5	0·0030		1	5	0·99	
Arm circumference (m)	16	2, 4, 6, 7, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24	0·0026	0·001–0·006	9	2, 6, 7, 10, 11, 22, 23, 24	0·95	0·85–0·99
Waist circumference (m)	2	25, 26	0·013	0·010–0·016	4	6, 7, 25, 26	0·97	0·97–0·98
Hip circumference (m)	2	25, 26	0·013	0·012–0·014	4	6, 7, 25, 26	0·97	0·96–0·99
Calf circumference (m)	11	2, 7, 12, 13, 14, 15, 18, 19, 20, 21	0·0031	0·001–0·008	3	2, 7	0·85	0·73–0·95
Biceps skinfold (mm)	3	6, 15, 24	0·17	0·1–0·2	2	6, 24	0·97	0·95–0·97
Triceps skinfold (mm)	21	6, 7, 10, 16, 22, 23, 24, 25, 26, 27, 28, 29	0·84	0·1–3·7	12	6, 7, 10, 16, 22, 23, 24, 25, 26, 27, 28, 29	0·93	0·81–0·99
Subscapular skinfold (mm)	19	4, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 26, 27, 28, 29	1·26	0·1–7·4	11	6, 7, 10, 16, 22, 23, 25, 26, 27, 28, 29	0·94	0·81–0·99
Suprailiac skinfold (mm)	10	4, 12, 13, 14, 15, 18, 19, 20, 25, 29	1·16	0·1–3·2	3	6, 25, 29	0·89	0·79–0·96
Medial calf skinfold (mm)	9	4, 13, 14, 15, 18, 19, 20, 21, 29	1·03	0·2–2·7	2	7, 29	0·88	0·82–0·95

* Sources: 1. Kaur & Singh, 1994; 2. Benefice & Malina, 1996; 3. Malina & Morayama, 1991; 4. Rosique *et al.* 1994; 5. Dangour, 1998; 6. Mueller & Kaplowitz, 1994; 7. Mueller *et al.* 1996; 8. Lampl, 1993; 9. S. J. Ulijaszek, unpublished results; 10. Pelletier *et al.* 1991; 11. Malina *et al.* 1973; 12. Malina & Buschang, cited in Malina, 1995; 13. Mueller, 1975; 14. Rocha Ferreira, 1987; 15. Meleski, 1980; 16. Zavaleta & Malina, 1982; 17. Gordon-Larsen *et al.* 1997; 18. Brown, 1984; 19. Chumlea *et al.* 1990; 20. Wellens, 1989; 21. Malina & Buschang, 1984; 22. Ulijaszek & Lourie, 1994; 23. Martorell *et al.* 1975; 24. Martine *et al.* 1997; 25. Mueller & Malina, 1987; 26. Ferrario *et al.* 1995; 27. Branson *et al.* 1982; 28. Johnston *et al.* 1972; 29. Pham *et al.* 1995.

the upper limits for TEM at two levels of reliability, for males and females respectively (Table 6). These use total TEM, which in the case of a single observer is equivalent to intra-observer TEM, and give some idea of the acceptability of measurement error.

Training, and measurement problems associated with multi-observer anthropometry

The services of a number of anthropometrists are required in a variety of contexts, including nutritional screening, surveillance, and both clinical and epidemiological studies involving large cross-sectional or extensive longitudinal design. This increases the possible extent of measurement error. Even where experienced anthropometrists are employed, small differences in technique can occur over time, and this should be controlled for. It is therefore important to assess, where possible, inter-observer differences between anthropometrists and there should be a designated criterion anthropometrist engaged both in training new anthropometrists, and subsequently in the course of work, to maintain quality of measurement. This serves to identify and correct systematic errors in newly trained anthropometrists and maintain quality of measurement among already trained anthropometrists. The determination of accuracy is problematic, since the correct value of any

anthropometric measure is impossible to know. Operationally, accuracy is determined by comparison of measures made against those of a criterion anthropometrist, an individual who has internalized, as far as is humanly possible, the rules of anthropometric measurement as delineated in the literature (e.g. Cameron *et al.* 1981; Cameron, 1984, 1986; Lohman *et al.* 1988; Gibson, 1990; Norton & Olds, 1996) and has received training to the highest level and compares well in anthropometric measurement against another criterion anthropometrist.

Imprecision as a measure of measurement error variance is easier to determine than accuracy, and is obtained by calculating TEM, R and/or %TEM. Ideally, duplicate measurement of at least ten subjects should be carried out for the calculation of intra- and inter-observer TEM, and R. It is important to carry out pilot tests of protocols before engaging in any study involving nutritional anthropometry. This gives an indication of the number of subjects that can be measured within a specified time period, and allows the individuals best suited to specific tasks to be identified. For example, one anthropometrist may be preferred above another on the basis of any of the following: accuracy relative to a criterion anthropometrist, low imprecision, and/or speed of measurement. Before data collection, one or more criterion anthropometrists should be appointed, their role being to over-see measurement and to verify any

Table 5. Reported values for inter-observer technical error of measurement (TEM) and coefficient of reliability (R)

Measurement	TEM				R			
	No. of studies	Sources*	Mean	Range	No. of studies	Sources*	Mean	Range
Weight (kg)	12	1, 4, 5, 6, 7, 8, 9, 10, 11, 12	1·28	0·1–4·1	14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	0·98	0·94–1·00
Length (m)	3	5, 13, 14	0·0027	0·001–0·005	2	5, 14	0·99	0·99
Height (m)	21	1, 5, 6, 9, 10, 11, 15, 16, 17, 18, 19 20, 21, 22, 23, 24, 25	0·0038	0·002–0·008	14	1, 2, 3, 5, 6, 9, 10, 11, 15, 17, 23, 25	0·99	0·95–1·00
Demispan (m)	2	6, 26	0·0030	0·001–0·005	2	6, 26	0·98	0·96–1·00
Arm circumference (m)	13	4, 6, 9, 10, 14, 18, 19, 21, 22, 23, 24	0·0037	0·001–0·013	9	2, 3, 4, 6, 9, 10, 14, 27	0·97	0·94–1·00
Waist circumference (m)	10	6, 7, 10, 11, 12, 30, 32, 33	0·0234	0·006–0·042	13	2, 3, 6, 7, 10, 11, 31, 32, 33	0·94	0·86–0·99
Hip circumference (m)	10	6, 7, 10, 11, 12, 30, 32, 33	0·0280	0·007–0·061	12	2, 3, 6, 7, 10, 11, 12, 30, 32, 33	0·89	0·68–0·99
Calf circumference (m)	8	4, 6, 9, 19, 21, 22, 28, 29	0·0029	0·002–0·004	4	4, 6, 9, 29	0·99	0·98–0·99
Biceps skinfold (mm)	8	4, 5, 6, 9, 10, 29, 32	0·84	0·2–2·1	9	2, 4, 5, 6, 9, 10, 29, 32	0·84	0·49–0·98
Triceps skinfold (mm)	28	4, 5, 6, 7, 9, 10, 14, 18, 19, 21, 22, 23, 24, 25, 30, 32, 33, 34, 35, 36, 37, 38, 39, 40	1·06	0·2–4·7	24	2, 3, 4, 5, 6, 7, 9, 10, 14, 18, 19, 21, 22, 23, 24, 25, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40	0·88	0·48–0·99
Subscapular skinfold (mm)	28	4, 5, 6, 7, 9, 10, 14, 18, 19, 21, 22, 23, 24, 25, 30, 32, 33, 34, 35, 36, 37, 39, 40	1·21	0·1–3·3	24	2, 3, 4, 5, 6, 7, 9, 10, 14, 23, 25, 30, 31, 32, 33, 34, 35, 36, 37	0·91	0·60–0·99
Suprailiac skinfold (mm)	11	4, 5, 6, 9, 19, 29, 30, 31, 32	2·28	0·3–6·4	10	2, 4, 5, 6, 9, 19, 21, 22, 29, 30, 31, 32	0·85	0·56–0·97
Medial calf skinfold (mm)	12	4, 6, 9, 10, 19, 21, 22, 29, 32, 38	1·51	0·3–3·9	8	3, 4, 6, 9, 10, 29, 32	0·88	0·81–0·99

* Sources: 1. Ohsawa *et al.* 1997; 2. Mueller & Kaplowitz, 1994; 3. Mueller *et al.* 1996; 4. Strickland & Tuffrey, 1997; 5. E. Karim, unpublished results; 6. Ross *et al.* 1994; 7. Gerber *et al.* 1995; 8. Marks *et al.* 1989; 9. Strickland *et al.* 1993; 10. Klipstein-Grobusch *et al.* 1997; 11. Sonnenschein *et al.* 1993; 12. Rimm *et al.* 1990; 13. Lampl, 1993; 14. Pelletier *et al.* 1991; 15. Voss *et al.* 1990; 16. Lopez-Blanco *et al.* 1995; 17. Lohman *et al.* 1975; 18. Buschang, 1980; 19. Mueller, 1975; 20. Gordon-Larsen *et al.* 1997; 21. Chumlea *et al.* 1990; 22. Wellens, 1989; 23. Ulijaszek & Lourie, 1994; 24. Chumlea *et al.* 1984; 25. Marks *et al.* 1989; 26. Smith *et al.* 1995; 27. Callaway, 1988; 28. Malina & Buschang, in Malina, 1995; 29. S. J. Ulijaszek and J. A. Lourie, unpublished results; 30. Mueller & Malina, 1987; 31. Jackson *et al.* 1978; 32. Williamson *et al.* 1993; 33. Ferrario *et al.* 1995; 34. Branson *et al.* 1982; 35. Johnston & Mack, 1985; 36. Lohman *et al.* 1975; 37. Johnston *et al.* 1972; 38. Blade *et al.* 1995; 39. Johnston *et al.* 1974; 40. Sloan & Shapiro, 1972.

questionable landmarks or measurements, at least in the first instance. The working environment for data collection should also be planned so that there is adequate space for each measurement station. If the measurement stations are too close or poorly lit, additional error can occur as a consequence of crowding, misrecording, or both.

Target training values

Targets for anthropometric assessment have been put forward by Zerfas (1985) using a repeat-measures protocol. The trainee and trainer measure the same subjects until the difference between the two of them is good, or at the very least, fair (Table 7). This scheme does not allow for the possibility that the trainee may have greater accuracy than the trainer, in which case any improved accuracy across the course of training by the trainee would not be detected. Furthermore, the Zerfas (1985) target values should not be used uncritically, since differences between trainer and trainee at the upper level of 'goodness' for height, weight, arm circumference and skinfolds represent different proportions of the absolute measure according to the size of the

measurement. Thus, although Zerfas (1985) gives values for differences that are possible given the techniques available, a 5 mm difference in height measurement is more accurate than the same difference in arm circumference. The proportion of the total measurement of a model young child and model adult which would be represented by measurement differences between trainee and trainer at the maximum level considered good in the Zerfas (1985) scheme has been estimated to be less than 1 % of the size of measurement of length, height and weight, 3·2 % and 1·7 % for arm circumference of the model child and adult respectively, and in excess of 10 % for both triceps and subscapular skinfolds (Ulijaszek, 1997). Thus the Zerfas recommendations for acceptable measurement error are good for length, height and weight, acceptable for arm circumference, but poor for skinfolds. This problem is greater in the youngest age groups, and among the smallest children within any age group (Ulijaszek, 1997). The use of the Zerfas scheme is appropriate for the training of anthropometrists, but for measurements other than length, height and weight, it should be used with care. Furthermore, the repeat-measurement protocol should report on any systematic biases in

Table 6. Reference values for total technical error of measurement (TEM) (adapted from Ulijaszek & Lourie, 1994)

Age group (years)	Height (m)	Weight (kg)	Arm circumference (mm)	Triceps skinfold (mm)	Subscapular skinfold (mm)
R = 0·095, males					
1–4·9	0·0103	0·21	3·1	0·61	0·43
5–10·9	0·0130	1·20	5·2	0·97	0·87
11–17·9	0·0169	5·94	7·5	1·45	1·55
18–64·9	0·0152	13·06	7·3	1·38	1·79
65+	0·0152	10·80	7·4	1·29	1·74
R = 0·99, males					
1–4·9	0·0046	0·04	1·4	0·28	0·19
5–10·9	0·0058	0·24	2·3	0·43	0·39
11–17·9	0·0076	1·19	3·3	0·65	0·69
18–64·9	0·0068	2·61	3·3	0·62	0·80
65+	0·0068	2·16	3·3	0·58	0·78
R = 0·95, females					
1–4·9	0·0104	0·22	3·0	0·65	0·47
5–10·9	0·0138	1·61	5·4	1·05	1·08
11–17·9	0·0150	8·66	7·8	1·55	1·74
18–64·9	0·0139	16·74	9·8	1·94	2·39
65+	0·0135	11·70	9·8	1·86	2·27
R = 0·99, females					
1–4·9	0·0047	0·04	1·3	0·29	0·21
5–10·9	0·0062	0·32	2·4	0·47	0·48
11–17·9	0·0067	1·73	3·5	0·69	0·78
18–64·9	0·0062	3·35	4·4	0·87	1·07
65+	0·0060	2·34	4·4	0·83	1·02

R, coefficient of reliability.

measurement between trainer and trainee, include evaluation of the technique of the trainee relative to the criterion anthropometrist, and of both trainee and criterion anthropometrist relative to standard techniques given in reference manuals (e.g. Cameron *et al*. 1981; Cameron, 1984, 1986; Lohman *et al*. 1988; Gibson, 1990; Norton & Olds, 1996).

Alternative anthropometric training targets have been set for the accreditation scheme of sports anthropometrists (Gore *et al*. 1996), which could also be considered for nutritional anthropometrists. Expected %TEM associated with different levels of completed training are: skinfolds 7·5 (level 1), 5·0 (levels 2 and 3); circumferences 1·0 (all levels) (Gore *et al*. 1996). Level 3 anthropometrists are at the instructor level and have previously completed level 1 and level 2 training. It would be expected that the TEM for these anthropometrists should be lower than for those achieved by level 1 anthropometrists. However, %TEM is not a consistent measure of imprecision across different types of

anthropometric measurement, and for accreditation, anthropometrists must submit TEM, %TEM and ICC data for each measurement. This scheme is still being developed, and thus far only acceptable levels for %TEM have been set.

Use and interpretation of measurement error

Estimates of anthropometric measurement can be used to improve measurement technique, and to identify measures with high levels of error. They can also be used in critical evaluation of longitudinal anthropometric change in individuals. In cross-sectional studies a proportion of the total variance for any group measure of nutritional anthropometry observed is due to measurement error. The total observed variance is composed of both biological and error variance and can be summarized thus:

$$V_t = V_b + V_{e1} + V_{e2} + V_{e3}, \qquad (9)$$

where V_t is the total variance observed, V_b is the biological, or true variance, V_{e1} is the variance due to intra-observer measurement error, V_{e2} is the variance due to inter-observer error and V_{e3} is the variance due to instrument error. Usually, anthropometric data are reported as though V_t were V_b. Although the two are often so close that for all practical purposes the slight difference does not matter, this is not always the case. If the variances due to errors of one sort or another are large, they may mask true biological differences in anthropometric characteristics between groups, when statistical comparisons are being made.

In longitudinal studies, both biological variance and anthropometric measurement error may change with time, and estimation of measurement error should be carried out longitudinally. In such studies, knowledge of TEM can give some estimate of the proportion of difference between two longitudinal measurements which might be attributed to measurement error. This is of potential diagnostic value, since most child survival, health and development programmes involve regular growth monitoring (Tomkins, 1994) and/or monitoring of nutritional status using anthropometry (World Health Organization, 1995). The proportion of difference between two measures which can be attributed to measurement error is illustrated in the following example. With a TEM of 0·3 for a given anthropometric variable, the TEM for the difference between two measurements is:

$$\sqrt{(0·3)^2 + (0·3)^2} = 0·42, \qquad (10)$$

since the two TEM values combine as variances. Thus, one can have confidence of 5 % probability of a measurement difference of $(2 \times 0·42) = 0·84$ being due to measurement

Table 7. Evaluation of measurement error among trainees (after Zerfas, 1985)

Measurement	Difference between trainee and trainer			
	Good	Fair	Poor	Gross error
Height or length (m)	0–0·005	0·006–0·009	0·010–0·019	≥0·020
Weight (kg)	0–0·1	0·2	0·3–0·4	≥0·5
Arm circumference (mm)	0–5	6–9	10–19	≥20
Skinfolds (any) (mm)	0–0·9	1·0–1·9	2·0–4·9	≥5·0

Table 8. Proportion of expected 6-month gain of an hypothetical 10-year-old 50th centile child (Frisancho, 1990) represented by 95 % confidence intervals for imprecision (technical error of measurement (TEM))

	6-month gain	Best case		Worst case	
		TEM	% of gain represented by ± 2TEM	TEM	% gain represented by ± 2TEM
Weight (kg)	1·9	0·1*	15		
Height (m)	0·033	0·002	17	0·007	60
Arm circumference (mm)	5	2	113	6	339
Triceps skinfold (mm)	0·5	0·3	170	1·9	1075
Subscapular skinfold (mm)	0·5	0·4	226	1·5	848

* From Ohsawa *et al.* 1997 (the only value reported for this age).

error alone. In the measurement of stature, a TEM of 0·001 m means that the probability of differences in excess of:

$$2 \times \sqrt{(0{\cdot}001)^2 + (0{\cdot}001)^2} = 0{\cdot}0028 \, \text{m} \qquad (11)$$

being due to measurement error alone are less than 5 %. Thus, TEM can be used in a study-specific way to evaluate longitudinal growth.

An hypothetical example is given for a 10-year-old male child who is on the 50th centile of National Health and Nutrition Examination Survey reference values (Frisancho, 1990) for weight, height, arm circumference, triceps and subscapular skinfolds (Table 8). Using reported inter-observer TEM values for children about this age to present best- and worst-case measurement error possibilities for measures of height, arm circumference, triceps and sub-scapular skinfolds, and an only-case measurement error possibility for weight, it is clear that weight measurement error is sufficiently small for 6-month gain to be observed without undue imprecision. The same is true for height, given the best case for measurement error. However, the worst case possibility for stature is that 60 % of the 6-month gain in this example might be attributed to measurement error. For arm circumference, the 6-month gain might be interpreted as consisting of 113 % and 339 % measurement error, in the best and worst cases for TEM, respectively. That is, the 6-month gain in arm circumference cannot be measured with precision given either the best or worst case TEM values. This is also true for triceps and subscapular skinfolds. In the best case, weight gain or loss of more than 0·3 kg can be detected with 95 % confidence, while height gain of 0·006 m, arm circumference change of 6 mm and triceps skinfold change of 0·8 mm and subscapular skinfold change of 1·1 mm can be detected with the same level of confidence. In the worst case, changes in excess of 0·3 kg (weight), 0·02 m (height), 17 mm (arm circumfer-ence), 5·4 mm (triceps skinfold) and 4·2 mm (subscapular skinfold) can be detected with 95 % confidence. In this way, knowledge of measurement error can give an estimate of the degree of anthropometric change that can be measured and accepted as real. This estimate is weakened if biological variation also changes with time. Although the exact extent of true biological variation and its change across time cannot be known, this method allows variation due to measurement error to be considered when interpreting change in anthropometric variables across time within any individual.

Concluding remarks

Anthropometric measurement error is unavoidable, and should be minimized by paying close attention to every aspect of the data collection process. This includes ensuring that there is good lighting in which to take measurements, regular calibration of equipment, and the prevention of tiredness among personnel to reduce the possibility of mistakes. It is important to follow a standard protocol which includes the double measurement of a sub-sample of the group or population under study, so that some measure of imprecision can be calculated. This can take one or more of several forms, including TEM, %TEM, R and ICC. Imprecision can be minimized by seeking comparison with reference values in the training process, and in the course of data collection. The knowledge that a large degree of imprecision exists in an anthropometric variable can be used to advantage in either analysis or interpretation. In longitudinal studies, knowledge of the TEM allows 95 % CI to be determined for change in anthropometric measures, such that differences across time can be realistically determined.

A comparison of studies reveals that there is a clear hierarchy in precision of different nutritional anthropo-metric measures. Weight and height are the most precisely measured, and it is entirely appropriate that they continue to be the predominant measure of choice in the vast majority of nutritional anthropometric studies. Waist and hip circum-ference show strong between-observer differences, and should, where possible, be carried out by one observer. Skinfolds remain problematic, and while valuable, can be associated with such large measurement error that interpre-tation is difficult. Regardless of the measurement made and the size of the error, it is better to know the size of error, since this will determine the confidence one has in the different measurements made, and will influence the inter-pretation of anthropometric data collected.

Acknowledgements

References

Becque MD, Katch VL & Moffatt RJ (1986) Time course of skin-plus-fat compression in males and females. *Human Biology* **58**, 33–42.

Benefice E & Malina RM (1996) Body size, body composition and motor performances of mild-to-moderately undernourished Senegalese children. *Annals of Human Biology* **23**, 307–321.

Blade L, Ward R & Martin A (1995) Dissociation between skinfold thickness changes and growth of adipose tissue volume in children and youth. *American Journal of Human Biology* **7**, 529–534.

Branson RS, Vaucher YE, Harrison GG, Vargas M & Thesis C (1982) Inter- and intra-observer reliability of skinfold thickness measurements in new born infants. *Human Biology* **54**, 137–143.

Brown KR (1984) Growth, physique and age at menarche of Mexican American females age 12 through 17 years residing in San Diego County, California. PhD Thesis, University of Texas at Austin.

Buschang PH (1980) Growth status and rate in school children 6 to 13 years of age in a rural Zapotec-speaking community in the Valley of Oaxaca, Mexico. PhD Thesis, University of Texas at Austin.

Callaway CW, Chumlea WC, Bouchard C, Himes JH, Lohman TG, Martin AD, Mitchell CD, Mueller WH, Roche AF & Seefeldt VD (1988) Circumferences. In *Anthropometric Standardization Reference Manual*, pp. 39–54 [TG Lohman, AF Roche and R Martorell, editors]. Champaign, IL: Human Kinetics Books.

Cameron N (1984) *The Measurement of Human Growth*. London: Croom Helm.

Cameron N (1986) The methods of auxological anthropometry. In *Human Growth: a Comprehensive Treatise*, vol. 3, pp. 3–46 [F Falkner and JM Tanner, editors]. New York, NY: Plenum Press.

Cameron N, Hiernaux J, Jarman S, Marshall WA, Tanner JM & Whitehouse RH (1981) Anthropometry. In *Practical Human Biology*, pp. 27–52 [JS Weiner and JA Lourie, editors]. London: Academic Press.

Carter JEL & Schmidt PK (1990) A simple method for calibrating skinfold calipers. In *Proceedings of the Commonwealth and International Conference on Physical Education, Sport, Health, Dance, Recreation and Leisure*, vol. 3, part 1, pp. 49–53. Auckland, New Zealand: ?????.

Chumlea WC, Guo, S, Kuczmarski RJ, Johnson CL & Leahy CK (1990) Reliability for anthropometric measurements in the Hispanic Health and Nutrition Examination Survey (HHANES 1982–1984). *American Journal of Clinical Nutrition* **51**, 902S–907S.

Chumlea WC, Roche AF & Rogers E (1984) Replicability for anthropometry in the elderly. *Human Biology* **56**, 329–337.

Dangour A (1998) Growth of body proportion in two Amerindian tribes in Guyana. PhD Thesis, University of London.

Ferrario M, Carpenter MA & Chambless LE (1995) Reliability of body fat distribution measurements. The ARIC study baseline cohort results. *International Journal of Obesity* **19**, 449–457.

Frisancho AR (1990) *Anthropometric Standards for the Assessment of Growth and Nutritional Status*. Ann Arbor, MI: University of Michigan Press.

Gerber L, Schwartz JE, Schnall PL & Pickering TG (1995) Body fat and fat distribution in relation to sex differences in blood pressure. *American Journal of Human Biology* **7**, 173–182.

Gibson RS (1990) *Principles of Nutritional Assessment*. Oxford: Oxford University Press.

Gordon-Larsen P, Zemel BS & Johnston FE (1997) Secular changes in stature, weight, fatness, overweight, and obesity in urban African American adolescents from the mid-1950's to the mid-1990's. *American Journal of Human Biology* **9**, 675–688.

Gore C, Norton K, Olds T, Whittingham N, Birchall K, Clough M,

Dickerson B & Downie L (1996) Accreditation in anthropometry: an Australian model. In *Anthropometrica*, pp. 395–411 [K Norton and T Olds, editors]. Sydney: University of New South Wales Press.

Gore CJ, Woolford SM & Carlyon RG (1995) Calibrating skinfold calipers. *Journal of Sports Sciences* **13**, 355–360.

Gruber JJ, Pollock ML, Graves JE, Colvin AB & Braith RW (1990) Comparison of Harpenden and Lange calipers in predicting body composition. *Research Quarterly* **61**, 184–190.

Habicht J-P, Yarbrough C & Martorell R (1979) Anthropometric field methods: criteria for selection. In *Nutrition and Growth*, pp. 365–387 [DB Jelliffe and EFP Jelliffe, editors]. New York, NY: Plenum Press.

Heymsfield SB, McManus CB, Seitz SB, Nixon DW & Andrews JS (1984) Anthropometric assessment of adult protein–energy malnutrition. In *Nutritional Assessment*, pp. 27–81 [RA Wright, S Heymsfield and CB McManus, editors]. Boston, MA: Blackwell Scientific Publications Inc.

Himes JH, Roche AF & Siervogel RM (1979) Compressibility of skinfolds and the measurement of subcutaneous fatness. *American Journal of Clinical Nutrition* **32**, 1734–1740.

Jackson AS, Pollock ML & Gettman LR (1978) Intertester reliability of selected skinfold and circumference measurements and percent fat estimates. *Research Quarterly* **4**, 546–551.

Jelliffe DB & Jelliffe EFP (1989) *Community Nutritional Assessment*. Oxford: Oxford University Press.

Johnston FE, Dechow PC & MacVean RB (1975) Age changes in skinfold thickness among upper class school children of differing ethnic backgrounds residing in Guatamala. *Human Biology* **47**, 251–262.

Johnston FE, Hamill PVV & Lemeshow S (1972) Skinfold thickness of children 6–11 years, United States. *National Health Survey Series II*, no. 120, *DHEW Publication* no. (HSM) 73-1602. Rockville, ML: National Center for Health Statistics.

Johnston FE, Hamill PVV & Lemeshow S (1974) Skinfold thicknesses in a national probability sample of U.S. males and females aged 6 through 17 years. *American Journal of Physical Anthropology* **40**, 321–324.

Johnston FE & Mack RW (1985) Interobserver reliability of skinfold measurements in infants and young children. *American Journal of Physical Anthropology* **67**, 285–289.

Kaur B & Singh R (1994) One year follow-up study of stature, weight, emergence of dentition, and sexual maturation of well-nourished Indian girls from birth to 20 years. *American Journal of Human Biology* **6**, 425–436.

Klipstein-Grobusch K, Georg T & Boeing H (1997) Interviewer variability in anthropometric measurements and estimates of body composition. *International Journal of Epidemiology* **26**, S174–S180.

Lampl M (1993) Evidence of saltatory growth in infancy. *American Journal of Human Biology* **5**, 641–652.

Lee MMC & Ng CK (1967) Postmortem studies of skinfold caliper measurement and actual thickness of skin and subcutaneous tissue. *Human Biology* **37**, 91–103.

Lohman TG, Boileau RA & Massey BH (1975) Prediction of lean body mass in young boys from skinfold thickness and body weight. *Human Biology* **47**, 245–262.

Lohman TG, Roche AF & Martorell R (1988) *Anthropometric Standardisation Reference Manual*. Champaign, IL: Human Kinetics Books.

Lopez-Blanco M, Izaguirre-Espinoza I, Macias-Tomei C & Saab-Verardy L (1995) Growth in stature in early, average, and late maturing children of the Caracas mixed-longitudinal study. *American Journal of Human Biology* **7**, 517–527.

Lourie JA & Ulijaszek SJ (1992) Observer error in anthropometry: age dependence of technical error of measurement. *American Journal of Physical Anthropology*, Suppl. 14, 113.

Malina RM (1995) Anthropometry. In *Physiological Assessment of Human Fitness*, pp. 205–219 [PJ Maud and C Foster, editors]. Champaign, IL: Human Kinetics Books.

Malina RM & Buschang PH (1984) Anthropometric asymmetry in normal and mentally retarded males. *Annals of Human Biology* **11**, 515–531.

Malina RM, Hamill PVV & Lemeshow S (1973) Selected body measurements of children 6–11 years, United States. *Vital and Health Statistics*, series 11, no. 123. Washington, DC: U.S. Government Printing Office.

Malina RM & Moriyama M (1991) Growth and motor performance of black and white children 6–10 years of age: A multivariate analysis. *American Journal of Human Biology* **3**, 599–611.

Marks GC, Habicht J-P & Mueller WH (1989) Reliability, dependability, and precision of anthropometric measurements: The Second National Health and Nutrition Examination Survey 1976–1980. *American Journal of Epidemiology* **130**, 578–587.

Martin AD, Drinkwater DT, Clarys JP, Daniel M & Ross WD (1992) Effects of skin thickness and skinfold compressibility on skinfold thickness measurement. *American Journal of Human Biology* **4**, 453–460.

Martine T, Claessens AL, Vlietinck R, Marchal G & Beunen G (1997) Accuracy of anthropometric estimation of muscle cross-sectional area of the arm in males. *American Journal of Human Biology* **9**, 73–86.

Martorell R, Habicht J-P, Yarbrough C, Guzman G & Klein RE (1975) The identification and evaluation of measurement variability in the anthropometry of preschool children. *American Journal of Physical Anthropology* **43**, 347–352.

Meleski BW (1980) Growth, maturity, body composition and familial characteristics of competitive swimmers 8 to 18 years of age. PhD Thesis, University of Texas at Austin.

Mueller WH (1975) Parent-child and sibling correlations and heritability of body measurements in a rural Colombian population. PhD Thesis, University of Texas at Austin.

Mueller WH & Kaplowitz HJ (1994) The precision of anthropometric assessment of body fat distribution in children. *Annals of Human Biology* **21**, 267–274.

Mueller WH & Malina RM (1987) Relative reliability of circumferences and skinfolds as measures of body fat distribution. *American Journal of Physical Anthropology* **72**, 437–439.

Mueller WH & Martorell R (1988) Reliability and accuracy of measurement. In *Anthropometric Standardisation Reference Manual*, pp. 83–86 [TG Lohman, AF Roche and R Martorell, editors]. Champaign, IL: Human Kinetics Books.

Mueller WH, Taylor WC, Chan W, Sangi-Haghpeykar H, Snider SA & Hsu H (1996) Precision of measuring body fat distribution in adolescent African American girls from the 'Healthy Growth Study'. *American Journal of Human Biology* **8**, 325–329.

Norton K & Olds T (editors) (1996) *Anthropometrica*. Sydney: University of New South Wales Press.

Ohsawa S, Ji C-Y & Kasai N (1997) Age at menarche and comparison of the growth and performance of pre- and post-menarcheal girls in China. *American Journal of Human Biology* **9**, 205–212.

Pederson D & Gore C (1996) Anthropometry measurement error. In *Anthropometrica*, pp. 77–96 [K Norton and T Olds, editors]. Sydney: University of New South Wales.

Pelletier DL, Low JW & Msukwa LAH (1991) Sources of measurement variation in child anthropometry in the Malawi maternal and child nutrition survey. *American Journal of Human Biology* **3**, 227–237.

Pham CL, Mueller WH, Wear ML, Emerson JB, Hanis CL & Schull WJ (1995) Precision of the one- versus two-handed method of skinfold measurement in the obese. *American Journal of Human Biology* **7**, 617–621.

Pheasant S (1988) *Bodyspace. Anthropometry, Ergonomics and Design*. London: Taylor & Francis.

Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB & Willett W (1990) Validity of self-reported waist and hip circumferences in men and women. *Epidemiology* **1**, 466–473.

Rocha Ferreira MB (1987) Growth, physical performance and psychological characteristics of eight year old Brazilian school children from low socio-economic background. PhD Thesis, University of Texas at Austin.

Roebuck JA, Kroemer KHE & Thomson WG (1975) *Engineering Anthropometry Methods*. New York, NY: Wiley.

Rosique J, Rebato E, Apraiz AG & Pacheco JL (1994) Somatotype related to centripetal fat patterning of 8- to 19-year old Basque boys and girls. *American Journal of Human Biology* **6**, 171–181.

Ross WD, Kerr DA, Carter JEL, Ackland TR & Bach TM (1994) Anthropometric techniques: precision and accuracy. In *Kinanthropometry in Aquatic Sports. A Study of World Class Athletes*, *Human Kinetics Sport Science Monograph* Series vol. 5, pp. 158–173 [JEL Carter and TR Ackland, editors]. Champaign, IL: Human Kinetics Books.

Schmidt PK & Carter JEL (1990) Static and dynamic differences among five types of skinfold calipers. *Human Biology* **62**, 369–388.

Sloan AW & Shapiro M (1972) A comparison of skinfold measurements with three standard calipers. *Human Biology* **44**, 29–36.

Smith WDF, Cunningham DA, Paterson DH & Koval JJ (1995) Body mass indices and skeletal size in 394 Canadians aged 55–86 years. *Annals of Human Biology* **22**, 305–314.

Sonnenschein EG, Kim MY, Pasternack S & Toniolo PG (1993) Sources of variability in waist and hip measurements in middle-aged women. *American Journal of Epidemiology* **138**, 301–309.

Strickland SS & Tuffrey VR (1997) *Form and Function. A Study of Nutrition, Adaptation and Social Inequality in Three Gurung Villages of the Nepal Himalayas*. London: Smith-Gordon.

Strickland SS, Tuffrey VR, Gurung GM & Ulijaszek SJ (1993) *Micro-economic Consequences of Physique and Physical Performance in South Asia. Overseas Development Administration Project Report* R4568. London: Overseas Development Administration.

Tomkins AM (1994) Growth monitoring, screening and surveillance in developing countries. In *Anthropometry: the Individual and the Population*, pp. 108–116 [SJ Ulijaszek and CGN Mascie-Taylor, editors]. Cambridge: Cambridge University Press.

Ulijaszek SJ (1997) Anthropometric measures. In *Design Concepts in Nutritional Epidemiology*, pp. 289–311 [BM Margetts and M Nelson, editors]. Oxford: Oxford University Press.

Ulijaszek SJ & Lourie JA (1994) Intra- and inter-observer error in anthropometric measurement. In *Anthropometry: the Individual and the Population*, pp. 30–55 [SJ Ulijaszek and CGN Mascie-Taylor, editors]. Cambridge: Cambridge University Press.

Ulijaszek SJ & Mascie-Taylor CGN (editors) (1994) *Anthropometry: the Individual and the Population*. Cambridge: Cambridge University Press.

Voss LD, Bailey BJR, Cumming K, Wilkin TJ & Betts PR (1990) The reliability of height measurement (the Wessex growth study). *Archives of Disease in Childhood* **65**, 1340–1344.

Ward R & Anderson G (1993) Examination of the skinfold compressibility and skinfold thickness relationship. *American Journal of Human Biology* **5**, 541–548.

Weiner JS & Lourie JA (1981) *Practical Human Biology*. London: Academic Press.

Wellens RE (1989) Activity as a temperamental trait: relationship to physique, energy expenditure and physical activity habits in young adults. PhD Thesis, University of Texas.

Williamson DF, Kahn HS, Worthman CM, Burnette JC & Russell CM (1993) Precision of recumbent anthropometry. *American Journal of Human Biology* **15**, 159–167.

Wong WW, Cochran WJ, Klish WJ, Smith EO, Lee LS, Fiorotto ML & Klein PD (1988) Body fat in normal adults estimated by oxygen-18- and deuterium-dilution and by anthropometry: a comparison. *European Journal of Clinical Nutrition* **42**, 233–242.

World Health Organization (1995) *Physical Status: the Use and Interpretation of Anthropometry. Report of a WHO Expert Committee. WHO Technical Report Series* 854. Geneva: WHO.

Zavaleta AN & Malina RM (1982) Growth and body composition of Mexican-American boys 9 through 14 years of age. *American Journal of Physical Anthropology* **57**, 261–271.

Zerfas AJ (1985) *Checking Continuous Measures: Manual for Anthropometry*. Los Angeles, CA: Division of Epidemiology, School of Public Health, University of California, Los Angeles.

Zillikens MC & Conway JM (1990) Anthropometry in blacks: applicability of generalised skinfold equations and differences in fat patterning between blacks and whites. *American Journal of Clinical Nutrition* **52**, 45–51.