Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference

Mercedes de Onis, Adelheid W. Onyango, Jan Van den Broeck, Wm. Cameron Chumlea, and Reynaldo Martorell, for the WHO Multicentre Growth Reference Study Group

Abstract

Thorough training, continuous standardization, and close monitoring of the adherence to measurement procedures during data collection are essential for minimizing random error and bias in multicenter studies. Rigorous anthropometry and data collection protocols were used in the WHO Multicentre Growth Reference Study to ensure high data quality. After the initial training and standardization, study teams participated in standardization sessions every two months for a continuous assessment of the precision and accuracy of their measurements. Once a year the teams were restandardized against the WHO lead anthropometrist, who observed their measurement techniques and retrained any deviating observers. Robust and precise equipment was selected and adapted for field use. The anthropometrists worked in pairs, taking measurements independently, and repeating measurements that exceeded preset maximum allowable differences. Ongoing central and local monitoring identified anthropometrists deviating from standard procedures, and immediate corrective action was taken. The procedures described in this paper are a model for research settings.

Mercedes de Onis and Adelheid W. Onyango are affiliated with the Department of Nutrition, World Health Organization, Geneva. Jan Van den Broeck is affiliated with the Africa Centre for Health and Population Studies, Mtubatuba, South Africa. Wm. Cameron Chumlea is affiliated with the Department of Community Health, Lifespan Health Research Center, Wright State University School of Medicine, Kettering, Ohio, USA. Reynaldo Martorell is affiliated with the Department of International Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA.

Please direct queries to: Mercedes de Onis, Study Coordinator, Department of Nutrition, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland. Telephone: 41-22-791 3320; fax: 41-22-791 4156; e-mail: deonism@who.int.

Members of the WHO Multicentre Growth Reference Study Group and Acknowledgments are listed at the end of the first paper in this supplement (pp. S13–S14).

Mention of the names of firms and commercial products does not imply endorsement by the United Nations University.

Key words: Anthropometry, growth curves, growth references, height, length, methods, skinfold, weight

Introduction

The World Health Organization (WHO) Multicentre Growth Reference Study (MGRS) was undertaken to generate new growth curves for assessing the growth and development of infants and young children from around the world. The children included in the study came from six countries: Brazil, Ghana, India, Norway, Oman, and the United States. The methodology and eligibility criteria for the study have been described elsewhere in this supplement [1]. Identical, rigorous data collection procedures were followed in all sites in order to minimize measurement error and to avoid systematic differences among sites.

Variability in infant and child measurements can result from a number of influences: the setting in which the measurements are taken; stomach and bladder volume (in the case of weight); diurnal variation (in length/height); the behavior and cooperation of the child being measured; the accuracy and precision of the instruments; the anthropometrist's technical capacity (training, experience, and reliability), fitness, and mood; and the methods of data recording (reading, writing down). Appropriate training and continued standardization, adherence to specified methods and procedures, and monitoring of data quality are essential to reduce measurement error and minimize bias in multisite studies. The purpose of this article is to describe the measurement protocols and routine standardization sessions that were used in the MGRS. The study protocols and quality control procedures can be applied in research settings without substantially increasing costs or complicating logistics.

Selection and training of anthropometrists

The field staff collecting anthropometric data in the

MGRS (referred to herein as observers or anthropometrists) had to have at least secondary school education, be motivated, write legibly, speak the local language, and be able to interact appropriately with the high-socioeconomic-status families that were targeted for the study. All candidates received standardized training, and only those who met the MGRS performance criteria were retained for the study.

The measurement procedures and training guidelines were prepared by the MGRS Coordinating Centre at WHO in Geneva, based on best practices recommended in anthropometry manuals and in the literature [2–8]. The initial training of anthropometrists at each site was carried out by an experienced anthropometrist following the procedures detailed in the MGRS protocol. All anthropometrists were trained to interview mothers, complete the study questionnaires, measure children as described in the protocol, avoid digit preference or transposition of numbers, record measurement values immediately after reading them, and write legibly to reduce mistakes during data transfer. Strict adherence to the measuring techniques and recording procedures was emphasized. Instructions were also provided on handling uncooperative children, taking into account cultural factors and individual mothers' sensitivity to their babies' crying.

Early in the study, four anthropometrists were trained and standardized against an expert designated by WHO as lead anthropometrist for the MGRS (W.C.C.) in a cross-site session held in Rotterdam, the Netherlands. Two of the participants were study supervisors from sites, one was a member of the Coordinating Centre, and the fourth (J.V.dB.) became the second WHO-designated lead anthropometrist for the MGRS.

Following the initial training and before the start of data collection, the anthropometrists in each site were standardized against one of the two WHO lead anthropometrists. The anthropometrist with the best performance at this session was designated "local lead anthropometrist" and was responsible for retraining teammates who deviated from MGRS techniques, and for training newly recruited anthropometrists later in the study. A WHO lead anthropometrist visited each site annually to ensure that identical methods were followed throughout the seven years of the study. The measurement procedures followed in the MGRS were documented on videotape (available to readers on request from the first author) and viewed during training and regularly thereafter to reinforce the key features of the measurement protocols.

Standardization

Given the objectives of the MGRS, standardization within and among sites was a key aspect of the study [1]. An important goal of standardized training is to enable observers to measure accurately, that is, without bias. To achieve this, observers need to be trained to obtain measurements that are on average equal to the values measured by an expert anthropometrist who is considered the "gold standard." The degree of accuracy can be assessed in a test–retest study in which several children are measured by both the expert and the observer, and bias is calculated as the average deviation of the observer's mean measurement values from those of the expert.

It is equally important that the measurements taken be precise, that is, reproducible. High precision is possible only if measurement procedures are highly standardized. Precision is assessed on the basis of differences between replicate measurements taken on several subjects in the test–retest study. The most commonly used parameter for lack of precision is the technical error of measurement (TEM) [9].

Following the initial standardization session and throughout the data collection phase, each site conducted standardization sessions bimonthly (every two months) that coincided once a year with the visit of the WHO lead anthropometrist. The purpose of these sessions was to identify anthropometrists deviating from the MGRS procedures. Corrective actions, such as retraining, were taken whenever deviations in measurement techniques were noted.

The initial standardization session used groups of 20 children for each set of measurements and took five or six days to complete, whereas the bimonthly sessions required only 10 children and could be accomplished in two or three days. At the initial session, the observers were standardized against the WHO lead anthropometrist, who served as the gold standard, whereas the bimonthly sessions used the observers' overall mean of each anthropometric variable as the gold standard. The longitudinal screening and follow-up teams were standardized separately because of the different age groups and settings involved: the screening teams measured newborns in maternity wards, whereas the follow-up teams measured infants and older children during home visits.

Analyses of accuracy and precision were performed soon after the standardization sessions using a centrally prepared Excel spreadsheet program with standard formulas for calculating relevant statistics [9–14]. To illustrate how the observers' performance was assessed, table 1 presents length data from the Rotterdam session, in which 25 children participated. For precision (TEM), the observers' performance compared well with that of the lead anthropometrist and the overall mean. This demonstrated that the participants in the session followed consistent techniques in measuring length and obtained reproducible values. The sign test for precision assesses the "measurement effect," where an observer's retest measurements may be systematically lower or higher than his or her own first measurements

TABLE 1. Precision and accuracy from the standardization session in Rotterdam: length data

Variable	WHO lead anthropo-metrist	Observer 1	Observer 2	Observer 3	Observer 4	Overall mean
TEM[a] (cm)	0.34	0.48	0.33	0.39	0.35	0.38
F test						
Lead anthropometrist[b]	—	$.10 < p < .25$	$p > .25$	$p > .25$	$p > .25$	
Overall mean[b]	$p > .25$	$.10 < p < .25$	$p > .25$	$p > .25$	$p > .25$	
Sign test[c]	$p > .05$	$p > .05$	$p > .05$	$p > .05$	$p > .05$	
Bias (cm)						
Lead anthropometrist[d]	—	−0.49	−0.21	−0.15	−0.15	
F test[e]	—	$p < .01$	$.01 < p < .05$	$.05 < p < .10$	$p < .01$	
Sign test[f]	—	$p < .05$	$p > .05$	$p > .05$	$p > .05$	
Bias (cm)						
Overall mean[d]	0.21	−0.33	−0.00	0.08	0.07	
F test[e]	$.10 < p < .25$	$p > .25$	$p > .25$	$p > .25$	$p > .25$	
Sign test[f]	$p > .05$	$p < .05$	$p > .05$	$p > .05$	$p > .05$	

a. Technical error of measurement: $\sqrt{(\sum d_i^2/2n)}$; where d_i is the difference between the ith subject's test and retest measurements by the observer and n is the number of measured subjects.

b. F ratio for precision: Observer $\sum d_i^2$/Lead anthropometrist $\sum d_i^2$. When overall mean is the gold standard, d_i in the denominator is the difference between the ith subject's overall mean of test and overall mean of retest measurements.

c. Precision sign test: binomial proportion p, where $p = x/n$, and x is the frequency of the observer's retest scores that are higher (or lower) than the corresponding test scores. Significance is based on exact confidence limits for proportions when $n \leq 75$ (see Table B.11 in Daly and Bourke [10]).

d. Average bias: Observer $\sum \Delta_i/n$; where Δ_i is the difference between the observer's mean and the lead anthropometrist's (or overall) mean measurement for the ith subject.

e. F ratio for bias: Observer $\sum \Delta_i^2$/lead anthropometrist's or overall means') $\sum d_i^2$ (same denominator as the precision F ratio).

f. Bias sign test: binomial proportion p, where $p = x/n$, and x is the frequency of the observer's means that are above (or below) the lead anthropometrist's or overall mean. Significance is based on exact confidence limits for proportions when $n \leq 75$ (see Table B.11 in Daly and Bourke [10]).

[15]. No such measurement effect was evident for any participant in this session. For accuracy, the observers showed a systematic tendency toward negative bias compared with the lead anthropometrist; consequently, the techniques for measuring length were reviewed. As expected, the negative bias was not evident when compared with the overall mean, except for observer 1. Both the F test and the sign test for accuracy are useful. The sign test checks whether poor accuracy results from systematic or occasional bias [10, 15]. For example, the average bias could be low and nonsignificant when a large deviation overwhelms smaller but systematic differences. In this case, the sign test, but not the F test, would indicate bias. For this session, only one observer's bias was systematic, and this was corrected by retraining.

The results of the bimonthly standardization sessions were sent to the Coordinating Centre soon after their completion. The average TEMs for each site were plotted to monitor overall performance over time, as figure 1 illustrates (for length). In general, the TEMs were highest at the start, and following a pattern that is consistent for all the other measurements, precision improved as the observers gained experience. Once stabilized, the average TEMs remained low, reflecting the high precision of the measurements taken by the study teams. When sending the bimonthly results to the Coordinating Centre, sites reported on extraneous circumstances that affected the observed performance. For example, figure 1 shows a peak in TEM for the eighth bimonthly session in Brazil, when the children involved were particularly uncooperative. On the rare occasions when a problem identified in the sites needed external assistance, the WHO lead anthropometrist visited the affected site to retrain the observers. This was the case for triceps skinfold measurements at one site.

Anthropometric procedures

Measuring equipment

All study sites used the same measuring equipment. The instruments needed to be highly accurate and precise, yet sturdy and portable enough to be carried back and forth on home visits.

Length was measured with the Harpenden Infantometer (range, 30–110 cm for portable use, with digit counter readings precise to 1 mm). Because the MGRS protocol specified measuring length in the cross-sectional study for children aged 18 to 30 months (to allow a precise estimation of the systematic difference between length and height), a longer-than-usual infantometer was specially built for the study.

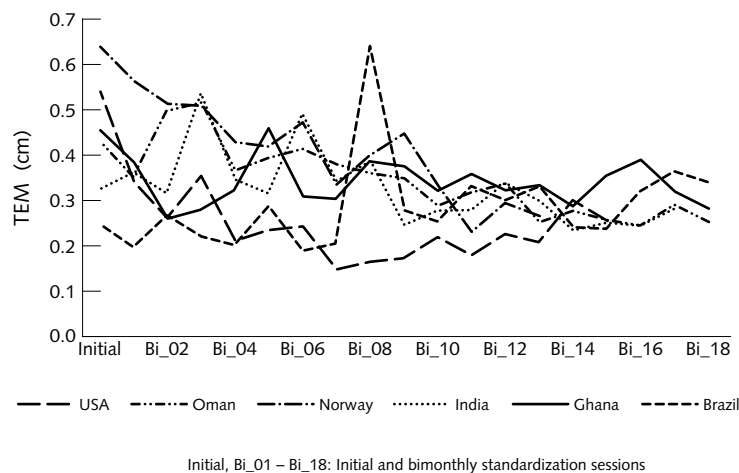Initial, Bi_01 – Bi_18: Initial and bimonthly standardization sessions

FIG. 1. Technical error of measurement (TEM) for length at initial session and up to 18 bimonthly (every two months) standardization sessions in the study sites

The Harpenden Portable Stadiometer (range, 65–206 cm, digit counter reading) was used to measure both adult and child heights. At the request of WHO, the manufacturer designed a wooden base to replace the heavy carrying case that serves as a mount for the traditional portable stadiometer. This adaptation decreased the weight of the packaged stadiometer by about 7 kg and reduced the time required to assemble it.

A self-retracting, 0.7-cm-wide, flat metal tape with blank lead-in strip (range, 0–200 cm, calibrated to 1 mm) was used to measure circumferences. Metal tapes were chosen because they are more robust and accurate and stay in a single plane around the head. They were replaced on a regular basis when the grading marks faded. The Holtain/Tanner-Whitehouse Skinfold Caliper (jaw face area, 35 mm^2; pressure between the jaws, 10 ± 2 g/mm^2; range, 0–40 mm; calibrated to 0.2 mm) was used to measure skinfolds.

To measure weight, we used portable electronic scales that have taring capability and are calibrated to 0.1 kg (UNICEF Electronic Scale 890 or Uniscale). Ideally, newborns should be measured with a scale of higher precision (within 10 g). However, the advantages of the Uniscale greatly outweighed the disadvantage of its lower precision for young babies. The scales were satisfactorily pilot tested in the Brazilian site; they were easy to use and transport, and tared weighing allowed the infants to remain in their mothers' arms where they were more calm and relaxed. This was important for the mothers' positive perception of the study and, thus, participation. The scale's electronic display decreased the observer measurement error. In cold climates, the infants could be wrapped up in a blanket for weighing after the weight of the blanket had been tared. Another advantage of the Uniscale was that it allowed the mother's weight to be recorded at each visit, thus permitting the collection, at no extra cost, of weight

data for lactating women.

The equipment was calibrated regularly, usually daily before the home or hospital visits. The scales were calibrated with locally available standard weights over the full weight range, and tared weighing was simulated. The infantometer and stadiometer were calibrated by using metal rods of known lengths. The skinfold calipers, being particularly fragile, were checked before each use with calibration blocks of various widths for accuracy and to ensure that the needle moved smoothly and continuously with the opening of the caliper jaws.

Anthropometric data collection

Measurements were taken and recorded by two trained and standardized anthropometrists. Both the questionnaire forms and the standard procedures were designed to ensure that each observer read and recorded measurements independently of the other. At each session, the two exchanged roles as "leading" and "assisting" observers. The role of the assisting observer was to help position the child correctly while the leading observer took and recorded measurements. The first observer measured and immediately recorded each of the measurements, and they then exchanged roles so that the second observer could also take the full set of measurements. They then compared their values to ensure that duplicate measurements were within the maximum allowed differences. Any measurements falling outside the maximum allowed differences were repeated by both observers and entered in designated boxes on the data recording sheet. No more than two remeasurements were allowed (i.e., a maximum of three duplicate measurement sets for a given anthropometric parameter at any one visit). All recorded measurements were entered into the computer. The final value to be

used for the construction of the growth curves will be the average of the last pair of measurements. In the rare cases (< 0.1%) when the child was judged to be too agitated for reliable duplicate measurements to be taken, only one set of measurements was recorded. In practice, it was observed that large differences owing to reading or recording errors were resolved by a first repeat measurement. However, when the babies were uncooperative, measurements became increasingly difficult, and hence the decision to discontinue measuring and use unpaired measurements in the few cited cases.

The maximum allowable differences for acceptable precision used in the study for the various anthropometric variables were based on the TEM obtained in the initial standardization session conducted at the Brazilian site. To achieve a rate of remeasurement of around 5%, the maximum allowed differences were set at 2.8 times the TEM achieved during the session, i.e., 7 mm for length, 5 mm for circumferences, and 1.2 mm for skinfolds (table 2). The maximum allowable difference for weight was set at 100 g to allow for rounding off within the smallest calibration unit of the scale. Because skinfold thicknesses were the measurements with which mothers and children were least familiar and felt most uncomfortable, the decision was taken to raise the maximum allowable difference for skinfolds to 2 mm. This was considered to be a more appropriate limit, as a narrower margin might lead to too many repeat measurements, with negative implications for the anthropometrists' morale and the participants' responsiveness.

Measurement schedule

The MGRS anthropometric measurements are weight, recumbent length, standing height, head and arm circumferences, and triceps and subscapular skinfold thickness. For the longitudinal study, newborns were measured at birth (usually within the first 12 hours of life, and never after 24 hours) and visited at home 21 times: at weeks 1, 2, 4, and 6; monthly from 2 to 12 months; and every other month during the second year (table 3). Data collection was more frequent at younger ages so that these early phases of rapid growth could be adequately described. The week 1 visit was done by the lactation counselor, and only weight was measured, following the standard MGRS procedure (using the Uniscale and weighing the baby twice). The mother's weight was recorded at each visit, and the father's weight and both parents' heights were measured once.

In the cross-sectional study, children aged 18 to 71 months were measured once, except in the two sites that used a mixed-longitudinal design [1], in which some children were measured two or three times, at 3-month intervals. All children aged 18 to 30 months had both recumbent length and standing height measured, and parental weights and heights were measured once.

Concerted efforts were made to collect the anthropometric data on scheduled visit dates. Theoretically, the maximum delay or advance of measurements allowed by the protocol was 10% of the child's age (e.g., 3 days at 1 month, 18 days at 6 months), but in practice, teams worked with more restricted tolerable delay or advance targets (0, 1, 2, 4, and 5 days for visits at weeks 1, 2, 4, and 6 and at 2 months, respectively; 7 days for visits taking place at 3 months onwards). Of more than 32,000 home visits completed by April 2003, only 217 (0.7%) were done outside the maximum allowable delay, out of these, 58 (26.7%) exceeded the limit by less than one day.

Measurement techniques

A comprehensive description of the techniques used for the measurements is found in the MGRS Measurement and Standardization Protocols and documented in the anthropometric training video (available on request from the first author). The anthropometrists explained to the mothers all procedures to be undertaken and emphasized that these were harmless. Infants and young children were held by their mothers to foster a sense of

TABLE 2. Maximum allowable differences between the measurements of two observers

Measurement	Brazil TEM[a] from pilot study	Maximum allowable difference
Weight	Not available	100 g
Length	2.5	7.0 mm
Head circumference	1.4	5.0 mm
Arm circumference	1.8	5.0 mm
Triceps skinfold	0.44	2.0 mm
Subscapular skinfold	0.43	2.0 mm

a. TEM, Technical error of measurement (see formula in footnote to table 1).

TABLE 3. Time schedule for the collection of anthropometric measurements in the longitudinal study

Measurement and time frame	Frequency	No. of visits
Weight,[a] length, head circumference		
Birth	Once	1
2–6 wk	Every 2 wk	4
2–12 mo	Monthly	10
14–24 mo	Every 2 mo	6
Arm circumference, skinfold thickness (triceps, subscapular)		
3–12 mo	Monthly	10
14–24 mo	Every 2 mo	6

a. Weight was also measured at week 1 by the lactation counselor.

security for the baby. The anthropometrist's confidence and poise was important for reassuring both mother and child, and included maintaining eye contact and talking to the child in a calm, reassuring voice.

Arm circumference and skinfold measurements were taken on the left side of the body. The choice of which side to measure (right or left) matters little to accuracy and precision [6]; however, the left-hand side is used more often. Length, height, circumferences, and skinfolds were recorded to the last completed unit rather than the nearest unit. To correct for the systematic negative bias introduced by this practice, half of the smallest measurement unit (i.e., half of 0.2 mm for skinfolds and half of 0.1 mm for circumferences) was added to each measurement before analysis. This correction did not apply to weight, which was rounded off to the nearest 100 g.

For measurement of weight, the mother removed all the child's clothes, but as noted earlier, use of a blanket to cover the baby was encouraged in cold weather. The parents took off their shoes, heavy clothing, and other heavy objects before being weighed. They wore light clothing of known weight that was recorded and later subtracted from the subject's weight. This was done by using a list of weights of local clothes. In the longitudinal study, the mother was weighed first, and after her weight was recorded, the scale was tared and the baby was given to her. She was asked to stand still until the baby's weight had been displayed and recorded. When children could not be undressed, they also wore standard light clothing of known weight that was recorded and subtracted to obtain the child's weight. Children aged two years or more in the cross-sectional study were weighed on their own, standing with their feet slightly apart in the center of the platform of the scale.

To measure recumbent length, braids were undone and hair ornaments were removed if they interfered with positioning of the head. Diapers were also removed, because they made it difficult to hold the infant's legs together and straighten them. The leading observer stood on one side of the board to hold down the baby's legs with one hand and move the foot board with the other. The assisting observer stood at the headboard to help position the child's head. The head was positioned so that the crown touched the headboard and a vertical line from the ear canal to the lower border of the eye socket was perpendicular to the horizontal board (i.e., the Frankfort plane positioned vertically). The leading observer positioned the child's shoulders and hips at right angles to the long axis of the body. Gentle pressure was applied to the knees to straighten the legs. To avoid causing injury, minimal but prolonged pressure was applied to the knees of newborns. To take the measurement, the footboard was positioned against the child's feet with the soles flat on the board and the toes pointing upwards. The measurement was recorded to the last completed 1 mm.

To measure standing height, hair ornaments were removed and braids were undone. The child stood on the stadiometer with bare feet placed slightly apart and the back of the head, shoulder blades, buttocks, calves, and heels touching the vertical board. The assisting observer held the child's knees and ankles to keep the legs straight and the feet flat. The leading observer got down to a face-to-face level with the child and positioned the child's head so that a horizontal line drawn from the ear canal to the lower edge of the eye socket ran parallel to the baseboard (i.e., the Frankfort plane positioned horizontally). Because young children have difficulty standing to full stature, a gentle push applied to the tummy was used to help them stand to full height. The headboard was pulled down to rest firmly on top of the head and compress the hair, and the reading was taken to the last completed 1 mm.

To measure head circumference, hairpins or headbands were removed and braids were undone. An infant or child below the age of two years was held on the mother's lap, and older children could stand or sit unassisted. The leading observer stood or sat at the left side of the child, passed the tape around the head, and anchored it just above the eyebrows and over the fullest protuberance of the skull at the back of the head. The assisting observer stood or sat in front of the child and helped by positioning the tape correctly on the side away from the lead observer. Once positioned correctly, the tape was pulled tight to compress the hair and skin, and the reading was recorded to the last completed 1 mm.

The mid-upper-arm point is half the distance between the acromion process (the most lateral bony protuberance of the back of the shoulder) and the olecranon (the bony structure that stands out when the elbow is bent). The midpoint was located and marked for measurement of the mid-upper-arm circumference (MUAC) and triceps skinfold thickness. One observer palpated the shoulder to find the acromion and marked it with a felt-tip pen or cosmetic pencil. The child's forearm was then bent 90° at the elbow, palm facing up, so that the olecranon stood out at the elbow. The observer placed the zero point of the tape on the mark over the acromion process and ran it downward along the back of the arm to the tip of the elbow. The other observer made a small horizontal mark at the midpoint on the posterior aspect of the arm before the tape was removed.

For measurement of the MUAC, the child's arm hung in a relaxed position or was held in the extended position by the assisting observer; care was taken not to flex or tighten the muscles. The tape was then wrapped around the arm over the marked midpoint. The tape had to lie flat around the arm, without compressing the skin or underlying tissue; the assisting observer checked to ensure that there was no gap or compression on the inner part of the arm before the measurement was

recorded to the last completed 1 mm.

A skinfold consists of a double fold of skin and subcutaneous fat, excluding the underlying muscle. The teams were trained to grasp the skinfold gently to avoid causing unnecessary discomfort to the child and compressing the fat. Skinfolds were recorded to the last completed 0.2 mm. For measurement of triceps skinfold thickness, young babies were held by their mothers; older children sat or stood on their own. The left arm hung relaxed at the side or was held down by the mother or assisting observer. The leading observer stood behind the child and picked up the skinfold about 1 cm above the midpoint mark over the triceps muscle, with the fold running downward along the midline of the back upper arm. The caliper jaws were applied at right angles to the "neck" of the fold just below the finger and thumb over the midpoint mark. While maintaining a grip on the skinfold, the observer gently released the caliper handles and allowed the jaws to close on the fat fold for two seconds before taking the reading to the last completed 0.2 mm.

The measurement point for the subscapular skinfold located immediately below the inferior angle of the scapula was identified by palpating and marking the inferior angle of the scapula. The child stood or sat with shoulders relaxed or gently held down to prevent movement of the scapula. The skinfold was picked up 1 cm above and medial to the subscapular mark, and the caliper was applied to the "neck" of the fold over the mark so that the fold ran diagonally down toward the left elbow. The same procedure as described for the triceps skinfold was followed to read and record the measurement.

Quality control during data collection

The observers' performance was monitored in several ways during the study:

The requirement to take and record all measure-

ments independently by the two observers and to compare their measurement values for maximum allowable differences was a key strategy for detecting errors and remeasuring the child on the spot.

The proportion of repeated measurements at each site was closely monitored as an important quality control measure. Low levels of remeasurement signal a possible lack of independence between the observers, whereas high levels might indicate poor measurement techniques on the part of at least one of the observers. The levels of maximum allowable differences selected anticipated repeat rates of about 5%. The observed rates according to site are reported for newborns (fig. 2), young children in the longitudinal study (fig. 3), and older children in the cross-sectional study (fig. 4). For the overall study, the rate of repeated length measurements in the cross-sectional sample was 5%, as expected, but it was double this percentage in the longitudinal study (11%)(table 4). The lowest proportions of repeat measurements were observed for the skinfolds (3% for triceps and 1% for subscapular
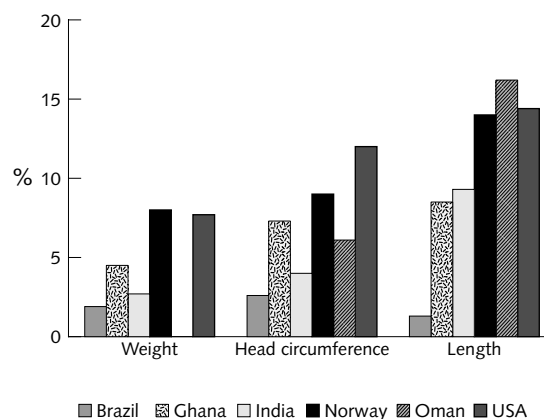

FIG. 2. Percentage of newborn measurements repeated for exceeding the maximum allowable difference between observers
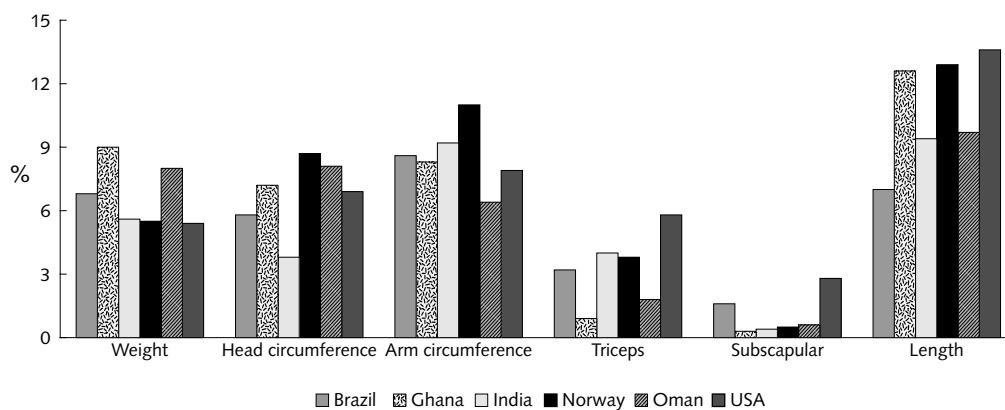

FIG. 3. Percentage of longitudinal follow-up measurements repeated for exceeding the maximum allowable difference between observers
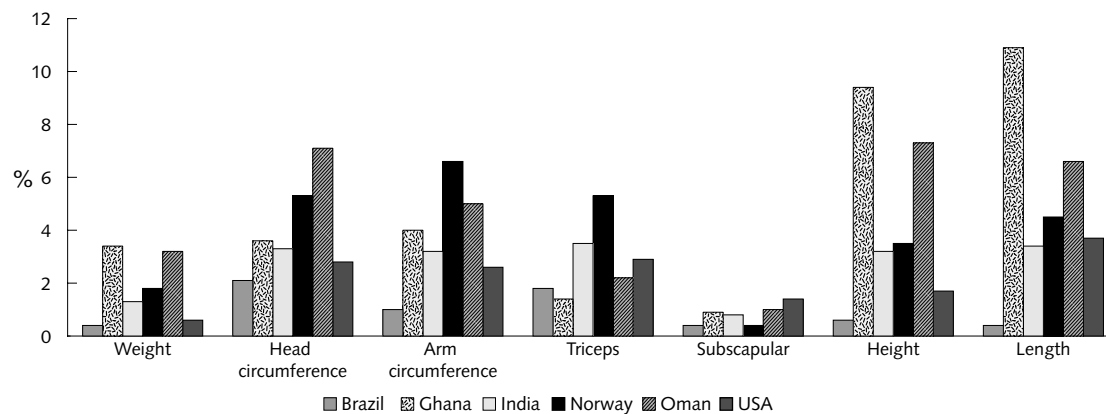
FIG. 4. Percentage of cross-sectional study measurements repeated for exceeding the maximum allowable difference between observers

TABLE 4. Summary of measurements repeated for exceeding the maximum allowable difference between observers

| Measurement | No. (%) of measurements | | |
	Newborns (n = 1,746)	Longitudinal study (n = 31,248)	Cross-sectional study (n = 8,254)
Weight	69 (4.0)	2,194 (6.8)	148 (1.8)
Length	180 (10.8)	3,450 (10.7)	81 (4.9)[a]
Head circumference	114 (6.5)	2,173 (6.8)	335 (4.1)
Arm circumference	N/A[b]	2,761 (8.6)	309 (3.7)
Triceps skinfold	N/A	982 (3.1)	236 (2.9)
Subscapular skinfold	N/A	300 (0.9)	69 (0.8)
Height	N/A	N/A	354 (4.3)

a. n = 1,653: only children aged 18–30 months in the cross-sectional study were measured for length.
b. N/A, Indices not measured.

skinfolds in both longitudinal and cross-sectional components), probably as a result of the adoption of a wider margin of allowable differences than the initial standard set in Brazil (table 4).

The completed questionnaires were delivered soon after the home visit, usually within one or two days, to the local coordination center, where they were checked by the supervisor for completeness and consistency using procedures described elsewhere in this supplement [16]. For anthropometry, the data entry system included built-in range and consistency checks that flagged measurements exceeding ±2 standard deviations of age- and sex-specific reference values for attained size. Flagged values were then checked for consistency between the two observers, consistency with other anthropometric variables measured on the same visit, consistency with previous measurements of the same child, and possible data entry errors.

Periodic computer checks were also done for each observer to detect digit preferences and unusual values. For example, because the skinfold caliper reads to 0.2-mm units, there should be no odd decimal values (e.g., 0.1 mm, 0.3 mm) recorded for skinfolds. Table 5 is a sample digit preference table for triceps skinfold measurements taken by one site team. The output from this analysis was examined for terminal digit preference and avoidance. According to table 5, observer 1 tended to avoid digit 6, but there was no pattern to suggest bias in observed proportions of the other digits. On the other hand, for observer 2, the proportions of digit 0 (8.4%) versus 2 (34.4%) suggested a tendency to overestimate measurements. When the imbalance between two consecutive digits was particularly large, the differences between measurement pairs were analyzed to determine whether the affected observer was biased in relation to others that had been paired with him or her.

Quality control checks were performed by randomly calling approximately 10% of the mothers to repeat a selection of the questions on the study forms and to ask the mother whether the child had been measured twice by the interviewers. These calls also provided the opportunity to monitor participant responsiveness and satisfaction with the study.

Bimonthly (every two months) standardization sessions served to ensure that the observers were not departing from the measuring techniques of the study,

TABLE 5. Sample table of terminal digit preference analysis in longitudinal follow-up study (triceps skinfold data from Oman)

Observer	No. of measure-ments	Terminal digit % (95% confidence interval)					Probability of equal proportions
		0	2	4	6	8	
1	773	22.6 (19.7, 25.6)	21.6 (18.7, 24.5)	20.6 (17.7, 23.4)	14.2 (11.8, 16.7)	21.0 (18.1, 23.8)	0.002
2	1,051	8.4 (6.7, 10.0)	34.4 (31.6, 37.3)	20.9 (18.5, 23.4)	16.8 (14.6, 19.1)	19.4 (17.0, 21.8)	< 0.0001
3	866	19.7 (17.1, 22.4)	25.1 (22.2, 27.9)	22.4 (19.6, 25.2)	6.6 (4.9, 8.2)	26.2 (23.3, 29.1)	< 0.0001
4	996	23.0 (20.4, 25.6)	20.2 (17.7, 22.7)	23.7 (21.1, 26.3)	15.5 (13.2, 17.7)	17.7 (15.3, 20.0)	< 0.0001
5	839	16.5 (13.9, 19.0)	20.5 (17.8, 23.2)	22.2 (19.4, 25.0)	19.2 (16.5, 21.9)	21.7 (18.9, 24.5)	0.065
6	702	16.1 (13.4, 18.8)	20.2 (17.3, 23.2)	18.0 (15.1, 20.8)	21.9 (18.9, 25.0)	23.8 (20.6, 26.9)	0.01
7	1,123	13.7 (11.7, 15.7)	23.2 (20.7, 25.6)	26.0 (23.4, 28.6)	14.0 (12.0, 16.0)	23.2 (20.7, 25.6)	< 0.0001
8	785	29.9 (26.7, 33.1)	18.6 (15.9, 21.3)	18.5 (15.8, 21.2)	15.8 (13.2, 18.4)	17.2 (14.6, 19.8)	< 0.0001
9	657	19.8 (16.7, 22.8)	20.1 (17.0, 23.2)	21.2 (18.0, 24.3)	16.4 (13.6, 19.3)	22.5 (19.3, 25.7)	0.1514

to monitor precision and accuracy, and to corrective measures (e.g., retraining) when required.

To maintain a good rapport with the families, each participant in the longitudinal study had one "fixed" fieldworker for the duration of the follow-up. The other fieldworkers were rotated every two months in order to distribute error terms, avoid boredom, and prevent complicity that might undermine the measurement protocol.

Discussion

The rigorous anthropometric protocols described in this paper were set in place to ensure high data quality. These MGRS procedures serve as a model for research settings. The methods and procedures reviewed will be applicable to multi- and single-site studies. It will not be possible to be as rigorous in nonresearch settings, such as child clinics. At the very least, the procedures should be carefully documented in training manuals, staff members collecting anthropometric data should be trained and refresher sessions should be held period-ically, weighing scales and any other instruments used should be maintained in good order and calibrated before use, and fieldworkers should be supervised.

The standardization sessions were effective in identi-fying factors that contribute to low accuracy and preci-sion in anthropometric measurements. Training and retraining opportunities were available to help keep the anthropometrists' skills sharp, as were printed and videotaped reference materials. These were particularly useful when reserve staff were preparing to take part

in data collection and when new team members were recruited in the course of the study. In general, new staff began taking anthropometric measurements for the MGRS only after being standardized against the WHO lead anthropometrist.

Factors that affected measurement accuracy and precision included the identification of landmark fea-tures when measuring soft tissues (arm circumference and skinfolds). In some sites, the teams experienced difficulties in taking measurements because they did not mark the upper-arm midpoint or the subscapular point. In this respect, the Coordinating Centre's ongo-ing monitoring of anthropometric data and the regular participation of the WHO lead anthropometrists in site standardization sessions were extremely important for detecting and correcting problems.

For research and programmatic activities, it is relevant to note that that the child's age could affect the precision of some measurements, judging by the differences in repeat rates for arm circumference (9% versus 4%) and head circumference (7% versus 4%) in the longitudinal and cross-sectional studies, respectively. Users who adopt the same limits of maximum allowable differences between independently recorded duplicate measurements could evaluate performance with the MGRS-observed proportions as a reference. Thus, for children below the age of two years, about 11% of length measurement pairs will differ by more than 7 mm. In the same age group, 7% of duplicate head circumferences will differ by more than 5 mm, as will 9% of duplicate arm circumference measurements. Overall, the rates of repeated measurement are expected to be lower in older children, who tend to be

calmer and more cooperative. A team that exceeds these proportions may be in need of further training, and a team that has substantially lower rates may be taking nonindependent measurements.

References

1. de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J, for the WHO Multicentre Growth Reference Study Group. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. Food Nutr Bull 2004;25(1)(suppl 1):S15–26.
2. Lohman TG, Roche AF, Martorell R, eds. Anthropometric standardization reference manual. Champaign, Ill, USA: Human Kinetics Books, 1988.
3. Cameron N. Anthropometric measurements. In: Cameron N, ed. The measurement of human growth. London: Croom Helm, 1984:56–99.
4. Gordon CC, Cameron Chumlea W, Roche AF. Stature, recumbent length, and weight. In: Lohman TG, Roche AF, Martorell R, eds. Anthropometric standardization reference manual. Champaign, Ill, USA: Human Kinetics Books, 1991:3–8.
5. Habicht J-P, Yarbrough C, Martorell R. Anthropometric field methods: criteria for selection. In: Alfin-Slater RB, Kritchevsky D, series eds; Jelliffe DB, Jelliffe EFP, volume eds. Human growth—a comprehensive treatise. Vol 2. Nutrition and growth. New York: Plenum Publishing Company, 1979:365–87.
6. Martorell R, Mendoza F, Mueller WH, Pawson IG. Which side to measure: right or left? In: Lohman TG, Roche AF, Martorell R, eds. Anthropometric standardization reference manual. Champaign, Ill, USA: Human Kinetics Books, 1988:87–91.
7. Tanner JM, Hiernaux J, Jarman S. Growth and physique studies. In: Weiner JS, Lourie JA, eds. Human biology. A guide to field methods. IBP Handbook No. 9. Oxford, UK: Blackwell Scientific Publications, 1969:1–42.
8. World Health Organization. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. Technical Report Series No. 854. Geneva: World Health Organization, 1995.
9. Malina RM, Hamill PVV, Lemeshow S. Selected measurements of children 6–11 years. Vital Health and Statistics Series 11, No. 123, USDHHS. Washington, DC: US Government Printing Office, 1973.
10. Daly LE, Bourke GJ. Interpretation and uses of medical statistics, 5th ed. Oxford, UK: Blackwell Science, 2000.
11. Johnson TS, Engstrom JL, Gelhar DK. Intra- and interexaminer reliability of anthropometric measurements of term infants. J Pediatr Gastroenterol Nutr 1997;24:497–505.
12. Marks GC, Habicht JP, Mueller WH. Reliability, dependability, and precision of anthropometric measurements. Am J Epidemiol 1989;130:578–87.
13. Martorell R, Habicht JP, Yarbrough C, Guzman G, Klein RE. The identification and evaluation of measurement variability in the anthropometry of preschool children. Am J Phys Anthropol 1975;43:347–52.
14. Villar J, Kestler E, Pareja G. Measurement error in clinical perinatal data. Am J Obstet Gynecol 1989;160:380–2.
15. World Health Organization. Measuring change in nutritional status. Annex 1: Standardization procedures for the collection of weight and height data in the field. Geneva: World Health Organization, 1983.
16. Onyango AW, Pinol AJ, de Onis M, for the WHO Multicentre Growth Reference Study Group. Managing data for a multicountry longitudinal study: experience from the WHO Multicentre Growth Reference Study. Food Nutr Bull 2004;25(1)(suppl 1):S46–52.