Published in final edited form as:

Am J Clin Nutr. 2022 July 06; 116(1): 97–110. doi:10.1093/ajcn/nqac064.

Accuracy of a handheld 3D imaging system for child anthropometric measurements in population-based household surveys and surveillance platforms: an effectiveness validation study in Guatemala, Kenya, and China

Karim Bougma[1], **Zuguo Mei**[2], **Mireya Palmieri**[3], **Dickens Onyango**[4], **Jianmeng Liu**[5], **Karla Mesarina**[3], **Victor Akelo**[6], **Rael Mwando**[4], **Yubao Zhou**[5], **Ying Meng**[5], **Maria Elena Jefferds**[2]

[1]Centers for Disease Control and Prevention Foundation, Atlanta, GA, USA

[2]Centers for Disease Control and Prevention, Atlanta, GA, USA

[3]Nutrition and Micronutrients Unit, Institute of Nutrition of Central America and Panama (INCAP), Guatemala City, Guatemala

[4]Kisumu County Department of Health, Kisumu, Kenya

[5]Institute of Reproductive and Child Health, Peking University, Peking, China;

[6]Office of the Director, Center for Global Health, Centers for Disease Control and Prevention, Kisumu, Kenya

Abstract

Background: An efficacy evaluation of the AutoAnthro system to measure child (0–59 months) anthropometry in the United States found 3D imaging performed as well as gold-standard manual measurements for biological plausibility and precision.

Objectives: We conducted an effectiveness evaluation of the accuracy of the AutoAnthro system to measure 0- to 59-month-old children's anthropometry in population-based surveys and surveillance systems in households in Guatemala and Kenya and in hospitals in China.

Methods: The evaluation was done using health or nutrition surveillance system platforms among 600 children aged 0–59 months (Guatemala and Kenya) and 300 children aged 0–23 months (China). Field team anthropometrists and their assistants collected manual and scan anthropometric measurements, including length or height, midupper arm circumference (MUAC), and head circumference (HC; China only), from each child. An anthropometry expert and assistant later collected both manual and scan anthropometric measurements on the same child. The expert manual measurements were considered the standard compared to field team scans.

Address correspondence to KB: kbougma@cdc.gov; kbougma@gmail.com.

The authors' responsibilities were as follows—MEJ and ZM: designed the evaluation; MEJ, ZM, and KB: trained the field teams; KB: carried out the primary analyses; ZM: completed data checks; MP, DO, JL, KM, VA, RM, YZ, and YM: supervised data collection; MEJ and KB: wrote the first draft of the manuscript; and all authors: interpreted the data, edited the report, and read and approved the final manuscript.

Author disclosures: The authors report no conflicts of interest.

Supplemental Figures 1–7 and Supplemental Table 1–8 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

Results: Overall, in Guatemala, Kenya, and China, for interrater accuracy, the average biases for length or height were −0.3 cm, −1.9 cm, and −6.2 cm, respectively; for MUAC were 0.9 cm, 1.2 cm, and −0.8 cm, respectively; and for HC was 2.4 cm in China. The inter-technical errors of measurement (inter-TEMs) for length or height were 2.8 cm, 3.4 cm, 5.5 cm, respectively; for MUAC were 1.1 cm, 1.5 cm, and 1.0 cm, respectively; and for HC was 2.8 cm in China. For intrarater precision, the absolute mean difference and intra-TEM (interrater, intramethod TEM) were 0.1 cm for all countries for all manual measurements. For scans, overall, absolute mean differences for length or height were 0.4–0.6 cm; for MUAC were 0.1–0.1 cm; and for HC was 0.4 cm. For the intra-TEM, length or height was 0.5 cm in Guatemala and China and 0.7 cm in Kenya, and other measurements were ≤0.3 cm.

Conclusions: Understanding the factors that cause the many poor scan results and how to correct them will be needed prior to using this instrument in routine, population-based survey and surveillance systems.

Keywords

children; anthropometry; 3D imaging system; accuracy; precision; validation

Introduction

Anthropometry is used to assess the prevalence of malnutrition, to identify at-risk populations or individuals, to monitor changes in nutritional status over time, and to evaluate impacts of interventions (1). In population-based household surveys and surveillance systems, anthropometry usually has been manually performed by trained personnel using dedicated instruments, such as weighing scales, tapes, calipers, and stadiometers (1). However, the manual anthropometric measurements may be subject to errors from equipment malfunction and/or human errors, even with well-trained measurers (2–4). These errors affect the data quality and may have negative consequences for screening, program evaluation, policy-making, accountability, advocacy, and global reporting. The WHO, UNICEF, and others have developed multiple resources and tools to support high-quality, manual anthropometric data collection (5, 6), but errors may still occur with manual methods despite the application of these tools and resources. Various technologies exist to automatically capture anthropometric dimensions; interest in 3D technologies focused on the ability to capture data indirectly without touching the person and the potential for faster data capture compared to manual methods (7, 8). Nonportable, expensive 3D technologies exist that document better performance than manual methods, primarily in studies of adults (9, 10). To alleviate the challenges related to manual anthropometric measurements, low-cost, portable, automatic body measurement systems that use 3D body scan data (11) have emerged as a potential new approach for anthropometry assessments in population-based surveys (12). An efficacy evaluation study conducted under ideal, controlled conditions was carried out on the AutoAnthro system developed by Body Surface Translations, Inc. (BST) among children 0–59 months in different facilities in metro Atlanta, Georgia. The efficacy results found that 3D imaging performed as well as gold-standard manual measurements for child anthropometry for biological plausibility and precision, but there was systematic bias compared to manual measurements (13); thereafter, BST adjusted the system to address the systematic bias identified in the efficacy study.

The primary objectives of the study described in this publication were to carry out an effectiveness evaluation under usual, real-world conditions of the accuracy and precision of the AutoAnthro imaging system in measuring young children's length or height, midupper arm circumference (MUAC), and head circumference (HC) in population-based household surveys and surveillance systems in Guatemala, Kenya, and China (HC measurements were only done in China). The study also assessed the experiences and acceptability of caregivers of the measured children and the field teams that collected the measurements. These results will be reported elsewhere.

Methods

Study design, settings, and participants

The effectiveness evaluation study was carried out through surveys in Guatemala, Kenya, and China. In all 3 countries, the evaluation was done using health or nutrition surveillance system platforms. For Guatemala and China, the evaluation was embedded within routine data collection procedures, whereas Kenya carried out a stand-alone, population-based anthropometric survey using the surveillance system sampling frame to determine the prevalence of malnutrition among children ≤59.9 months in the population catchment area. While the survey was designed to be representative of children in the catchment area, the effectiveness evaluation was not designed to collect data representative of the populations, and evaluation data collection stopped upon reaching required sample sizes, which might be less than the target survey or surveillance system sample size.

The surveillance system in Guatemala is an annual, nationally representative, cross-sectional, continuous household health and nutrition survey that uses multistage sampling to select 100 enumeration areas and 30 households within each enumeration area. Anthropometric measurements for children usually include weight and length or height. The MUAC is not routinely measured but was added for the evaluation. All children ≤59.9 months in the surveyed households were invited to participate in the evaluation.

In Kenya, the study was conducted within a health and demographic surveillance system located in an urban informal settlement that is primarily used for surveillance of under-5 mortality. The county-based health surveillance system carries out a census and survey of all households in the catchment area every 6 months. The most recent census of villages and households was used to carry out 2 stages of sampling where clusters within villages randomly selected children ≤59.9 months, who were invited to participate. Weight, length or height, and MUAC were collected.

In China, the data collection platform was not a population-based household survey; instead, data collection occurred at 2 county maternal and child health (MCH) hospitals that carry out routine growth monitoring and physical examinations and follow-up visits to children <6 years. The children under 2 years old were recruited when they came for these routine health services. Weight and length or height for all children are routinely measured; HC is measured only for children <1 year old but was expanded to those <2 years old for this evaluation. MUAC is not routinely measured but was added for the evaluation. We originally planned to conduct the evaluation only in Guatemala and Kenya, but the China site was opportunistically added last to include data from Asia. Funding only allowed for data collection among children <24 months through MCH hospitals.

Sample size

For power calculations, we assumed an α of 0.05, β of 0.1, and power (1-β) of 0.9 (90%). For a paired *t*-test to detect a 0.2 cm difference with 90% power, the minimum sample size is 263. Accounting for potential nonresponses, the sample size designed for each country was 300 children per age group (<24.0 months and 24.0–59.9 months), with a total of 600 children per evaluation site in Guatemala and Kenya and 300 children in China.

Prepilot among 100 children in Guatemala and software adjustments

In May 2018, the Centers for Disease Control and Prevention Foundation (CDCF) paid BST to travel to Atlanta to train CDC and CDCF staff on use of the AutoAnthro system over 2 days. The CDC and CDCF then traveled to Guatemala for approximately 2 weeks in May 2018 to carry out training and pilot testing for the collection of manual and scan anthropometry data among 50 children <24 months and 50 children 24 to 59.9 months in the department of Huehuetenango to prepare for the main effectiveness evaluation. The 100-child evaluation data collection was integrated into the Guatemala surveillance platform and occurred over approximately 1 month. As agreed in the Bill and Melinda Gates Foundation (BMGF) funding to CDCF, BST observed the training and pilot testing, helped troubleshoot use of the AutoAnthro system, and later was given access to the manual and scan anthropometry data from the 100 children. The data from the 100-child pilot are not included in this paper.

Under separate grants awarded to BST by the BMGF, BST was expected to use the final manual and scan data collected from the 100 Guatemalan children to resolve the bias identified in the Conkle et al. (13) study. The revision to the software took approximately 2 months to complete, and BST considered the AutoAnthro system ready for the main effectiveness evaluation by August 2018; thereafter, the CDC and CDCF independently started the main evaluation training in Guatemala in late August 2018. The CDCF paid BST to process the scans collected for the main evaluation in the 3 countries reported in this paper. After BST had already provided the final scan results for each country, in September 2019 BST identified problems with their software and reprocessed the evaluation data with the updated software. The revised data were received mid-November 2019 and used for the analyses reported in this paper. BST did not have access to the manual anthropometry data or analyses reported in this study prior to publication.

Training

The CDC and CDCF carried out all evaluation training of field teams, including on traditional manual anthropometry, use of the AutoAnthro system, administering the caregiver acceptability questionnaire, filling out the anthropometrist questionnaire, obtaining informed consent, collecting the date of birth of the child for estimating age, and recording the date of data collection.

A manual was developed for use by field teams on the traditional manual anthropometry methods, and BST provided the CDC and CDCF with a manual on the use of the AutoAnthro system. The training included theoretical classroom training; practical, hands-on training for manual and scan methods; and standardization exercises with children <5 years (China <2 years) for manual anthropometry methods. The standardization data were analyzed using a Microsoft Excel spreadsheet from the Micronutrient Survey Toolkit (14). Classroom training covered all the different aspects of measuring anthropometry and data collection procedures to conduct the effectiveness evaluation component [manual anthropometric measurement procedures following WHO guidance (15, 16); communication management during measurements; referral procedures; validating calibration of stadiometers, measuring tapes, and scales; manual anthropometric standardization tests; AutoAnthro data capture and review of scans; administering a caregiver acceptability questionnaire; informed consent and sampling].

Selection criteria for the "expert" anthropometrist and assistant included receipt of prior anthropometric measurement training; previously experience measuring anthropometry data in population-based surveys or surveillance systems; and meeting all required levels of precision and accuracy through a manual standardization exercise conducted as part of this evaluation. These 3 criteria are based on the WHO growth standard study (16). To assess the acceptability of scan captures, an acceptability questionnaire was administered to caregivers after observing data collection by manual and scan methods. In Guatemala and Kenya only, the expert also had to show acceptable performance administering the acceptability questionnaire to caregivers in order to be selected.

The field team make-ups were as follows: Guatemala had 15 team members (1 expert and 9 field anthropometrists; 5 assistants); Kenya had 8 (1 expert anthropometrist+1 assistant; 3 field anthropometrists + 3 assistants); and China had 6 (1 expert anthropometrist+1 assistant; 2 field anthropometrists+2 assistants). In Guatemala, 1 expert anthropometrist was assisted by 2 assistants, each at a different time during the survey, and 9 field anthropometrists were assisted by 3 supervisors. Kenya and China had independent, 2-person teams.

Data collection

3D imaging.—The AutoAnthro system, developed by BST, uses off-the-shelf hardware and a manufacturer-provided Software Development Kit installed on a smartphone or tablet to assess anthropometry among children 0–59.9 months. This evaluation used an iPad tablet and a Structure Sensor 3D camera (Occipital) that is connected to the tablet and was the same equipment used in the efficacy study (13). The system did not produce weight measurements or immediate results. All measurements were collected by an anthropometrist with the help of an assistant, and in some cases also the caregivers.

Each data collection team had their own AutoAnthro system that was charged nightly with electricity. Teams used a foam mat approximately 1 meter by 1 meter to create a plain on the ground, or on top of a table or bed, for the AutoAnthro software to identify the area to measure. Anthropometrists stood approximately 3 to 6 feet away from the child, depending on child age and size, and pointed the tablet and camera sensor at the child. On the screen, a cubic box appeared and the anthropometrist changed the size of the box until the child was fully within the cubic frame. Children <24 months were measured lying down on the mat with their arms and legs spread so they were not touching the body. Children 24 months or older were measured standing up on the mat with their legs spread apart and arms held away from the body. They were instructed to place their arms and legs apart and hold their arms in 3 specific positions. Six scans were taken of the front of the child and another 6 scans of the back. Children had to remain still during the scanning process with little movement. If they were unable to remain still or in the correct position, then the assistant and caregiver would hold the child's finger tips and toes to keep the child still, and keep their arms and legs spread, while simultaneously keeping their own body out of the scan as much as possible. BST later removed the assistant and caregiver from the scans during scan processing. The possibility of touching children varied from the efficacy study, where children were not touched during scan capture (13).

The child had to wear minimal clothing (e.g., diaper, underwear, no socks) or tight clothing (e.g., leggings, shorts, T-shirt with sleeves rolled up over shoulders, or tank top, all of which could be wrapped tightly around the child's body). Hats were removed but otherwise no other hair adjustments were needed. Although the AutoAnthro System does not create a photographic image, for ethical reasons all children had to wear a diaper, underwear, or shorts for the scans. Teams had new, packaged clothing to give to families in cases when the family did not have suitable clothing. Scans had to be collected away from bright natural light (e.g., in shaded area or indoors) and away from walls and furniture. For the 12 scans, anthropometrists were able to check each scan for a few seconds before the initial scan image disappeared from the screen, and could decide whether it was acceptable or should be deleted and redone (e.g., body part of child not in cubic frame or too much movement). Once uploaded to the cloud, supervisors reviewed the scans for quality to provide feedback to the field teams. Extra AutoAnthro systems and tablets were available to field teams, if needed, for malfunction. BST processed the measurements over a period of several weeks or months, depending on the country.

Manual anthropometry.—New manual anthropometry equipment was purchased for each country evaluation, and each data collection team had their own full set of equipment. All manual anthropometric measurements were collected by a measurer and an assistant, and this team completed all measurements on 1 child before starting another. Shorr Boards were used to collect recumbent length for children <24 months or standing height for children 24 months or older to the nearest 0.1 cm, following WHO guidance (15, 16). The MUAC and HC were measured to the nearest 0.1 cm using QM2000 Circumference Measuring Tape (QuickMedical), following WHO guidance (15, 16). Weight was measured to the nearest 50 grams using a Scale 874 DR (Seca). Stadiometers and scales were placed on sturdy, flat surfaces with adequate light for reading the measurements. Caregivers were seated and held younger children for the MUAC and HC measurements. Children wore minimal, light clothing with no socks, and hair accessories were removed. Sleeves were rolled up, if necessary, so they did not interfere with MUAC measurements. Polyvinyl chloride (PVC) pipe and known weights were provided to each team for daily validation of the calibration of the stadiometer (PVC pipe), measuring tape (PVC pipe), and scales (known weights). Extra tapes were purchased, and teams routinely replaced their tapes when they showed evidence of wear or failed the validation test; stadiometer latches were repaired if they loosened or fell off. Data collection teams noted after each manual measurement whether the child was calm, crying, or actively fighting or resisting during each measurement.

Collection of anthropometric data.—Field team anthropometrists and their assistants collected from the same child the manual anthropometric measurements, including weight, length or height, MUAC, and HC (China only) and scans for measuring length or height, MUAC, and HC. Manual anthropometric measurements were collected 2 times each for each child consecutively. If differences between the first and second manual anthropometric measurements exceeded WHO guidance (15, 16), then they were carried out a third time and the 2 closest measurements were used in the analysis. The field teams also collected 1 session of 12 scans (6 front and 6 back) on the same child. Following the same protocols as the field teams, later that day or the next day (China and Kenya) or up to 7 days later (Guatemala), an anthropometry expert and assistant collected both manual anthropometric measurements and scans on the same child as the field teams. We decided to include expert team data collected in Guatemala within 7 days from the field teams based on a review of differences between manual measurements collected within 3 days and within 7 days that varied little (data not shown), and to minimize the loss of sample size (removing 79 at >3 days compared with 4 at >7 days). Supplemental Figure 1A–C shows the flowcharts for enrollment and final sample sizes in Guatemala, Kenya, and China.

Two steps were taken to minimize potential bias arising from the order of measurements and method of measurements. Manual anthropometric measurements were always collected following the order of *1*) length or height; *2*) MUAC; and *3*) HC. Each scan simultaneously generates length or height, MUAC, and HC measurements. For each country, the order of manual compared with scan measurements was switched either every day or for every cluster so that either manual anthropometric measurements were collected first and the scans second, or the scans were collected first and the manual anthropometric measurements second.

Data management and analysis.—The data were analyzed using SAS 9.4 (SAS Institute, Inc.). The main variables for the evaluation were length or height and MUAC in Guatemala and Kenya, and length or height, MUAC, and HC in China. BST processed the 12 scans for each child using a single anatomic model [in the efficacy study (13), 2 models were used: 1 for children less than 1 month old and another for children 1 to 59 months] and provided 2 sets of measurements for length or height, MUAC, and HC per child using different scans for each measurement calculation. That is, per child, 6 scans were used to calculate a single height measurement and the other 6 scans were used to create the second height measurement. Analyses were conducted separately for each country. While this evaluation in Guatemala and Kenya was embedded in surveys to assess the population prevalence of malnutrition, this was not the design of the effectiveness evaluation, so our analysis does not account for complex survey design because it does not attempt to draw inferences to make generalizations about the population. Data quality tests examined the percentages of missing data by age and sex, end-digit preference of the first measurement, and biologically implausible values (BIV) and SDs for each measurement and method (5). We examined the BIV and SDs using the weights measured by the expert, as well as the anthropometrist teams. Overall, the differences in the indices computed using either weight measurement were minor for the weight-for-length *z*-score (WLZ), weight-for-height *z*-score (WHZ), and BMI-for-age *z*-score (BMIZ). Therefore, the analysis used the weight measured by each team: that is, the weight measured by the expert for the expert results and the weight measured by the field teams for the field results.

For the accuracy and precision assessments, we used the mean of the 2 manual measurements and the mean of the 2 scan measurements. To assess the agreement between methods, we compared the manual measurements from the expert team with the scan measurements from the field teams. The expert anthropometrist team values collected using manual methods with a stadiometer and measuring tape served as the "true" reference value. Accuracy is defined as the closeness of measurements to the true reference value (17); the difference between the scan measurements from the team of field anthropometrists and the true reference value from the manual measurements from the expert anthropometrist (scan – manual) is referred to as the bias. Precision is defined as the closeness of 2 measurements to each other. An intrarater difference is defined as an observer's (expert or field anthropometrist) first measurement compared to the same observer's second measurement, and an interrater difference is defined as the field anthropometrist's measurement compared to the expert measurement. An intramethod difference is defined as a manual measurement compared to a manual measurement and a scan measurement compared to a scan measurement. An intermethod difference is defined as a manual measurement compared to a scan measurement.

For our primary objective to examine how close the mean scan measurements by field teams are to the true reference value of the mean manual measurements of the expert teams, we assessed interrater, intermethod accuracy (expert manual measurements compared with field scan measurements). We calculated the mean, SD, absolute mean difference, inter-TEM [interrater, intermethod technical error of measurement (TEM)] and relative inter-TEM (18), intraclass correlation (ICC) (19, 20), and the agreement based on the Bland-Altman method (21, 22). The inter-TEM is an indicator of accuracy, and the TEM examines the SDs between repeated measurements obtained from the 2 methods (manual and scan) on the same subjects. A lower inter-TEM reflects better accuracy, and we considered an acceptable TEM threshold to be a 0.7-cm difference for length or height measurements (23). The relative inter-TEM is a percentage that takes into consideration the denominator (mean of the measurements), and a value below 1.5% (expert) to 2.5% (novice) for interrater comparisons indicates better accuracy (24, 25, 26). The ICC assesses accuracy and precision, and values range from 0 to 1, with values close to 1.0 reflecting low error (26), and values <0.5, 0.5–0.75, 0.75–0.9, and >0.9 reflecting poor, moderate, good, and excellent accuracy, respectively (20). When the ICC is 0.0, there is no variance, and the CI cannot be calculated. Bland-Altman plots examine the differences and agreement between measurements from 2 methods, with 95% of the observations falling within ±1.96 (upper and lower LOAs) of the mean difference (21, 22). We used a Pitman test to examine the

Author Manuscript
Author Manuscript
Author Manuscript

correlation between accuracy and age (or size) of the child (24) and a Levene test of variance homogeneity to examine the SDs of the differences between methods (27). In addition to the widely used ICC, we examined the concordance correlation coefficient (CCC) (10, 28), which does not have ANOVA assumptions, unlike the ICC (29). The mean absolute difference is the average of the absolute difference between two measurements from the same subject.

To assess whether the difference in time between when the field teams and the expert teams collected evaluation data (from minutes to up to 7 days apart) influenced accuracy, we examined intrarater, intermethod agreement (mean of manual measurements from the expert teams compared with mean of scan measurements from the expert teams, and mean of manual measurements from the field teams compared with mean of scan measurements from the field teams). We calculated the mean, SD, absolute mean difference, intra-TEM (interrater, intramethod TEM), relative intra-TEM, and ICC for the interrater, intramethod precision.

To examine repeatability and how close the first manual measurement was to the second manual measurement and how close the first scan measurement was to the second scan measurement, we examined the intrarater, intramethod precision agreement within each method (the repeated manual measurements from the expert team for manual precision and the repeated scan measurements from the field teams for scan precision). We calculated the mean, SD, absolute mean difference, intra-TEM, relative intra-TEM, and ICC.

Each child was measured manually twice, consecutively, by a field team and later by the expert team. To minimize the potential for precision bias, we examined interrater, intramethod comparisons of the mean manual measurements from the field teams to the mean manual measurements from the expert teams, and the 2 scan measurements from the field teams to the 2 scan measurements from the expert teams. We calculated the mean, SD, absolute mean difference, intra-TEM, relative intra-TEM, and ICC for the interrater, intramethod precision.

Last, we excluded BIV and calculated the prevalences of stunting and wasting (5) by expert manual and field scan measurements.

Ethical approval was obtained for the evaluation from the following institutions for the data collection occurring in each respective country: Ministry of Public Health and Social Assistance National Health Ethics Committee, Guatemala; Maseno University Ethics Review Committee, Kenya; and Peking University Institutional Review Board, China. The CDC determined this evaluation was research involving human subjects with no CDC investigators.

Results

Descriptive

The sample characteristics are presented in Table 1 and Supplemental Table 1. The order of the method of measurement was alternated; manual measurements were first in 42% to

56% of children across countries. The proportion of children reported as calm during manual measurements varied from 51% to 86%, depending on the measurement, across countries.

Data quality control

Missing data.—The results for missing data are presented in Supplemental Table 2 by age and sex, and for length or height, MUAC, and HC by age and sex. Overall across countries, the results showed little or no missing age or sex data for manual or scan data and little or no missing manual data for length or height, MUAC, or HC. While overall scans from China had little missing anthropometric data, field team data in Guatemala and Kenya were missing 6.0% and 8.1%, respectively, of anthropometric scan measures.

Digital preference.—The results for terminal digit preferences of the first manual and scan measurements are presented in Supplemental Table 3 for length or height, MUAC, and HC for each expert team and the average of the field teams. The expectation is that each terminal digit value should be close to 10%, and deviation above or below suggests rounding. The results overall and by age groupings showed more statistically significant ($P < 0.05$) evidence of a terminal digit preference for the manual measurements compared with the scan measurements.

BIVs.—The results for BIVs are presented in Supplemental Table 4 for length-for-age *z*-scores (LAZs) and height-for-age *z*-scores (HAZs), WLZs and WHZs, MUAC-for-age *z*-scores (MUACZs), HC-for-age *z*-scores (HCZs), and BMIZs. A WHO expert committee recommended that BIV of 1.0% or higher reflect data quality problems (1). Across countries, manual measurement BIV were rare and only exceeded the 1.0% threshold 2 times among an age grouping in Guatemala by the expert team. With few exceptions, overall and age grouping results showed that when BIV occurred there were fewer for the manual measurements compared with the scan measurements, and the BIV were highest among children <24 months. Overall, there were no manual HC BIV, while scan BIV ranged from 12.3% to 28.9% across teams.

SDs of z-scores.—The SDs of the *z*-scores are presented in Supplemental Table 5 for the LAZ, HAZ, WLZ, WHZ, MUACZ, BMIZ, and HCZ. WHO recommendations expect a SD of 1.0, and values larger than 1.5 SD suggest poor anthropometry quality (1). The SDs for MUACZs for both manual and scan measurements, and the SDs for manual measurements of WLZ, WHZ, HCZ, and BMIZ, were generally around 1.0 or less across countries. Overall, across countries the SDs from scan measurements are larger than the SDs of manual measurements (e.g., LAZ and HAZ overall manual measurement SD ranges are 1.05–1.24 and scan measurement SD ranges are 1.57–4.71), and many SD scan values are higher than 1.5 (e.g., HCZ overall SD range, 1.64–2.87). Across countries, the SD scans were highest in the youngest age groupings, except for HCZ scans for Guatemala and Kenya.

Accuracy

Interrater and intermethod agreement.—The interrater and intermethod agreement compared the expert manual measurements with field scan measurements. The mean measurements, mean absolute differences, absolute and relative inter-TEMs, and ICCs for length or height, MUAC, and HC are presented in Table 2, and the biases, Bland-Altman upper and lower limits of agreement (LOAs), and Pitman test *P* values for the measurements are in Table 3 and Supplemental Figure 2 for the Bland-Altman LOAs. The results in Tables 2 and 3, presented by 6-month or 12-month age intervals, are in Supplemental Tables 6 and 7. The CCC results are in Supplemental Figures 5–7.

For length or height, the mean absolute differences (mean expert manual—mean field scan measurements) were 2.9 cm (SD, 2.8 cm) in Guatemala, 3.5 cm (SD, 3.2 cm) in Kenya, and 6.6 cm (SD, 4.0 cm) in China. The overall absolute inter-TEMs were above the acceptable TEM threshold (<0.7 cm), at 2.8 cm in Guatemala, 3.4 cm in Kenya, and 5.5 cm in China; the overall relative inter-TEMs were also above acceptable thresholds, at 3.5% in Guatemala, 4.0% in Kenya, and 7.5% in China. For height, relative inter-TEMs were below the novice threshold for children >24 months. In Guatemala and Kenya, the values for children <24 months were higher than those for older children for absolute mean (SD) differences and absolute and relative inter-TEMs. Overall, the ICCs were 0.58 in Guatemala (moderate), 0.47 in Kenya (poor), and 0.00 in China (poor). In Guatemala and Kenya, the ICC values for children <24 months were <0.1 (poor), while those for older children were around 0.8 (good; Table 2). The overall average bias was −0.3 cm in Guatemala, −1.9 cm in Kenya, and −6.2 cm in China (Table 3). For Guatemala, the overall length or height Bland-Altman lower and upper LOAs were −8.2 cm and 7.6 cm, respectively; for Kenya they were −10.6 cm and 6.9 cm, respectively; and for China they were −15.4 cm and 3.0 cm, respectively (Supplemental Figure 2). The results for Kenya and China overall and for children in Kenya <24 months had variances that were significantly different. The Pitman test was significant for all measurements across countries, except for length or height in Guatemala (Table 3). We then used the Levene test of variance homogeneity to test the SDs of the differences between methods, which had a *P* value < 0.05. Therefore, the difference in the SDs could not be ruled out as the reason for the inconsistency of the accuracy across age groups.

For MUAC, the mean absolute differences were 1.2 cm (SD, 0.9 cm) in Guatemala, 1.7 cm (SD, 1.3 cm) in Kenya, and 1.0 cm (SD, 0.9 cm) in China (Table 2). We considered a 0.5-cm difference an acceptable TEM threshold for MUAC measurements (23). The overall absolute inter-TEMs and relative inter-TEMs were 2 to 3 times or more over the acceptable thresholds (TEM <0.5 cm; novice relative inter-TEM <2.5%) across countries. Overall, the ICCs were poor across countries, at 0.13 in Guatemala, 0.00 in Kenya, and 0.00 in China. The overall average biases were 0.9 cm in Guatemala, 1.2 cm in Kenya, and −0.8 cm in China (Table 3). The overall MUAC Bland-Altman lower and upper LOAs for Guatemala were −1.6 cm and 3.3 cm, respectively; for Kenya were −2.3 cm and 4.7, respectively; and for China were −3.0 cm and 1.4 cm, respectively (Supplemental Figure 3), and the variance was significantly different both overall and by <24-month ≥24-month age groups.

For HC, in China, the mean absolute difference was 3.1 cm (SD, 2.5 cm). We considered a 0.5-cm difference an acceptable TEM threshold for HC measurements (23). The overall absolute and relative inter-TEMs were 2.8 cm and 6.1%, respectively, and the ICC was 0.0 (Table 2). The overall average bias was 2.4 cm (Table 3); the overall HC Bland-Altman lower and upper LOAs were −3.9 cm and 8.7 cm, respectively (Supplemental Figure 4); and the variance was statistically significant.

The differences in overall stunting and wasting prevalences between methods ranged from 8 to 50 percentage points and <1 to 5 percentage points, respectively (Supplemental Table 8).

Intrarater and intermethod agreement.—The intrarater, intermethod results which compared the expert manual measurements with expert scan measurements and field manual measurements with field scan measurements (Table 4) are similar to the comparisons between the expert manual measurements and the field scan measurements (interrater, intermethod), and suggest the differences in time between data collection by the 2 teams in each country (from minutes to 7 days apart) do not explain the differences in results by each method.

Precision

Intrarater, intramethod precision.—The intrarater, intramethod precision compared the expert manual first measurement with expert manual second measurement, the field scan first measurement with field scan second measurement. Across countries and age groups, the absolute mean difference between first and second manual measurements was 0.1 cm, and the absolute mean differences ranged between 0.1 and 0.8 cm for scan measurements (Table 5). The intrarater TEM was low, at 0.1 cm, for all manual measurements across countries. The intrarater TEMs were similar (0.1–0.2 cm) for scan-derived MUAC measurements, but the intrarater TEMs for length or height and HC were higher (less precise) for scan-derived measurements compared with manual measurements; however, all length or height relative intra-TEM and HC intra-TEM and relative intra-TEM measurements were within acceptable thresholds. The ICCs for manual measurements and scan measurements met the criteria for excellent.

Interrater, intramethod precision.—We also examined the interrater, intramethod results to assess the agreement of the mean measurements by method and team (expert mean manual measurements compared with field mean manual measurements; expert scan measurements compared with field scan measurements). Across countries and age groups, the absolute mean differences between the means of the expert measurements and the means of the field measurements ranged from 0.3 cm to 0.8 cm for manual measurements and from 0.3 cm to 3.9 cm for scan measurements (Table 6). Across all countries and measurements, the intrarater TEMs were lower (higher precision) for all manual measurements compared with scan measurements. For example, length or height intrarater TEMs ranged from 0.5 cm to 1.0 cm for manual measurements compared with 2.2 to 3.2 cm for scan measurements. The ICCs ranged from 0.91 to 0.99 for manual measurements (all excellent) and from 0.50 to 0.81 for scan measurements (moderate or good) among age groupings.

Discussion

Overall, this effectiveness evaluation of the use of AutoAnthro scanning technology to measure anthropometric measurements of children <60 months in Guatemala, Kenya, and China through population-based survey and surveillance systems found many poor results for anthropometric data quality, accuracy, and precision compared to the expert manual anthropometric measurements of length or height, MUAC, and HC, especially for children <24 months, but with some acceptable results among children ≥24 months. These results vary from the efficacy evaluation of the AutoAnthro system (13) and other evaluations of 3D technology collected in controlled settings using different 3D devices to measure anthropometry among various population groups, which found better accuracy and precision (9, 10, 30, 31).

Most anthropometrists were experienced with manual anthropometry methods, and for this evaluation had comprehensive classroom and practical training. Standardization exercises for scan measurements were not possible, as scan results were not immediately available. Based on the missing, BIV, and SD results, the manual anthropometry data were of high quality both overall and by <24-month and ≥24-month age groups. The scanned data for digit preference were of higher quality, but scans had a higher percentage of missing data, similar to other 3D technology studies (32). Multiple *z*-score indicators across countries for children <24 months had scan BIV above 1% or SDs >1.5, indicating data quality problems (1); however, except for HCZ, the scan BIV and SD data suggested good data quality for older children ≥24 months. It is unknown why the scan data quality was poor primarily for younger children. It was frequently necessary to have adults hold the child's hands and feet apart, and children often cried and struggled to move during scan measurements. From a practical and cost perspective, acceptable data quality is needed for children 0–59 months, not just for older children.

All anthropometrists carried out daily quality control checks to validate the calibration for stadiometers, measuring tapes, and weighing scales, and carried extra equipment to use as replacements when needed. There was no way to check the AutoAnthro scan results against known lengths or circumferences, because scan results were not available for several months. Other 3D evaluation studies have identified ways to validate the calibration when results are immediately available (32). During a scanning session, anthropometrists could quickly decide to discard a scan if they thought it was unacceptable based on their training. Although the anthropometrists thought the scans were suitable and accepted them, the data quality results suggest otherwise.

For precision, the intrarater, intramethod TEMs for manual measurements were more precise compared with the scans for length or height and HC, but were similar for MUAC. Among children ≥24 months, precision results were mostly acceptable for scans for all measurements. Compared with the efficacy study (13), the scan intrarater precision improved for MUAC measurements (0.4 cm efficacy compared with 0.1–0.2 cm across all countries), while the scans were less precise, with intrarater TEMs higher across countries compared to the efficacy study.

The interrater, intermethod accuracy results from this effectiveness study were generally poor both overall and for the <24-month and ≥24-month age groups, with some exceptions. For example, the overall inter-TEMs were 4 to almost 8 times beyond the threshold for length or height; 2 to 3 times the threshold for MUAC; and more than 5 times the threshold for HC. However, for children ≥24 months, relative inter-TEMs for height were acceptable below the novice threshold. The ICCs showed poor performance, except for height measurements in children ≥24 months in Guatemala and Kenya, which were categorized as good. The average bias was within WHO criteria (23) for overall and <24-month and ≥24-month length or height data for Guatemala and <24-month MUAC data for Guatemala and Kenya, but beyond thresholds for all other measurements across countries.

These poor results could be because high accuracy and precision are more challenging in effectiveness field settings due to contextual differences (e.g., more anthropometrists, less controlled setting). However, in all countries the teams were trained as typically done for anthropometry surveys, used a mat to confirm the scanning plain, and carried extra clothing in case the children's clothing was not suitable. In Guatemala and Kenya, the field teams usually were allowed inside households and moved furniture and closed window dressings to limit lighting and resolve furniture issues affecting scans. In China, the setting was more similar to controlled efficacy settings with dedicated rooms, and allowed team members to easily control the lighting, temperature, and equipment placement. There were more field anthropometrists and expert anthropometrists included in the effectiveness study in Guatemala and Kenya compared to the efficacy study (13), but the numbers of field anthropometrists (range, 4–12) in these evaluations were smaller than those used in many Demographic Health Surveys or Standardized Monitoring and Assessment of Relief and Transitions nutrition surveys (33, 34), for example.

Other potential factors that could influence the results in Guatemala and Kenya are the time differences between data collection by field and expert teams, as natural diurnal variation has been documented in the literature (35, 36), even though the same variations would influence both the manual and the scan measurements. The intrarater, intermethod accuracy results were similar to the interrater, intermethod results, suggesting differences of minutes, hours, or up to 7 days were unlikely to be the main causes of poor results.

The BST algorithm is proprietary; it is unknown whether changes after the efficacy study resolved the bias or improved accuracy and precision. The AutoAnthro training manual suggests 3 specific positions for the arms and legs (unpublished, BST manual), which were challenging to adhere to during fieldwork, and were often impossible with the youngest children. Body positioning and inconsistent scan posture were noted to effect scan results in other studies (37). The CDC and CDCF were informed by BST that a scan would be acceptable if the arms and legs were not touching the body and there was minimal movement, appropriate clothing, and adults only holding the fingers and toes, with the child's entire body inside the box on the tablet screen. It is unknown whether lack of adherence to these instructions in terms of positions, movement, clothing, or holding children affected the scan calculations. Landmarking software for data capture was noted as a potential source of error influencing the body dimensions captured and results (38). Touching children during scans was a variation from the efficacy study (13).

Data collection lasted from a few weeks to 7 months; it is unknown whether all evaluation staff retained fidelity to the training for the duration of fieldwork. Only 1 anthropometry expert team was used in each country evaluation, the teams alternated measuring first with the manual compared with the scan method, and the order of manual measurements was fixed. The similarities of findings across countries and age groups, despite differences in the contexts, strengthens our findings, as the contexts varied from a complex, comprehensive, multitopic survey that lasted several hours in each household (Guatemala) to faster, 1-component anthropometry surveys (household-based in Kenya and hospital-based in China). The China setting was more controlled and more similar to the efficacy study than the Guatemala or Kenya surveys, suggesting that differences between this evaluation and the efficacy study may not be based on the setting and controlled environments, compared with potentially the equipment or proprietary software changes. Overall, this effectiveness study found many poor results and suggests that information on the factors causing the poorer quality data and what steps could be taken to correct these factors is needed before using the AutoAnthro system in population-based survey and surveillance system settings to assess young child anthropometry.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the field teams and the parents and children who participated in the study. We are grateful to Yaw Addo and Rafael Flores-Ayala for statistical consultations, to Jose Aponte and John Copeland for EpiInfo software support, and to Usha Ramakrishnan for training on manual head circumference measurements.

The Bill and Melinda Gates Foundation funded the CDC Foundation, Atlanta, GA, to conduct the evaluation (grant number OPP1179307).

The funding bodies had no involvement in the design; the collection, analysis, or interpretation of data; the writing of the report; or the decision to submit the paper for publication. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC. The CDC does not directly or indirectly endorse any particular product.

Data Availability

Data described in the manuscript, code book, and analytic code will be made public and freely available without restriction at Dryad (https://datadryad.org/stash)).

Abbreviations used:

BIV	biologically implausible values
BMGF	Bill and Melinda Gates Foundation
BMIZ	BMI-for-age *z*-scores
BST	Body Surface Translations, Inc.
CCC	concordance correlation coefficient
CDCF	Centers for Disease Control and Prevention Foundation
HAZ	height-for-age *z*-score
HC	head circumference

HCZ	head-circumference-for-age *z*-score
ICC	intraclass correlation
inter-TEM	interrater or intrarater, intermethod technical error of measurement
intra-TEM	interrater or intrarater, intramethod technical error of measurement
LAZ	length-for-age *z*-score
LOA	limits of agreement
MCH	maternal and child health
MUAC	midupper arm circumference
MUACZ	midupper-arm-circumference-for-age *z*-score
PVC	polyvinyl chloride
TEM	technical error of measurement
WHZ	weight-for-height *z*-score
WLZ	weight-for-length *z*-score

References

1. WHO Expert Committee. Physical status: the use and interpretation of anthropometry. WHO Technical Report Series No 854 Geneva (Switzerland): World Health Organization; 1995.
2. Bubb H Challenges in the application of anthropometric measurements. Theor Issues Ergon Sci 2004;5(2):154–68.
3. Gordon CC, Bradtmiller B. Interobserver error in a large scale anthropometric survey. Am J Hum Biol 1992;4(2):253–63. [PubMed: 28524349]
4. Assaf S, Kothari MT, Pullum T. An assessment of the quality of DHS anthropometric data, 2005–2014. DHS Methodological Reports No 16. Rockville (MD): ICF International; 2015.
5. WHO, UNICEF. Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old Geneva (Switzerland): World Health Organization; 2019.
6. Perumal N, Namaste S, Qamar H, Aimone A, Bassani DG, Roth DE. Anthropometric data quality assessment in multisurvey studies of child growth. Am J Clin Nutr 2020;112(Suppl 2): 806S–15S. [PubMed: 32672330]
7. Nariño Lescay R, Alonso Becerra A, Hernández González A. Anthropometry. Comparative analysis of technologies for the capture of anthropometric dimensions. Revista EIA 2016:47–59.
8. Bartol K, Bojanić D, Petković T, Pribanić T. A review of body measurement using 3D scanning. IEEE Access 2021;9:67281–301.
9. Medina-Inojosa J, Somers VK, Ngwa T, Hinshaw L, Lopez-Jimenez F. Reliability of a 3D body scanner for anthropometric measurements of central obesity Obes Open Access 2016;2(3).
10. Koepke N, Zwahlen M, Wells JC, Bender N, Henneberg M, Ruhli FJ, Staub K. Comparison of 3D laser-based photonic scans and manual anthropometric measurements of body size and shape in a validation study of 123 young Swiss men. PeerJ 2017;5: e2980. [PubMed: 28289559]
11. Han H, Nam Y. Automatic body landmark identification for various body figures. Int J Ind Ergon 2011;41(6):592–606.
12. Zancanaro C, Milanese C, Lovato C, Sandri M, Giachetti A. Reliability of three-dimensional photonic scanner anthropometry performed by skilled and naïve operators. Int J Ergon 2015;5:2015.

13. Conkle J, Suchdev PS, Alexander E, Flores-Ayala R, Ramakrishnan U, Martorell R. Accuracy and reliability of a low-cost, handheld 3D imaging system for child anthropometry. PLoS One 2018;13(10):e0205320. [PubMed: 30356325]
14. Centers for Disease Control and Prevention, World Health Organization, Nutrition International, UNICEF. [Internet] Micronutrient survey manual Geneva: World Health Organization; 2020. Available from: https://mnsurvey.nutritionintl.org/.
15. WHO Multicentre Growth Reference Study Group. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development Geneva (Switzerland): World Health Organization; 2006.
16. WHO Multicentre Growth Reference Study Group, de Onis M. Reliability of anthropometric measurements in the WHO multicentre growth reference study. Acta Paediatrica 2006;450(95):38–46. [PubMed: 16817677]
17. Viviani C, Arezes PM, Bragança S, Molenbroek J, Dianat I, Castellucci HI. Accuracy, precision and reliability in anthropometric surveys for ergonomics purposes in adult working populations: A literature review. Int J Ind Ergon 2018;65:1–16.
18. Perini TA, Oliveira GLd, Ornellas JdS, Oliveira FPd. Technical error of measurement in anthropometry. Rev Bras Med Esporte 2005;11(1): 86–90.
19. Ayele B, Aemere A, Gebre T, Tadesse Z, Stoller NE, See CW, Yu SN, Gaynor BD, McCulloch CE, Porco TC, et al. Reliability of measurements performed by community-drawn anthropometrists from rural Ethiopia. PLoS One 2012;7(1):e30345–e. [PubMed: 22291939]
20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–63. [PubMed: 27330520]
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet North Am Ed 1986;327(8476):307–10.
22. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8(2):135–60. [PubMed: 10501650]
23. de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. Food Nutr Bull 2004;25(Suppl 1):S27–36. [PubMed: 15069917]
24. Olds T, Norton KI, Australian Sports Commission. Anthropometrica: a textbook of body measurement for sports and health courses Sydney (Australia): UNSW Press; 1996.
25. Goto R, Mascie-Taylor CGN. Precision of measurement as a component of human variation. J Physiol Anthropol 2007;26(2):253–6. [PubMed: 17435374]
26. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 1999;82(3):165–77. [PubMed: 10655963]
27. Gastwirth JL, Gel YR, Miao W. The impact of Levene’s test of equality of variances on statistical theory and practice. Stat Sci 2009;24(3): 343–60.
28. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45(1):255–68. [PubMed: 2720055]
29. Chen C-C, Barnhart HX. Comparison of ICC and CCC for assessing agreement for data without and with replications. Comp Stat Data Anal 2008;53(2):554–64.
30. Kuehnapfel A, Ahnert P, Loeffler M, Broda A, Scholz M. Reliability of 3D laser-based anthropometry and comparison with classical anthropometry. Sci Rep 2016;6(1):26672. [PubMed: 27225483]
31. Lovato C, Milanese C, Giachetti A, Zancanaro C. 3D digital anthropometry using the BodySCAN Paper presented at the International Conference on 3D Body Scanning Technologies, Lugano (Switzerland), 19–20 October 2010, pp 259–263; 2010. doi:10.15221/10.259.
32. Simenko J, Cuk I. Reliability and validity of NX-16 3D body scanner. Int J Morphol 2016;34(4):1506–14.
33. Ministere de la Sante. Enquête nutritionnelle nationale SMART Burkina Faso: Minstere de la Sante; 2016:53.
34. Nairobi County. SMART survey report Nairobi (Kenya): Nairobi County; 2020.

35. Siklar Z, Sanli E, Dallar Y, Tanyer G. Diurnal variation of height in children. Pediatr Int 2005;47(6):645–8. [PubMed: 16354217]

36. Tillmann V, Clayton PE. Diurnal variation in height and the reliability of height measurements using stretched and unstretched techniques in the evaluation of short-term growth. Ann Hum Biol 2001;28(2): 195–206. [PubMed: 11293727]

37. Lu J-M, Wang M-JJ, Mollard R. The effect of arm posture on the scan-derived measurements. Appl Ergon 2010;41(2):236–41. [PubMed: 19640502]

38. Kouchi M, Mochimaru M. Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry. Appl Ergon 2011;42(3):518–27. [PubMed: 20947062]

TABLE 1

Child age and sex, manual anthropometry measurement first, child calm during measurement by age in Guatemala, Kenya, and China for 3D body imaging evaluation

	Guatemala		Kenya		China	
Descriptive by age group, months	***n***	**Mean (SD) or % (95% CI)**	***n***	**Mean (SD) or % (95% CI)**	***n***	**Mean (SD) or % (95% CI)**
Age group						
All	641	27.6 (17.4)	548	26.7 (16.3)	301	12.8 (6.5)
0.0–23.9	307	11.9 (7.0)	279	13.0 (6.3)	301	12.8 (6.5)
24.0–59.9	334	42.0 (10.1)	269	41.0 (9.9)	—	—
0.0–5.9	76	3.0 (1.5)	39	3.3 (1.7)	51	3.1 (1.5)
6.0–11.9	73	8.3 (1.8)	83	8.6 (1.7)	73	8.2 (1.8)
12.0–23.9	158	17.9 (3.4)	157	17.8 (3.5)	177	17.5 (3.4)
24.0–35.9	103	29.7 (3.3)	85	29.2 (3.4)	—	—
36.0–47.9	120	41.9 (3.3)	98	41.0 (3.3)	—	—
48.0–59.9	111	53.5 (3.4)	86	52.6 (3.4)	—	—
Sex: female						
All	641	47.1 (43.2–51.1)	548	52.9 (48.6–57.2)	147	48.8 (43.1–54.6)
0.0–23.9	307	44.0 (38.3–49.7)	137	49.1 (43.1–55.1)	147	48.8 (43.1–54.6)
24.0–59.9	334	50.0 (44.5–55.5)	153	56.9 (50.7–62.9)	—	—
Order of measurement: manual first						
All	641	55.5 (51.6–59.4)	548	49.6 (45.4–53.9)	125	41.5 (35.9–47.3)
0.0–23.9	307	59.3 (53.6–64.8)	144	51.6 (45.6–57.6)	125	41.5 (35.9–47.3)
24.0–59.9	334	52.1 (46.6–57.6)	128	47.6 (41.5–53.7)	—	—
Expert reported child calm during manual measurement of height or length						
All	641	51.3 (47.4–55.3)	548	70.3 (66.2–74.1)	219	72.8 (67.4–77.7)
0.0–23.9	307	23.8 (19.1–28.9)	134	48.0 (42.0–54.1)	219	72.8 (67.4–77.7)
24.0–59.9	334	76.7 (71.7–81.1)	251	93.3 (89.6–96.0)	—	—
Expert reported child calm during manual measurement of midupper arm circumference						
All	641	71.3 (67.6–74.8)	472	86.1 (83.0–88.9)	203	67.4 (61.8–72.7)
0.0–23.9	307	55.4 (49.6–60.0)	213	76.3 (70.9–81.2)	203	67.4 (61.8–72.7)
24.0–59.9	334	85.9 (81.7–89.5)	259	96.3 (93.3–98.2)	—	—

	Guatemala		Kenya		China	
Descriptive by age group, months	***n***	**Mean (SD) or % (95% CI)**	***n***	**Mean (SD) or % (95% CI)**	***n***	**Mean (SD) or % (95% CI)**
Expert reported child calm during manual measurement head circumference						
All	—	—	—	—	228	75.8 (70.5–80.5)

TABLE 2

Interrater and intermethod accuracy for length or height, MUAC, and head circumference (HC): mean, SD, mean absolute difference, inter-TEM, relative TEM, and ICC by country and age in Guatemala, Kenya, and China for 3D body imaging evaluation[1]

Age groups, months	*n*	Mean[2] (SD), cm		Mean (SD) absolute difference,[3] cm	Inter-TEM (95% CI), cm	Relative inter-TEM, %	ICC[4] (95% CI)
		Manual	Scan				
Length or height							
Guatemala							
All (0.0–59.9)	641	80.8 (13.5)	80.4 (12.9)	2.9 (2.8)	2.8 (2.6–3.0)	3.5	0.58 (0.57–0.59)
0.0–23.9	307	69.7 (8.8)	69.1 (6.7)	3.8 (3.3)	3.6 (3.2–3.9)	5.1	0.09 (−0.03 to 0.16)
24.0–59.9	334	91.0 (7.7)	90.8 (7.3)	2.0 (1.7)	1.9 (1.6–2.1)	2.1	0.81 (0.81–0.82)
Kenya							
All (0.0–59.9)	548	84.5 (13.4)	82.6 (13.9)	3.5 (3.2)	3.4 (3.0–3.7)	4.0	0.47 (0.44–0.48)
0.0–23.9	279	73.9 (8.2)	71.0 (6.3)	4.5 (3.9)	4.2 (3.6–4.7)	5.8	0.05 (−0.10 to 0.13)
24.0–59.9	269	95.4 (7.7)	94.6 (8.3)	2.5 (1.8)	2.2 (2.0–2.3)	2.3	0.78 (0.77–0.78)
China							
All (0.0–23.9)	301	75.7 (8.9)	69.5 (6.7)	6.6 (4.0)	5.5 (5.1–5.8)	7.5	0.0
Midupper arm circumference							
Guatemala							
All (0.0–59.9)	641	15.0 (1.2)	15.8 (1.5)	1.2 (0.9)	1.1 (1.0–1.1)	6.9	0.13 (0.07–0.18)
0.0–23.9	307	14.4 (1.0)	14.8 (0.5)	0.9 (0.8)	0.8 (0.8–0.9)	5.7	0.0
24.0–59.9	334	15.5 (1.1)	16.8 (1.4)	1.4 (1.0)	1.2 (1.1–1.3)	7.7	0.16 (0.07–0.22)
Kenya							
All (0.0–59.9)	548	15.2 (1.2)	16.4 (2.0)	1.7 (1.3)	1.5 (1.4–1.6)	9.6	0.0
0.0–23.9	279	14.7 (1.1)	14.8 (0.6)	1.0 (0.8)	0.9 (0.8–1.0)	6.2	0.0
24.0–59.9	269	15.8 (1.0)	18.1 (1.6)	2.4 (1.4)	2.0 (1.8–2.1)	11.5	0.0
China							
All (0.0–23.9)	301	15.4 (1.1)	14.7 (0.4)	1.0 (0.9)	1.0 (0.9–1.0)	6.3	0.0
Head circumference							
Guatemala							
All (0.0–59.9)	641	—	46.6 (3.6)	—	—	—	—
0.0–23.9	307	—	46.6 (3.6)	—	—	—	—

Age groups, months	*n*	Mean[2] (SD), cm		Mean (SD) absolute difference,[3] cm	Inter-TEM (95% CI), cm	Relative inter-TEM, %	ICC[4] (95% CI)
		Manual	Scan				
24.0–59.9	334	—	46.5 (3.7)	—	—	—	—
Kenya							
All (0.0–59.9)	548	—	47.8 (3.6)	—	—	—	—
0.0–23.9	279	—	48.2 (3.3)	—	—	—	—
24.0–59.9	269	—	47.3 (3.8)	—	—	—	—
China							
All (0.0–23.9)	301	44.7 (2.9)	47.2 (2.6)	3.1 (2.5)	2.8 (2.6–3.0)	6.1	0.0

[1]Abbreviations: HC, head circumference; ICC, intraclass correlation coefficient; inter-TEM, interrater, intermethod technical error measurement; MUAC, midupper arm circumference; TEM, technical error measurement.

[2]Based on repeated manual measurement by reference anthropometrists (experts anthropometrists; mean of 2 manual measurements) compared with repeated scan measurements by team anthropometrists (field anthropometrists; mean of 2 scan measurements). $\text{Mean manual} = \sum_{i=1}^{N} M_{ei}/N$ and $\text{Mean scan} = \sum_{i=1}^{N} S_{ai}/N$, where M_{ei} = Mean manual measurement value of child i (experts), S_{ai} = mean scan measurement value of child i (field) and N = sample size.

[3]$\text{Mean absolute difference} = \sum_{i=1}^{N} (|M_{ei} - S_{ai}|)/N$, where M_{ei} = mean manual measurement value of child i (expert anthropometrist), S_{ai} = mean scan measurement value of child i (field anthropometrist), and N = sample size.

[4]When the ICC is 0.0, there is no variance, and the CI cannot be calculated.

TABLE 3

Interrater and intermethod accuracy, Bland-Altman bias and upper and lower limits of agreement, and Pitman test *P* value for length or height, midupper arm circumference, and head circumference by country and age in Guatemala, Kenya, and China for 3D body imaging evaluation

Age group, months	*n*	Mean (SD) absolute difference,[1] cm	Average bias,[2] cm	Bland-Altman 95% upper and lower limits of agreement	Pitman test *P* value[3]
Length or height					
Guatemala					
All (0.0–59.9)	641	2.9 (2.8)	−0.3	−8.2 to 7.6	0.620
0.0–23.9	307	3.8 (3.3)	−0.5	−10.6 to 9.5	0.360
24.0–59.9	334	2.0 (1.7)	−0.2	−5.5 to 5.1	0.760
Kenya					
All (0.0–59.9)	548	3.5 (3.2)	−1.9	−10.6 to 6.9	0.023
0.0–23.9	279	4.5 (3.9)	−2.9	−13.4 to 7.6	0.001
24.0–59.9	269	2.5 (1.8)	−0.8	−6.7 to 5.1	0.248
China					
All (0.0–23.9)	301	6.6 (4.0)	−6.2	−15.4 to 3.0	0.001
Midupper arm circumference					
Guatemala					
All (0.0–59.9)	641	1.2 (0.9)	0.9	−1.6 to 3.3	0.001
0.0–23.9	307	0.9 (0.8)	0.4	−1.8 to 2.6	0.001
24.0–59.9	334	1.4 (1.0)	1.3	−1.1 to 3.7	0.001
Kenya					
All (0.0–59.9)	548	1.7 (1.3)	1.2	−2.3 to 4.7	0.001
0.0–23.9	279	1.0 (0.8)	0.2	−2.4 to 2.7	0.032
24.0–59.9	269	2.4 (1.4)	2.3	−0.6 to 5.3	0.001
China					
All (0.0–23.9)	301	1.0 (0.9)	−0.8	−3.0 to 1.4	0.001
Head circumference, cm					
China					
All (0.0–23.9)	301	3.1 (2.5)	2.4	−3.9 to 8.7	0.001

[1]Mean absolute difference $= \sum_{i=1}^{N} (|M_{ei} - S_{ai}|)/N$, where M_{ei} = mean manual measurement value of child *i* (expert anthropometrist), S_{ai} = mean scan measurement value of child *i* (field) and *N* = sample size.

[2] Average bias, $\sum_{i=1}^{N}(S_{ai} - M_{ei})/N$.

[3] A Pitman test was used to examine variance. When the Pitman test was significant, we used the Levene test of variance homogeneity to test the SD of the differences between methods across age groups.

TABLE 4

Intrarater and intermethod accuracy for length or height, MUAC, and HC: mean, SD, mean absolute difference, inter-TEM, relative TEM, and ICC by country and age in Guatemala, Kenya, and China for 3D body imaging evaluation[1]

		Mean[2] (SD), cm					
Age group, months	***n***	**Manual**	**Scan**	**Mean (SD) absolute difference,[3] cm**	**Inter-TEM (95% CI), cm**	**Relative TEM, %**	**ICC[4] (95% CI)**
Length and/or height							
Guatemala							
Expert							
All (0.0–59.9)	641	80.8 (13.5)	82.6 (13.0)	3.7 (3.3)	3.5 (3.2–3.7)	4.3	0.44 (0.42–0.46)
0.0–23.9	307	69.7 (8.8)	71.5 (7.0)	4.8 (3.8)	4.3 (3.9–4.7)	6.1	0.07 (−0.06 to 0.14)
24.0–59.9	334	91.0 (7.7)	92.9 (7.8)	2.7 (2.3)	2.5 (2.1–2.8)	2.7	0.69 (0.69–0.70)
Field							
All (0.0–59.9)	641	80.8 (13.6)	80.4 (12.9)	2.9 (2.8)	2.8 (2.6–3.1)	3.5	0.58 (0.56–0.59)
0.0–23.9	307	69.5 (8.8)	69.1 (6.7)	3.8 (3.4)	3.6 (3.2–4.0)	5.2	0.07 (−0.05 to 0.15)
24.0–59.9	334	91.2 (7.7)	90.8 (7.3)	2.0 (1.7)	1.9 (1.7–2.0)	2.0	0.82 (0.81–0.82)
Kenya							
Expert							
All (0.0–59.9)	548	84.5 (13.4)	83.1 (13.9)	3.1 (3.0)	3.0 (2.7–3.4)	3.6	0.52 (0.50–0.53)
0.0–23.9	279	73.9 (8.2)	71.3 (5.8)	4.3 (3.6)	3.9 (3.4–4.4)	5.1	0.0
24.0–59.9	269	95.4 (7.7)	95.4 (7.9)	1.9 (1.5)	1.7 (1.5–1.8)	1.8	0.86 (0.86–0.86)
Field							
All (0.0–59.9)	548	84.9 (13.4)	82.6 (13.9)	3.7 (3.1)	3.4 (3.1–3.7)	4.1	0.44 (0.42–0.46)
0.0–23.9	279	74.2 (7.9)	71.0 (6.3)	4.7 (3.8)	4.2 (3.8–4.7)	5.9	0.44 (0.42–0.46)
24.0–59.9	269	95.9 (7.7)	94.6 (8.3)	2.6 (1.8)	2.3 (2.1–2.5)	2.4	0.76 (0.75–0.76)
China							
Expert							
All (0.0–23.9)	301	75.7 (8.9)	70.6 (6.5)	5.9 (4.0)	5.0 (4.7–5.4)	6.9	0.0
Field							
All (0.0–23.9)	301	75.7 (8.9)	70.6 (6.5)	6.4 (3.9)	5.3 (5.0–5.7)	7.3	0.0
Midupper arm circumference							
Guatemala							

Age group, months	*n*	Mean[2] (SD), cm		Mean (SD) absolute difference,[3] cm	Inter-TEM (95% CI), cm	Relative TEM, %	ICC[4] (95% CI)
		Manual	Scan				
Expert							
All (0.0–59.9)	641	15.0 (1.2)	15.9 (1.5)	1.3 (0.9)	1.1 (1.0–1.2)	7.2	0.07 (−0.01 to 0.12)
0.0–23.9	307	14.4 (1.0)	14.9 (0.6)	1.1 (0.8)	0.9 (0.8–1.0)	6.5	0.0
24.0–59.9	334	15.5 (1.1)	16.9 (1.4)	1.4 (1.0)	1.2 (1.1–1.3)	7.6	0.13 (0.03–0.20)
Field							
All (0.0–59.9)	641	15.3 (1.2)	15.8 (1.5)	1.1 (0.9)	1.0 (0.9–1.0)	6.3	0.24 (0.19–0.27)
0.0–23.9	307	14.6 (1.1)	14.8 (0.5)	0.9 (0.7)	0.8 (0.7–0.9)	5.5	0.05 (−0.09 to 0.13)
24.0–59.9	334	15.9 (1.1)	16.8 (1.4)	1.2 (0.9)	1.1 (1.0–1.2)	6.8	0.28 (0.22–0.32)
Kenya							
Expert							
All (0.0–59.9)	548	15.2 (1.2)	16.6 (2.2)	1.8 (1.4)	1.6 (1.5–1.7)	10.4	0.0
0.0–23.9	279	14.7 (1.1)	14.8 (0.5)	1.0 (0.7)	0.9 (0.8–1.0)	6.0	0.0
24.0–59.9	269	15.8 (1.0)	18.5 (1.6)	2.7 (1.4)	2.2 (2.0–2.3)	12.7	0.0
Field							
All (0.0–59.9)	548	15.5 (1.2)	16.4 (2.0)	1.6 (1.2)	1.4 (1.3–1.5)	9.0	0.0
0.0–23.9	279	15.0 (1.2)	14.8 (0.6)	1.0 (0.8)	0.9 (0.8–1.0)	6.3	0.0
24.0–59.9	269	16.0 (1.0)	18.1 (1.6)	2.4 (1.4)	1.8 (1.7–1.9)	10.6	0.0
China							
Expert							
All (0.0–23.9)	301	15.4 (1.1)	14.8 (0.5)	1.0 (0.9)	0.9 (0.8–1.0)	6.2	0.0
Field							
All (0.0–23.9)	301	15.6 (1.1)	14.7 (0.4)	1.1 (0.9)	1.0 (0.9–1.1)	6.9	0.0
Head circumference							
Guatemala							
Expert							
All (0.0–59.9)	641	—	47.8 (4.3)	—	—	—	—
0.0–23.9	307	—	49.8 (4.0)	—	—	—	—
24.0–59.9	334	—	46.0 (3.7)	—	—	—	—
Field							
All (0.0–59.9)	641	—	46.6 (3.6)	—	—	—	—

Age group, months	n	Mean[2] (SD), cm		Mean (SD) absolute difference,[3] cm	Inter-TEM (95% CI), cm	Relative TEM, %	ICC[4] (95% CI)
		Manual	Scan				
0.0–23.9	307	—	46.7 (3.6)	—	—	—	—
24.0–59.9	334	—	46.5 (3.7)	—	—	—	—
Kenya							
Expert							
All (0.0–59.9)	548	—	48.6 (3.5)—	—	—	—	—
0.0–23.9	279	—	49.0 (3.1)	—	—	—	—
24.0–59.9	269	—	48.3 (3.8)	—	—	—	—
Field							
All (0.0–59.9)	548	—	47.8 (3.6)	—	—	—	—
0.0–23.9	279	—	48.2 (3.3)	—	—	—	—
24.0–59.9	269	—	47.3 (3.8)	—	—	—	—
China							
Expert							
All (0.0–23.9)	301	44.7 (2.9)	49.1 (2.7)	4.5 (3.1)	3.8 (3.6–4.1)	8.2	0.0
Field							
All (0.0–23.9)	301	44.5 (2.9)	47.2 (2.6)	3.3 (2.6)	2.9 (2.7–3.1)	6.4	0.0

[1] Abbreviations: HC, head circumference; ICC, intraclass correlation coefficient; inter-TEM, interrater, intermethod technical error measurement; MUAC, midupper arm circumference; TEM, technical error measurement.

[2] Manual: reference anthropometrist (expert anthropometrists; mean of 2 manual measurements) and team anthropometrists (field anthropometrists; mean of 2 manual measurements). Scan: reference anthropometrist (expert anthropometrists; mean of 2 scan measurements) and team anthropometrists (field anthropometrists; mean of 2 scan measurements. $\text{Mean manual} = \sum_{i=1}^{N} M_i/N$ and $\text{Mean scan} = \sum_{i=1}^{N} S_i/N$, where M_i = mean manual measurement value of child i, S_i = mean scan measurement value of child i (field anthropometrist), and N = sample size.

[3] $\text{Mean absolute difference} = \sum_{i=1}^{N} (|M_i - S_i|)/N$, where M_i = mean manual measurement value of child i, S_i = mean scan measurement value of child i, and N = sample size.

[4] When the ICC is 0.0, there is no variance, and the CI cannot be calculated.

TABLE 5

Intrarater and intramethod precision: mean, SD, mean absolute difference, intra-TEM, relative intra-TEM, and ICC for length or height, MUAC, and HC by country and by age in Guatemala, Kenya, and China for 3D body imaging evaluation[1]

		Mean[2] (SD), cm		Mean (SD) absolute difference, cm		Intra-TEM, cm		Relative intra-TEM, %		ICC (95% CI)	
Age group, months	***n***	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**
Length or height											
Guatemala											
All (0.0–59.9)	641	80.8 (13.5)	80.4 (12.9)	0.1 (0.1)	0.4 (0.6)	0.1	0.5	0.1	0.6	1.0 (1.0–1.0)	0.98 (0.98–0.98)
0.0–23.9	307	69.7 (8.8)	69.1 (6.7)	0.1 (0.1)	0.6 (0.7)	0.1	0.7	0.2	1.0	1.0 (1.0–1.0)	0.97 (0.97–0.97)
24.0–59.9	334	91.0 (7.7)	90.8 (7.3)	0.1 (0.1)	0.3 (0.3)	0.1	0.3	0.1	0.3	1.0 (1.0–1.0)	0.99 (0.99–0.99)
Kenya											
All (0.0–59.9)	548	84.5 (13.4)	82.6 (13.9)	0.1 (0.1)	0.6 (0.8)	0.1	0.7	0.1	0.8	1.0 (1.0–1.0)	0.98 (0.98–0.98)
0.0–23.9	279	73.9 (8.2)	71.0 (6.3)	0.1 (0.1)	0.8 (1.0)	0.1	0.9	0.1	1.2	1.0 (1.0–1.0)	0.96 (0.96–0.96)
24.0–59.9	269	95.4 (7.7)	94.6 (8.3)	0.1 (0.1)	0.3 (0.3)	0.1	0.3	0.1	0.4	1.0 (1.0–1.0)	0.99 (0.99–0.99)
China											
All (0.0–23.9)	301	75.7 (8.9)	69.5 (6.7)	0.1 (0.1)	0.4 (0.5)	0.1	0.5	0.1	0.7	1.0 (1.0–1.0)	0.98 (0.98–0.98)
Midupper arm circumference											
Guatemala											
All (0.0–59.9)	641	15.0 (1.2)	15.8 (1.5)	0.1 (0.1)	0.1 (0.1)	0.1	0.1	0.5	0.7	0.99 (0.99–0.99)	0.99 (0.99–0.99)
0.0–23.9	307	14.4 (1.0)	14.8 (0.5)	0.1 (0.1)	0.1 (0.1)	0.1	0.1	0.6	0.7	0.99 (0.99–0.99)	0.96 (0.96–0.96)
24.0–59.9	334	15.5 (1.1)	16.8 (1.4)	0.1 (0.1)	0.1 (0.1)	0.1	0.1	0.5	0.7	0.99 (0.99–0.99)	0.99 (0.99–0.99)
Kenya											
All (0.0–59.9)	548	15.2 (1.2)	16.4 (2.0)	0.1 (0.1)	0.1 (0.1)	0.1	0.1	0.4	0.8	1.00 (1.00–1.00)	0.99 (0.99–0.99)
0.0–23.9	279	14.7 (1.1)	14.8 (0.6)	0.1 (0.1)	0.1 (0.1)	0.1	0.1	0.5	0.6	0.99 (0.99–0.99)	0.98 (0.98–0.98)
24.0–59.9	269	15.8 (1.0)	18.1 (1.6)	0.0 (0.1)	0.2 (0.2)	0.1	0.2	0.4	0.4	1.00 (1.00–1.00)	0.99 (0.99–0.99)
China											
All (0.0–23.9)	301	15.4 (1.1)	14.7 (0.4)	0.1 (0.1)	0.1 (0.1)	0.1	0.1	0.6	0.5	0.99 (0.99–0.99)	0.97 (0.97–0.97)
Head circumference											
China											
All (0.0–23.9)	301	44.7 (2.9)	47.2 (2.6)	0.1 (0.1)	0.4 (0.3)	0.1	0.3	0.2	0.7	1.00 (1.00–1.00)	0.98 (0.98–0.98)

[1] Abbreviations: HC, head circumference; ICC, intraclass correlation coefficient; intra-TEM, interrater, intramethod technical error measurement; MUAC, midupper arm circumference; TEM, technical error measurement.

[2] Based on repeated manual measurements by reference anthropometrist (expert anthropometrists; first manual measurement compared to second manual measurement) and repeated scan measurements by team anthropometrists (field anthropometrists; first scan measurement compared to second scan measurement).

TABLE 6

Interrater and intramethod precision: mean, SD, mean absolute difference, inter-TEM, relative intra-TEM, and ICC for length or height, MUAC, and HC by country and age in Guatemala, Kenya, and China for 3D body imaging evaluation[1]

		Mean[2] (SD), cm		Mean (SD) Absolute difference, cm		Intra-TEM, cm		Relative intra-TEM, %		ICC (95% CI)	
Measures by age group, months	***n***	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**	**Manual**	**Scan**
Length or height											
Guatemala											
All (0.0–59.9)	641	80.8 (13.5)	81.5 (12.8)	0.6 (1.0)	3.4 (2.9)	0.8	3.2	1.0	3.9	0.98 (0.98–0.98)	0.64 (0.62–0.65)
0.0–23.9	307	69.6 (8.8)	70.3 (6.5)	0.7 (1.4)	3.9 (3.4)	1.1	3.7	1.6	5.3	0.94 (0.93–0.94)	0.50 (0.47–0.53)
24.0–59.9	334	91.1 (7.7)	91.8 (7.4)	0.5 (0.4)	2.9 (2.3)	0.5	2.6	0.5	2.8	0.99 (0.99–0.99)	0.75 (0.73–0.76)
Kenya											
All (0.0–59.9)	548	84.7 (13.4)	82.9 (13.8)	0.8 (1.2)	2.8 (2.5)	1.0	2.7	1.2	3.2	0.97 (0.97–0.97)	0.72 (0.71–0.73)
0.0–23.9	279	74.1 (8.0)	71.1 (5.6)	0.8 (1.5)	3.0 (2.9)	1.2	3.0	1.7	4.2	0.91 (0.90–0.91)	0.57 (0.54–0.59)
24.0–59.9	269	95.7 (7.7)	95.0 (7.9)	0.7 (0.5)	2.6 (2.0)	0.6	2.3	0.6	2.4	0.99 (0.99–0.99)	0.81 (0.79–0.82)
China											
All (0.0–23.9)	301	75.6 (8.8)	70.0 (6.4)	0.5 (0.4)	2.2 (2.1)	0.5	2.2	0.6	3.1	0.98 (0.98–0.98)	0.73 (0.71–0.75)
Midupper arm circumference											
Guatemala											
All (0.0–59.9)	641	15.1 (1.2)	15.7 (1.3)	0.4 (0.3)	0.7 (0.7)	0.3	0.7	2.2	4.2	0.92 (0.91–0.92)	0.76 (0.74–0.76)
0.0–23.9	307	14.5 (1.0)	14.8 (0.4)	0.4 (0.3)	0.5 (0.4)	0.3	0.5	2.3	3.1	0.92 (0.91–0.92)	0.51 (0.47–0.53)
24.0–59.9	334	15.7 (1.1)	16.6 (1.2)	0.4 (0.3)	0.9 (0.8)	0.3	0.8	2.1	4.9	0.92 (0.91–0.92)	0.75 (0.73–0.76)
Kenya											
All (0.0–59.9)	548	15.4 (1.2)	16.3 (1.7)	0.4 (0.3)	0.7 (0.7)	0.3	0.7	2.2	4.4	0.93 (0.92–0.93)	0.79 (0.78–0.80)
0.0–23.9	279	14.8 (1.1)	14.8 (0.4)	0.4 (0.3)	0.4 (0.4)	0.4	0.4	2.5	3.0	0.91 (0.90–0.92)	0.55 (0.52–0.57)
24.0–59.9	269	15.9 (1.0)	17.8 (1.2)	0.4 (0.2)	1.1 (0.8)	0.3	0.9	1.9	5.1	0.94 (0.93–0.94)	0.73 (0.70–0.74)
China											
All (0.0–23.9)	301	15.5 (1.1)	15.0 (0.4)	0.4 (0.3)	0.3 (0.3)	0.4	0.3	2.4	2.2	0.92 (0.92–0.93)	0.63 (0.60–0.65)
Head circumference											
China											
All (0.0–23.9)	301	44.6 (2.9)	48.1 (2.3)	0.3 (0.2)	2.6 (2.0)	0.3	2.3	0.6	4.8	0.98 (0.98–0.98)	0.54 (0.51–0.56)

[1] Abbreviations: HC, head circumference; ICC, intraclass correlation coefficient; inter-TEM, interrater, intermethod technical error measurement; intra-TEM, interrater, intramethod technical error measurement; MUAC, midupper arm circumference; TEM, technical error measurement.

[2] Based on repeated manual measurement by reference anthropometrist (expert anthropometrists; 2 manual measurements) compared to team anthropometrists (field anthropometrists; 2 manual measurements) and repeated scan measurements by reference anthropometrist (expert anthropometrists; scan 2 measurements) compared to team anthropometrists (field anthropometrists; 2 scan measurements).