



Published in final edited form as:

Am J Hum Biol. 2016 May ; 28(3): 364–371. doi:10.1002/ajhb.22796.

Anthropometric Measurement Standardization in the US-Affiliated Pacific: Report from the Children's Healthy Living Program

FENFANG LI^{1,*}, LYNNE R. WILKENS², RACHEL NOVOTNY¹, MARIE K. FIALKOWSKI¹, YVETTE C. PAULINO⁴, RANDALL NELSON⁵, ANDREA BERSAMIN⁶, URSULA MARTIN⁷, JONATHAN DEENIK³, and CAROL J. BOUSHEY²

¹Department of Human Nutrition, Food and Animal Sciences, University of Hawai'i at Mānoa, Honolulu, Hawai'i ²Cancer Center, University of Hawai'i at Mānoa, Honolulu, Hawai'i ³Department of Tropical Plants and Soil Science, University of Hawai'i at Mānoa, Honolulu, Hawai'i ⁴School of Nursing and Health Sciences, University of Guam, Mangilao, Guam ⁵Land Grant Program, Northern Marianas College, Saipan, Northern Mariana Islands ⁶Department of Alaska Native Health, University of Alaska, Fairbanks, Alaska ⁷Land Grant Program, American Samoa Community College, Pago Pago, American Samoa

Abstract

Objectives—Anthropometric standardization is essential to obtain reliable and comparable data from different geographical regions. The purpose of this study is to describe anthropometric standardization procedures and findings from the Children's Healthy Living (CHL) Program, a study on childhood obesity in 11 jurisdictions in the US-Affiliated Pacific Region, including Alaska and Hawai'i.

Methods—Zerfas criteria were used to compare the measurement components (height, waist, and weight) between each trainee and a single expert anthropometrist. In addition, intra- and inter-rater technical error of measurement (TEM), coefficient of reliability, and average bias relative to the expert were computed.

Results—From September 2012 to December 2014, 79 trainees participated in at least 1 of 29 standardization sessions. A total of 49 trainees passed either standard or alternate Zerfas criteria and were qualified to assess all three measurements in the field. Standard Zerfas criteria were

*Correspondence to: Fenfang Li, Department of Human Nutrition, Food and Animal Sciences, University of Hawaii at Manoa, 1950 East West Road, AgSci 302H, Honolulu, Hawaii 96822. ; Email: fenfang@hawaii.edu.

Author Contribution

F. Li participated in study design, led the data analysis, interpretation of data and writing of the manuscript, and had primary responsibility for the integrity of the data, accuracy of the data analysis and the final article. L.R. Wilkens led the study analytic strategy, interpretation of data, and took responsibility for the integrity and the accuracy of the data analysis. R. Novotny led the study, acquisition of the data and interpretation of data; M.K. Fialkowski co-led the study design, acquisition of the data and interpretation of data. Y.C. Paulino participated in design, data acquisition and led interpretation of Guam data. R. Nelson participated in design, led in data acquisition and interpretation of Commonwealth of the Northern Marianas data. A. Bersamin participated in design, led in data acquisition and interpretation of Alaska data. U. Martin participated in design, led in data acquisition and interpretation of American Samoa data. J. Deenik participated in design and led in interpretation of data of Freely Associated States (Republic of Palau, Republic of Marshall Islands, and Federated States of Micronesia (in this study Yap, Kosrae, and Pohnpei)). C.J. Boushey participated in study design and data interpretation. All authors critically reviewed and approved the final manuscript.

difficult to achieve: only 2 of 79 trainees passed at their first training session. Intra-rater TEM estimates for the 49 trainees compared well with the expert anthropometrist. Average biases were within acceptable limits of deviation from the expert. Coefficient of reliability was above 99% for all three anthropometric components.

Conclusions—Standardization based on comparison with a single expert ensured the comparability of measurements from the 49 trainees who passed the criteria. The anthropometric standardization process and protocols followed by CHL resulted in 49 standardized field anthropometrists and have helped build capacity in the health workforce in the Pacific Region.

Anthropometric measurements of height, weight, and waist circumference components are the most frequently used techniques for the assessment of growth and nutritional status among children (Lohman et al., 1988; Ulijaszek, 1997). Like all human measures, anthropometry is subject to measurement error. For large studies where multiple sites are involved and a number of anthropometrists are needed, the degree of measurement error due to interobserver variation increases. Therefore, standardization of measurement procedures is essential to obtain reliable and comparable anthropometric data from different geographical/jurisdictional regions (Lohman et al., 1988; Ulijaszek, 1997; Ulijaszek and Kerr, 1999).

The importance of training new staff in anthropometry using standardized rules of measurement and lead/expert anthropometrists has been delineated in the literature and adopted by large, national and international studies (de Onis et al., 2004; Johnson et al., 1997; National Health and Nutrition Examination Survey (NHANES), 2014; Zerfas, 1985). For example, the NHANES used a consultant anthropometry expert to verify that the anthropometry protocol was being implemented properly and consistently (NHANES, 2014). In the World Health Organization (WHO) Multicentre Growth Reference Study, two anthropometrists were designated as lead anthropometrists, and anthropometrists at each participating site were standardized against one of the two lead anthropometrists initially and yearly during follow-up (de Onis et al., 2004).

Nevertheless, protocols and procedures on how anthropometry standardization was conducted and, in particular, what criteria were used in the evaluation of anthropometrists or anthropometry technologists against the expert anthropometrist, were not reported, even in large-scale, national studies. For example, the anthropometry procedures manual for NHANES states that the anthropometry was measured by trained health technicians and recorders. Nevertheless, how they were trained and whether the training included an anthropometry standardization session were not included in the manual (NHANES, 2014). In addition, there is little information in the peer-reviewed literature, about applying guidelines and criteria of anthropometry standardization in areas where education and experience of measurement technicians is lacking and may be inadequate. This is particularly true in the US-Affiliated Pacific Region (USAP), where professional anthropometrists are rare and childhood obesity rates are high (Bruss et al., 2010; de Onis et al., 2010; Novotny et al., 2013a; Novotny et al., 2014; Ogden et al., 2014; Paulino et al., 2015).

The purpose of this paper is to describe the anthropometric standardization process, protocol, lessons learned, and findings of the Children's Healthy Living (CHL) Program for Remote Underserved Minority Populations of the Pacific. CHL is a multicomponent program in the USAP designed to prevent child obesity by building regional capacity, including a multicenter trial testing a community-based environmental intervention to prevent childhood obesity and to promote healthful behaviors (Novotny et al., 2013b; Wilkens et al., 2013). The 11 jurisdictions of CHL in the USAP region include Alaska, American Samoa, the Commonwealth of the Northern Mariana Islands (CNMI), Guam, Hawai'i, and the Freely Associated States of Micronesia (FAS), which include the Republic of Palau, the Republic of the Marshall Islands (RMI), and the Federated States of Micronesia (Chuuk, Kosrae, Pohnpei, and Yap) (Fig. 1).

METHODS

To ensure a unified, standardized protocol for field data collection, 2–5 days of measurement and standardization trainings were held for each jurisdiction before measurement and data collection were started. The guidelines for the conduct of anthropometric standardization training were adapted from Lohman et al. (1988) and were incorporated into CHL standard operating procedures, which were then compiled into a field manual and provided to all trainees. In addition to CHL staff, local agencies (e.g., Head Start, The Supplemental Nutrition Program for Woman, Infants and Children [WIC]) in some jurisdictions had staff attend these trainings as well. The CHL expert measurement team, which included the Program Director (RN) and Assistant Program Director (MKF), were in charge of training and reviewing the measurements in each jurisdiction's training sessions. The Program Director (RN) served as the expert anthropometrist and the Assistant Program Director (MKF) served as the standard recorder for each training session. The expert anthropometrist for the CHL standardization process is a professor in human nutrition who was trained and standardized by a physical anthropologist and has conducted more than 15 studies with anthropometry in the past three decades and has trained hundreds of measurers (Guerrero et al., 2008; Novotny et al., 2004; Novotny et al., 2007a; Novotny et al., 2007b; Novotny et al., 2013b).

All applicable institutional and governmental regulations regarding the ethical issues of human volunteers were followed. The standardization trainings for measurement were approved by the University of Hawai'i Institutional Review Board (IRB). Jurisdiction-level approvals were also obtained from the University of Alaska Fair-banks and the University of Guam institutional review boards. Parent consent forms were obtained for all children before each training session, and assents were obtained from all children before measurement occurred.

Anthropometric measurement

Child height was measured by a Portable Adult/Infant Measuring Unit stadiometer (Model PE-AIM-101, Perspective Enterprises, Portage, MI). This device has a flat vertical surface on which a measuring rule is attached. The stadiometer also has an attached movable headpiece with a screw to hold it in place. Per the protocol, the portable stadiometer was to

be positioned on a level floor without carpeting, and flush against a wall or other flat surface. Height was measured with the children barefoot, or with light socks (Alaska), to the nearest 0.1 cm. The instrument was calibrated with a 140 cm aluminum rod before each measurement session or every time the stadiometer was moved.

Child weight was measured using a portable SECA 876 scale (SECA 876, Hamburg, Germany). The scales came equipped with four adjustable feet and a bull's eye spirit level. The scales were placed on a noncarpeted floor and leveled as needed to ensure accuracy. Children were weighed with bare feet, or with light socks (Alaska), and wearing only light weight clothing (e.g., shorts and t-shirt). Weight was measured to the nearest 0.1 kg. The scale was calibrated with a 4.5 kg weight before each measurement session or every time the scale was moved.

Child waist circumference was measured with a SECA 201 circumference measuring tape (SECA 201, Hamburg, Germany) to the nearest 1 mm. The measuring tape needed to be flush to the skin at the child's umbilicus, which provides similar measures in children as the iliac crest (Lohman et al., 1988). Measurers waited to make sure the child was breathing normally prior to taking the measurement and was in standard position (standing on both feet, looking ahead, arms folded across chest). The measurement was taken at the midpoint between inspiration and expiration since breath control is difficult in young children. The measuring tape was calibrated against the 140 cm aluminum rod before every measurement session.

For each training session, the expert team explained the anthropometry protocol, which the trainees had reviewed beforehand. The NHANES III Anthropometric Procedure video was shown and trainees practiced measuring each other to familiarize themselves with the procedure and equipment before measuring children. The trainees then measured the designated anthropometry components of the child volunteers. Two trainees were paired as a team to measure and record each anthropometric component for each child participant. One trainee took the measurements and the other recorded the values, followed by providing the values to the measurer to view and confirm agreement or to correct or redo the measure. Then, the two trainees traded roles. Every trainee eventually assumed the role of measurer and recorder for each child. The goal was for each trainee to measure eight children aged 2–8 years old at a session; effort was made to recruit a minimum of 10 children, as a few children at each session would become fatigued before being measured by all trainees. Upon completion of the session, measurements were then compared against those of the expert anthropometrist.

Three anthropometric components were standardized at each training session: weight, height, and waist circumference. Each child was measured at least three times for each component by each measurer, including the expert anthropometrist. If no two measurements were within two units (0.2 kg for weight and 0.2 cm for height and waist circumference), the measurer was instructed to repeat the measurement until there were at least two measurements within two units. Staff were allowed to cross out measurements they felt the least confident about as long as at least three measurements remained. The average of all

available measurements of an anthropometric component was used in the assessment (de Onis et al., 2004).

Anthropometry standardization procedures and protocols

Zerfas criteria, developed to compare the measurements of trainees against the measurements of the expert anthropometrist, were used to assess each trainee (Zerfas, 1985). Initially, the trainee measurement for each child was rated (Step 1) and then the trainee performance across children was rated (Step 2). In Step 1, the trainee measurements for each component and each child were rated based on the bias, defined as the difference (d) in the mean values of the component for the expert and for the trainee. Zerfas provided two tiers of rating: the Standard (more stringent) criteria or the Alternative (more lenient) criteria (Table 1). Each rating system categorizes each measurement based on magnitude of the bias into one of the following four categories: Good, Fair, Poor, and Blunder Errors. In the Zerfas manual, there is no specific criteria given for waist circumference; however, as other anthropometric components that are given in mm units (height and arm circumference) have the same Zerfas criteria, those Zerfas criteria were applied to waist circumference in this study.

In Step 2, the individual measurement ratings for an anthropometric component were aggregated across children to assess the trainee's overall performance. The Zerfas criteria at this step categorizes the overall performance as Pass, Borderline Pass, No Pass, or No Pass by Blunder. One or more blunders at Step 1 resulted in an overall evaluation of "No Pass by Blunder" for that anthropometric component. For components where there was no "blunder," the overall evaluation was determined by the number of unsuccessful (n_F) measurements at Step 1, which were those that fell into the category of "Fair" or "Poor." The number of unsuccessful measurements is compared to the expected number based on a binomial distribution with parameters P =the probability of 0.05 and M =number of children measured. "No Pass" is assigned when n_F is greater than or equal to n_1 , the minimum number meeting the criteria $\text{Prob}(X \geq n_1 | M, P) < 0.05$. "Pass" is assigned when n_F is less than or equal to n_2 , the maximum number meeting the criteria $\text{Prob}(X \leq n_2 | M, P) < 0.95$. "Borderline Pass" is assigned when $n_2 < n_F < n_1$.

Table 2 gives the number of unsuccessful measurements that would be considered a Pass, Borderline Pass, or No Pass by the number of children measured. The number of unsuccessful measurements for a trainee are based either on the standard or alternate Zerfas criteria in Table 1, respectively, depending on whether a more stringent or lenient approach is desired. The numbers in Table 2 mirror those in Zerfas (1985) and were expanded to larger values of M . It is advantageous to have more than eight children measured at a training session, as one unsuccessful measurement with five to seven children will usually result in a no pass rating at Step 2 for that component.

We chose to accept a rating of "Pass" or "Borderline Pass" as our criterion for qualifying to measure the anthropometric component in the field and a rating of "No Pass" or "No Pass by Blunder" as our criterion for not qualifying for that component. Only trainees who measured a minimum of five children for each component at the training session were evaluated. We also used the more stringent standard Zerfas criteria at a trainee's first session and the more

lenient alternate Zerfas criteria at subsequent sessions, in order to ensure that trainees received a sufficient level of experience with measurement (i.e., at least two sessions unless the trainee was very experienced at the outset). If a trainee failed the standardization for any of the three anthropometric components, he/she had to participate in another training session. Only anthropometric components that were not passed were required to be repeated. At an individual's second or any other subsequent training session, the Alternate Zerfas criteria were used. Trainees passing any of the three anthropometric components by the Standard Zerfas criteria at his/her first training session or by the Alternate Zerfas criteria at any subsequent training session were qualified and permitted to measure in the field. No one was permitted to measure in the field that did not pass by at least the Alternate Zerfas criteria at his/her second or subsequent training sessions. Individuals who underwent the training but did not meet the evaluation criteria could only record measurements in the field.

Data analysis

Anthropometric standardization results from each training session were sent to the CHL Coordinating Center (Honolulu, HI) to be analyzed. A SAS program (SAS Institute Inc., Cary, NC) was developed to evaluate each standardization training session. An evaluation report was produced for each training session that provided information on each trainee's ratings for each measure based on the Zerfas criteria.

The technical error of measurement (TEM) is commonly reported as a measure of imprecision in anthropometric assessment (Chumlea et al., 1990; de Miguel-Etayo, et al. 2014; Johnson et al., 1997; Marks et al., 1989; Moreno et al., 2003; Perini et al., 2005; Stomfai et al., 2011; Uljaszek and Kerr, 1999; WHO Multicentre Growth Reference Study Group, 2006). It is the square root of the measurement error variance. Intra-rater TEM is a measure of the imprecision about the rater-specific means and is computed as

$$TEM_{intra} = \sqrt{\sum_{c=1}^C \sum_{t=1}^{T_c} \sum_{r=1}^{R_{ct}} (M_{ctr} - \bar{M}_{ct.})^2 / CT_c (R_{ct} - 1)},$$

where M_{ctr} is the measure for child c ($c = 1 \dots C$) for trainee t ($t = 1 \dots T_c$) for replicate r ($r = 1 \dots R_{ct}$), and $\bar{M}_{ct.}$ is the mean of the R_{ct} replicates for the T_c trainees that measured child c : $\bar{M}_{ct.} = \sum_{r=1}^{R_{ct}} M_{ctr} / R_{ct}$. Therefore, there is one $\bar{M}_{ct.}$ per trainee and child combination.

Inter-rater TEM is a measure of the imprecision of the group-specific means and is

$$TEM_{inter} = \sqrt{\sum_{c=1}^C \sum_{t=1}^{T_c} (\bar{M}_{ct.} - \bar{M}_{c..})^2 / CT_c},$$

where $\bar{M}_{ct.}$ is defined as above and $\bar{M}_{c..}$ is the mean across the T_c trainees of $\bar{M}_{ct.}$: $\bar{M}_{c..} = \sum_{t=1}^{T_c} \bar{M}_{ct.} / T_c$. Therefore, there is one $\bar{M}_{c..}$ per child.

Total TEM is the aggregate measure of imprecision and is computed as

$$TEM_{total} = \sqrt{TEM_{intra}^2 + TEM_{inter}^2}.$$

A related statistic, the coefficient of reliability, R , estimates the proportion of the total variance that is not due to measurement error (Chumlea et al., 1990; Johnson et al., 1997; Marks et al., 1989; Moreno et al., 2003; Perini et al., 2005; Stomfai et al., 2011; Uljaszek

and Kerr, 1999; WHO Multicentre Growth Reference Study Group, 2006). It is defined as

$$R=1-TEM_{\text{total}}^2/SD^2, \text{ where } SD^2=\sum_{c=1}^C \sum_{t=1}^{T_c} \sum_{r=1}^{R_{ct}} (M_{\text{ctr}}-\bar{M}_{\dots})^2/(CT_cR_{ct}-1).$$

Average bias was assessed in terms of magnitude and whether or not the trainees systematically overestimated or underestimated measurements compared to the expert anthropometrist (WHO Multicentre Growth References Study Group, 2006).

RESULTS

Number of training sessions, number of participating children, and trainees

From September 2012 to December 2014, a total of 29 training sessions were conducted among the 11 CHL participating jurisdictions. A total of 280 children and 79 trainees participated. The total number of children at each training session varied, with a minimum of 5 and a maximum of 19. Of the 79 trainees, 20 were from Alaska, 9 were from American Samoa, 13 were from CNMI, 8 were from Guam, 11 were from Hawai'i, and 18 were from FAS. The distribution of the 18 FAS staff was: one from Palau, five from RMI, one from Chuuk, four from Kosrae, five from Pohnpei, one from Yap, and one who was posted at the Coordinating Center. Among the 79 trainees, 57 participated in at least two sessions, 29 participated in at least three sessions, and seven participated in more than three sessions.

Anthropometric standardization results

First session results—Among the 79 trainees, 58% ($n=46$) passed weight and another 19% ($n=15$) passed weight at borderline using the Standard Zerfas criteria. For height, only 13% ($n=10$) passed and another 23% ($n=18$) passed at borderline. For waist, only 6% ($n=5$) passed and another 10% ($n=8$) passed at borderline.

Although the Alternate Zerfas criteria were not used for the evaluation of an individual's first training session, the results would be much improved had the Alternate Zerfas criteria been used. For example, for weight, 73% ($n=58$) would have passed and another 20% ($n=16$) would have passed at borderline. For height, 49% ($n=39$) would have passed and an additional 27% ($n=21$) would have passed at borderline. For waist, 21% ($n=17$) would have passed and an additional 25% ($n=20$) would have passed at borderline.

Second session results—Among the 79 trainees, 57 (72%) participated in a second training session. Measurement improvement was seen in all three components (Fig. 2). The percentage of trainees that passed or passed at borderline for waist by the Standard Zerfas criteria increased by 50%, from 14% in the first session to 21% in the second training. For height, there was a 59% increase in the pass rate (from 32% to 51%). For weight, there was a 24% increase (from 74% to 92%). In contrast, increase in rates of passing or passing at borderline for all three components was substantially smaller when the Alternate Zerfas criteria was used, e.g., 3% for weight, 26% for height, and 45% for waist.

Overall session results—Among the 79 trainees, a total of 49 (62%) passed the criteria and were qualified to measure all three anthropometric components in the field. Among those 49 trainees, 2 (4%) passed at their first training session, 23 (47%) passed by their

second session, and 17 (35%) passed at their third session. The rest, seven (16%), passed at a session/s subsequent to their third session.

Intra- and inter-rater TEM, average bias, and reliability score

Intra- and inter-rater TEMs, average bias, and the coefficient of reliability score (R) statistics were calculated first for all 79 trainees, then for the 49 trainees who passed the criteria for all three anthropometric components and the 30 who did not pass (Table 3). Blunder types of errors were removed from these calculations. The intra-rater TEMs for the components were relatively similar between the 49 who passed the criteria and the 30 who did not pass, indicating that the variability of measurements within trainees about their respective means was small, regardless of whether they passed the Zerfas criteria. However, the inter-rater TEMs were much larger for the 30 who did not pass than for the 49 who did pass for all components, indicating that the measurements for the trainees who passed were close to a common mean for each child, while those that did not pass were not converging to a common mean. Intra-rater TEMs of the expert anthropometrist were 0.20 for height, 0.17 for waist, and 0.02 for weight.

The coefficient reliability scores were higher than 0.99 for all three components among those 49 who passed, compared to 0.95 for height, 0.80 for waist, and 0.82 for weight among those 30 trainees who did not pass. Average bias for both the height and waist was negative, -0.32 for height and -0.19 for waist among those 49 trainees who passed the criteria, indicating that in general, trainees tended to underestimate height and waist measures. The average absolute biases for both height and waist were larger for those that did not pass than those who did, as would be expected based on the Zerfas criteria. Average bias for weight was small at 0.01 among all trainees.

Lessons learned

Measurement from a minimum of eight children allows for the most efficient situation to evaluate the Zerfas criteria among trainees. To achieve the minimum number of measures, children were often measured multiple times; therefore, techniques to engage children in the training sessions were important. Successful strategies included starting the measurement with children who were more eager to be measured, or starting the measurement with an anthropometric component that required minimum touching (e.g., weight), or letting the child pick a measure to start. In addition, giving each child a small reward, e.g., a colorful sticker, after each measure, was a useful strategy.

Although having at least eight children for a training session was found to be beneficial, increasing the number of trainees in each session beyond 10 was difficult to manage. With more trainees, it became a challenge to ensure that all trainees measured the minimum number of children that were also measured by the expert anthropometrist within a reasonable time period. The development of a grid that trainees could use to track the children they had measured in comparison to the expert anthropometrist was found to be useful and motivating both to children and trainees. Lastly, practicing recording, with verification by the measurer, was just as important as measuring. Most of the blunders were due to recording errors.

DISCUSSION

CHL's anthropometric standardization process and protocols were innovative and unique in several ways. The process included the use of standardized measurement protocols and equipment at 11 US-Affiliated Pacific jurisdictions, the evaluation of measurement teams from all jurisdictions against a single expert anthropometrist, and the use of Zerfas criteria. Due to the multicenter design of CHL and the unique geographical distances, the risk of measurement error due to variation in technique and skill was high, and adherence to the study protocol between members of the field staff was important. For this reason, an emphasis was placed on the standardization of anthropometric measurements to minimize measurement error.

Criteria for anthropometric assessment have been put forward by Zerfas, using a repeated-measure protocol (Ulijaszek, 1997; Zerfas, 1985). The trainee and trainer measure the same individuals until the difference between the trainee and trainer is "good" by standard Zerfas criteria, or at the very least, "fair" by the alternate Zerfas criteria (Ulijaszek, 1997; Zerfas, 1985) (Table 1). Our experience was that the standard Zerfas criteria were very strict and few trainees could pass their first standardization training using this criteria, e.g., only two of the 79 trainees passed all three anthropometric components at their first training session by the standard Zerfas criteria.

Waist was the most difficult measurement for the trainees to pass, and weight was the least difficult. This is in accordance with findings from other studies (WHO Multi-centre Growth Reference Study Group, 2006; de Miguel-Etayo et al., 2014) which reported that the "problem" measurements were those that require careful positioning of the child. Accurate weight measurement required that the scale be level and the child stand still in the middle of the digital scale to provide accurate measures, a technique which was easily learned by the trainees. Accurate height measurement required that the child's posture be monitored and the headboard of the stadiometer be manipulated, a skill that took practice. Accurate waist measurement required that the trainee manage the posture and breathing of the child, manipulate the tape to ensure it is flat, level, and neither too lax nor too taut, a technique that required the most practice to learn. Most trainees had to learn to loosen the cinching of the tape and ensure that children were not holding their breath or holding in their abdomens.

Measurements from the 49 trainees who passed all criteria to become field anthropometrists were precise, as measured by the intra-rater TEM, which were less than twice the expert's TEM in all three components and were comparable to what were reported in the literature (Lohman et al., 1988). This is also shown in the *R* coefficients for all three anthropometric components, which were higher (all > 99%) than the 90% reliability threshold suggested by Marks and colleagues (1989) and what has been published in the literature (Chumlea et al., 1990; de Miguel-Etayo, et al., 2014; Johnson et al., 1997; Moreno et al., 2003; Perini et al., 2005; Stomfai et al., 2011; Ulijaszek and Kerr, 1999; WHO Multicentre Growth Reference Study Group, 2006). This might be explained by the CHL protocol that for each anthropometric component, the trainee measure the child at least three times with at least two measurements within two units (0.2 kg for weight and 0.2 cm for height and waist circumference).

The negative-signed bias for both height and waist indicated that trainees in general tended to underestimate those two measures. This may be explained by the lack in experience in positioning children or manipulating the instrument as compared to the expert anthropometrist, in particular, gaining full extension for height, with the head in the Frankfort plane, and waiting for the children to be sufficiently relaxed to not hold their breath and not cinching the tape too tight, for waist circumference. Nevertheless, the magnitude of biases for all three components in the team's measurements were much smaller than the maximum allowable difference set by the WHO Multicentre Growth Reference Study study at 2.8 times the expert's TEM (WHO Multicentre Growth Reference Study Group, 2006).

This is one of the first anthropometric standardization protocols that used a single expert anthropometrist. This protocol requires a great deal of effort, and in this case travel, on the part of the expert team. It was unclear at the outset whether this approach was necessary to ensure accurate measurement by anthropometrists in diverse settings with little direct supervision. Our experience demonstrates that this training is optimal, particularly for our large and diverse region and team composition. Precision and accuracy were high among those 49 trainees who passed the criteria for all three measures. The distinction between trainees who did and did not pass the Zervas criteria appears to be in accuracy more than in variability and precision of measurements. The small intra-rater TEMs for the 30 trainees who did not pass indicate that they generally obtained similar repeat measurements for a single child (i.e., low variability); however, the large inter-rater TEMs indicate that their measures were not close to a common mean for that child (i.e., low accuracy). Conversely, the 49 trainees who passed were able to achieve consistent repeat measurements about a common mean. The 49 trainees started with very different technique and skill levels and types of measurement errors, but all moved closer in their values to the experts' values and eventually passed, implying that a specific measurement technique can be achieved.

One of the goals of CHL's anthropometry measurement training was to build capacity and infrastructure for data system development and monitoring for nutrition and health (Novotny et al., 2013b; Wilkens et al., 2013). This was particularly important because there is currently no uniform anthropometry monitoring in the region, as NHANES does not include this region, and population obesity rates are higher than the contiguous US and known global prevalence rates (Bruss et al., 2010; Leon-Guerrero et al., 2008; Novotny et al., 2014; Ogden et al., 2014). The anthropometric standardization process and protocols followed by CHL have helped build capacity in the health workforce in the Pacific Region. The 49 standardized field anthropometrists are a first step of this capacity building.

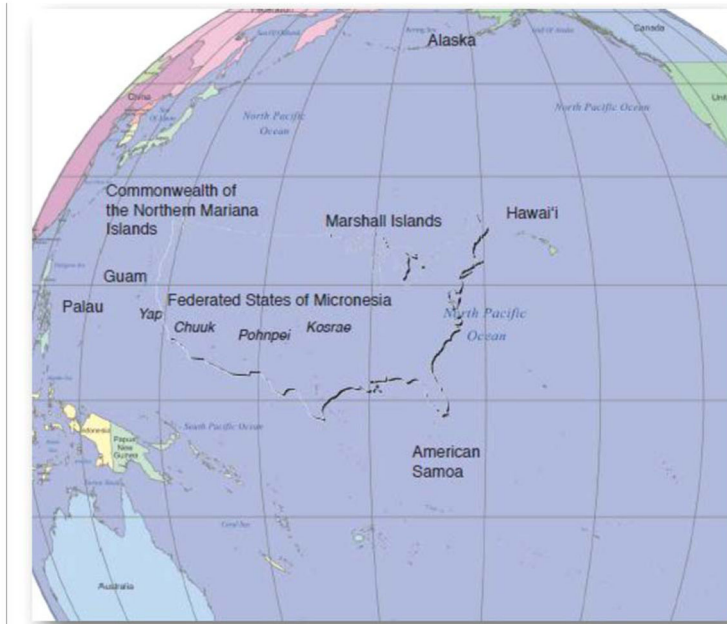
Acknowledgments

Grant sponsor: US Department of Agriculture/Agriculture and Food Research Initiative/National Institute of Food and Agriculture; Grant number: 2011-68001-30335; Grant sponsor: Children's Healthy Living Program for Remote Underserved Minority Populations of the Pacific (R. Novotny, PI).

References

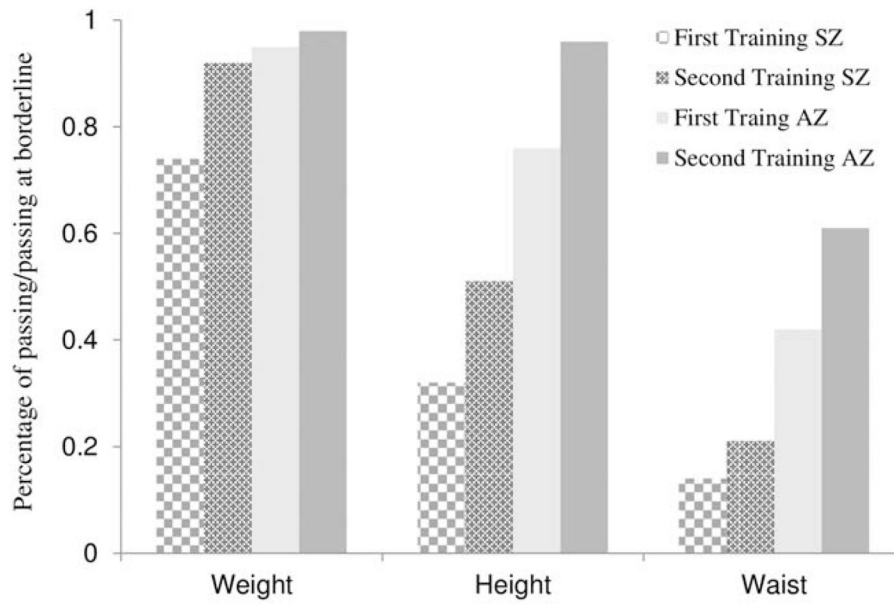
- Bruss MB, Michael TJ, Morris JR, Applegate B, Dannison L, Quitugua JA, Palacios RT, Klein DJ. Childhood obesity prevention: an intervention targeting primary caregivers of school children. *Obesity*. 2010; 18:99–107. [PubMed: 19424164]
- Chumlea WC, Guo S, Kuczmarski RJ, Johnson CL, Leahy CK. Reliability of anthropometric measurements in the Hispanic Health and *American Journal of Human Biology* Nutrition Examination Survey (HHANES 1982–1984). *Am J Clin Nutr*. 1990; 51:902S–907S.
- de Miguel-Etayo P, Measana MI, Cardon G, De Bourdeaudhuij I, Gozdz M, Socha P, Lateva M, Iotova V, Koletzko BV, Duvinage K, Androutsos O, Manios Y, Moreno LA. Reliability of anthropometric measurements in European preschool children: the ToyBox-Study. *Obes Rev*. 2014; 15(Suppl 3):67–73. [PubMed: 25047381]
- 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(1 Suppl):S27–S36. [PubMed: 15069917]
- de Onis M, Blossner M, Borghi E. Global prevalence and trends of overweight and obesity among preschool children. *Am J Clin Nutr*. 2010; 92:1257–1264. [PubMed: 20861173]
- Guerrero RT, Paulino YC, Novotny R, Murphy SP. Diet and obesity among Chamorro and Filipino adults on Guam. *Asia Pac J Clin Nutr*. 2008; 17:216–222. [PubMed: 18586639]
- Johnson TS, Engstrom JL, Gelhar DK. Intra- and interexaminer reliability of anthropometric measurements of term infants. *J Pediatr Gastroenterol Nutr*. 1997; 24:497–505. [PubMed: 9161941]
- Lohman, TG.; Roche, AF.; Martorell, R. Anthropometric standardization reference manual. Champaign, IL: Human Kinetics Books; 1988.
- Marks GC, Habicht JP, Mueller WH. Reliability, dependability, and precision of anthropometric measurements. The Second National Health and Nutrition Examination Survey 1976–1980. *Am J Epidemiol*. 1989; 130:578–587. [PubMed: 2764002]
- Moreno LA, Joyanes M, Mesana MI, Gonzalez-Gross M, Gil CM, Sarria A, Gutierrez A, Garaulet M, Perez-Prieto R, Bueno M, Marcos A. Harmonization of anthropometric measurements for a multicenter nutrition survey in Spanish adolescents. *Nutrition*. 2003; 19:481–486. [PubMed: 12781845]
- National Health and Nutrition Examination Survey (NHANES). [Accessed on July 8, 2014] Anthropometry procedures manual. 2014. Available at http://www.cdc.gov/nchs/data/nhanes/nhanes_07_08/manual_an.pdf
- Novotny R, Daida YG, Acharya S, Grove JS, Vogt TM. Dairy intake is associated with lower body fat and soda intake with greater weight in adolescent girls. *J Nutr*. 2004; 134:1905–1909. [PubMed: 15284374]
- Novotny R, Coleman P, Tenorio L, Davison N, Camacho T, Ramirez V, Vijayadeva V, Untalan P, Tudela MD. Breastfeeding is associated with lower body mass index among children of the Commonwealth of the Northern Mariana Islands. *J Am Diet Assoc*. 2007a; 107:1743–1746. [PubMed: 17904934]
- Novotny R, Nabokov V, Derauf C, Grove J, Vijayadeva V. BMI and waist circumference as indicators of health among Samoan women. *Obesity (Silver Spring)*. 2007b; 15:1913–1917. [PubMed: 17712106]
- Novotny R, Oshiro CE, Wilkens LR. Prevalence of childhood obesity among young multiethnic children from a health maintenance organization in Hawaii. *Childhood Obes*. 2013a; 9:35–42.
- Novotny R, Fialkowski MK, Areta AAR, Bersamin A, Braun K, DeBaryshe B, Deenik J, Dunn M, Hollyer J, Kim J, Leon Guerrero RT, Nigg CR, Takahashi R, Wilkens LR. The Pacific way to child wellness: the Children’s Healthy Living Program for remote underserved minority populations of the Pacific region (CHL). *Hawaii J Med Public Health*. 2013b; 72:406–408. [PubMed: 24251089]
- Novotny R, Fialkowski MK, Li F, Paulino Y, Vargo D, Jim R, Coleman P, Bersamin A, Nigg CR, Leon Guerrero RT, Deenik J, Kim JH, Wilkens LR. Systematic review of prevalence of young child overweight and obesity in the United States-Affiliated Pacific Region compared with the 48 contiguous states: The Children’s Healthy Living Program. *Am J Public Health*. 2014:e1–e14.

- Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011–2012. *JAMA*. 2014; 311:806–814. [PubMed: 24570244]
- Paulino YC, Guerrero RT, Uncangco AA, Rosadino MG, Quinene JC, Natividad ZN. Overweight and obesity prevalence among public school children in Guam. *J Health Care Poor Underserved*. 2015; 26(2 Suppl):53–62. [PubMed: 25981088]
- Perini TA, De Oliveira GL, Ornellas JS, De Oliveira FP. Technical error of measurement in anthropometry. *Rev Bras Med Esporte*. 2005; 11:86–90.
- Stomfai S, Ahrens W, Bammann K, Kovacs E, Marild S, Michels N, Moreno LA, Pohlabein H, Siani A, Tornaritis M, Veidebaum T, Molnar D. Intra- and inter-observer reliability in anthropometric measurements in children. *Int J Obes (Lond)*. 2011; 35(Suppl 1):S45–51. [PubMed: 21483422]
- Ulijaszek SJ. Anthropometric measures. In: Margetts, BM.; Nelson, M., editors. *Design Concepts in Nutritional Epidemiology*. New York: Oxford; 1997. p. 289-311.
- Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. *Br J Nutr*. 1999; 82:165–177. [PubMed: 10655963]
- WHO Multicentre Growth Reference Study Group. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. *Acta Paediatr Suppl*. 2006; 450:38–46. [PubMed: 16817677]
- Wilkens LR, Novotny R, Fialkowski MK, Boushey CJ, Nigg C, Paulino Y, Leon Guerrero R, Bersamin A, Vargo D, Kim J, et al. Children’s healthy living (CHL) program for remote underserved minority populations in the pacific region: rationale and design of a community randomized trial to prevent early childhood obesity. *BMC Public Health*. 2013; 13:944. [PubMed: 24107083]
- Zerfas, AJ. *Checking continuous measures: manual for anthropometry*. Los Angeles, CA: Division of Epidemiology, School of Public Health, University of California; 1985.



*The overlay of the contiguous U.S., set between Hawai'i and Guam, indicates that the distance width of the contiguous U.S. at its widest point, is 400 miles short of the distance between those two Pacific Islands (Guam and Hawaii) that participated in CHL.

Fig. 1. Map of the 11 jurisdictions across the Pacific Region participating in the Children's Healthy Living (CHL) Program.



*SZ: the standard Zerfas criteria; AZ: the alternate Zerfas criteria

Fig. 2. Improvement in anthropometric measures among trainees participating in the Children’s Healthy Living Program—first to second training sessions ($n=57$).

Evaluation of measurement error for the Children’s Healthy Living Program anthropometry trainees for each child by the Zerfas criteria (modified from Zerfas, 1985)

TABLE 1

Measurement	Zerfas Criteria ^a	Difference between trainee and the expert anthropometrist				Blunder (gross error)
		Good	Fair	Poor		
Height and arm circumference (cm) ^b	SZ	0-0.5	0.6-0.9	1.0-1.9	2.0	
	AZ	0-0.9	1.0-1.9	2.0-2.9	3.0	
Weight (kg)	SZ	0-0.1	0.2	0.3-0.4	0.5	
	AZ	0-0.2	0.3-0.4	0.5-0.9	1.0	

^aSZ, standard Zerfas; AZ, alternate Zerfas.

^bThe Zerfas criteria for height and arm circumference were applied to waist circumference in this study.

TABLE 2

Expanded Zerfas criteria for overall evaluation of each trainee as Pass, Borderline Pass, and No Pass in the Children’s Healthy Living (CHL) program for anthropometric standardization

M (number of children measured)	Pass		Borderline Pass		No Pass	
	Number of unsuccessful measurements (n_F) for Pass ^a	Probability of Pass = $\text{prob}(X \leq n_F M, P)$	Number of unsuccessful measurements (n_F) for Borderline pass	Probability of Borderline Pass = $1 - \text{probability of Pass} - \text{probability of No Pass}$	Number of unsuccessful measurements (n_F) for No Pass	Probability of No Pass = $\text{prob}(X > n_F M, P)$
5	0	0.7738	=1	0.2036	2	0.0226
6	0	0.7351	=1	0.2321	2	0.0328
7	0	0.6983	=1	0.2573	2	0.0444
8	1	0.9428	=2	0.0534	3	0.0038
9	1	0.9288	=2	0.0628	3	0.0084
10	1	0.9139	=2	0.0746	3	0.0115
11	1	0.8981	=2	0.0867	3	0.0152
12	1	0.8816	=2	0.0988	3	0.0196
13	1	0.8646	=2	0.1109	3	0.0245
14	1	0.8470	=2	0.1229	3	0.0301
15	1	0.8290	=2	0.1348	3	0.0362
16	1	0.8108	=2	0.1463	3	0.0429
17	2	0.9497	=3	0.0415	4	0.0088
18	2	0.9419	=3	0.0472	4	0.0109
19	2	0.9334	=3	0.0534	4	0.0132
20	2	0.9245	=3	0.0596	4	0.0159

^aThe number of unsuccessful measurements for a trainee are based either on the Standard Zerfas (SZ) or Alternate Zerfas (AZ) criteria in Table 1, respectively based on whether a more stringent or lenient approach is desired. In this study, SZ criteria were used at each trainee’s first training session and AZ criteria were used at his/her second or any subsequent training sessions.

Intra- and inter-rater (trainee in this study) technical error of measurement (TEM), bias relative to the expert anthropometrist, and reliability score from the 29 anthropometry standardization sessions from the Children's Healthy Living program^a

TABLE 3

Components	Reliability statistics	Expert anthropometrist	Among all 79 trainees	Among 49 passed the criteria	Among 30 not passing the criteria
Height	Intra-TEM	0.20 ^b	0.28 ^c	0.27	0.30
	Inter-TEM		1.44 ^d	0.30	2.58
	Average bias		-0.33 ^e	-0.32	-0.35
	R coefficient		0.984	0.999	0.951
Waist	Intra-TEM	0.17	0.32	0.30	0.38
	Inter-TEM		2.02	0.49	4.05
	Average bias		-0.20	-0.19	-0.23
	R coefficient		0.936	0.994	0.801
Weight	Intra-TEM	0.02	0.03	0.03	0.02
	Inter-TEM		1.69	0.06	3.11
	Average bias		0.01	0.01	0.01
	R coefficient		0.938	0.999	0.821

^aBlunder type of errors, defined as the mean difference between the expert anthropometrist and a trainee of 2 cm or larger for height and waist circumference, or 0.5 kg or larger for weight were excluded for the calculation of all statistics presented in this table.

^bThe intra-TEM of the expert anthropometrist was based on all measurements taken by the expert anthropometrist across the 29 training sessions for a specific measurement component.

^cThe intra-TEM among the trainees based on all measurements taken by all trainee across the 29 training sessions for a specific measurement component.

^dThe inter-TEM among all trainees, excluding the expert anthropometrist.

^eAverage biases relative to the expert were calculated using measurements from children measured by both the expert anthropometrist and a trainee.