USAID
FROM THE AMERICAN PEOPLE

ANTHROPOMETRIC DATA IN POPULATION-BASED SURVEYS

MEETING REPORT

WASHINGTON, DC
JULY 14–15, 2015

JANUARY 2016

This publication was produced for review by the United States Agency for International Development. It was prepared by the Food and Nutrition Technical Assistance III Project (FANTA) managed by FHI 360. The author's views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

This report is made possible by the generous support of the American people through the support of the Office of Health, Infectious Diseases, and Nutrition, Bureau for Global Health, U.S. Agency for International Development (USAID), under terms of Cooperative Agreement No. No. AID-OAA-A-12-00005, through the Food and Nutrition Technical Assistance III Project (FANTA), managed by FHI 360. The contents are the responsibility of FHI 360 and do not necessarily reflect the views of USAID or the United States Government.

Recommended citation: USAID. 2016. *Anthropometric Data in Population-Based Surveys, Meeting Report, July 14–15, 2015.* Washington, DC: FHI 360/FANTA.

Food and Nutrition Technical Assistance III Project (FANTA)
FHI 360
1825 Connecticut Avenue, NW
Washington, DC 20009-5721
T 202-884-8000
F 202-884-8432
fantamail@fhi360.org
www.fantaproject.org

ACKNOWLEDGMENTS

The USAID Anthropometric Data in Population-Based Surveys Meeting Report would not have been possible without the contributions of Dr. Reynaldo Martorell and Dr. Edward Frongillo, facilitators for the meeting; participants from the various organizations that participated in the meeting; the staff of the U.S. Agency for International Development (USAID)/Office of Health, Infectious Diseases, and Nutrition (HIDN) and in particular, Elizabeth (Betsy) Jordan-Bell who organized the meeting; and other USAID/Washington staff for their feedback and inputs into this meeting report. Many thanks also to FANTA staff Monica Woldt and Elisabeth Sommerfelt for developing the content of the report.

CONTENTS

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

CAPI	Computer assisted personal interviewing
CDC	United States Centers for Disease Control and Prevention
cm	centimeter(s)
DDL	Development Data Library
DHS	Demographic and Health Surveys
ENA	Emergency Nutrition Assessment
FANTA	Food and Nutrition Technical Assistance III Project
FEWS NET	Famine Early Warning Systems Network
GAM	global acute malnutrition
HH	household
MAR	mean absolute residual
MGRS	Multicentre Growth Reference Study
MICS	Multiple Indicator Cluster Surveys
MUAC	mid-upper arm circumference
NHANES	National Health and Nutrition Examination Survey
NCHS	National Center for Health Statistics
NNS	National Nutrition Survey
ODK	Open Data Kit
PAHO	Pan American Health Organization
PPS	probability proportional to size
SAM	severe acute malnutrition
SE	standard error
SMART	Standardized Monitoring and Assessment of Relief and Transitions
SDG	Sustainable Development Goal
TEAM	Technical Expert Advisory Group on Nutrition Monitoring
TEM	technical error of measurement
U.N.	United Nations
UNICEF	United Nations Children's Fund
USAID	United States Agency for International Development
USG	United States Government
WHO	World Health Organization

EXECUTIVE SUMMARY

Population-based surveys have been used in numerous countries to determine the anthropometric status of target groups, especially children under 5 years of age. These anthropometric data are used by host country governments, donors, and national and international partners to assess child malnutrition and monitor country progress and goals. In a number of countries where multiple types of surveys have been fielded, such as the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), and Standardized Monitoring and Assessment of Relief and Transitions (SMART) household surveys – used in both emergency situations and increasingly, to inform the development of national nutrition surveys (NNS), important differences in anthropometric results have occasionally been observed across survey types conducted in similar geographic locations and at close time points, causing confusion at country and global levels.

In response to these issues, the United States Agency for International Development (USAID) Nutrition Division hosted a technical meeting in Washington, DC in July 2015 to develop a shared understanding of the purposes, strengths, and challenges of these survey methodologies and provide recommendations for improving comparability of anthropometric data and accuracy of population estimates of nutritional status. The meeting provided an overview of the survey methodologies by experts in the various survey types; presentations on some of the differences in prevalence estimates of child malnutrition across the survey types and possible causes for these differences; considerations for moving toward harmonization of methodologies; and consensus on next steps (see Agenda, Appendix 1). Participants included United States Government (USG) agencies working in international nutrition (USAID, United States Centers for Disease Control and Prevention [CDC]), their key partners, representatives of United Nations (U.N.) agencies (UNICEF, the World Health Organization [WHO], and the Pan American Health Organization [PAHO]), and external nutrition experts (see participant list, Appendix 2). The meeting was facilitated by Dr. Reynaldo Martorell, Emory University and Dr. Edward Frongillo, University of South Carolina.

Participants shared an understanding that high-quality anthropometric assessment is required to produce credible, objective, valid, equivalent, and compelling information that can be used by decision makers at various levels. Accurate anthropometric data, while challenging to obtain, is critical for countries and other data users to focus programming appropriately to meet the needs of populations. Each survey system, whether it be DHS, MICS, SMART (for emergency contexts or NNS), or others, has strengths and opportunities for improvement. There is room for collaboration among the implementers of the various survey types and with national statistics offices to develop survey plans that consider the survey types themselves, country needs and constraints, and budget. There may also be some room to harmonize protocols and questionnaires across survey types, while some participants felt it was essential to harmonize indicator definitions.

Participants also agreed on the importance of collecting high-quality anthropometric data, especially length/height and the correct determination of age. There was a felt need to develop guidance on how to conduct good quality anthropometric assessment; improve training and supervision; and ensure representative sampling of clusters and within-cluster selection of households and individuals across geographic areas and socio-economic groups and over time (e.g., seasonality). It also became apparent that more detailed documentation of processes for training, field procedures, data cleaning, and reporting would provide data users with a greater understanding of the results and the context in which the data were collected, including any systematic differences in data quality. Measures to quantify data quality, including, for example, precision through reporting of the technical error of measurement (TEM), would be useful to gauge survey quality, in addition to other information such as the standard deviations of z-scores, uncertainty around sample estimates (standard errors [SEs]), the proportion of flagged cases, age heaping, and heaping on anthropometric measurements.

There was an acknowledged need for further guidance on a number of topics, including the appropriate use of various flags in different situations; the basis for switching from length to height (i.e., 24 months versus 87 cm) and in which situations; quality standards for equipment; statistics to report (mean, prevalence, standard deviation, confidence intervals) and for which age groups, demographic groups, and purposes; the importance and interpretation of standard deviations as a possible measure of data quality and/or heterogeneity; expectations regarding precision and accuracy; systematically reporting on seasonality and relevant contextual information and how to use these meta data; possible adjustments to data in surveys already conducted; and standard methods to evaluate the quality of surveys to flag when data fall below minimum standards and what should be done in such cases. Minimum standards for analysis and reporting will first need to be defined.

Proposed next steps to achieve consensus to improve quality of anthropometric data in population-based surveys included the following.

1. Develop guidance on the minimum technical documentation on how a survey was conducted.
2. Develop best practices for identification, selection, training, standardization, supervision, re-training, and reporting (e.g., TEM) of interviewers and their performance.
3. Examine whether breadth of surveys, large numbers of questions and duration of interviews, large sample sizes, large numbers of interviewers, and/or short survey periods may negatively impact quality of anthropometric assessment and how such impact might be reduced, considering interviewer training, stress, and fatigue; respondent burden and fatigue; and behavior of and interaction among interviewer, caregiver, and child.
4. Explore how supervisors may best incentivize and provide timely feedback to interviewers to do well and to prevent errors or correct them as they occur without introducing biases.
    - Review messages and/or eliminate information provided to interviewers when entering data that might bias their data-collection approach or reporting.
    - Consider taking duplicate measurements, with a trigger for a third if discrepancies exist, in all surveys, certain surveys, or sub-samples in surveys, if/when feasible.
    - Establish mechanisms to immediately identify fatigue so that appropriate action can be taken, e.g., developing alternative schedules.
5. Investigate possibilities and catalyze development of technology to help interviewers do their job more accurately and easily, e.g., improved equipment for measuring length and height, and tools to assist with age determination.
6. Develop setting-specific examples of best practices, which may be situation-dependent, for obtaining representative sampling of clusters for mapping and household listing in the sample clusters, for within-cluster selection of households and individuals, across physical and social gradients, and over time (e.g., seasonality).
7. Strengthen commitment and advocacy to ensure public access to raw data and develop a database (registry/repository) with survey data and protocols.
8. Review, and if needed, update the 1995 WHO guidelines on assessing survey data quality. This will ensure there are standardized approaches to assess data quality, with relevant indicators and thresholds, e.g., number of missing cases, digit preference, standard deviation of z-scores, proportion of extreme values, and other measures of quality.
9. Investigate whether and how best to adjust existing survey data for imprecision:
    - Shape of distributions
    - Heterogeneity across place, group, or time
    - Implications of providing revised estimates

WHO participants announced the forthcoming convening of a new WHO/UNICEF Technical Expert Advisory Group on Nutrition Monitoring (TEAM), which is expected to coordinate efforts in addressing issues around nutrition monitoring, including the collection and use of anthropometric data. As an entity convened by WHO and UNICEF, TEAM is uniquely positioned to move forward the next steps articulated during this meeting, given its function to advise on methods to improve the quality of nutrition monitoring; develop harmonized standards, tools and approaches; and identify emerging research questions and needs related to nutrition monitoring. USAID's Nutrition Division views the TEAM as the entity to provide leadership and, ultimately, global guidance on the issues that this meeting addressed and USAID will support UNICEF and WHO's TEAM as they assemble a relevant subcommittee and move forward with next steps.

1. BACKGROUND

Population-based surveys have been used in numerous countries to determine the anthropometric status of target groups, especially children under 5 years of age. These anthropometric data are used by host country governments, donors, and national and international partners to assess child malnutrition and monitor country progress and goals, and allow aggregation for deriving regional and global estimates and trends. In a number of countries where multiple types of surveys have been fielded, such as the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), and Standardized Monitoring and Assessment of Relief and Transitions (SMART) household surveys (for emergency situations and as part of national nutrition surveys [NNS]), important differences in anthropometric results have occasionally been observed across survey types conducted in similar geographic locations and at close time points, causing confusion at country and global levels. In response to these issues, the United States Agency for International Development (USAID) Nutrition Division hosted a technical meeting among United States Government (USG) agencies working in international nutrition (USAID, United States Centers for Disease Control and Prevention [CDC]), their key partners, representatives of UN agencies (UNICEF, the World Health Organization [WHO], and the Pan American Health Organization [PAHO]), and external nutrition experts to share and discuss methodologies and develop next steps to improve the quality of anthropometric data collection, analysis, and use.

The "Anthropometric Data in Population-Based Surveys" meeting was held July 14–15, 2015 at the FHI 360 Conference Center in Washington, DC. Dr. Reynaldo Martorell, Emory University and Dr. Edward Frongillo, University of South Carolina, recognized international experts in the subject of anthropometric survey methods, facilitated the meeting. The agenda for the meeting included an overview of the survey methodologies by representatives or expert users of the DHS, MICS, and SMART/NNS survey types; WHO's perspective on the use of population-based surveys for anthropometric data; presentations on the differences in prevalence estimates of child malnutrition across the survey types and possible causes for these differences; considerations for moving toward harmonization of methodologies; and consensus on next steps (see agenda, Appendix 1). Invited participants included individuals from CDC, the Famine Early Warning Systems Network (FEWS NET), the Food and Nutrition Technical Assistance III Project (FANTA)/FHI 360, Harvard University, DHS/ICF International, Nigerian National Bureau of Statistics, Ottawa Hospital Research Institute, PAHO, PATH, Tulane University, UNICEF, University of Aberdeen, USAID, and WHO (see participant list, Appendix 2). This report synthesizes the information shared through presentations and discussions during the meeting. A list of the presentations is provided in the references section of this document.

2. GOAL AND OBJECTIVES OF THE MEETING

The goal of the meeting was to develop a shared understanding of the purposes, strengths, and challenges of different survey methodologies (e.g., DHS, MICS, and SMART/NNS) to provide recommendations for improving comparability of anthropometric data and accuracy of population estimates of nutritional status. The specific objectives of the meeting included the following:

1. Develop a shared understanding of the methodologies and field work practices of the major surveys that collect anthropometric data including, for each survey, its purpose as well as its training, standardization, sampling, data collection, data cleaning, processing, analysis, reporting, and data-sharing methodologies.
2. Develop a shared understanding of how potential errors in measurement (weight, height, and age) and selection introduced during data collection may affect the prevalence of acute and chronic

malnutrition, and reach consensus on best methods to determine the validity and reliability of anthropometric data collected through household surveys.

3. Articulate best practices in data collection and reporting, as applicable, for different types of surveys in the different factors that influence the anthropometric results.
4. Define areas for potential harmonization among survey methodologies, as well as define when and how data produced by different surveys may be harmonized and mutually used.
5. Determine steps throughout the survey process that can be applied to ensure and improve the quality of the anthropometric data collected.
6. Define next steps to contribute to global guidelines for the appropriate use and interpretation of anthropometric data collected through different survey methodologies.

3. GENERAL OVERVIEW OF THE SURVEY METHODOLOGIES

The meeting focused primarily on anthropometric results obtained from the DHS, MICS, and SMART/NNS. Meeting participants representing each survey type provided brief overviews of each survey methodology. Participants shared when and why each survey type first came into use and general descriptions of the purpose and content of each survey type. The DHS surveys are implemented as part of the DHS Program, a USAID-funded program that provides technical assistance to countries to improve the collection, analysis, and presentation of population, health, and nutrition data and facilitates data use for planning, policy-making, and program management. Since 1984 there have been over 320 nationally representative surveys in more than 90 countries. The UNICEF MICS program was initiated to fill data gaps on children's and women's well-being, including tracking progress toward the World Summit for Children Goals in 1995 and 2000, when the MICS1 and MICS2 surveys were conducted, respectively. The MICS5 is currently in its fifth round and is now focused on the final data collection for the Millennium Development Goals and the Sustainable Development Goals (SDGs) baseline setting. While MICS1 had just over 100 questions, MICS5 has about 750 harmonized or validated questions. MICS6 will be pilot tested in 2016, pending finalization of the SDGs and indicators. The SMART methodology was initiated in 2002 based on a felt need to reform and harmonize assessments of and responses to emergencies and for surveillance to ensure that policy and programming decisions are based on reliable, standardized data and that humanitarian aid is provided to those most in need (SMART 2015). The SMART methodology is currently used for both emergency nutrition surveys and to generate figures for non-emergency settings. A total of 76 SMART/NNS surveys have been conducted in 18 of 24 west and central African countries from 2008 to 2014.

The DHS, MICS, and SMART/NNS vary in the breadth of data collected. The DHS and MICS are multi-topic surveys that collect data on household socio-economic status, reproductive health, child mortality, child health, nutrition, and water and sanitation, among other areas, with some unique topics within each survey type; for example MICS collects data on early childhood development, which most DHS surveys do not include, while DHS collects data on certain indicators of women's empowerment that MICS does not collect. SMART/NNS has generally been more nutrition-focused, collecting children's and women's anthropometric data and selected indicators of child health, nutrition, morbidity, and mortality.

All three surveys are conducted in collaboration with governments and partners; designed to be implemented in a fashion that strengthens local capacity to effectively collect, analyze, and use survey data; emphasize country ownership of the results; and are directed by a country-level steering or technical committee. The DHS, MICS, and NNS are intended to provide nationally-representative household level data based on standardized, internationally accepted indicators, and each survey also provides data at the sub-national level for certain indicators, as feasible depending on country-level need and budget, including anthropometric indicators. The SMART/NNS are often conducted every two years, annually, or two times a year, while DHS and MICS are generally conducted about once every three to five years.

Representatives from DHS, MICS, and SMART/NNS had very limited time to present on each survey type during the meeting, and it was noted that more details are needed for each survey to compare and contrast the surveys with regard to sampling, training, field work, data processing, data analysis, and reporting. WHO meeting participants developed a framework to collect more details from representatives of each survey type, and information was in the process of being collected at the time this report was written. The product is intended to be a table demonstrating the similarities and differences in approaches among these surveys.

4. DIFFERENCES IN PREVALENCE ESTIMATES ACROSS SURVEY TYPES AND POTENTIAL CAUSES OF DIFFERENCES

Divergent anthropometric results collected through DHS, MICS, and SMART/NNS between 2003 and 2014 in Nigeria provided the impetus for initial discussions, which ultimately led to the July 2015 meeting. In 2013 a DHS survey was implemented from about February through June while a large SMART/NNS survey was implemented from July through September. In most states the DHS survey results for wasting (weight for height < -2 z-score) and severe wasting (weight for height < -3 z-score) for children less than 5 years of age were much higher than the SMART/NNS results for severe acute malnutrition (SAM) and global acute malnutrition (GAM) for children 6–59 months of age.[1] Given that the surveys were conducted on the same populations and during months that would not be expected to show a seasonality effect, it was felt by both USAID and UNICEF staff in country that substantial quality problems must exist in either one or both surveys, since they could not both be right. Differences also existed in prevalence of stunting between the surveys.

Identifying the reasons for the differences in anthropometric results among surveys is important because the surveys often serve as important sources of information on population nutritional status. Governments, donors, and partners use the data from the surveys for planning, budgeting, and funding decisions. Data quality and consistency across surveys must be ensured to form a solid foundation for decision making.

The following paragraphs summarize the presentations and discussions by participants on the potential causes of the differences in anthropometric results across the survey types. Potential causes of differences are presented below in relation to sampling, interviewer training, aspects of data collection during field work, data analysis and reporting, and contextual factors related to the country or region where data may be collected. **Table 1** provides a summary description of each survey type for selected characteristics as shared during the meeting. The content below reflects what was presented or discussed during the meeting and is not an exhaustive description of each survey type in each area or possible causes of differences in anthropometric results across survey types. Although the information below does provide a

[1] The Nigeria DHS and MICS reported on the prevalence of wasting and severe wasting among children under 5 years of age, while the SMART/NNS reported on the prevalence of SAM (weight for height < -3 z-score or bilateral pitting edema) and GAM (SAM plus moderate acute malnutrition [weight for height ≥ -3 z-score and < -2 z-score]) among children 6–59 months of age. However, UNICEF Nigeria shared that the prevalence of edema was insignificant in the Nigeria SMART/NNS.

list of the many factors that may have led to differences in anthropometric results among the three survey types, further investigation is needed before consensus can be achieved.

Table 1. Summary Description of each Survey Type by Specified Characteristic

Characteristic	Survey type		
	DHS	MICS	SMART/NNS
Sampling	Cross-sectional household survey; 2-stage cluster sampling, cluster selection using PPS; HHs randomly pre-selected from HH listing; no replacement HHs		Cross-sectional household survey; 2-stage cluster sampling, cluster selection using PPS; large clusters may be segmented; HHs randomly or systematically selected, HH lists used when available or may be developed; for emergency situations – HHs selected in the field; protocol includes no replacement HHs
Training on anthropometry	3 days plus field practice	2.5 days plus one week of field practice	3–6 days including theory and practical sessions
Survey duration	3–5 months	2–3 months	1–2 months (< 2 weeks for smaller surveys)
Team composition	Supervisor, interviewers, 1–2 biomarker technicians, field editor	Supervisor, interviewers, measurer (for anthropometry), field editor	Supervisor (one for every 2 teams), team leader, two measurers (for anthropometry)
Field checks for data quality	Field-check tables		Cluster control form and ENA software
Age determination	Exact age in days based on year, month, and day of birth and visit date		Age based on year, month, and day of birth or if exact birth date unavailable, estimated in full months based on local calendar
Equipment	Seca 878 digital scales; Shorr boards, or boards similar to the Shorr board		
Flags used in data analysis	WHO flags		WHO or SMART/NNS flags
Public availability of survey data	De-identified survey data publicly available		Government endorses and releases survey results and authorizes release to individuals

PPS = probability proportional to size; HH = household; ENA = Emergency Nutrition Assessment; SMART/NNS flags use flexible exclusion ranges as described in the 1995 WHO Technical Report Series 854 (Physical Status, The Use and Interpretation of Anthropometry), and exclude cases that are greater or less than 3 standard deviations from the observed sample mean, rather than the reference mean.

Sampling. The DHS, MICS, and SMART/NNS are all cross-sectional household surveys that use two-stage cluster sampling, including selection of clusters using probability proportional to size (PPS) sampling in Stage 1. For Stage 2, DHS uses household listings and mapping to select households randomly within sampled clusters. Households are pre-selected from the household listing, with no replacement households, and the sample is implemented exactly as designed. The household listing and mapping of the cluster is carried out by survey staff in a separate operation from the data-collection activity. DHS recommends that households be pre-selected in the central office prior to the onset of field work and not by teams in the field. Design effects and weights are calculated and available, as are the non-response rates. Similarly, MICS uses simple random sampling for Stage 2, drawn on a census-based sampling frame of household listings, with no replacement, no departure from the sampling design, and normalized sample weights with full documentation.

SMART/NNS uses household listings to select households randomly within sampled clusters using simple or systematic random sampling. Large clusters (more than 250 households) may be segmented before systematic random sampling. If updated household listings are unavailable or too expensive to procure in advance, lists are developed in the cluster with key informants.[2] For SMART/NNS in both emergency and non-emergency settings, households are sampled with no replacement, the sample size is adjusted for expected non-response at the planning stage, and design effects are automatically calculated and included in survey reports.

Participants agreed on a need for transparency in reporting on sampling methods and household selection at the field level, and suggested the development of standardized guidance on sampling approaches as well as on documenting and archiving survey implementation at the field level, including clear descriptions of sampling frames, mapping procedures, selection of households and individuals, and when, how, and with whose support sampling frames are updated in urban and rural areas.

Training and re-training. Training time varies among survey types. DHS conducts four or more weeks of field training, with at least three days of training on taking anthropometric measurements plus field practice. The overall training for MICS depends on the size and content of the questionnaires, but overall the recommendation is for three weeks of training for paper-based surveys, with two and a half days for training on anthropometry and one week of practice taking anthropometric measurements. Trainees for the DHS break into pairs and practice setting up and checking the equipment, and conduct practice sessions in the classroom, health facilities/nurseries, and the field. DHS trainees are standardized in anthropometry against the facilitator (gold standard), assessed for inter- and intra-measurer variability, and taught how to determine child age. Equipment is also standardized.

Each field team in MICS and SMART/NNS includes a team member, or in the case of SMART/NNS, two members, exclusively for anthropometry, and most DHS teams now include one to two "biomarker technicians" who collect anthropometric data in addition to blood samples to measure other biomarkers such as anemia and HIV. The MICS team "measurer" receives a separate, additional training on measuring anthropometry and has a separate field manual. In the field the measurer functions exclusively for taking the anthropometric measurements, assisted by another team member who has also been trained in anthropometry. The "measurer" position was introduced into the MICS4 field team in response to concerns regarding data quality, and UNICEF participants shared that it resulted in significant improvements in data quality. MICS anthropometry training is conducted by experts in anthropometry. DHS and MICS include a pre-test of the questionnaires in advance of the main training to ensure that the questionnaires function as desired. After field practice with the interview teams, a pilot test is conducted, which serves as a full "dress rehearsal" prior to the start of the field work.

The SMART/NNS includes three to six days of instruction on the SMART methods, survey training with theoretical and practical sessions by experts, a standardization test, and a pilot test. The standardization test involves measuring a minimum of ten children twice as recommended by WHO. Technical error of measurement (TEM) is calculated to assess precision of the estimates against a standard threshold. Bias is calculated comparing measurers to the facilitator. Interviewers are retrained or replaced if they do not meet minimum requirements.

[2] NNS samples are based on the national sample frame, which may not be updated in times of emergency unless specifically requested by the national technical committee. Separate surveys or screening for malnutrition in emergency contexts are often conducted focusing on settlements of refugees and internally displaced people. These smaller surveys often select households in the field as these settings can experience significant population displacement and migration.

UNICEF felt that generally speaking, training materials were focused on the positioning of the child once they are on the board, but were missing the entire part that comes before getting the child on the board – which is key to ensuring a good measurement. Guidance is needed on how to explain to caregivers their role in the anthropometric measurement process and how measurers should best handle a child for measurement to make the experience less traumatic. Participants agreed on the need for solid training and standardization of anthropometric measurers and clear guidance on how sound training may be achieved, including identifying triggers for and guidance on retraining. Some participants felt strongly that an established national survey team would be needed, while others felt that it was important to consider new interviewers during each survey round, to ensure the most motivated and capable interviewers are selected for each survey.

Interviewer workload and respondent fatigue. The survey period for the DHS typically lasts three to five months, although the field work period has been longer in a few surveys, while MICS usually lasts two to three months, and SMART/NNS surveys one to two months, but less than two weeks for smaller surveys. The differences in survey duration are due to DHS and MICS collecting data on more variables than SMART/NNS. The DHS and MICS field teams include a supervisor, interviewers, and a field editor, while the SMART/NNS teams generally include a team leader responsible for the quality of his/her team's work, two measurers, and one supervisor for every two teams. As mentioned above, the DHS team may include one to two biomarker technicians, and the MICS teams include a measurer for the sole purpose of taking anthropometric measurements, although this individual may also conduct other modules like water quality testing.

Participants who developed and/or worked with SMART/NNS emphasized that teams were carefully organized to ensure a reasonable number of households could be visited each day to avoid excess work load and interviewer fatigue. However, the example of the 2011 SMART/NNS in Benin was shared where interviewer fatigue, as well as transportation difficulties in the field, may have resulted in large numbers of missing cases. The breadth of data collected in the DHS and MICS may also contribute to respondent fatigue and may affect the interaction between the interviewer and respondent when anthropometric data are collected. Participants agreed that more information is needed on the influence of interviewer and/or respondent fatigue on the quality of anthropometric data, as well as the most appropriate combination of questionnaire length, field team size, and work load to minimize interviewer and/or respondent fatigue and improve the quality of data that are collected.

Supervision and monitoring. The DHS, MICS, and SMART/NNS all include in-person supervision via observation. Supervisors observe anthropometric measurements in the field and children may be re-measured when values are outside a designated range. Participants who have worked with SMART/NNS shared that in addition to supervision by the supervisors and survey consultants, the SMART/NNS technical committee, which includes individuals from the ministry of health, UNICEF, and the national nutrition cluster, also make random visits to monitor survey teams. UNICEF participants said they recommend that local experts monitor the MICS data collection in the field, and while fieldwork monitoring does take place, it is not clear whether the monitoring is adequate. DHS shared that during the 2013 DHS survey in Nigeria the security situation in some of the northern states was very poor during the period of data collection, which decreased DHS staff capacity to provide the usual level of field supervision. Participants agreed that guidance would be useful regarding supervision and monitoring of field teams, including optimal supervisor-field team ratios, how supervisors and monitors can best support field teams to improve data quality, and how to address data collection, supervision, and monitoring when field situations may hinder quality data collection and possibly even place teams in harm's way.

Editing data in the field. DHS, MICS, and SMART/NNS surveys include checks for data quality. For those teams using computer assisted personal interviewing (CAPI) surveys or the Open Data Kit (ODK), data may be reviewed daily while teams are still in the field. DHS field-check tables for reviewing data quality include, per field team, information on the percent of eligible children a) measured, not present, or missing; b) with out-of-range length/height measures, z-scores, or incomplete date of birth; and c) since early 2015, the distribution of last digits for height and weight. Every household and/or respondent in the DHS survey must be visited at least three times before being treated as “not at home.” MICS field-check tables include the response rate, age distributions, flags, and information on heaping. SMART/NNS field teams use a cluster control form to review outcomes by household, such as missing or refused cases. Data are entered into the Emergency Nutrition Assessment (ENA) software either in the field or if not feasible, when teams return to the base. Team leaders meet with the survey manager to review the data or send data so that the manager can provide technical assistance and monitor the data for quality. The survey manager produces a plausibility check report either with data that has been entered during field work, or if that is not possible, after data collection is complete. The plausibility check report includes outliers (flags), age and sex distribution, age heaping, rounding (height, weight, mid-upper arm circumference [MUAC]), standard deviation, tests of normality, and missing anthropometric data. A composite data quality score based on this information is produced, ranging from 0–9 for excellent to more than 25 for problematic (10–14 is good and 15–24 is acceptable). Feedback on missing data, poor age estimation, age heaping, and rounding is shared with the survey teams.

Concerns were raised that the feedback SMART/NNS supervisors provide to interviewers may result in over-editing of data in the field and may suppress genuine variation within clusters or shift heaping from one digit to another during the course of data collection. However, representatives from SMART/NNS did not believe that over-editing in the field was taking place. Some meeting participants felt strongly that data collection teams should only learn to take the best possible measurement with a low TEM and that supervisors, rather than interviewers, should react to the data, provide guidance to improve quality of data collected by the field teams, and indicate when repeat measurements should be taken. There should be no incentives for teams to change the data to meet specific criteria. Participants also acknowledged the need to detect possible errors immediately so that children can be re-measured in the field, not at the end of the day when re-measurement may not be possible. The United States National Health and Nutrition Examination Survey (NHANES) includes a second anthropometric measurement and then a third when the deviation between the first two measurements is beyond a given range, which is a practice that increases the likelihood that erroneous measurements may be corrected and may also result in cleaner data. There was a suggestions that, similar to NHANES, double measurement and triggers for a third measurement should perhaps be considered for all surveys. However, many participants expressed that a third measurement may not be feasible in most settings. There is need for further clarification regarding when data checks are performed, at what frequency, by whom, and followed by what actions.

Age determination/age heaping. DHS and MICS survey teams collect data on child age and efforts are made to determine year, month, and day of birth to calculate exact age in days on the date of the household visit. If day of birth is missing, “98” (for “Don’t Know”) is recorded by the interviewer. For the purpose of calculating the age in days, the 15th day of the month is substituted for the “98” code in the central office. SMART/NNS survey staff also collect data on child age; however, it was not clear from the presentations and discussions if exact age or rounded age (up or down) is determined. Participants shared that child age can be one of the most difficult aspects of anthropometric data to collect in a survey, especially if the birth mother is no longer living or is not present. Errors in age other than date of birth later than date of visit are difficult to detect in the field. Ideally observers should be trained to probe as much as possible so that they obtain good and complete information on date of birth and date of visit. It is best for the actual age calculation to take place at the analysis stage. According to WHO, it is very important to determine age in days as accurately as possible, as the WHO child growth standards are in units of days and collecting age in months will provide inaccurate results.

Age heaping can be a problem in many surveys, especially for estimates of underweight and stunting. A simple histogram of survey data of age in months can be used to identify age heaping. Age heaping is usually associated with the problem of rounding age down, for example recording that a child is 24 months of age at any age from 24 to 30 months, but in some settings rounding up can also be a problem. Quantifying the extent to which age heaping takes place can help to decrease this problem, and can also be used as a factor to determine the comparability of data across surveys. For example, researchers from Tulane University presented histograms of age in months from DHS surveys in Ethiopia, and calculated the mean absolute residual (MAR) for age for each survey, which measures how far the frequencies of each age differ from the average, a higher number signifying a greater amount of age heaping. The MARs for the 2000, 2005, and 2011 DHS surveys were fairly similar (0.131, 0.146, and 0.151, respectively), while the MAR for the 2014 "mini DHS" (which was not conducted under the DHS Program) was 0.234, illustrating a greater amount of age heaping in the 2014 "mini DHS" and that it should perhaps not be compared to the previous three DHS surveys for trends.

Researchers from Harvard University also shared results of their assessment of the quality and validity of child anthropometric data for children 0–59 months of age from 45 DHS (1990–2012) and 28 MICS (2000–2011), selected from the WHO West and Central Africa Region data that were available in the public domain, and 27 NNS (2006–2012) provided to the researchers by UNICEF, given the NNS data are not in the public domain (Corsi et al. 2014). They found that digit preference for age was about the same for the MICS and SMART/NNS and slightly lower for the DHS. Simulation exercises they conducted with a sample of DHS and MICS datasets to induce heaping/digit preference in distributions for age found that inaccuracy in age could result in a 4.5 percent over-estimation in the prevalence of stunting and a 4.2 percent overestimation in the prevalence of underweight. The presenters felt that there were opportunities to catch errors in age while still in the field through improved training, supervision, and data quality checks in the field.

Digit preference in length/height or weight. The Harvard researchers noted above also presented results from the same analysis of DHS, MICS, and SMART/NNS on digit preference of the last digit in measures of length/height and weight. Digit preference in length/height or weight can often result in a disproportionately high number of length/height or weight values that end in ".0" or ".5." They found that digit preference was greater for height than for weight. Digit preference for height affected a higher percentage of cases for the DHS and MICS compared to NNS, but affected few cases for weight for any of the surveys. However, these findings should be considered in light of the fact that the NNS data provided to the researchers did not include the full household listings as did the DHS and MICS data, so it was not possible to clearly understand the extent of missing, implausible, or ineligible cases in the NNS data as it was in the DHS and MICS.

The Harvard researchers also conducted simulation exercises with the sample of DHS and MICS datasets to induce digit preference on distributions of height and weight, which indicated that digit preference for height (0.1 centimeters) was relatively unimportant in terms of its impact on prevalence, but that inaccuracy in weight (0.1 kilograms) was more important and could result in a 2 percent over-estimation of prevalence of underweight or wasting. The Harvard presenters felt that there was scope for improvement in training and supervision of field teams collecting anthropometric data, that implementation of consistency and range checks during field work could possibly catch errors, and that it may be beneficial to attempt additional re-visits to households prior to leaving a cluster. Participants noted that the errors in height introduced in the Harvard model may be smaller than common errors in height made in the field, and that it may be worthwhile to simulate the impact of larger errors in height similar to those commonly seen in the field.

Measurement of height when length should be measured, or vice versa. Researchers from Tulane also presented data from the 2011 Ethiopia DHS survey that showed that a substantial number of children 24 months of age or older had their length measured instead of their height, especially at 24, 25, 26, and 28 months of age, when more than half of the sample in these age groups was measured incorrectly. This was also a problem, but to a lesser degree, for children younger than 24 months, in which their height was measured instead of their length. A generally recognized "correction" for this is to add or subtract 0.7 cm to adjust for the gravitational/compressive forces which should make the same person measured lying be taller than if they were measured standing. Applying this factor to the 2011 DHS data in Ethiopia did not seem to correct the problem because at every age in months from 18 to 30, the average height of those measured standing was taller than the average height of those measured lying, ranging from a difference of about 2 centimeters to as much as 8 centimeters. This indicates that children were probably not just measured lying when they should have been measured standing, or vice versa, but that misreported age was also a likely problem. Given that nearly all of the children measured standing were taller at each age in months than those measured lying, this provides some evidence that recorded age could frequently have been rounded down.

Participants who implement SMART/NNS indicated that in their surveys a child less than 2 years of age or less than 87 cm is measured lying down and a child that is 2 years of age or older or equal to or taller than 87 cm is measured standing up. The SMART/NNS manual states that one of these criteria, either age or length/height, should be selected and used consistently throughout the survey. One participant recommended that actual height should be used as a criteria to determine if a child should be measured lying down or standing up, irrespective of age. This is an issue that could be discussed further. There was general agreement that it is important to quantify problems that surveys may have with correct measurement of standing height for children 24 months of age or older and recumbent length for children less than 24 months of age, and include better training and measurement tools to overcome these issues. However, much of the evidence presented points to the need to overcome problems in accurately determining child age.

Equipment. All three surveys use Seca 878 digital scales to measure weight and Shorr boards, or boards similar to the Shorr board, to measure length/height. Meeting participants from UNICEF felt very strongly that there is a need to improve equipment, for example, exploring the possible use of length boards with automatic numerical readout rather than difficult to read tape measures and light-weight materials that are easy to carry and assemble/disassemble, and reviewing whether locally-made length/height boards are acceptable. Some meeting participants discussed opportunities that may exist to reach out to colleagues in other disciplines, such as biomedical sciences or engineering, to explore solutions to the seemingly intractable problems around age determination or length/height measurement that were discussed during the meeting.

Seasonality. Seasonality can have a significant influence on prevalence of wasting. The research team from Tulane University found that the prevalence of wasting can differ by five percentage points depending on the season, emphasizing the importance of conducting surveys during the same period of the year and reporting on seasonal issues that may affect anthropometric results. A participant from UNICEF presented on the influence of seasonality on anthropometric results in Bangladesh, where wasting is consistently higher during the monsoon season compared to the dry season by about eight percentage points, influenced largely by poor food security. In the horn of Africa, seasonal effects typically result in a five percentage point increase in wasting, although it can be as much as 10 percentage points when situations deteriorate. The 2013 SMART/NNS in Mauritania found a 7.5 percentage point difference in GAM between December and July, and a difference of over 13 percentage points in some areas of the country. However, the SMART/NNS in Mauritania also at times showed very different prevalence of stunting over a relatively short period of time, for example, 21 percent in July 2013, 16 percent in August 2014, and 21 percent in December 2014, while differences in some regions were more extreme and clearly implausible (e.g., Trarza, 3 percent in August 2014 compared to 25 percent in December 2014). There was agreement among participants that more detailed reporting on the context could assist in better understanding these types of results, for example, reporting on movement of refugee populations, or other dynamics that could significantly influence results. It is important for data users to be able to distinguish between anthropometric results that are truly representative of the current situation, and those that may reflect an undue amount of error.

Shocks and crises. Shocks and crises can also influence the prevalence of wasting and underweight and result in differences across survey types. A participant from UNICEF shared an analysis of surveillance data collected in Bangladesh between 1992 and 2000 that showed that when rice prices increased, rice consumption remained unchanged, but child underweight increased because a higher expenditure on rice was accompanied by lower non-rice food expenditures, that is, decreased diet quality. The extent of declines or improvements in wasting and speed of response to changing situations depends in great part on the nature of the shock or crisis, as well as the response. More detailed reporting on shocks and crises can assist data users to better understand and use anthropometric data from various survey types.

Data entry and recording. CDC staff shared through simulations that data entry and recording mistakes can result in large non-directional errors in anthropometric measurement, that is, errors that result in z-scores outside the distribution that are biologically or statistically improbable. These types of errors can have a large influence on the estimated prevalence, as can the exclusion criteria/flags used (see below), especially for SAM. CDC showed that the Harvard analyses from West Africa suggest that there were systematic differences in the proportion of outliers for weight-for-height z-score by survey type (DHS 4.1 percent, MICS 4.4 percent, SMART/NNS 0.49 percent [Leidman 2015]), suggesting that these errors may have contributed to the differences in anthropometric results seen in the Nigeria case described earlier in this report. However, as indicated above, the findings should be considered in light of the lack of full household listings in the NNS data and lack of clarity on the extent of missing, implausible, or ineligible cases. During the discussion a caution was also raised that the CDC simulations focused primarily on weight-for-height z-score, and the conclusions pertain to weight-for-height, not necessarily to height-for-age or weight-for-age z-scores.

Systematic under- or over-estimation of weight, height, or age. CDC participants shared during the meeting that systematic under- or over-estimation in weight, height, or age can affect the mean z-score, and can occur, for example, if the scale is not correctly calibrated, weight of children is systematically measured with clothing on, ages are rounded down, etc. The effect of the error on prevalence depends on the direction of the error. Estimation of prevalence will be unreliable if data contain these types of systematic, directional errors. The CDC presenter also acknowledged that it is difficult to assess whether directional errors are present during analysis, and emphasized the need to prevent and correct these errors at the field level.

Measurement error. CDC demonstrated how small errors in measurement can affect the distribution of z-scores, resulting in wider distributions (see Section 5 below).

Selection of flags in data analysis. DHS and MICS use WHO flags, while SMART uses WHO flags or SMART/NNS flags.[3] There was concern among some participants that use of the significantly narrower SMART/NNS flags may suppress true variation in the data. It also seemed that in the case of Nigeria, SMART/NNS used different flags for different states, based on the central value of the specific anthropometric z-score for each state, which may also suppress variation from cluster to cluster and potentially exclude true values. SMART/NNS representatives shared that for the purpose of the meeting all of the presented data were analyzed using WHO flags, including the data from Nigeria.

Some meeting participants were concerned that WHO flags detect only the most extreme outliers, which they felt were usually a result of recording rather than measurement error. They indicated that very large measurement mistakes, e.g., 15–18 centimeters, would not be flagged by WHO criteria. WHO flags will exclude measurements that are biologically implausible, which often is not all measurements with errors. It was suggested that a group of independent experts on data quality should review the different flags for

[3] SMART/NNS flags use flexible exclusion ranges as described in the 1995 WHO Technical Report Series 854 (Physical Status, The Use and Interpretation of Anthropometry), and exclude cases that are greater or less than 3 standard deviations from the observed sample mean, rather than the reference mean.

cleaning biologically implausible measurement errors. CDC made the point in a presentation that if the anthropometric data are of high quality, then the selection of flags makes little difference on the prevalence estimate.

The Harvard researchers presented information on flagging and implausible values from their analysis of DHS, MICS, and SMART/NNS data. The researchers applied the same WHO flags to all the survey data. They found the percent of flagged or implausible values for height and weight were higher in DHS and MICS surveys compared to SMART/NNS. As noted above, the findings should be considered in light of the lack of full household listings for the NNS data and lack of clarity on the extent of missing, implausible, or ineligible cases.

Participants agreed that more guidance is needed regarding the selection and use of flags and their potential impact on prevalence estimates, standard deviations of z-scores, and their reflection on data quality, as well as reporting on use of flags, proportion of flagged cases, and response rates. One of the meeting facilitators commented that when looking at trends over time and results across survey types, in addition to looking at differences in prevalence estimates, it is also important to look at differences in mean z-scores, since the means will be less affected by use of different flags.

Reporting of anthropometric data. Although issues related to reporting of anthropometric data may not necessarily result in differences in anthropometric findings from different survey types, improved reporting can help data users better understand the results. WHO staff shared that WHO would prefer to see the results for national level surveys presented in a standardized manner, e.g., the percent of children with z-scores below or above standard cut-offs using WHO flags and age groups (0–5, 6–11, 12–23, 24–35, 36–47, and 48–60 months). Some meeting participants also shared the need for reports to include anthropometric results with confidence intervals. Means and standard deviations of z-scores should always be reported as well.

The adoption of the new 2006 WHO Growth Standards has resulted in a few changes in the determination of individual z-scores, as seen in **Table 2**. In addition, the new 2006 WHO growth standards have also resulted in a few differences in determination of population prevalence. For example, the National Center for Health Statistics (NCHS)/WHO reference used listwise deletion,[4] while the new WHO growth standards include all valid z-scores. The tabulation of all valid z-scores has been adopted by MICS and is being adopted for DHS-7 surveys.

Table 2. Some Differences between the NCHS/WHO Reference and the WHO Child Growth Standards with Regard to Individual Z-scores

NCHS/WHO Reference	WHO Child Growth Standards
Length up to 24 months or 85 cm	Length up to 24 months or 87 cm
Length-height conversion factor: ±1.0 cm Flags: WHZ -4 and +6 HAZ -6 and +6 WAZ -6 and +6	Length-height conversion factor: ± 0.7 cm Flags: WHZ -5 and +5 HAZ -6 and +6 WAZ -6 and +5 BAZ -5 and +5
Edema cases do not have WHZ or WAZ	Edema cases do not have WHZ, WAZ, or BAZ
n/a	Weight > 36 kg or < 0.9 kg set to missing Height > 138 cm or < 38 cm set to missing

Source: WHO, 2015. http://www.who.int/nutgrowthdb/software/Differences_NCHS_WHO.pdf. Note: Weight-for-height z-score = WHZ; height-for-age z-score = HAZ; weight-for-age z-score = WAZ; BMI-for-age z-score = BAZ.

[4] Listwise deletion is a method for handling missing data. In this method, an entire record is excluded from analysis if any single value is missing.

In addition to stratification by sex, age groups, urban/rural, and subnational regions, there are plans to expand the UNICEF-WHO-World Bank Group joint dataset under the new standards to also include standard errors (SEs), weighted and unweighted total number (N), stratification by wealth quintiles and mother's education, and measures of data quality.[5] WHO staff also shared that they felt that there was a need for more information in survey reports regarding training and how it is conducted; the number of interviewers trained and their workload; frequency of calibration of equipment in the field; data checking in the field during data collection and field work supervision; data cleaning procedures; deriving exact age; cases of edema (currently only consistently measured by SMART/NNS); seasonality issues; additional context such as natural or manmade disasters, epidemics, or other limitations encountered; missing data; and data quality. Quality measures found in the 1995 WHO Technical Report Series 854 (*Physical Status, The Use and Interpretation of Anthropometry*) included a review of histograms of age in months, incomplete date of birth and rounded age in months, height and weight digit heaping, missing data, and number of flagged records, which, as discussed above, are still relevant measures of quality today.

Release of data from anthropometric surveys. Releasing survey data for public use can also help users better understand the data and the context in which the data were collected, and allow for use of the data for the benefit of the population from which the data were collected. Both DHS and UNICEF make de-identified survey data publicly available. UNICEF participants shared that if a MICS survey produces poor quality results, UNICEF recommends that the data still be made public and that problems with data quality be addressed in the report (even if the report does not include mention of any nutritional status findings based on such data). For SMART/NNS data, the government is responsible for endorsement and release of the survey results, and provides the authorization to release the datasets to individuals or for access on the internet. All meeting participants agreed on the importance of producing data that are freely available. USAID released an open data policy in October 2014, and USAID-funded data collected after this date should be submitted to the Development Data Library (DDL).[6] The Harvard researchers also noted the importance of releasing minimally cleaned data, for example, pre- and post-application of flags, so that researchers may apply uniform flags to various data sets. There is a need for prescriptive guidance on what needs to be made available and in what form, including minimal data cleaning and documentation.

5. STANDARD DEVIATION OF Z-SCORES AND QUALITY AND/OR HETEROGENEITY OF ANTHROPOMETRIC DATA

The 1995 WHO Technical Report Series 854 recommended assessing quality of anthropometric data partly based on the standard deviation of the z-scores. A standard deviation greater than expected was associated with poorer quality data. Although some meeting participants placed a great deal of importance on having standard deviations of z-scores close to 1, others felt that too much emphasis was placed on the latter given a standard deviation greater than 1 could reflect heterogeneity in the population. The following summarizes several presentations and the discussion around this issue.

The Harvard researchers presented data on standard deviations from their analysis of the MICS, DHS, and SMART/NNS (as noted above) and shared that the standard deviations for height-for-age, weight-for-age, and weight-for-height z-scores were higher for the DHS and MICS compared to the NNS, applying the same exclusion criteria (flags), although as indicated above, the researchers were not clear on the extent of missing, implausible, or ineligible cases in the NNS data due to lack of full household listings. During

[5] More information regarding the UNICEF-WHO-World Bank Group joint dataset can be found at the following website: http://www.who.int/nutgrowthdb/data_entry_inf/en/.
[6] For more information about the USAID open data policy, see https://www.usaid.gov/data.

the discussion participants who have worked with SMART/NNS attributed the lower standard deviation to SMART methodology's emphasis on testing data quality with plausibility tests while the data collection teams are in the field, as well as to the SMART training, quality measurements, careful determination of the number of households to be visited each day to avoid team fatigue, and the limited additional data collected as a part of the survey.

A presenter for CDC shared that high quality anthropometric data should be normally distributed with a standard deviation of approximately 1, and that the standard deviation of the distribution is only affected by data quality and, to a limited extent, population heterogeneity. Examples were cited of the NHANES in the United States, which had a standard deviation for weight-for-height and height-for-age z-scores close to 1 from 2003 to 2010, despite the US being heterogeneous ethnically, and examples of DHS surveys in Brazil, a heterogeneous population, with a low standard deviation for weight-for-height z-score, and Cambodia, with a generally homogeneous population, that had a relatively high standard deviation for weight-for-height z-score. The case was also made that there is no relationship between the mean z-score and standard deviation, indicating that the shape of the distribution does not change as a population becomes more malnourished. However, during the discussion one of the meeting facilitators countered that it is not true that the shape of the distribution does not change as nutritional status of the population changes. Although height in a well-nourished population is normally distributed, more and more it is seen that weight will be a bit skewed and not necessarily normally distributed because of overweight and heterogeneity in countries. Other participants also pointed out that in terms of the factors that influence anthropometric indicators, such as water and sanitation and food security, the United States may be more homogeneous than other countries, such as India, that have greater inequality in these areas across population groups.

The CDC presenter also shared evidence that small non-directional errors in anthropometric measurement of weight and height included in weight-for-height z-scores have an effect on the standard deviation and overestimate wasting. The CDC presenter shared that the findings of the Harvard analysis suggested that there were systematic differences in the standard deviations for weight-for-height, weight-for-age, and height-for-age z-scores by survey type, that is, DHS, MICS, and SMART/NNS, suggesting that perhaps these kinds of errors may have contributed to the differences in anthropometric results seen among the surveys in Nigeria.[7] Using the data from the Harvard analysis, CDC further showed that higher standard deviations were highly correlated with other tests of data quality including digit preference and percent flagged values. The standard deviations for height-for-age, weight-for-age, and weight-for-height z-scores were positively correlated with the percent of flagged values, digit preference for height, and digit preference for weight ($p < 0.001$) for all tests using nonparametric analysis.

However, the DHS team also conducted an analysis of the Harvard data and did not find the data quality tests to be always significant when analyzed by survey type. The Harvard researchers agree with the latter and shared that the DHS did show the most variability in parameters such as standard deviation, but the DHS Program also had the largest number of surveys and covered the largest span of time. The standard deviations may have changed with time as nutritional status of the populations changed or improved. The Harvard researchers also shared that there is some ambiguity in the use of standard deviation as a measure of data quality. Given the standard deviation captures inherent population heterogeneity they felt there is no reason to assume that the standard deviation will be the same across all surveys. The Harvard researchers did acknowledge that poor data quality could inflate the standard deviation of anthropometric measures, but given that anthropometric z-scores are biologic parameters, they would anticipate that there would be some population heterogeneity both within and between countries even in situations of high quality data collection.

[7] Corsi and Subramanian, 2014. Mean standard deviation of weight-for-height z-score for DHS, 1.44; MICS, 1.45; and SMART NNS, 1.11; mean standard deviation for height-for-age z-score for DHS, 1.80; MICS, 1.82; and SMART NNS, 1.36.

DHS presenters expressed concern regarding the emphasis on standard deviations of height-for-age, weight-for-age, and weight-for-height z-scores close to 1 as an indication of quality. They felt this was based on two statistical fallacies: a) that the z-score distributions in the 2006 WHO growth standards are similar to standard normal distributions that arise in the context of, say, the Central Limit Theorem, when they are actually the result of fitting observations of height, weight, and age in a homogeneous well-nourished population, described thoroughly in the WHO documentation for the Multicentre Growth Reference Study (MGRS); this population is very different from the populations in the DHS and SMART/NNS surveys; and b) that the standard deviation of 1 should apply at all levels of aggregation, which is impossible for heterogeneous populations, given that a national standard deviation of 1 requires a standard deviation less than 1 at the state level and even lower values at the cluster level. In Kano state, Nigeria, for example, a majority of the within-cluster standard deviations were below 1, however, the average standard deviation in Kano state was more than 1. If the states are different, it is impossible for the standard deviation to be 1 in every state, and 1 for the country as a whole.

While there was no agreement on what is a reasonable standard deviation of z-scores to expect in heterogeneous populations, there was some agreement that 1 may be unrealistic in some situations and that very large standard deviations, for example greater than 2, might be a sign of poor quality. There is also a need to consider whether expectations around standard deviations of z-scores may vary by anthropometric indicator, e.g., height-for-age versus weight-for-age versus weight-for-height z-scores. Overall this is an important discussion that needs to be continued.

Adjusting data for imprecision -- A possible option?

One of the meeting facilitators shared that given large standard deviations of z-scores correspond to large TEMs, and excess TEM can cause substantial bias in estimation of stunting prevalence, a potential correction may be feasible by borrowing from measurement error models by adjusting the z-scores for the standard deviation of the distribution. This can be done by a) shifting the distribution to zero, b) dividing each z-score by the standard deviation, and c) re-shifting the distribution to restore the mean. This would preserve the shape of the distribution, does not bias the mean as does truncating z-scores below or above specified cut-offs using flags, and assists in obtaining accurate prevalence estimates for stunting. The method assumes large standard deviations result from large TEMs, and not biology. The presenter acknowledged that this potential option requires further development, but it evoked a great deal of interest among meeting participants. A representative of the SMART methodology also agreed that determination of the TEM was important to be able to adjust prevalence estimates.

During the discussion it was emphasized that regardless of this potential option, good quality data are still needed. A number of questions were raised, e.g., “What about situations where large standard deviations may be due to actual biological differences in the population rather than poor quality data?” Another option that was proposed was the use of a fitted model and analysis of residuals from fitted values, which may be feasible if the assumptions are correct; but the question remaining was, “What if the standard deviation is below 1; how is this interpreted?” In addition, there was concern expressed regarding day to day variability in weight that is not present when looking at length/height, and it was questioned how this should be considered.

6. MOVING TOWARD HARMONIZATION OF METHODOLOGIES: CONSENSUS AND NEXT STEPS

The meeting drew to a close with participants agreeing on a number of important points, as well as acknowledging areas where consensus was still necessary to achieve. There was agreement that high quality anthropometric assessment is required to produce credible, objective, valid, equivalent, and compelling information that can be used by decision makers at various levels. Accurate anthropometric data, while challenging to obtain, is critical for countries and other data users to focus programming appropriately to meet the needs of populations. Each survey system, whether it be DHS, MICS, SMART (for emergency contexts or NNS), or others, has strengths and opportunities for improvement. There is room for collaboration among the implementers of the various survey types and with national statistics offices to develop survey plans that consider the survey types themselves, country needs and constraints, and budget. There may also be some room to harmonize protocols and questionnaires across survey types, while some participants felt that harmonization of indicator definitions was essential.

Participants also agreed on the importance of collecting high-quality anthropometric data, especially length/height and the correct determination of age. There was a felt need to develop guidance on how to conduct good quality anthropometric assessment; improve training and supervision; and ensure representative sampling of clusters and within-cluster selection of households and individuals across geographic areas and socio-economic groups and over time (e.g., seasonality). It also became apparent that more detailed documentation of processes for training, field procedures, data cleaning, and reporting would provide data users with a greater understanding of the results and the context in which the data were collected, including any systematic differences in data quality. Measures to quantify data quality, including, for example, precision through reporting of the TEM, would be useful to gauge survey quality, in addition to other information such as the standard deviations of z-scores and uncertainty around sample estimates (SEs), the proportion of flagged cases, age heaping, and heaping on anthropometric measurements. In September of 2015 the DHS Program published "An Assessment of the Quality of DHS Anthropometric Data, 2005–2014" (Assaf et al. 2015). Although released after the meeting discussed in this report, some meeting participants shared that they found the document to be a comprehensive and useful review and that such analyses of anthropometric data should be encouraged in the future.

Although it is essential that all surveys use agreed upon good practice, how to implement them remained unanswered. There was an acknowledged need for further discussion and consensus on a number of topics, including the appropriate use of different flags in various situations; the basis for switching from length to height (i.e., 24 months versus 87 cm) and in which situations; quality standards for equipment; statistics to report (mean, prevalence, standard deviation, confidence intervals) and for which age groups, demographic groups, and purposes; the importance and interpretation of standard deviations as a possible measure of data quality and/or heterogeneity; expectations regarding precision and accuracy; systematically reporting on seasonality and relevant contextual information and how to use these meta data; possible adjustments to data in surveys already conducted; and standard methods to evaluate the quality of surveys to flag when data fall below minimum standards and what should be done in such cases. Minimum standards for analysis and reporting will first need to be defined.

Proposed next steps to achieve consensus to improve quality of anthropometric data in population-based surveys included the following:

1. Develop guidance on the minimum technical documentation on how a survey was conducted.
2. Develop best practices for identification, selection, training, standardization, supervision, re-training, and reporting (e.g., TEM) of interviewers and their performance.
3. Examine whether breadth of surveys, large numbers of questions and duration of interviews, large sample sizes, large numbers of interviewers, and/or short survey periods may negatively impact quality of anthropometric assessment and how such impact might be reduced, considering interviewer training, stress, and fatigue; respondent burden and fatigue; and behavior of and interaction among interviewer, caregiver, and child.
4. Explore how supervisors may best incentivize and provide timely feedback to interviewers to do well and to prevent errors or correct them as they occur without introducing biases.

   a. Review messages and/or eliminate information provided to interviewers when entering data that might bias their data-collection approach or reporting.
   b. Consider taking duplicate measurements, with a trigger for a third if discrepancies exist, in all surveys, certain surveys, or sub-samples in surveys, if/when feasible.
   c. Establish mechanisms to immediately identify fatigue so that appropriate action can be taken, e.g., developing alternative schedules.

5. Investigate possibilities and catalyze development of technology to help interviewers do their job more accurately and easily, e.g., improved equipment for measuring length and height, and tools to assist with age determination.
6. Develop setting-specific examples of best practices, which may be situation-dependent, for obtaining representative sampling of clusters for mapping and household listing in the sample clusters, for within-cluster selection of households and individuals, across physical and social gradients, and over time (e.g., seasonality).
7. Strengthen commitment and advocacy to ensure public access to raw data and develop a database (registry/repository) with survey data and protocols.
8. Review, and if needed, update the 1995 WHO guidelines on assessing survey data quality. This will ensure there are standardized approaches to assess data quality, with relevant indicators and thresholds, e.g., number of missing cases, digit preference, standard deviation of z-scores, proportion of extreme values, and other measures of quality.
9. Investigate whether and how best to adjust existing survey data for imprecision:
   a. Shape of distributions
   b. Heterogeneity across place, group, or time
   c. Implications of providing revised estimates

WHO participants announced the forthcoming convening of a new WHO/UNICEF Technical Expert Advisory Group on Nutrition Monitoring (TEAM), which is expected to coordinate efforts in addressing issues around nutrition monitoring, including the collection and use of anthropometric data. As an entity convened by WHO and UNICEF, TEAM is uniquely positioned to move forward the next steps articulated during this meeting, given its function to advise on methods to improve the quality of nutrition monitoring; develop harmonized standards, tools and approaches; and identify emerging research questions and needs related to nutrition monitoring. USAID’s Nutrition Division views the TEAM as the entity to provide leadership and, ultimately, global guidance on the issues that this meeting addressed and USAID will support UNICEF and WHO’s TEAM as they assemble a relevant subcommittee and move forward with next steps.

7. REFERENCES

Assaf, S; Kothari, MT; and Pullum, T. 2015. *An Assessment of the Quality of DHS Anthropometric Data, 2005–2014*. DHS Methodological Reports No. 16. Rockville, Maryland, USA: ICF International.

Blössner, M and Borghi, E. 2015. "Using national population-based household surveys as a source for child anthropometry data, DHS, MICS, LSMS, SMART and other national nutrition surveys, a WHO perspective." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Blössner, M and Borghi, E. 2015. "Opportunities to Inform Global Processes." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Cassard, F. 2015. "National Nutrition Surveys with SMART Methods." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Corsi, DJ and Subramanian, SV. 2014. *Report on the Quality of Anthropometric Data on Children's Height and Weight in 100 Health and Nutrition Surveys in the West Central Africa Region. New York: United Nations Children's Fund*. (forthcoming).

Corsi, DJ and Subramanian, SV. 2015. "Assessment of Child Anthropometric Data Quality in the West Central Africa Region." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Crowe, S; Seal, A; Grijalva-Eternod, C; and Kerac, M. 2014. "Effect of nutrition survey 'cleaning criteria' on estimates of malnutrition prevalence and disease burden: secondary data analysis." *PeerJ* 2:e380.

Frongillo, E. 2015. "Three thoughts about data quality in anthropometric survey systems." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Frongillo, E. 2015. "Toward building a consensus: Key areas of agreement and proposals for next steps." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Golden, M and Grellety, Y. 2002. *Population Nutritional Status during Famine*. Department of Medicine & Therapeutics, University of Aberdeen, Scotland.

Golden, M. 2015. "SMART – Emphasis on Data Quality." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Kishor, S. 2015. "Demographic and Health Surveys Program, A USAID Funded Project Implemented by ICF International." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Krasevec, J. 2015. "Issues that can impact rates of malnutrition." Presentation made at Anthropometric Data in Population-Based Surveys, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Leidman, E. 2015. "Anthropometry data quality." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

McDonald, C. 2015. "Factors Informing the Selection of a Survey Methodology." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Mei, Z and Grummer-Strawn, LM. 2007. "Standard deviation of anthropometric Z-scores as a data quality assessment tool using the 2006 WHO growth standards: a cross country analysis." *Bulletin of the World Health Organization.* 85:441–448. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17639241.

Olarewaju, I. 2015. "Assessment of anthropometric data quality and Impact of data quality on survey results: Reflection on Nigeria's experience." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Pedersen, B. 2015. "MICS – An Overview." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Pedersen, B. 2015. "Implementation: Considerations and Best Practices." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Potts, K and Mason, J. 2015. "Issues on national survey methods." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Pullum, T; Arnold, F; Assaf, S; Kishor, S; and Kothari, M. 2015. "A Comparison of DHS and SMART Results: Weight-for-Age in Kano State, Nigeria, 2012–13." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

Quinley, J. 2015. "Nigerian Anthropometry Data, Experience at USAID Nigeria 2009–2015." Presentation made at *Anthropometric Data in Population-Based Surveys*, FHI 360 Conference Center, Washington DC, July 14–15, 2015.

SMART. 2015. "About SMART." Available at: http://smartmethodology.org/about-smart/.

UNICEF. 2013. Evaluation of Community Management of Acute Malnutrition (CMAM), Global Synthesis Report, UNICEF, Evaluation Office, May 2013.

WHO. 1995. "Section 5.4.4 Data integrity or quality measures." *Physical Status, The Use and Interpretation of Anthropometry*, *WHO Technical Report Series 854*. Geneva, Switzerland: WHO. pp 217-219.

WHO. 2015. "The standard analysis and reporting for the WHO Global Database." Accessed November 2, 2015 at http://www.who.int/nutgrowthdb/software/Differences_NCHS_WHO.pdf.

APPENDIX 1. AGENDA

Anthropometric Data in Population-Based Surveys

July 14–15, 2015 FHI 360 Conference Center, Washington DC

Agenda

Tuesday, July 14th, 2015

8:00–8:30	Breakfast & coffee
8:30–8:45	**Welcome** Katie Taylor, Deputy Assistant Administrator, Bureau of Global Health, USAID
8:45–8:50	**Goal and objectives of the meeting and introduction of Chairs** Anne Peniston, Chief, Nutrition Division, Bureau of Global Health, USAID
8:50–9:05	**Overview of agenda and introduction of participants** Reynaldo Martorell
Topic 1: Overview of survey methodologies and WHO 1995 Guidelines: Overview of each survey will include purpose, scope, approaches toward sampling, collection, aggregation, cleaning, reporting—in the field and at central level	
9:05–9:15	DHS, Sunita Kishor
9:15–9:25	MICS, Bo Pedersen
9:25–9:40	SMART, Fanny Cassard
9:40–10:05	Using national population-based household surveys as a source for child anthropometry data, Monika Blössner
10:05–10:25	Facilitated Q&A
10:25–10:40	Break
Topic 2: Differences in prevalence estimates of child malnutrition across survey platforms and potential causes of these differences	
10:40–10:45	Introduction of topic, Reynaldo Martorell
Part A: Assessment of anthropometric data quality and impacts of data quality on survey results	
10:45–11:05	Differences in prevalence estimates across survey platforms: The case of Nigeria, John Quinley
11:05–11:15	Reflection on Nigeria's experience, Isiaka Olarewaju
11:15–11:55	The assessment of data quality in anthropometric surveys and potential impact on prevalence estimates Edward Frongillo (11:15-11:25)
	Daniel Corsi (11:25 – 11:45)
	Kaity Potts (11:45 – 11:55)
11:55–12:30	Q&A of presenters Isiaka Olarewaju, John Quinley, Edward Frongillo, Daniel Corsi, Kaity Potts
12:30–1:15	Lunch

Part B: Presentations of further hypotheses that may explain varying estimates	
1:15–1:20	Introduction and framing, Reynaldo Martorell
1:20–1:55	The effect of editing on estimates of stunting, underweight, overweight, and wasting, Tom Pullum
1:55–2:15	Data quality and SMART surveys, Michael Golden
2:15–3:05	Errors in measurement of weight, height, and age best explain the differences in estimates, Eva Leidman
3:05–3:35	The effect of other factors (equipment, seasonality) in influencing data outcomes, Julia Krasevec
3:35–3:50	Break
3:50–4:15	Reflections from panel, Edward Frongillo, Christine McDonald, and Subu Subramanian
4:15–5:15	Facilitated discussion, including consensus on suggestions for further analysis Reynaldo Martorell and Edward Frongillo
5:15–5:30	Wrap-up and looking forward to Wednesday morning Reynaldo Martorell and Edward Frongillo

Wednesday, July 15th, 2015

8:00–8:30	Breakfast & coffee
8:30–8:40	Welcome and agenda/priorities for Wednesday, Reynaldo Martorell
8:40–9:00	Toward building a consensus: Key areas of agreement and proposals for next steps, Edward Frongillo
9:00–9:10	Q&A
Topic 3: Moving toward harmonization of methodologies: guidance on survey selection, training and implementation, analyses and interpretation, and appropriate use of results (Each presenter/key informant will use 10–15 minutes)	
9:10–10:00	Framing the topic: Opportunities to inform global processes on survey standards/guidelines and use of results from a global perspective, Monika Blössner/Elaine Borghi
	Factors informing selection of survey method, Christine McDonald
	Implementation: Considerations and best practices, e.g., survey design, training and supervision of enumerators, equipment, data handling- collecting, cleaning, and reporting, Bo Pedersen
	The interpretation and use of results from the country perspective, Isiaka Olarewaju
10:00–10:15	Break
10:15–12:00	Facilitated discussion and consensus building on next steps, Reynaldo Martorell and Edward Frongillo
12:00–12:15	Concluding remarks and farewell, Reynaldo Martorell and Edward Frongillo
12:15	Lunch

APPENDIX 2. LIST OF PARTICIPANTS

Anthropometric Data in Population-Based Surveys

July 14–15, 2015 FHI 360 Conference Center, Washington DC

	Name	Email address	Position	Affiliation
1	Sally Abbott	sabbott@usaid.gov	Nutrition Advisor, Bureau for Food Security	USAID
2	Lindsey Anna	lanna@usaid.gov	Monitoring & Evaluation Advisor, Bureau for Food Security	USAID
3	Fred Arnold	fred.arnold@icfi.com	Senior Fellow, ICF International; Technical Deputy Director, DHS	ICF International
4	Shireen Assaf	shireen.assaf@icfi.com	Senior Research Associate	ICF International
5	Bo Beshanski-Pedersen	bpedersen@unicef.org	Household Survey Specialist, MICS	UNICEF
6	Mounkaila Abdou Billo	mbillo@usaid.gov	Senior Health Advisor	USAID/Nigeria
7	Oleg Bilukha	obb0@cdc.gov	Associate Director of Science, Emergency Response and Recovery Branch	CDC
8	Monika Blössner	bloessnerm@who.int	Technical Officer	WHO
9	Elizabeth Bontrager	ebontrager@usaid.gov	Nutrition Advisor, Nutrition Division, Office of Health, Infectious Diseases, and Nutrition	USAID
10	Elaine Borghi	borghie@who.int	Statistician	WHO
11	Erin Boyd[1]	eboyd@usaid.gov	Nutrition Advisor, Office of US Foreign Disaster Assistance (OFDA)	USAID
12	Judy Canahuati	jcanahuati@usaid.gov	Senior Technical Advisor, Office Food for Peace (FFP)	USAID
13	Fanny Cassard	fcassard@gmail.com	Nutrition Consultant	Consultant
14	Daniel Corsi	daniel.corsi@gmail.com	Senior Epidemiologist	Ottawa Hospital Research Institute
15	Omar Dary	odary@usaid.gov	Health Science (Nutrition) Specialist, Bureau for Global Health	USAID

16	Madeleine Short Fabic	mshort@usaid.gov	Public Health Advisor, Office of Population and Reproductive Health	USAID
17	Rafael Flores-Ayala	rnf2@cdc.gov	Team Lead, International Micronutrient Malnutrition Prevention and Control Program	CDC
18	Edward Frongillo	efrongillo@sc.edu efrongil@mailbox.sc.edu	Professor	University of South Carolina
19	Michael Golden	mike@pollgorm.net	Emeritus Professor of Medicine	University of Aberdeen
20	Ruben Grajeda	grajedar@paho.org	Regional Technical Advisor, Nutrition	PAHO
21	Attila Hancioglu[1]	ahancioglu@unicef.org	Global MICS Coordinator	UNICEF
22	Elizabeth (Betsy) Jordan-Bell	ejordanbell@usaid.gov	Nutrition Advisor, Nutrition Division, Office of Health, Infectious Diseases, and Nutrition, Bureau of Global Health	USAID
23	Sunita Kishor	sunita.kishor@icfi.com	Project Director, The DHS Program	ICF International
24	Monica Kothari	monica.kothari@icfi.com	Senior Program Officer	PATH (DHS)
25	Julia Krasevec	jkrasevec@unicef.org	Statistics and Monitoring Specialist	UNICEF
26	Eva Leidman	wzu0@cdc.gov	Epidemiologist, Emergency Response and Recovery Branch	CDC
27	Raphael Makonnen	rmakonnen@usaid.gov	Nutrition Advisor, Nutrition Division, Bureau for Global Health	USAID
28	Reynaldo Martorell	rmart77@emory.edu	Professor	Emory University
29	Christine McDonald	cmcdonald@fews.net	Nutrition Advisor	FEWS NET
30	Zuguo Mei	zam0@cdc.gov	Epidemiologist	CDC
31	Isiaka Olarewaju	iolarewaju@nigerianstat.gov.ng ialarewaju@yahoo.com	Statistician, Programmer/Analyst	National Bureau of Statistics, Nigeria
32	Anne Peniston	apeniston@usaid.gov	Chief, Nutrition Division, Office of Health, Infectious Diseases and Nutrition, Bureau for Global Health	USAID
33	Kaity Potts	kstorck@tulane.edu	Public Health Consultant	Tulane University

34	Tom Pullum	tom.pullum@icfi.com	Senior Advisor for Research and Analysis, DHS	ICF International
35	John Quinley	jquinley@unicef.org	Senior Monitoring Officer, A Promised Renewed Secretariat	UNICEF
36	Elisabeth Sommerfelt	esommerf@fhi360.org	Scientist, PROFILES and MCHN Advocacy	FANTA/FHI 360
37	S (“Subu”) V Subramanian	svsubram@hsph.harvard.edu	Professor	Harvard University
38	Katherine (Katie) Taylor		Deputy Assistant Administrator, Bureau for Global Health	USAID
39	Melanie Thurber[1]	mthurber@usaid.gov	Nutrition Advisor, Office of Food for Peace (FFP)	USAID
40	Sonia Walia	swalia@usaid.gov	Public Health and Nutrition Advisor, OFDA	USAID
41	Monica Woldt	mwoldt@fhi360.org	Technical Advisor, MCHN	FANTA/FHI 360

[1] Please note these individuals participated by phone for part or all of the meeting.