Technical Notes from the background document for country consultations on the 2021 edition of the UNICEF-WHO-World Bank Joint Malnutrition Estimates

SDG Indicators 2.2.1 on stunting, 2.2.2a on wasting and 2.2.2b on overweight

unicef

World Health Organization

WORLD BANK GROUP

Contents

1. Introduction

The UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, in its role as joint custodian, is consulting government authorities on the three Sustainable Development Goal (SDG) indicators for child malnutrition that fall under Target 2.2 (stunting, wasting and overweight) of Goal 2, Zero Hunger.

This **background document** aims to provide national SDG focal points with details on the inter-agency efforts to gather and validate anthropometric data for the joint child malnutrition database, updated methods for **SDG indicators 2.2.1** (child stunting) and **2.2.2b** (child overweight). This document has three sections:

- **Section 1** provides an introduction, including details about Target 2.2, the JME and the country consultation process.
- **Section 2** provides the rationale for enhancing the methodology for deriving estimates for child stunting and overweight, including details on the previous and updated models and approaches.
- **Section 3** summarizes the methods employed to review and generate the database of national malnutrition estimates from primary data sources (e.g., from household surveys).

A. History of the Joint Malnutrition Estimates

The UNICEF-WHO-World Bank JME was established in 2011 to address the call for harmonized child malnutrition estimates. The first edition of the JME was released in 2012, providing estimates for stunting, wasting, severe wasting, overweight and underweight, as well as a detailed description of the methodology (1). The first JME was used to report against Millennium Development Goal 1 and has since been used to report on the progress towards the World Health Assembly 2025 Global Nutrition Targets (from 2012) and the SDGs (from 2015), as well as other global efforts. Since its inception, the JME outputs have been comprised of a harmonized country-level dataset based on primary data (e.g., national estimates based on household surveys), as well as regional and global model-based estimates.

The JME group releases annual updates of the estimates, which are disseminated through various knowledge products, including a key-findings report, an interactive dashboard and a series of databases (2). The JME group also maintains an expanded country dataset, with estimates disaggregated by sex, age, type of residence, maternal education, wealth and sub-national geographical regions (2). The JME group also undertakes methodological work, and to this end, initiated a process of enhancing the methods used to estimate child stunting and overweight in 2016. This work was recently finalized (*see Section 2*) and will be used for SDG reporting in 2021.

B. Malnutrition in the Sustainable Development Goals

Nutrition is central to the 2030 Agenda. The need for better nutrition is specifically highlighted through the Nutrition SDG Target (target 2.2): "by 2030 end all forms of malnutrition, including achieving by 2025 the internationally agreed targets on stunting and wasting in children under 5 years of age, and address the nutritional needs of adolescent girls, pregnant and lactating women, and older persons". Target 2.2 also lays the foundation for

Box 1. Malnutrition indicator descriptions and definitions for SDG Target 2.2

Stunting (Indicator 2.2.1)

Child stunting refers to a child who is too short for his or her age and is the result of chronic or recurrent malnutrition. Stunting is a contributing risk factor to child mortality and is also a marker of inequalities in human development. ***Stunting prevalence is defined as the percentage of under-fives falling below minus 2 standard deviations (moderate and severe) from the median height-for-age of the WHO Growth Standard population.***

Wasting (Indicator 2.2.2a)

Wasting refers to a child who is too thin for his or her height. Wasting is the result of recent rapid weight loss or the failure to gain weight. A child who is moderately or severely wasted has an increased risk of death, but treatment is possible. ***Wasting prevalence is defined as the percentage of under-fives falling below minus 2 standard deviations (moderate and severe) from the median weight-for-height of the WHO Growth Standard population.***

Overweight (Indicator 2.2.2b)

Child overweight refers to a child who is too heavy for his or her height. This form of malnutrition results from expending too few calories for the amount of food consumed and increases the risk of noncommunicable diseases later in life. ***Overweight prevalence is defined as the percentage of under-fives falling above 2 standard deviations (moderate and severe) from the median weight-for-height of the WHO Growth Standard population.***

achieving many of the SDGs, as good nutrition is a leading factor in sustainable development.

The indicators used to monitor the SDG Nutrition Target are stunting (2.2.1), wasting (2.2.2a) and overweight (2.2.2b) among children under 5 years of age (**Box 1**). (Anaemia among women of reproductive age has been added as a new indicator under SDG Target 2.2 this year; however, it is not included in the current country consultation.)

C. Why is there a different consultation process this year?

Country consultations on estimates of stunting, wasting and overweight for SDG Target 2.2 were previously undertaken in 2018 and 2019. During these consultations, national SDG focal points were given one month to review the most recent estimates based on primary data sources (e.g., estimates based on a household survey) for each of the three indicators for their country, and provide feedback on their most recent estimates. Given that the current country consultation includes estimates for stunting and overweight that are based on an updated methodology encompassing a country-level model, the consultation will be open for a two-month period, between 23 November 2020 and 15 January 2021. Further research on data for wasting is required before a country-level model (trends) are proposed, and thus, the estimates used to monitor wasting (indicator 2.2.2a) continue to be based on primary data sources at the country level and previous methods applied for regional and global estimates (see Annex 1).(1,4) *

2. Revised JME estimation methods for stunting and overweight

A. Rationale and summary of previous and current methods

In many global health analyses, country progress is examined using indicators of socio-economic, nutrition, or health conditions based on national surveys. A series of surveys in a country results in data over time that are summaries of individual data at the country level. National surveys are administered sporadically, resulting in sparse data for many countries. Furthermore, the trend of the indicators over time is usually not a straight line and varies by country. Tracking the current level and progress of indicators helps determine if countries are on track to meet certain thresholds, such as those indicated in the SDGs. In addition, estimating confidence and prediction intervals is vital to determine true changes in prevalence and identify where data are low in quantity and/or quality. For these reasons, the use of statistical models is essential to provide information needed for country and global monitoring.

The model used over the past 15 years for producing estimates of child malnutrition focuses on the sub-regional (for United Nations/SDG classifications), regional, and global levels (3). For each region, the estimates of prevalence at the sub-regional level are obtained from regression models of prevalence in relation to time in years. The models include terms called random effects that are at country level, allowing for flexibility in the variability within countries over time and among countries. The model is fit on a scale called the logit or log-odds scale to ensure that prevalence estimates (i.e., proportions) are bounded between zero and one, and the estimates are then back-transformed to the prevalence.

In generating a regional trend, the level of influence of the data from each country is proportional to its population. A robust estimator for the confidence interval is obtained using a well-accepted statistic called a sandwich estimator, also known as empirical covariance matrix estimator. Sub-regional level prevalence estimates are used to derive numbers affected, which are then aggregated to compute estimates at the regional and global levels. Confidence intervals at the aggregated levels are approximated using a standard (i.e., delta) method. This currently used model has three main disadvantages. First, the model assumes relative homogeneity within sub-regions and is dependent on how countries are grouped. Second, the model does not account for the precision of country prevalence estimates, which are the data points used in the analysis. Third, the model does not focus on country level, and thus estimates at that level are not generated.

To overcome these disadvantages, under a World Bank Knowledge for Change Trust Fund grant, the Joint Malnutrition Estimates group collaborated with a group from the University of South Carolina. The two groups worked together to propose models for stunting and overweight that would be understood by country offices and ministries of health, account for precision of country prevalence estimates, and allow for robust estimation of malnutrition trends at country level, recognizing the infrequency of prevalence estimates available for many countries. The models were initially developed for Africa (4) and have been expanded to all regions and updated based on a global technical consultation (5).

The models use robust methodology while being reasonably simple. They are flexible enough to allow for curvature in trends and variation, provide enough structure to run with moderate sample sizes, and are reproducible in standard software. Moreover, they can be understood conceptually by non-statisticians, incorporate difficulties in the data, and use covariates (i.e., other related variables with useful information) to improve prediction. The new models account for variation among countries and incorporate information from covariates and the region to which they belong (*see Annex 2*). They incorporate measures of primary data source sampling errors (SSE) and use methods to account for data sources with missing age categories. Country-level trends from 2000–2020 are produced for each region separately using a method called B-splines that allows for flexible curves along with uncertainty intervals that have been tested and validated.

B. Model details

i. Data preparation

The JME Working Group carefully reviewed new data available from countries (as well as the existing data in the 2020 Edition of the JME) to update and develop the country dataset on stunting and overweight (*see Section 3*). Where microdata are available, the JME Working Group recalculates estimates to provide a point estimate and an SSE based on the standard definition (*See Section 3*). As such, there can be discrepancies between point estimates found in reports and ones that are recalculated by the JME Working Group. These discrepancies may result from differences in how the ages were calculated (e.g., completed age in months versus exact age in months), the inclusion or exclusion of oedema data or the use of different flags to exclude biologically implausible z-scores (*see Section 3 for more details*).

Before the country level stunting and overweight models were run, two procedures were implemented on the primary source data to: (i) make the input data comparable to the age group of the standard indicator definition; and (ii) to cover gaps where SSE was not available. First, malnutrition estimates from primary data sources based on a non-standard age-range (e.g., 6–59 months instead of 0–59 months), and for which an age-adjustment was not previously carried out (Section 3d), were adjusted to be representative of the full 0–59-month age-range. To complete this step, a dataset with prevalence by age group[1] was created for all surveys with disaggregated data in the JME, which was used to perform the adjustment. This data set was also used to estimate patterns between the age group specific prevalence rates and overall prevalence via a linear mixed model. The modeled estimates and the observed age-group data were then aggregated to estimate the full 0–59-months prevalence for sources with a non-standard age range (*see Annex 3 for more information*). Second, SSE estimates for the primary data sources were predicted where missing, using either the reported confidence intervals (preferred) or by modeling the SSE when the first option was not available. The SSE model used the reported prevalence, (unweighted) sample size (if available) and data source type to estimate the missing SSE values. To prevent data sources with missing SSE from being overly influential to the analyses, a conservative estimate of the missing SSE was used. These aspects are discussed in more detail in **Annex 3**.

ii. Overview of the child stunting and child overweight country-level models

The technical details of the statistical models are provided in **Annex 3**. The paragraphs that follow outline the key aspects of the models for stunting and overweight. For both indicators, prevalence was modeled by closely fitting the country data points, with due attention to unwarranted variability. The model uses a logit (log-odds) scale and is a penalized longitudinal mixed-model with a heterogeneous error term. The quality of the models was quantified with model-fit criteria that balance the complexity of the model with the closeness of the fit to the observed data. Additionally, the fit of the model was examined by the JME Working Group on a country-by-country basis. The final models had good measures of model-fit, taking onto account JME expertise and lessons learned on nutrition-related survey and indicator patterns globally.

The proposed method has important characteristics, including non-linear time trends, regional trends, country-specific trends, covariate data and a heterogeneous error term. The model

1 The age groups for this database were: 0 to 5; 6 to 11; 12 to 23; 24 to 35; 36 to 47; 48 to 59 months.

can be broken down into portions that are estimated on a global, regional and country level. On a global level, an overall time-trend of the indicator and the impact of the covariate data are available. That is, the overall trend of stunting/ overweight on a global scale, which is allowed to be non-linear. All countries with data contribute to estimates of the overall time trend and the impact of covariate data on the prevalence.

For stunting, the covariate data consisted of a linear and quadratic socio-demographic index[2] (SDI) (6), the average health system access over the previous five years, and an indicator about data source type (Standardized Monitoring and Assessment of Relief and Transitions (SMART) survey (7) or other type of data source). An all-risk-factors summary exposure value for unsafe sanitation was also tested but it was not included it in the final model. For overweight, the covariate data consisted of linear and quadratic SDI, and the same data source types as applied for stunting. Data source type was considered as a covariate because the prevalences of overweight and stunting from SMART surveys tended to be lower than what was observed from other sources. In addition, the SMART (standardized monitoring and assessment of relief and transition) methodology was initially designed to provide a simple and harmonized assessment for responses to emergencies (8) with a strong focus on acute malnutrition. Based on these considerations, this covariate was used to decrease the model error. Note that the SMART methodology has been transitioning towards harmonization with global guidance (9), which is expected to impact future updates of the child malnutrition estimates.

The quantities that are estimated on a regional level vary by indicator. A regional-level intercept and linear time trend are included for stunting, while a non-linear regional trend was added to the model for overweight. The latter allows for more noticeable differences in the trends from region to region (and was also tested for stunting and found to diminish the model goodness-of-fit). For both indicators, the nine regions used in the modeling were Eastern and Central Africa, Southern Africa, Western Africa, the Middle East and North Africa, East Asia and Pacific, South Asia, Europe and Central Asia, High-Income Western Countries, and Latin America and the Caribbean (*Annex 2*).

At the country level, random intercepts, slopes (for stunting only) and non-linear functions are included to account for the difference between the observed data and the trend predicted by the global and regional portions of the model. The country-level model parameters are predicted for all countries that have at least one primary data point. For countries with no primary data, the values of the country-level parameters are set to zero (the mean and median of all country-level parameters). As a result, the predictions for countries with no data are driven by the global and regional portions of the model. These predictions will differ between countries within the same region since the inputted covariate data differs by country. (Modelled estimates for countries without any input data will only be used to generate regional and global aggregates but will not be included in the JME country database, and thus, will not be used for country-level SDG reporting).

Lastly, adjustments are made for the reported (or predicted) SSE for each data source. Data sources with larger SSE have more uncertainty. To account for this, a residual variance function was used that relates the SSE to the model error. The impact of this step is that data sources with lower SSE will have more influence on model estimates than those with large SSE (though both have influence). The residual variance function has two parameters: one that estimates the overall error and one that is used to offset how much error is due to sampling (SSE) and non-sampling error. For more details on the modeling approach, see **Annex 3**. All models were fitted in R software.

2 SDI is a summary measure that identifies where countries or other geographic areas sit on the spectrum of development. Expressed on a scale of 0 to 1, SDI is a composite average of the rankings of the incomes per capita, average educational attainment, and fertility rates of all areas in the Global Burden of Disease study.

C. Model outcomes

This section outlines the outputs of the model, provides examples of how the model fits, and considers why the predictions appear the way they do. The new JME method estimates the prevalence of stunting and overweight by country up to the current year of available data and includes the uncertainty around these estimates. The results can be used to calculate estimates and uncertainty intervals for any groups of countries aggregated. The uncertainty intervals are important in monitoring trends, especially for countries with sparse data and where primary data sources present large SSEs. When only sparse data are available in the most recent period, the inclusion of a survey can affect a substantial change in the predicted trajectory. For this reason, uncertainty intervals are needed to enhance trend interpretability in terms of the caution level employed. The uncertainty intervals for the new JME method have been tested and validated with various data types (3). The outputs have two types of uncertainty intervals: confidence intervals and prediction intervals; however, only the confidence intervals have been included in the country estimates file to be reviewed by the SDG focal points during this consultation. The confidence intervals are designed to contain (95% of the time) the true prevalence for the given country/year. The true prevalence can be thought of as the prevalence if the whole population was surveyed. The prediction intervals are designed to contain (95% of the time) estimates from primary data sources, such as surveys (input data). Since estimates from primary data sources have more variability than the true prevalence, prediction intervals are wider than confidence intervals.

Figure 1 provides six examples of predicted stunting prevalence, which demonstrate various aspects of the model. The plots for Burkina Faso and Peru show how the model fits for countries with many data points (both countries have over 10 primary data points since the year 2000). As can be seen, the model is able to fit the non-linear pattern apparent in the data. Since 2010, most of the Burkina Faso surveys have been SMART surveys (the light blue dots), which the model found to be slightly lower for stunting prevalence than other surveys, given the other inputs of the model are equal. As a result, the predicted prevalence is slightly higher than what would be predicted based on the SMART surveys. Similarly, the plot for Mauritania shows three SMART surveys and three 'other' surveys since 2007 (the 'other' surveys are all higher than the SMART surveys). The predicted prevalence is closer to the 'other' surveys, but also takes the SMART data into account. The 2017 Burkina Faso SMART survey appears to have little impact on the predictions. The proposed model is not flexible enough to capture quick fluctuations in the prevalence (i.e., it will bend but not break). This is because a model that is flexible enough to capture quick fluctuations would have deleterious impacts elsewhere in the results (e.g., for countries with little data) and increased uncertainty.

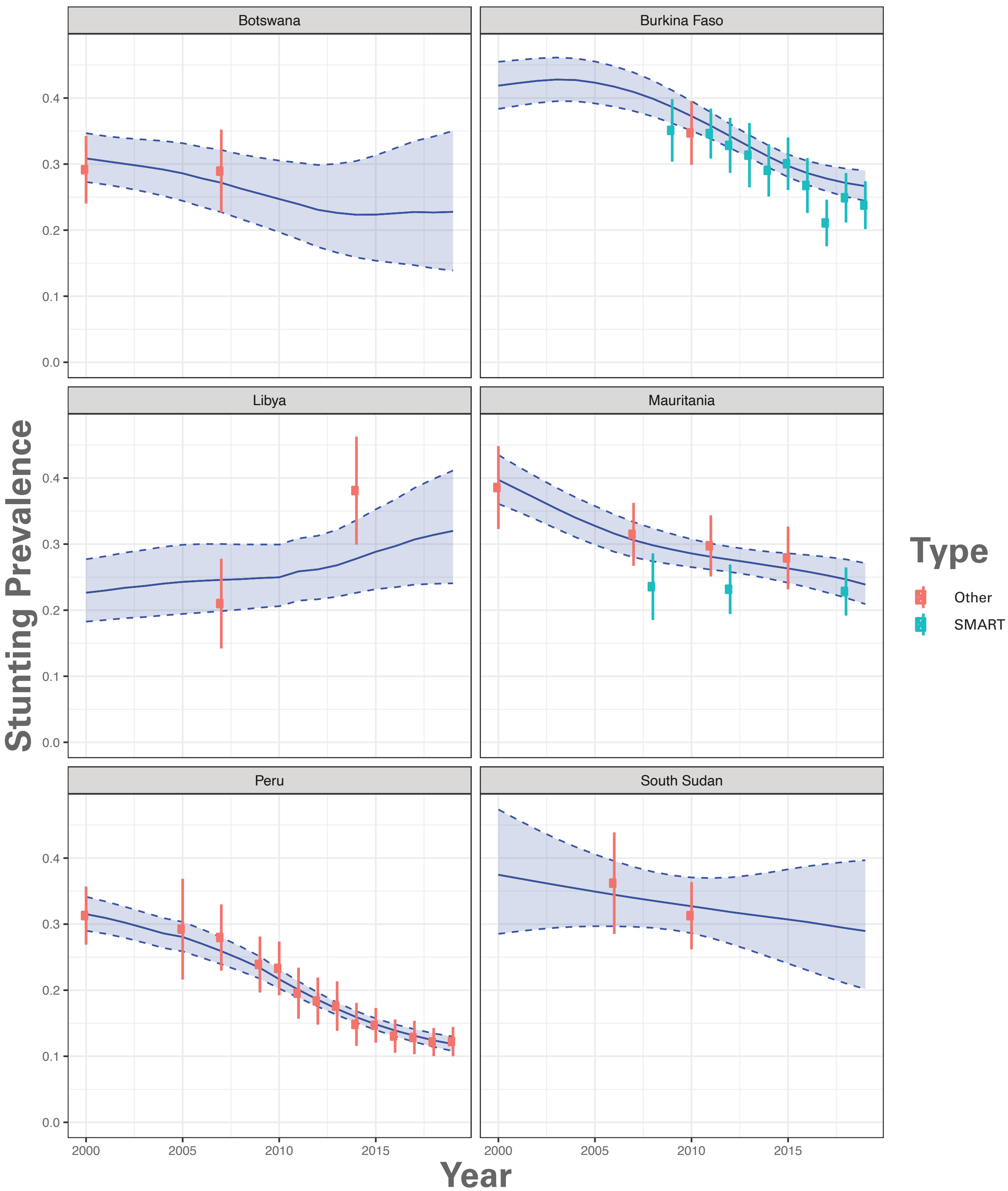
The predictions for Botswana and South Sudan are examples of countries with sparse data. The periods without data have increased uncertainty, which is substantial for the 2019 estimates. For example, Botswana in 2019 could have a stunting prevalence anywhere from 15 to 35%, and South Sudan anywhere from 20 to 40%. There are still trends in the predictions where there are no data (particularly evident in Botswana since 2014). These trends are driven by the covariate data and the overall regional trend.

The example for Libya demonstrates a unique case of a large increase in stunting prevalence over a relatively short time period. The predictions for Libya increase by more than 3 percentage points between 2007 and 2014 (when the two surveys occurred) and by more than 9 percentage points over the entire prediction period between 2000 and 2020. The increase in Libya is truly unique as no other country has had its prediction increase by more than 1 percentage point or its input data increase as much during such a short period of time. Of note is that the modeled estimates do not follow the input data points closely for Libya. In fact, given that the input data estimate was much higher than the modelled estimate (i.e., the survey estimate was 38.1 per cent in 2014 while the modeled estimate was 27.5 per cent in that same year), a larger increase in the prediction may have been warranted. One reason for the attenuated predicted increase is that the covariate data imply stunting prevalence is decreasing from 2000 to 2013, while remaining relatively flat thereafter. Additionally, the model used here tries to find an optimal mix of flexibility (so that it can fit to the observed data) and stability (so that predictions can be made for countries with sparse data); however, there will always be cases where the fit does not optimally capture a country's pattern.

Figure 2 provides examples of predicted overweight prevalence. Similar to the stunting model, the overweight model has non-linear trends that are driven by the primary source data points. For some countries, particularly those with many data points and smooth trends, the model follows the input data closely. For example, the predicted trends for Nigeria and the United States of America closely mirror the observed trends in the data. The United States of America example illustrates how the model can use many surveys with relatively high SSE to predict a trend with lower uncertainty.

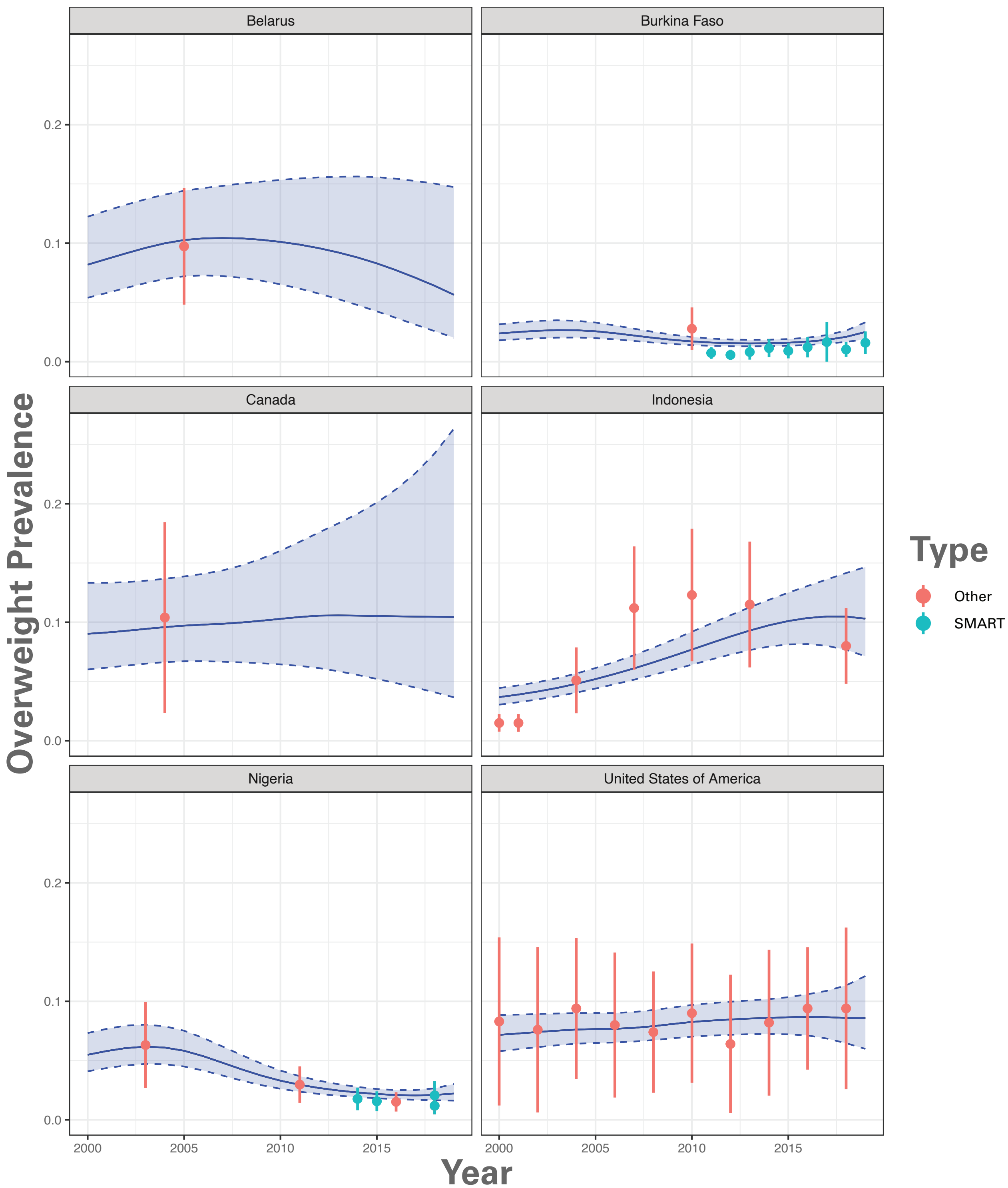
There are, however, some differences in how the stunting and overweight models fit the data. First, the covariate data are not as strong a predictor for overweight versus stunting. As a result, the overweight model has used region trends and the observed survey data more heavily. Second, the overweight model had a larger adjustment for SMART survey data source type than the stunting model. In both cases, the prevalence from SMART surveys was lower than what was expected based on the other factors in the model. The impact of this adjustment for primary data source type for overweight can be seen in the graph for Burkina Faso, where the predicted prevalence is consistently about 1 percentage point above what is reported in primary data sources based on SMART surveys since 2011. Third, the overweight primary source data have more variability between points than is seen for stunting, as shown by the example from Indonesia. Lastly, the regional trends are stronger. For example, both Canada (High-Income Western Countries region) and Belarus (Europe And Central Asia region) have only one primary data point around 2005 with a prevalence of about 10%. However, their trends in periods with no data are very different. This is because the High-Income Western Countries region has a relatively flat increasing trend and the Europe and Central Asia region trend increases until ~2006, then decreases thereafter. (The predictions for countries with no primary data points are driven by the regional trend and the covariate data, and have much higher uncertainty. However, as noted above, they have only been used to generate regional and global aggregates and will thus not be included in the JME country database or used for country-level SDG reporting.

Figure 1: Examples of the predicted stunting prevalence for Botswana, Burkina Faso, Libya, Mauritania, Peru, and South Sudan.



Note that the blue line shows the proposed JME estimates and the blue dotted line is the 95% confidence interval. The dots show estimates, and vertical lines show the standard error from primary data sources (blue for SMART surveys and red for 'other' data source types).

Figure 2: Examples of the predicted overweight prevalence for Belarus, Burkina Faso, Canada, Indonesia, Nigeria and United States of America.



Note that the blue line shows the proposed JME estimate and the blue dotted line is the 95% confidence interval. The dots show estimates, and vertical lines show the standard error from primary data sources (blue for SMART and red for 'other' data source types).

3. Primary data sources of stunting, wasting and overweight prevalence

A. Primary data sources

As of the latest review closure on 2 October 2020, the primary source dataset contains 957 data sources, including surveys – e.g., Multiple Indicator Cluster Surveys (MICS), Demographic and Health Survey (DHS), SMART surveys, Living Standards Measurement Study (LSMS) – and administrative data sources (e.g., from surveillance systems). The dataset contains the point estimate, and where available, the standard error, the 95 per cent confidence bounds and the unweighted sample size. Where microdata are available, this dataset contains estimates that have been recalculated to ensure that they adhere to the global standard definition. Where microdata are not available, reported estimates are used.

B. Data Compilation

The UNICEF-WHO-World Bank JME Working Group jointly review new data sources regularly, to update the country-level estimates. Each agency uses its existing mechanisms for obtaining data. WHO data gathering strongly relies on the organization's structure and network established over the past 30 years, since the creation of its global database, the WHO Global Database on Child Growth and Malnutrition, in late 1980's (4). For UNICEF, the cadre of dedicated data and monitoring specialists working at national, regional and international levels in 190 countries routinely provide technical support for the collection and analysis of data. UNICEF also relies on a data source catalogue that routinely collects data sources from other catalogues of other international organizations and national statistics offices. This data collection is done in close collaboration with UNICEF regional offices with the purpose of ensuring that UNICEF global databases contain updated and internationally comparable data. The regional office staff work with country offices and local counterparts to ensure the most relevant data are shared. Updates sent by the country offices are then reviewed by sector specialists at UNICEF headquarters to check for consistency and overall data quality, and re-analyzed where possible. This review is based on a set of objective criteria (see Section 3.E). The World Bank Group provides estimates available through the LSMS, which usually requires re-analysis of datasets given that the LSMS reports do not usually tabulate the malnutrition estimates. The World Bank's extensive Microdata Library provides access to the required datasets.

C. Reanalysis

Where microdata are available, the JME Working Group recalculates the estimates for stunting, wasting and overweight, taking the following parameters into consideration:

i. Calculation of dates

The standard approach (9) is to use a child's exact age in days (as opposed to age in completed months). This approach requires the full birth information (day, month and year). In cases where dates are missing or invalid (e.g., 31st September 2000), the day of birth is set to 15. Surveys that use a local events calendar and thus provide a child's age in months, require recovery of the date of birth using the approach outlined in **Annex 4**.

ii. Oedema

In 2019, the JME Working Group updated its analysis methodology to exclude oedema data when assessing a child's nutrition status. The reasoning for this decision is that malnutrition cases presenting as bilateral pitting oedema are uncommon in many countries, and more importantly, can be easily misdiagnosed (9). A child's weight, height and sex are thus exclusively used to assess child wasting and overweight. All reanalyzed estimates are derived without taking into consideration the reported cases of oedema. However, oedema is still considered for reported estimates (where no microdata are available) as there is no way to adjust them. Those sources have been noted in the 'Additional Notes' column of the Country Estimates file.

iii. Common sources of discrepancy between reported and re-analyzed estimates

Discrepancies between results reported by countries, and those reanalyzed applying the standard approach when microdata are available, may occur for various reasons. For example, the use of:

- different standards for z-score calculations (e.g. the National Center for Health Statistics (NCHS) growth reference) instead of the 2006 WHO Growth Standard used by the JME.
- a different method for imputation of the day of birth when missing (e.g., randomly assigning a day to each missing case) instead of the JME approach of using day 15.
- rounded/completed age in months (e.g., if a child was aged 2.9 months, the JME compares the child's height with the median height at 2.9 months). Other data sources could compare the child's height with the median height at 2 months (completed age in months) or 3 months (rounded age in months).
- different z-score flags to exclude children as being biologically implausible. The JME standard approach (9) considers all children greater than +6 or less than -6 z-scores for height-for-age as biologically implausible. This range is greater than +5 or less than -5 z-scores for weight-for-height. Some data sources exclude children from the calculation of all indicators if they have been flagged as implausible for any of the three indices of height-for-age, weight-for-age or weight-for-height. This standard approach only excludes a child if they were flagged as implausible for that specific index (e.g., stunting only considers height-for-age, wasting and overweight only consider weight-for-height).

D. Adjustments

Adjustments to reported values are made in cases where raw data are not available for re-analysis and it is known from the report that the estimates were derived based on indicators that do not adhere to the standard definition used for monitoring of the SDGs (e.g., they are based on different growth references, different age groups, etc.). The three types of adjustments that have been applied to the JME country dataset include adjustments to standardize for: (i) age, specifically for data sources that did not include the full 0–59-month age group (e.g., data sources reporting on 2–4-year-olds); (ii) area of residence, specifically for data sources that were only nationally representative at the rural level; and (iii) growth reference, specifically for data sources that used the 1977 NCHS/WHO Growth Reference instead of the 2006 WHO Growth Standards to generate the child malnutrition estimates. These three types of adjustments are described further in this section.

i. Age-adjustment

There are several nationally representative surveys that report on age groups that do not cover the entire 0–59-month age range in the standard definition for stunting, wasting and overweight. Adjustment for age is needed as malnutrition prevalence can vary by sub-age group. For example stunting prevalence among 24–59-month olds in the latest surveys with age-disaggregations available for Angola (10), India (11) and Lao (12), were more than two times higher than the stunting prevalence among 0–5-month olds. Surveys that omit part of the full age range might thus not be comparable with a survey that did cover all 0–59-month olds. Age adjustment can thus help to properly assess the country trend.

The adjustment method used by the JME group is to apply the relative proportions of malnutrition prevalence for each sub-age group from the closest survey in the country's JME dataset that covers the full age range and that includes disaggregated estimates by sub-age group, to the survey that covers only the smaller age range (simple rule of three). This is done under the assumption that the proportion of children in each sub-age group is equal throughout the 0–59-month age range (e.g., the number of children aged 0–11 months in the country is the same as the number of children aged 23–35 months) and also that relative prevalence of malnutrition across sub-age groups in the survey with the missing data is the same as in the survey with full information. This adjustment was applied to 63 surveys of the 957 included in the JME database 2021 edition (March 2021). There are 30 additional data sources in the dataset that do not cover the complete age range and have not been adjusted by the methods described here. The age discrepancy in these 30 data sources was addressed by the data preparation procedure described in section 2.b.i.

ii. Adjustment from national rural to national
A number of surveys cover only rural areas (26 out of 957 sources included in the JME database 2021 edition), and, while they have been sampled to be nationally representative for the rural parts of the country, they did not sample any urban areas. Given that malnutrition prevalence generally varies between urban and rural areas (i.e., stunting prevalence was reported to be two times higher in rural areas compared to urban areas at the global level (13)), a rural-only survey would not be comparable with a national survey that covered both urban and rural areas. To allow rural-only estimates to be comparable across time for the specific country, it is necessary to account for urban populations in estimates from these surveys. The method applied by the JME group is similar to that used for age adjustment – that is, to use another nationally representative survey that has disaggregated urban and rural estimates and apply relative proportions of malnutrition prevalence for each sub-group from that data source to the rural-only survey to adjust its estimate. Similar to the age adjustment, the proportion of children with malnutrition in rural and urban areas is assumed to be the same in the survey years in question.

iii. Adjustment to use the 2006 WHO Growth Standard (converted estimates):
The indicators of stunting, wasting and overweight used to track SDG Target 2.2 require a standard deviation (SD) score (z-score) to be calculated for each child who is measured for a data source; and the z-score requires a growth reference against which it can be calculated. Prior to the release of the WHO Child Growth Standards in 2006, the 1977 NCHS/WHO reference was recommended for international comparisons. The WHO Growth Standard results in estimates of stunting and wasting prevalence that are higher as well as estimates of overweight that are lower than estimates generated using the NCHS/WHO growth reference. (14); it was therefore necessary to account for these differences and standardize estimates across data sources. As such, data sources published prior to the release of the new growth standard in 2006 had to be re-analyzed to obtain comparable estimates across time and location. Raw data were not available for 178 of the 957 data sources included in the 2021 JME country dataset, and as such, algorithms were developed to convert estimates from surveys based on the NCHS reference to estimates based on the WHO Growth Standards (15). An Excel spreadsheet available online[3] can be used to implement the conversion.

E. Review of data sources for the dataset

UNICEF, WHO and the World Bank undertake a joint review for each potential data source. The group conducts a review when (at minimum) a final report with full methodological details and results is available, as well as (ideally) a data quality assessment flagging potential limitations. When the raw data are available, they are analyzed using the Anthro Survey Analyzer software to produce a standard set of results and data quality outputs against which the review is conducted (*see Sample Anthro Survey Analyzer Data Quality Report in Annex 5*). Comments are documented in a standard review template (*see Sample JME Review in Annex 6*) extracting methodological details (e.g., sampling procedures, description of anthropometrical equipment), data quality outputs (e.g., weight and height distributions, percentage of cases that were flagged as implausible according to the WHO Child Growth Standards) and the malnutrition prevalence estimates from the data source under review generated based on the standard recommended methodology outlined in section 3 c. These estimates are compared against the reported values, as well as against those from other data sources already included in the JME dataset, to assess the plausibility of the trend before including the new point. Reports that are preliminary, or that lack key details on methodology or results, cannot be reviewed and are left pending until full information is available.

3 https://www.who.int/nutgrowthdb/publications/algorithms/en/

References

1. United Nations Children's Fund, World Health Organization, The World Bank. UNICEF-WHO-World Bank Joint Child Malnutrition Estimates. (UNICEF, New York; WHO, Geneva; The World Bank, Washington, DC; 2012).

2. JME dissemination products can be accessed on: data.unicef.org/resources/jme, www.who.int/nutgrowthdb/estimates, data.worldbank.org/child-malnutrition

3. McLain AC, Frongillo EA, Feng J, Borghi E. Prediction intervals for penalized longitudinal models with multisource summary measures: an application to childhood malnutrition. *Statistics in Medicine* 38:1002-1012, 2019.

4. de Onis M, Blössner M, Borghi E, Morris R, Frongillo EA. Methodology for estimating regional and global trends of child malnutrition. *International Journal of Epidemiology* 33:1260-1270, 2004.

5. United Nations Children's Fund (UNICEF), World Health Organization, International Bank for Reconstruction and Development/The World Bank. *Meeting report on Technical Consultation on a Country-level model for SDG2.2*. December 2019

6. Murray, C. J., Aravkin, A. Y., Zheng, P., Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., ... & Abegaz, K. H. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258), 1223-1249.

7. SMART Survey Website. Accessed online in 26 Oct 2020, url: <https://smartmethodology.org/>.

8. Standardized Monitoring and Assessment of Relief and Transitions website. Accessed on 3 Nov 2020, url:<https://smartmethodology.org/about-smart/?doing_wp_cron=1604429438.7111239433288574218750>

9. *Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old. Geneva: World Health Organization and the United Nations Children's Fund* (UNICEF), 2019. Licence: CC BY-NC-SA 3.0 IGO. Accessed online 27 October 2020, url: <https://www.who.int/nutrition/publications/anthropomentry-data-quality-report-chapter3.pdf?ua=1>

10. *Inquérito de Indicadores Múltiplos e de Saúde em Angola 2015-2016.* Luanda, Angola e Rockville, Maryland, EUA: INE, MINSA, MINPLAN e ICF International.

11. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS. 2017

12. Lao Social Indicator Survey II 2017, Survey Findings Report. Vientiane, Lao PDR: Lao Statistics Bureau and UNICEF

13. UNICEF. *Progress for Children Beyond Averages: Learning from the MDGs.* June 2015.

14. De Onis, M., Onyango, A., Borghi, E., Garza, C., and Yang, H. (2006). Comparison of the World Health Organization (WHO) Child Growth Standards and the National Center for Health Statistics/WHO international growth reference: Implications for child health programmes. *Public Health Nutrition*, 9(7), 942-947. doi:10.1017/PHN20062005

15. Yang H, de Onis M. Algorithms for converting estimates of child malnutrition based on the NCHS reference into estimates based on the WHO Child Growth Standards. *BMC Pediatr* 2008;8:19.

Annexes

Contents

Annex 1: Regional and global aggregates

For the 2021 edition of the JME, different methods were applied to generate regional and global estimates for stunting and overweight compared to wasting and severe wasting, as described below. In short, results from the new country-level model were used to generate the regional and global estimates for stunting and overweight, while the JME sub-regional multi-level model[*] was used to generate the global and regional estimates for wasting and severe wasting.

Stunting and overweight:

Annual global and regional estimates from 2000 to 2020[†] were derived as the respective country averages weighted by the countries' under-5 population from The United Nations World Population Prospects, 2019 Revision, using annual country model-based estimates for the 204 countries listed in Annex 2. This includes 155 countries with national data sources (e.g. household surveys) included in the JME country dataset described in section 3 of the background document. It also includes 49 countries with modelled estimates generated for development of regional and global aggregates but for which country modelled estimates are not shown because they did not have any household surveys in the JME country dataset or because the modelled estimates remained pending final review at the time of publication. Confidence intervals were generated based on bootstrapping methodology.

Wasting and severe wasting:

The wasting and severe wasting prevalence data from national data sources (e.g. household surveys) described in section 3 of the background document were used to generate the regional and global estimates for the year 2020[†] using the JME sub-regional multi-level model*, applying population weights for children under 5 years of age from The United Nations World Population Prospects, 2019 Revision.

[*] The JME sub-regional model is described in the following publications:

United Nations Children's Fund, World Health Organization, The World Bank. UNICEF-WHO-World Bank Joint Child Malnutrition Estimates. (UNICEF, New York; WHO, Geneva; The World Bank, Washington, DC; 2012).

And

de Onis M, Blössner M, Borghi E, Morris R, Frongillo EA. Methodology for estimating regional and global trends of child malnutrition. *International Journal of Epidemiology* 33:1260-1270, 2004. <https://doi.org/10.1093/ije/dyh202>, accessed April 2021.

[†] Household survey data on child height and weight were not collected in 2020 due to physical distancing measures, with the exception of four surveys. These estimates are therefore based almost entirely on data collected before 2020 and do not take into account the impact of the COVID-19 pandemic.

Annex 2: Country classification and regions used in modelling

Eastern and Central Africa	East Asia and Pacific	Europe and Central Asia	High Income Countries	Latin America and the Caribbean	Middle East and North Africa	South Asia	Southern Africa	West Africa
Angola	American Samoa	Albania	Andorra	Antigua and Barbuda	Algeria	Afghanistan	Botswana	Benin
Burundi	Cambodia	Armenia	Australia	Argentina	Egypt	Bangladesh	Eswatini	Burkina Faso
Central African Republic	China	Azerbaijan	Austria	Belize	Iran (Islamic Republic of)	Bhutan	Lesotho	Cabo Verde
Comoros	Cook Islands	Belarus	Barbados	Bolivia (Plurinational State of)	Iraq	India	Namibia	Cameroon
Congo	Democratic People's Republic of Korea	Bosnia and Herzegovina	Bahamas	Brazil	Jordan	Maldives	South Africa	Chad
Democratic Republic of the Congo	Fiji	Bulgaria	Bahrain	Chile	Lebanon	Nepal	Zimbabwe	Côte d'Ivoire
Djibouti	Indonesia	Georgia	Belgium	Colombia	Libya	Pakistan		Gambia
Equatorial Guinea	Kiribati	Hungary	Bermuda	Costa Rica	Morocco	Sri Lanka		Ghana
Eritrea	Lao People's Democratic Republic	Kazakhstan	Brunei Darussalam	Cuba	Palestine			Guinea
Ethiopia	Malaysia	Kyrgyzstan	Canada	Dominica	Syrian Arab Republic			Guinea-Bissau
Gabon	Marshall Islands	Latvia	Croatia	Dominican Republic	Tunisia			Liberia
Kenya	Micronesia (Federated States of)	Lithuania	Cyprus	Ecuador	Yemen			Mali
Madagascar	Mongolia	Montenegro	Czechia	El Salvador				Mauritania
Malawi	Myanmar	North Macedonia	Denmark	Grenada				Niger
Mauritius	Nauru	Republic of Moldova	Estonia	Guatemala				Nigeria
Mozambique	Niue	Romania	Finland	Guyana				Sao Tome and Principe
Rwanda	Palau	Russian Federation	France	Haiti				Senegal
Seychelles	Papua New Guinea	Serbia	Germany	Honduras				Sierra Leone
Somalia	Philippines	Tajikistan	Greece	Jamaica				Togo
South Sudan	Samoa	Turkey	Greenland	Mexico				
Sudan	Solomon Islands	Turkmenistan	Guam	Nicaragua				
Uganda	Taiwan (Province of China)	Ukraine	Iceland	Panama				
United Republic of Tanzania	Thailand	Uzbekistan	Ireland	Paraguay				
Zambia	Timor-Leste		Israel	Peru				
	Tokelau		Italy	Saint Lucia				

Eastern and Central Africa	East Asia and Pacific	Europe and Central Asia	High Income Countries	Latin America and the Caribbean	Middle East and North Africa	South Asia	Southern Africa	West Africa
	Tonga		Japan	Saint Vincent and the Grenadines				
	Tuvalu		Republic of Korea	Suriname				
	Vanuatu		Kuwait	Uruguay				
	Viet Nam		Luxembourg	Venezuela (Bolivarian Republic of)				
			Malta					
			Monaco					
			Netherlands					
			New Zealand					
			Northern Mariana Islands					
			Norway					
			Oman					
			Poland					
			Portugal					
			Puerto Rico					
			Qatar					
			San Marino					
			Saudi Arabia					
			Singapore					
			Spain					
			Saint Kitts and Nevis					
			Slovakia					
			Slovenia					
			Sweden					
			Switzerland					
			Trinidad and Tobago					
			United Arab Emirates					
			United Kingdom					
			United States of America					
			United States Virgin Islands					

Annex 3: Additional details on the stunting and overweight country-level model

This Annex provides additional details to Section 2 of the Country Consultation background Document.

Data cleaning

Some surveys do not cover the entire age interval 0 to 59 months and thus are not aligned with the standard definition for the child malnutrition indicators (e.g., 0-5 or 0-12 months not covered). To incorporate these surveys, we ran a linear mixed model on the difference between the 0-59-month prevalence estimate and the estimates at 0-5-, 6-11-, 12-23-, 24-35-, 36-47- and 48-59-month age-groups, using data from surveys with both the 0-59-month prevalence and separate age-group prevalence values. Specifically, this difference was modeled as a function of the full prevalence, age-group and a full prevalence by age-group interaction. Model diagnostics showed that the linearity assumption was upheld. With the estimated mixed model, the data for missing age groups were then imputed using the data from the observed age groups. The prevalence estimate for the full age range was then aggregated using the estimated and observed age-group prevalence rates, for sources with at least one missing age group.

An essential component of the statistical methods applied to generate the country prevalence estimates is the incorporation of the sample standard error (SSE) of the survey to adjust for the survey uncertainty. However, SSE's were not reported for all surveys. Some surveys reported 95% confidence intervals (CI), but not the SSE. CI's are generally of the form $Estimate \pm 1.96 * StandardError$, however, for proportions this calculation can be made on different scales (e.g., log or logit scale) and then transformed to the scale of prevalence. The scale of calculation varies by study and is not always known. To identify the scale of calculation, we tried various transformations (no transformation, log, logit) of the survey prevalence and 95% CI limits, to identify which transformation resulted in the most symmetric 95% CI. Once the transformation was identified, the $StandardError$ (on the transformed scale) was estimated using the formula (range of the 95%CI)/(2*1.96). The SSE on the scale of prevalence was obtained using the delta method.[‡]

For surveys without SSE's or 95% CI's, they were predicted using the model

$$\log\,(SSE_{ij}) = b_{i0} + \beta_1 Type_{ij} + \beta_2 \log\big(Y_{ij}(1 - Y_{ij})\big) + \beta_3 \log\big(n_{ij}\big) + \epsilon_{ij},$$

[‡] Xu, Jun; Long, J. Scott (August 22, 2005). "Using the Delta Method to Construct Confidence Intervals for Predicted Probabilities, Rates, and Discrete Changes" (PDF). Lecture notes. Indiana University.

where SSE_{ij}, Y_{ij}, and n_{ij} are the estimated SSE, prevalence, and (non-weighted) sample size of the *j*-th survey in the *i*-th country, respectively, $Type_{ij}$ was an indicator of the survey type (MICS, NNS, DHS, SMART or Other) and $Var(\epsilon_{ij}) = \sigma_t^2$ where σ_t^2 varied by survey type. The motivation for this form is the standard formula for the standard error of a proportion $SE(\hat{p}) = \sqrt{np(1-p)}$, i.e., the log of the standard error linearly related to the log of the sample size and the log of the prevalence times 1-prevalence. Given the sample size and prevalence, we found that different survey types had different overall levels of SSE and heterogeneity between them. This could be due to different survey types using different sampling methodologies and some survey types being more homogeneous in their survey designs than others (e.g., one would think the 'other' category would have more heterogeneous survey designs than MICS).

The SSE model was fitted using the surveys that had complete data on SSE, survey type, prevalence and sample size. The fitted model was then used to estimate the SSE for surveys that were missing SSE. To limit the impact studies with missing SSE on the analysis, a conservative prediction of SSE was used where one standard error was added to the predicted SSE values. Some surveys with missing SSE were also missing the sample size (n_{ij}). For these surveys an identical process was ran without the n_{ij} in the model.

Additional details of the statistical methods

The model used in these analyses is similar to that proposed in McLain et al. (2019)[§] and discussed in more detail therein. The general statistical model is a penalized longitudinal mixed-model with a heterogeneous error term. The non-linear longitudinal patterns in the outcomes were captured using penalized cubic B-splines, with country-specific intercepts and random cubic B-splines. To fit the models, we exploited the connection between penalized spline smoothing and linear mixed-effects models.[**] That is, the fitted model is a linear mixed model where the random effects are chosen such that they penalize the likelihood when the B-spline coefficients are not constant (i.e., not a straight line). The number of random B-spline coefficients was chosen based on corrected Akaike information criterion (AICc) as has been suggested for penalized models. Along with penalized B-splines, the model contains country specific random effects and spline coefficients. To select the covariance matrix of the random splines we tested variance components, compound symmetric and unstructured specifications, and selected the type that minimized AICc. As demonstrated in McLain et al (2019)[†], accurate confidence intervals and prediction intervals can be obtained from this model.

For stunting we used the following statistical model

$$Y_{ij} = b_{i0} + b_{i1}t_{ij} + \beta X_{ij} + \sum_{k=1}^{K} \gamma_k B_k(t_{ij}) + \sum_{k=1}^{K^r} b_{ik+1} B_k^r(t_{ij}) + \epsilon_{ij} \qquad (1)$$

[§] McLain AC, Frongillo EA, Feng J, Borghi E. Prediction intervals for penalized longitudinal models with multisource summary measures: an application to childhood malnutrition. Statistics in Medicine 38:1002-1012, 2019.

[**] Currie, I. D. and Durban, M. (2002). Flexible smoothing with p-splines: a unified approach. Statistical Modelling 2, 333–349.

where Y_{ij} is the logit transform of stunting prevalence for country *i* in year *j*, X_{ij} and β vectors of covariate data and regression parameters, respectively, $\gamma = (\gamma_1, \dots, \gamma_K)$ are a vector of penalized B-spline coefficients, $b_i = (b_{i0}, \dots, b_{kK^r+1})$ is a vector of country-specific random effects which follow a multivariate normal distribution, and ϵ_{ij} is a residual term with variance $Var(\epsilon_{ij}) = \sigma^2(\delta + S_{ij}^2)$ where S_{ij} is the SSE of the stunting prevalence estimate. The random portion of the model included a random intercept, slope and two B-spline functions (corresponding to the first two components from a 4-dimensional cubic B-spline denoted by B_k^r in equation (1)). The knots of the penalized were separated by 1-year, with K=27. The covariate data consisted of linear and quadratic socio-demographic index (SDI), a five-year average of health system access, a dummy variable indicating whether the survey was based on the SMART methodology and another dummy variable indicating the region the country belongs to (we also tested an all risk factors summary exposure value for unsafe sanitation). For both indicators, the 9 regions used in the modeling were: Eastern and Central Africa, Southern Africa, Western Africa, the Middle East and North Africa, East Asia and Pacific, South Asia, Europe and Central Asia, High-income Western Countries, and Latin America and the Caribbean. The high-income group of countries were defined by those countries that were consistently classified as such for at least nine of the last ten years (that is, since 2010).

The overweight model had similar aspects to the stunting model discussed above (i.e., heterogeneous error term, penalized and random splines, etc.). Specifically, we used the following statistical model

$$Y_{ijl} = b_{i0} + \beta X_{ij} + \sum_{l=1}^{9} \sum_{k=1}^{K} \gamma_{lk} B_k(t_{ijl}) + \sum_{k=1}^{K^r} b_{ik} B_k^r(t_{ij}) + \epsilon_{ij}$$

where Y_{ijl} is the logit transformed overweight prevalence for country *i* in region *l* year *j*. The penalized B-spline portion of the model is different for all 9 regions, which allows the linearity of the regional curves to vary according to the regional data. The location and number ($K = 12$) of penalized and random B-splines were the same for each region. The random portion of the model included a random intercept, and three B-spline functions (from a 3-dimensional cubic B-spline). The covariate data consisted of linear and quadratic SDI, one dummy variable indicating whether the survey was based on the SMART methodology, another dummy variable indicating the region of the country and a region by time interaction factor.

As mentioned above, all models used a heterogeneous error term $Var(\epsilon_{ij}) = \sigma^2(\delta + S_{ij}^2)$. The δ term is used to offset how much of the error variability is due to non-sampling error. Non-sampling errors occur due to measurement error in the responses that are not present in the SSE estimates, and deviations in country-specific trends that cannot be fit into the smooth trend induced by cubic B-splines. If $\delta = 0$ then the error is entirely due to the sampling error, while if δ is large, then the error is mostly due to non-sampling error.

Annex 4: Improving age determination in children in settings where obtaining documents stating the date of birth is challenging

Background

Proper age determination is essential in national household surveys including anthropometric indicators which require age, such as "stunting" (height-for-age) or "underweight" (weight-for-age). Improper determination of age in children can lead to over or underestimating prevalence estimates for these indicators.

Determining age in household surveys is done by requesting the family head or the caretaker to show a document stating the child's date of birth (e.g., civil registration, health card), and confirming this date through questions to the parents. However, in some contexts, getting a written proof of the date of birth can be challenging. This is the case in instable settings, for example, where there is a high rate of displacements or, quite often, in isolated or rural areas. Parents uncertainty about date of birth has been also linked to low degrees of schooling[††]. When obtaining a written proof of the date of birth is challenging, what has been recommended is determination of month and year of birth using 'local calendars' that referenced local events and festivals, closely related to the individual's personal life, has proven to be relatively successful[††].

In many of the national household surveys including anthropometry, the use of the event calendar has been used to provide the age of the child in months (instead of providing the month and year of birth based on an event). However, this approach is not precise enough unless survey field work is less than one month in duration.

The UNICEF-WHO-WB Joint Child Malnutrition Estimates group carried out a study to evaluate the usefulness of using an approach to recover approximate month and year of birth for those children missing the date of birth in the survey sample, but for whom an age in months as well as date of visit were available, along with the seasonal calendar(s) used to obtain the age in months. In this study we aimed to translate the age provided in months to the estimated month and year of birth, in general, imputing 15 as the day of birth as this is one the methods used by JME during survey re-analysis for cases with missing day of birth (see section 3 of the country consultation background document).

While it is not possible to assess the accuracy of either method (i.e. use of reported age in months or adjusted ages using the seasonal calendar(s) to obtain a date of birth), the JME working group believes that surveys where field work duration lasted for longer than one month, and where only one seasonal calendar with age in months was used for all the field work, that ages will be inaccurate and require adjustment (e.g. for a survey with fieldwork implemented in June and July, using one seasonal calendar which indicated that births in May of the survey year were 1

[††] Trends in Demographic and Health Survey data quality: an analysis of age heaping over time in 34 countries in Sub Saharan Africa between 1987 and 2015. Mark Lyons-Amos, Tara Stones. BMC Res Notes. 2017; 10: 760.: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5738749/

month of age, would be accurate enough for dates of visit in June, but not for dates of visit in July).

Methodology

National household anthropometric surveys with a high percentage of cases missing the complete date of birth were identified: Mauritania NNS 2014, Mali NNS 2011. Raw data were obtained from the countries.

The event calendars were analyzed to identify their starting point and this date was included in the raw database, which also included others variables: date of visit, sex, weight, height, sampling weight factor, cluster and strata and the geographical stratification (region).

For those cases where date of birth was incomplete, the age in months was included in the raw database. A new variable called “imputed DoB” was created. For cases with complete date of birth, “imputed DoB” was equal to the complete date of birth. For those with missing date of birth, the date was imputed as follows: the age in months was subtracted from the date using the event calendar to estimate the month and year of birth in the format MM/YYYY, and the day was then imputed as day 15 of the month. However, for those cases when the imputed date of birth is after the date of visit, it is necessary to impute as day 1 of the month instead of day 15.

In the Standard Analysis, the software estimates the age in days by subtracting to the Date of Visit the “imputed DoB”. Then, prevalence estimates for stunting and underweight are calculated. Prevalence estimates obtained were then compared to those published in the official report and to the “reanalysis” done using only the cases having a complete date of birth.

Results

Tables 1 and 2 include the prevalence estimates using 3 different approaches in the two surveys:
A: estimates were calculated using the software ENA nutrisurvey (www.smartmethdology.org) , which use the age in months.
B: estimates were calculated using the WHO Anthro Survey Analyser (https://worldhealthorg.shinyapps.io/anthro/) , based only on data with complete date of birth
C: estimates were calculated using the WHO Anthro Survey Analyser, using only the data with estimated date of birth (converted from the age in months to age in days).

Example 1: Mali NNS 2011

The raw data for this survey included data for 8231 children aged 0 to 59 months old. Complete date of birth was provided for 3842 children (53% only had an age in months).

After discussions with the country, we learned that, as the surveys took place in June and July, two different event calendars were used (the first one starting on 1st of June for assessments done in that month, the second one starting in July for assessments done in that month). Prevalence estimates estimating DoB using both calendars are shown in column D.

<u>Table 1.</u> Anthropometric indicators results estimates: using age in months (official report) (A), reanalysis using only data with completed DoB (B), reanalysis converting age in months to imputed Date of Birth (C and D)

MALI NNS 2011	**Estimates using age in months (official report) (A)**	**Estimates using only data with complete DoB (B)**	**Estimates converting age in months to DoB and imputing day as 15 (C-event calendar starting in June)**	**Estimates converting age in months to DoB and imputing day as 15 using 2 event calendars (D)**
Stunting	27.0 (24.7-29.0) n=7927	29.2 (26.9-31.5) n=3745	29.9 (28.0-31.7) n=8001	27.9 (26.2-29.6) n=8002
Underweight	19.7 (18.0-21.7) n=7921	19.6 (18.0-21.4) n=3752	21.1 (19.8-22.6) n=8012	20.0 (18.7-21.3) n=8012
Wasting	10.4 (9.3-12.3)* n=7753	10.1 (9.1-11.3)* n=7995	10.5 (9.4-11.6)* n=7992	10.4 (9.3-11.6)* n=7994

*including edema (n=1)

Example 2: Mauritania NNS 2014

The raw data for this survey included data for 5708 children aged 0 to 59.9 months old. Complete date of birth was provided for 2117 children (63% included only the age in months).

<u>Table 2.</u> Anthropometric indicators results estimates: using age in months (official report) (A), reanalysis using only data with completed DoB (B), reanalysis converting age in months in imputed Date of Birth (C)

MAURITANIA NNS 2014	**Estimates using age in months (official report) (A)**	**Estimates using age in months with ENA (B)-excluding WHO flags**	**Estimates using age in months with the WHO Anthro Survey Analyser (C)**
Stunting	15.9 (14,4 – 17,5) n=?	22.6 (21.0-24.3) n=5698	22.7 (21.0-24.6) N=5698
Underweight	14.3 (12,9 - 15,8) n=?	13.9 (12.7-15.2) n=5707	13.7 (12.4-15.1) N=5707
Wasting	9.8 (8,9-10,9)* n=?	6.3 (5.6-7.1) n=5703	5.8 (5.1-6.5) N=5703

*including edema (n=0)

Annex 5: Annotated Sample Anthro Survey Analyser Data Quality Report

DATA QUALITY REPORT

Title Page

Table of Contents

Recommended citation:

Data quality assessment report template with results from WHO Anthro Survey Analyser

Analysis date: 2020-08-20 10:16:08

Link: https://worldhealthorg.shinyapps.io/anthro/

This report is a template that includes key data quality checks that can help to identify issues with the data and considerations when interpreting results. Other outputs that can be relevant to your analyses can be saved directly from the tool interactive dashboards and added to the report.

For guidance on how to interpret the results, user should refer to the document "Recommendations for improving the quality of anthropometric data and its analysis and reporting" by the Working Group on Anthropometric Data Quality, for the WHO-UNICEF Technical Expert Advisory Group on Nutrition Monitoring (TEAM). The document is available at www.who.int/nutrition/team, under "Technical reports and papers".

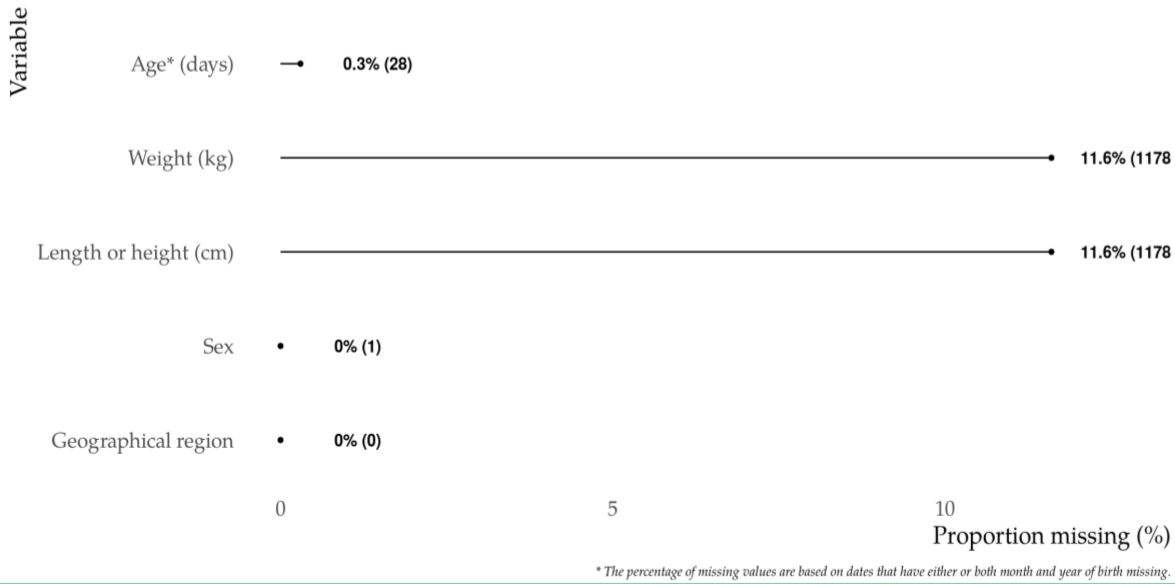
Section

Please note that the text in green is not part of the outputs from Anthro Analyser; rather, it has been added here to provide further information and clarification to the reader of the Technical Notes from the background document for country consultations on the 2021 edition of the UNICEF-WHO-World Bank Joint Malnutrition Estimates.

1. Missing data

Percentage (number of cases) of children missing information on variables used in the analysis

Total number of children: 10175.



Interpretation notes: This graphic provides information about the percentage of children in the dataset who are missing key variables needed to calculate the z-scores for malnutrition, such as age in days, weight, height, and sex. The graphic also summarizes the percentage of children in the dataset missing other background variables, such as their geographic region or mother's education level, which are used to generate disaggregated estimates, but are not required to generate the z-scores.

These estimates of missing data are only based on children covered in the survey; surveys that do not undertake a complete listing of all eligible children in interviewed households may underestimate these values. Demographic and Health Surveys and Multiple Indicator Cluster Surveys have full household listings and values represented here for those surveys are for all eligible children.

When the value for 'proportion missing' is high (e.g., >10 per cent), consider assessing the variable by geographic region (next table), when available, to identify any biases with respect to where children with missing data live.

2. Missing data by geographical region

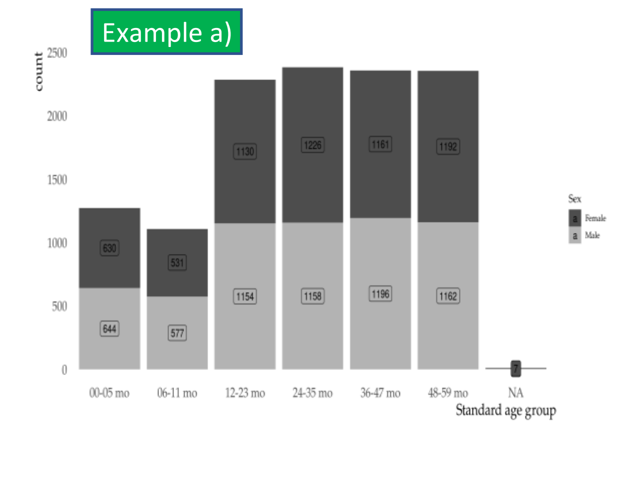
Geographical region	N	Age* (days)	Weight (kg)	Length or height (cm)	Sex
Area 1	600	3 (0.5%)	71 (11.8%)	71 (11.8%)	0 (0%)
Area 2	577	2 (0.3%)	38 (6.6%)	38 (6.6%)	0 (0%)
Area 3	637	2 (0.3%)	78 (12.2%)	78 (12.2%)	0 (0%)
Area 4	612	0 (0%)	78 (12.7%)	78 (12.7%)	0 (0%)
Area 5	613	1 (0.2%)	97 (15.8%)	97 (15.8%)	1 (0.2%)
Area 6	728	2 (0.3%)	97 (13.3%)	97 (13.3%)	0 (0%)
Area 7	764	5 (0.7%)	131 (17.1%)	131 (17.1%)	0 (0%)
Area n	781	0 (0%)	49 (6.3%)	49 (6.3%)	0 (0%)

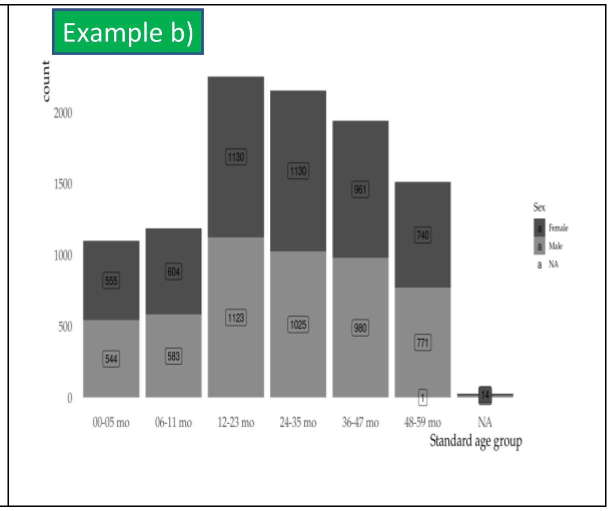
The percentage of missing values for age are based on dates that are missing either or both the month and year of birth.

Interpretation notes: This table is only produced when the variable of subnational geographic regions (e.g., districts) is available in the dataset. The table should be reviewed when the percentage of children missing at least one key variable at the national level exceeds a certain threshold (e.g., >10 per cent). This can provide insight as to whether any geographic areas have a higher percentage of children missing key variables, which could indicate a bias in the national estimates of malnutrition (i.e., missing cases may be concentrated in one geographic region; high in some and low in others; or similar across all geographic regions). If the percentage of children missing key variables is similar across all geographic regions, it can be assumed that there is no subnational geographic bias for missing cases.

3. Data distribution

i) Distribution by standard age grouping and sex



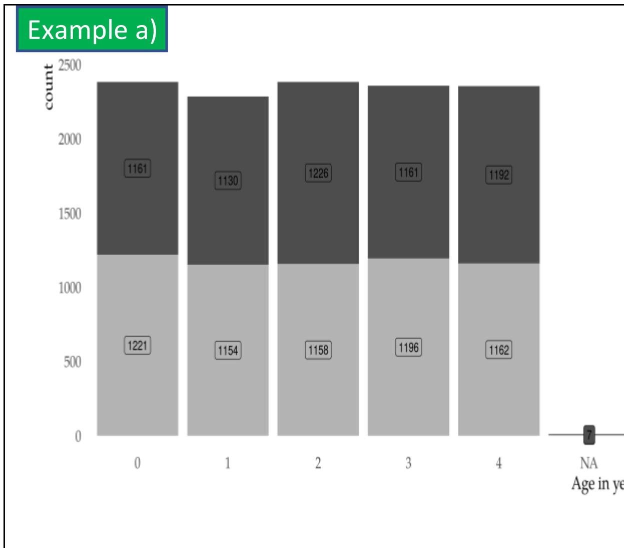


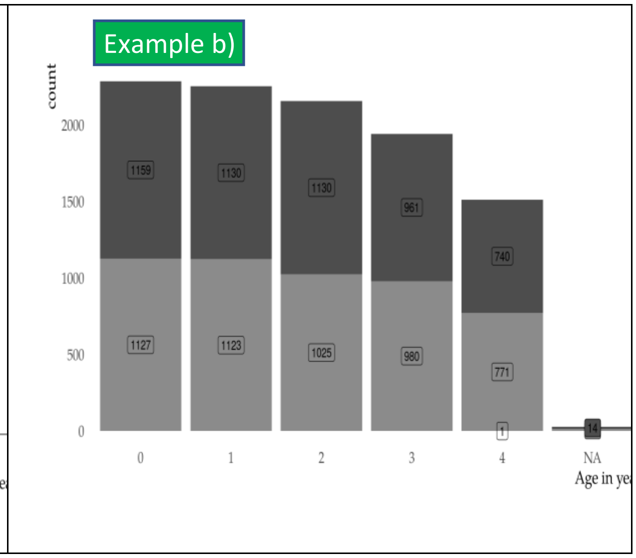
Interpretation notes: The output in this report contains one graph and is based on eligible children in the household for which an age in months (which requires that at minimum that the month and year of birth were collected during the survey interview) was available. [Note: two graphs are shown above to provide different examples]. The graph is meant to provide insight into whether any age groups were underrepresented in the overall sample of eligible children under 5 years of age for the survey (whether or not they were weighed or had their lengths or heights measured). Some of the children represented in this graphic may not be represented in the estimates of malnutrition. For example, if a child was missing height in the dataset, but had at least a month and year of birth recorded, he or she would show up as 'missing' under Section 1 on missing data, and would therefore not be included in the stunting estimate; however, he or she would be included in this graph as an eligible interviewed child in the appropriate age column.

In an ideal graph, the bars for ages 12–23 months, 24–35 months, 36–47 months and 48–59 months would be similar and have about equal numbers of boys and girls. The bars for ages 0–5 months and 6–11 months (covering 6-months intervals) should be about half the height of the other bars (which cover 12-month intervals). Regarding the distribution of children by sex, the bars should ideally split into half boys and half girls, as expected in the majority of the populations. Ideally, the N/A (not available) bar would be near 0.

This first graph above (Example a) illustrates the ideal distribution: the bars for all age groups with 12-month intervals are about the same height and the bars for the two age groups with 6-month intervals are about half the size. The second graph (Example b) illustrates a situation where older children may have been aged out by interviewers (e.g., when a child close to turning 5 years is labelled as a 5-year-old in the household roster so the interviewer can avoid completing the detailed module pertaining to children under 5 years, and thereby relieve daily work overload). This practice should be avoided, as it may bias the sample and void results. When the workload is too heavy for interviewers, the survey coordinator must find solutions to spread out workload or engage more field workers.

ii) Distribution by age in years and sex



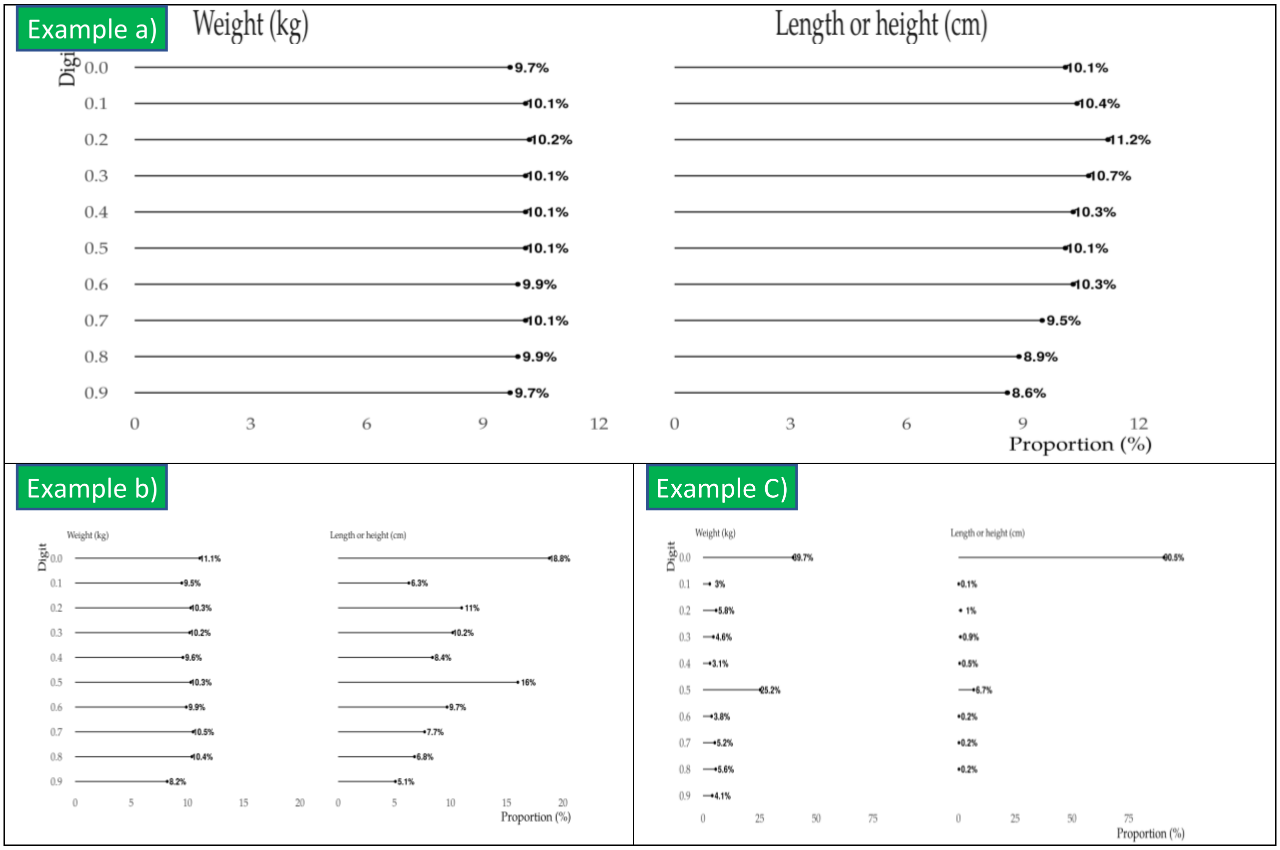


Interpretation notes: The output in this report contains one graph and is based on eligible children in the household for which an age in months (which requires that at minimum that the month and year of birth were collected during the survey interview) was available. [Note: two graphs are shown above to provide different examples]. The graph is meant to provide insight into whether any age groups were underrepresented in the overall sample of eligible children under 5 years of age for the survey (whether or not they were weighed or had their lengths or heights measured). Some of the children represented in this graphic may not be represented in the estimates of malnutrition. For example, if a child was missing height in the dataset, but had an age recorded in years, he or she would show up as 'missing' under Section 1 on missing data, and would therefore not be included in the stunting estimate; however, he or she would be included in this graph as an eligible interviewed child in their appropriate age column.

The graphs used as examples here are for the same countries as the graphs in 3 a, but here, each bar represents a 1-year age grouping (i.e., there is no disaggregation of <6-month-old children and 6-11-month-old children; both groups are included together in the 0-year bar). Ideally, all bars would be about the same height, with about half boys and half girls each and the N/A (not available) bar near 0. **Example a** represents an ideal scenario. **Example b** suggests that older children may have been aged out by interviewers (*see notes for graph 3 i; Example b on the previous page*).

4. Digit preference charts

i) Decimal digit preference for weight and length/height - NATIONAL



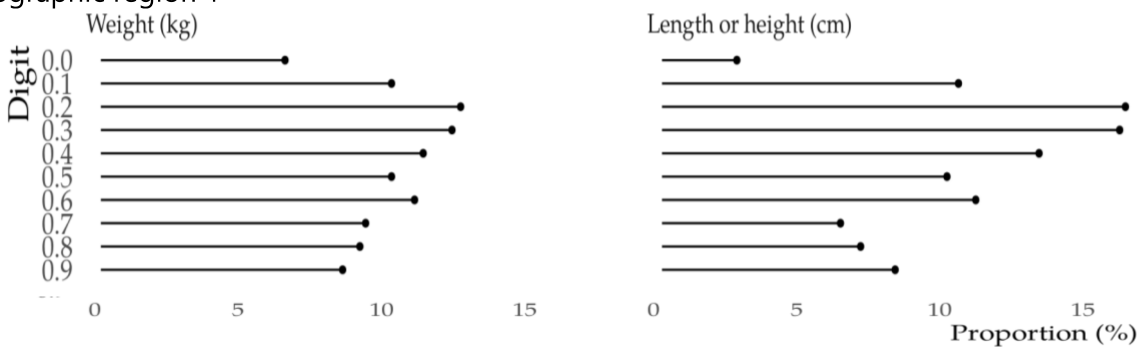
Interpretation notes: The output in this report contains one graph showing the distribution of the decimal digits (e.g., the '0.1' in the measurement 12.1 kg or the '0.5' in the measurement 94.5 cm) in the survey set of reported measurements. [Note: three graphs are shown above to provide different examples]. Decimal (or terminal) digit preference refers to an unexpected distribution of the final digit of the weight and length/height measurements. If survey teams use the recommended equipment of a digital weighing scale and wooden height board with a centimetre tape graduated for each millimetre, each weight and length/height measurement in the survey should display one terminal digit: namely, one tenth of a kilogram for weight and millimetres for length/height.

There are 10 possible terminal digits ranging from 0 to 9. In a survey where the length/height and weight of each child has been measured and recorded correctly on properly functioning equipment, the digits should be uniformly distributed. In other words, each terminal digit should represent approximately 10 per cent of all terminal digits. In general, when digital scales are used, rounding is less likely to occur for weights than for heights given that the equipment used in most surveys to take heights does not have a digital output and requires reading off a centimetre tape. Terminal digit preference refers to the process whereby data heaping (i.e., a distribution distinctive peak) occurs because the final digit of the number has been rounded off. For example, 10.0 kg is recorded instead of the actual reading of 10.3 kg, or 75.0 cm is recorded instead of the actual reading of 74.9 cm. Common terminal digit preference patterns include: (i) a preference for the terminal digits 0 and 5; or (ii) a preference for a terminal digit(s) other than 0 and 5.

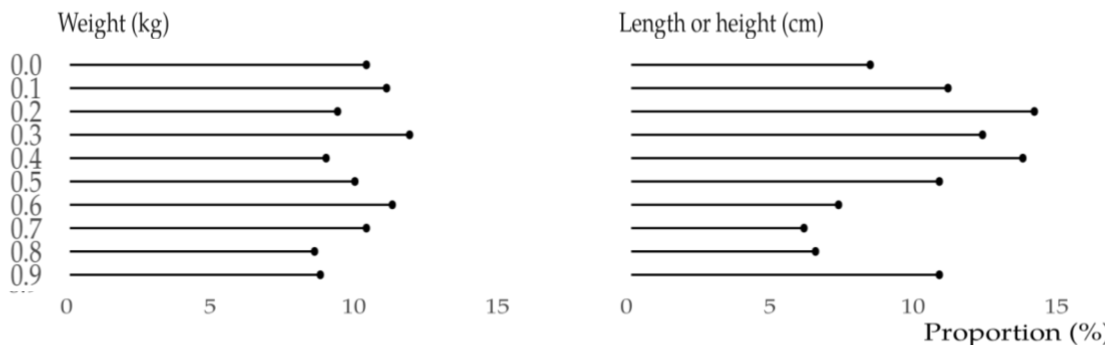
Example a shows an expected pattern where all terminal digits from 0 to 9 make up about 10 per cent of the total each, for weight and length/height. **Example b** represents a survey with moderate rounding on heights for both 0 and 5, but no rounding on weights. While this may indicate deficiency in fieldworker training or inadequate supervision, rounding on heights of this degree is generally less detrimental to the accuracy of estimates than if the same degree of terminal digit preference had occurred for weights. In other words, an error between 1 mm and 5 mm for length or height in this age group with lengths/heights typically spanning from 45 cm to 125 cm has less impact on the accuracy of z-scores than of that caused by an error between 100 g and 500 g for weight in this age group with weights typically spanning from 2 kg to 25 kg. **Example c** shows a country where there was rounding for both weights and heights, with rounding on weights at 0 and 5 and on heights mainly at 0. This represents a case where the malnutrition estimates would likely not be very accurate due to the degree of error in weights and heights, which in turn affects z-score values. Terminal digit preference on digits other than 0 and 5 can occur but is not shown among the examples and would most likely not be due to rounding.

ii) Decimal digit preference by geographical region

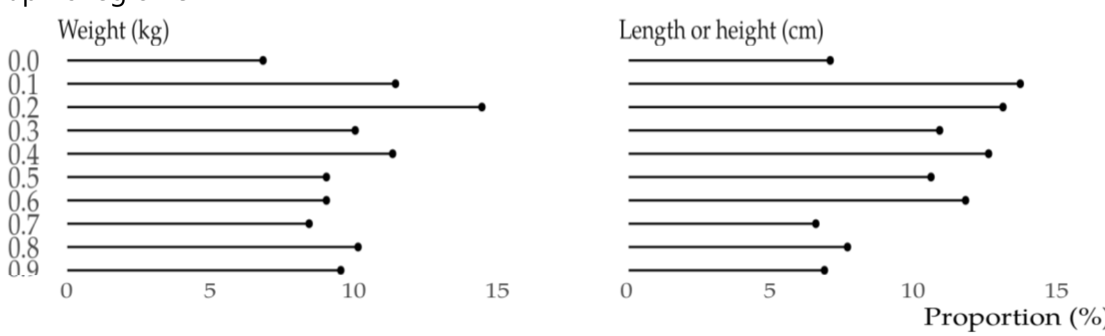
Geographic region 1



Geographic region 2

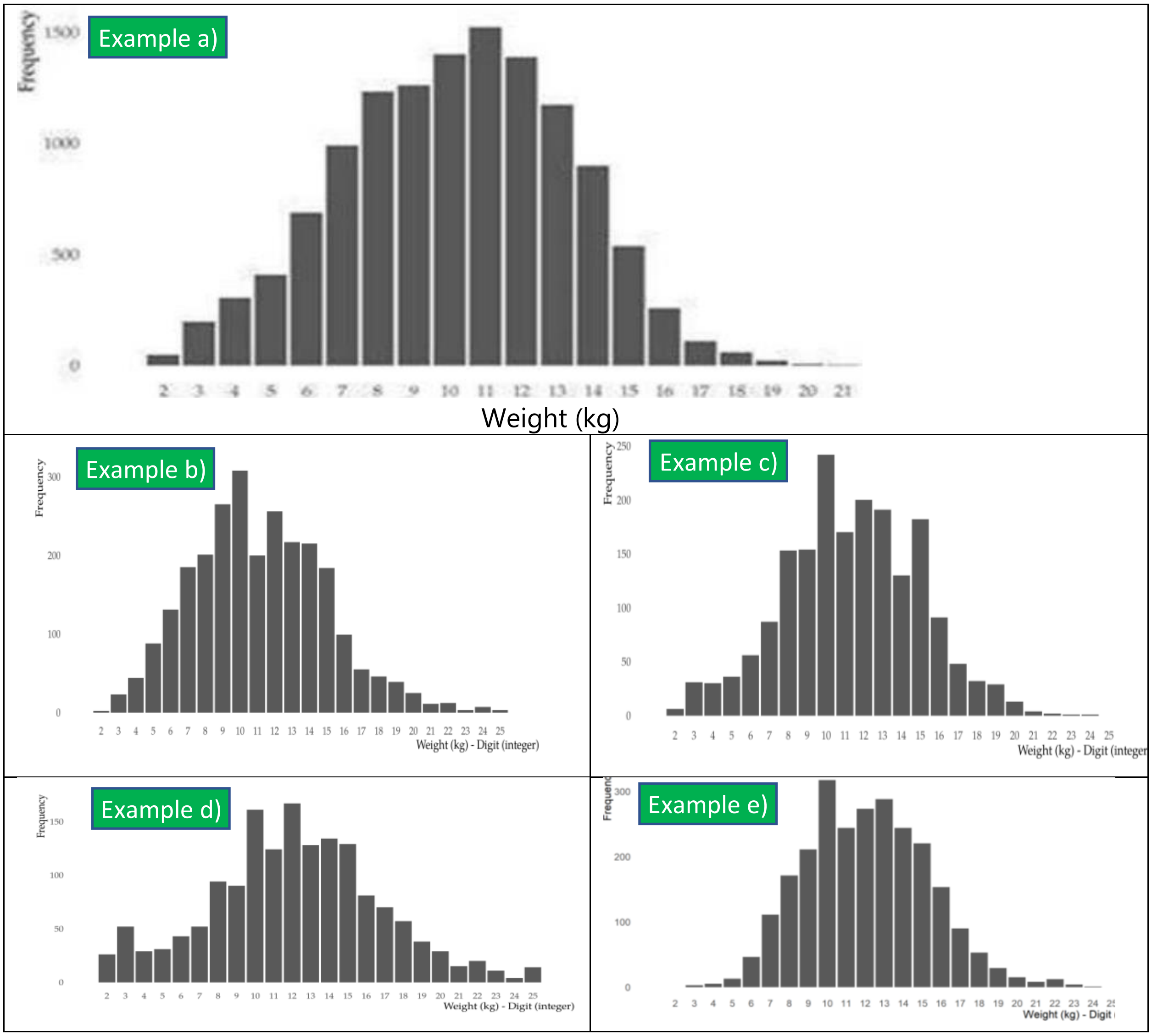


Geographic region 3



Interpretation notes: These graphics are only produced when the variable of subnational geographic regions (e.g., districts) is available in the survey analysis dataset. When available, they are useful to review for each subnational area using the same notes on interpretation from Section 4.i.

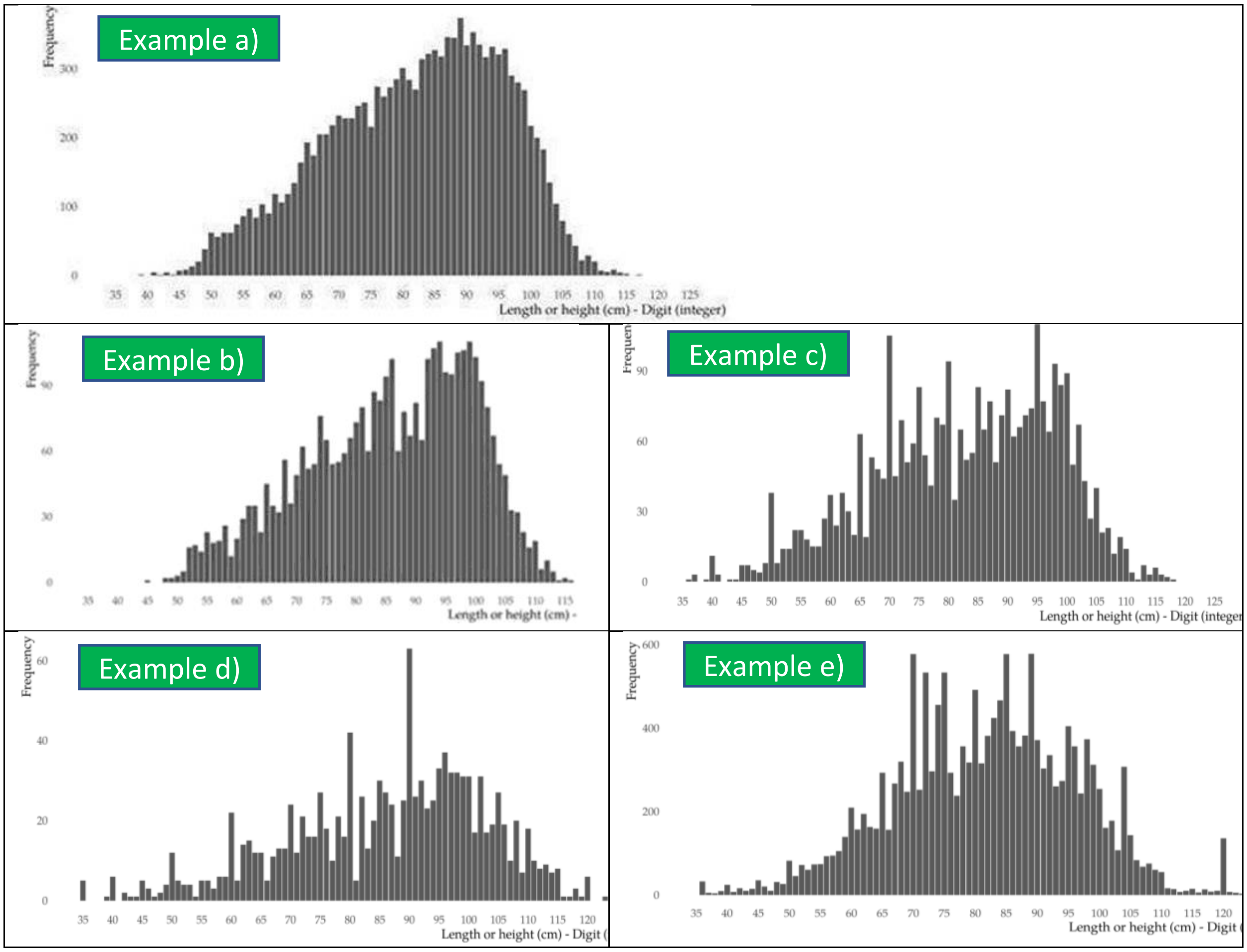
iii) Whole number digit preference for weight



Interpretation notes: The output in this report contains one graph (but the five examples above illustrate different example scenarios). In general, a whole number weight distribution with a shape similar to that of **Example a** is expected, with no peaks or troughs in the distribution. Whole number preference for weight refers to a situation where data heaping occurs because the integer part of the number has been rounded off (e.g., recording of 10 kg instead of 9 kg or 11 kg). While less common than for heights, whole number heaping of weight can seriously affect the accuracy of malnutrition estimates.

Example a shows a pattern with no indication of whole number preference. **Example b** presents a scenario with a depression at 11 kg, which may suggest that children weighing 11 kg were rounded to be either 10 kg or 12 kg, leading to inaccurate estimates. **Example c** shows several peaks and troughs that would be unlikely in a population of children under 5 years of age: notably, the peaks at 10 kg and 15 kg and the troughs at 11 kg and 14 kg. **Example d** shows peaks at 10 kg and 12 kg, but also on the tails at 3 kg and 25 kg. Malnutrition estimates from a survey with a pattern similar to those in **Examples b, c, d and e** may be inaccurate.

iv) Whole number digit preference for length/height



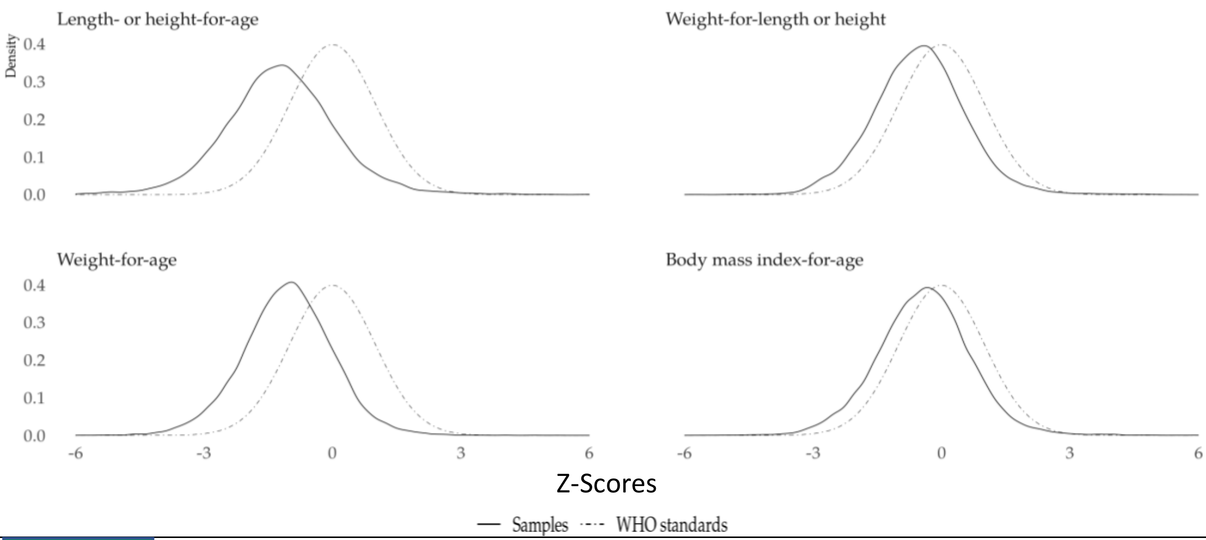
Interpretation notes: The output in this report contains one graph (but the five examples above illustrate different example scenarios). In general, for a population of children under 5 years of age, a whole number weight distribution with a shape similar to that of **Example a** is expected – that is, a curve with no peaks or throughs in the distribution, and which is skewed to the left (due to the higher growth velocity in the first six months of life). Whole number preference for length or height refers to a situation where data heaping occurs because the integer part of the number has been rounded off, (e.g., recording 75 cm instead of 74 cm or 76 cm). This is generally less common than terminal digit heaping but can occur and generally affects the accuracy of malnutrition estimates to a larger degree than terminal digit heaping of heights.

Example a shows a pattern with no indication of whole number preference. **Example b** presents a scenario with some unexpected peaks and troughs that could lead to inaccurate estimates – but at this moderate degree, would not likely have that large of an impact. **Example c** shows a distribution with several peaks and troughs that would be unlikely in a population of children under 5 years of age (e.g., with some notable peaks at 50 cm, 70 cm). **Example d** shows even more concerning heaping than Example c, with peaks approximately every 10 cm through some ranges of the distribution (e.g., at 40 cm, 50 cm, 60 cm, 70 cm, 80 cm, 90 cm and 120 cm). **Example e** has some large peaks at heights that would be too high for so many children to have in a population of children under 5 (e.g., 105 cm, 120 cm). Malnutrition estimates from surveys with height frequency distribution patterns similar to those in **Examples c, d or e** are likely to be inaccurate.

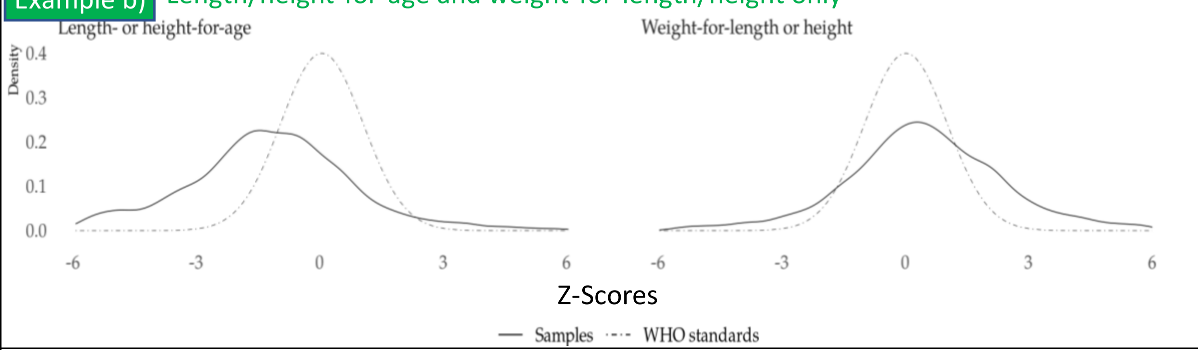
5. Z-score distribution of indicators

i) Z-score distribution by index (whole population)

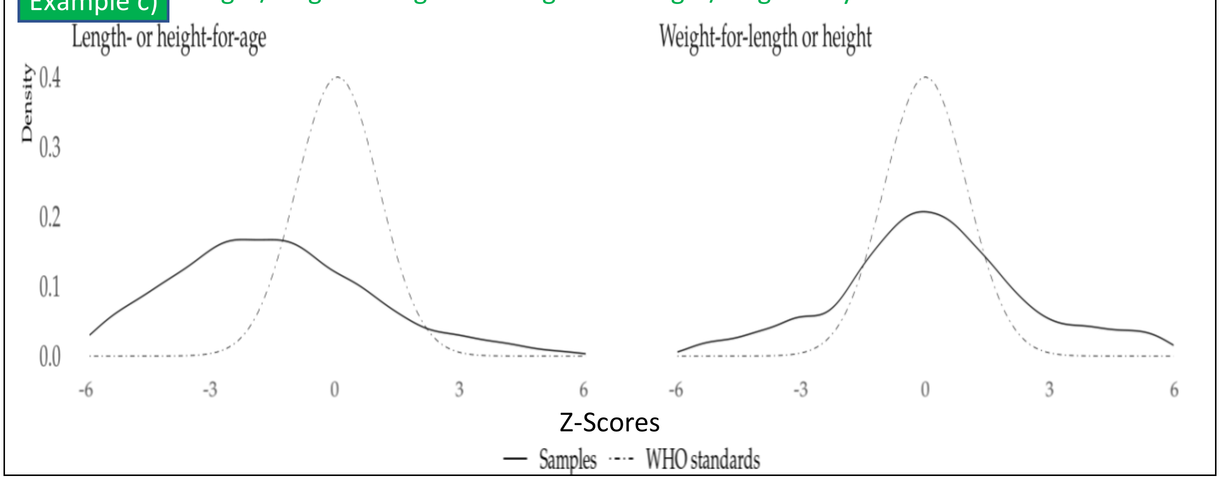
Example a) **Standard output with 4 graphs**



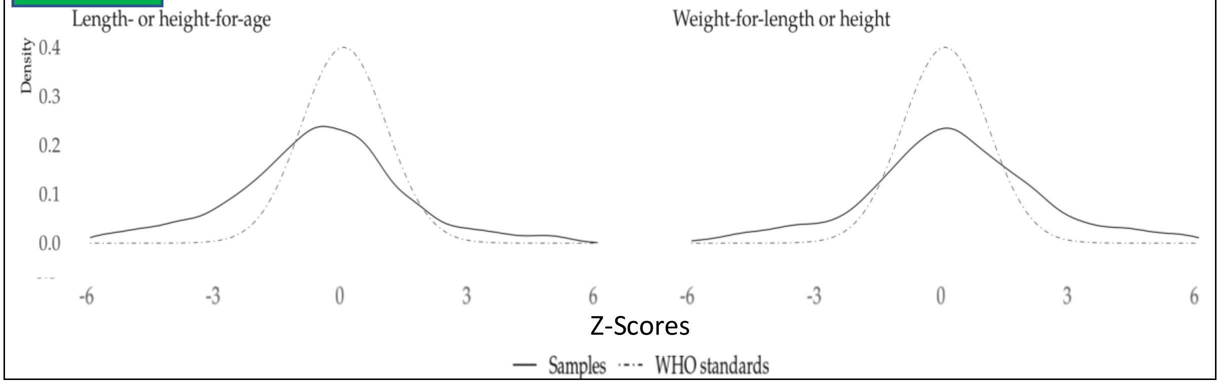
Example b) Length/height-for-age and weight-for-length/height only



Example c) Length/height-for-age and weight-for-length/height only



Example d) Length/height-for-age and weight-for-length/height only



Interpretation notes: The z-score distribution graphs provide information about the distribution of malnutrition in the survey population. They are limited to the range -6 to +6 z-scores for simplicity, but there could be extreme values in the sample beyond this range.[‡‡] The survey data are shown by a solid line that can be compared with the WHO Child Growth Standards (the light dotted line that has a normal distribution with a mean of 0 and a standard deviation of 1).

The output in the standard Anthro Analyser report contains four graphs, as shown in **Example a**: one for each of (i) length/height-for-age; (ii) weight-for-length/height; (iii) weight-for-age; and (iv) BMI-for-age. For **Examples b, c and d**, only the length/height-for-age and weight-for-length/height are provided, as these are the indices associated with the Sustainable Development Goal (SDG) indicators stunting, wasting and overweight. The expected distribution has a smooth bell shape, but the shape of the tails would vary by the malnutrition levels of the population and the quality of the data. The centre of the distribution also departs from 0 according to the level of malnutrition of the population, not only the tails.

The distributions of all indices shown in **Example a** represent a moderately malnourished population – i.e., one that is stunted, as well as wasted, indicated by the distributions shifted to the left. However, all distributions in **Example a** have a smooth bell shape, consistently declining and with no visible bumps (as expected) at the right tail, with a negligible area beyond +2 z-scores. The smooth shape of the distributions in this example does not suggest any issue with data quality, such as poor measurements or systematic data entry issues.

Example b, however, shows quality issues on both distributions; notably, the weight-for-length/height distribution right tail and the length/height-for-age left tail are bumpy. Also, the right tail for length/height-for-age should flatten out beyond +2 because any children in that tail would be beyond the 97.5th percentile and any children beyond +3 would be beyond the 99.9th percentile. This means that z-score values beyond those limits are extremely rare for any population, let alone a population with high levels of stunting. Such inflation in the right tail for length/height-for-age z-score is most likely associated with major error in the height and/or age

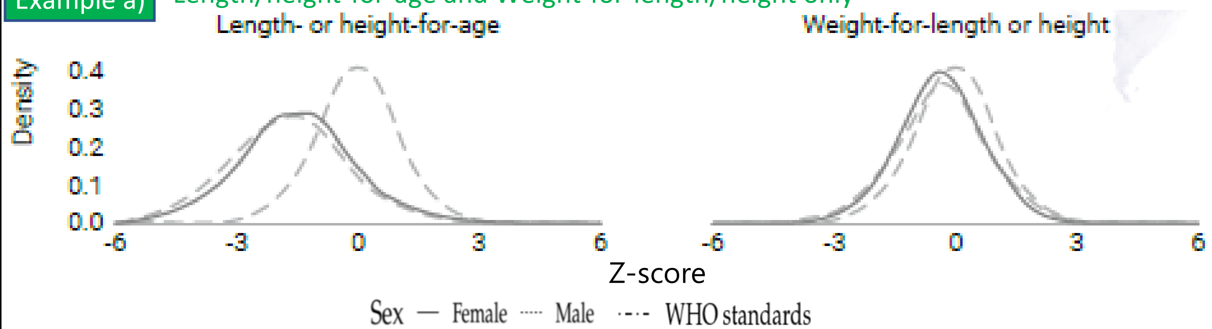
[‡‡] Note that these graphs aim to show the distribution within plausible ranges. While one axis range is used for simplicity (-6 to +6), the actual implausible ranges for each index are: <-6 and >+6 for length/height-for-age z-scores, <-5 and >+6 for weight-for-age z-scores and <-5 and >+5 for weight-for-length/height z-scores.

assessments, or data entry errors. The inaccuracy of length/height measurements will likely lead to inaccuracy of wasting and overweight estimates, as indicated by the bumpiness of both left and right tails of the weight-for-length/height distribution.

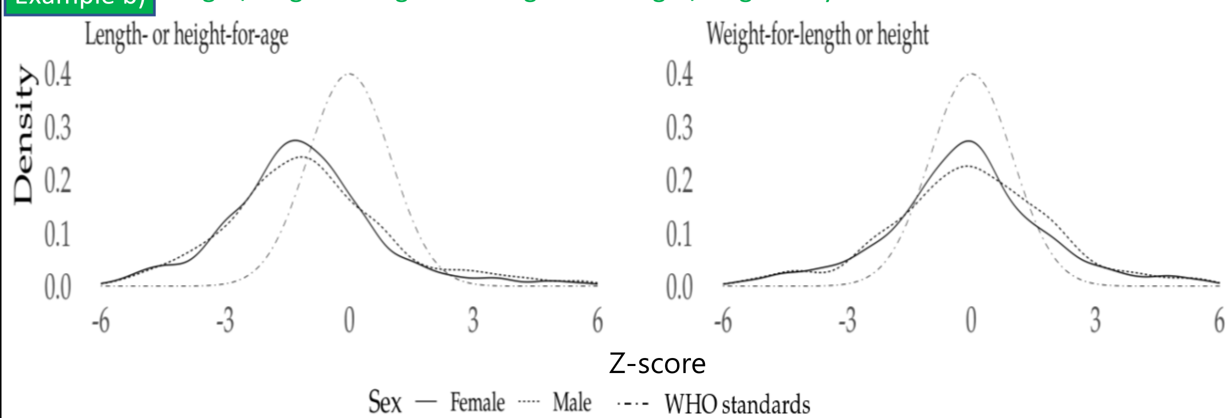
Example c shows an even more concerning inflation in the right side of the tail for length/height-for-age, as well as on the left tail, where there is an abrupt cessation with about 0.03 density at -6. In this same example, both tails of the weight-for-length/height distribution are equally heavier than expected and bumpy, indicating potential data quality issues, likely due to the issues with length/height measurements, age determination or data entry. The large inflation at the right tail for the length/height-for-age distribution in **Example d** suggests substantial errors occurred for length/height and/or age determination during data collection and/or entry, which may have also affected the weight-for-length/height distribution, which has inflation at both tails. **Examples b, c and d** would all raise a flag for poor data quality and should be further investigated to identify potential causes and lessons learned.

ii) Z-score distribution by index and sex

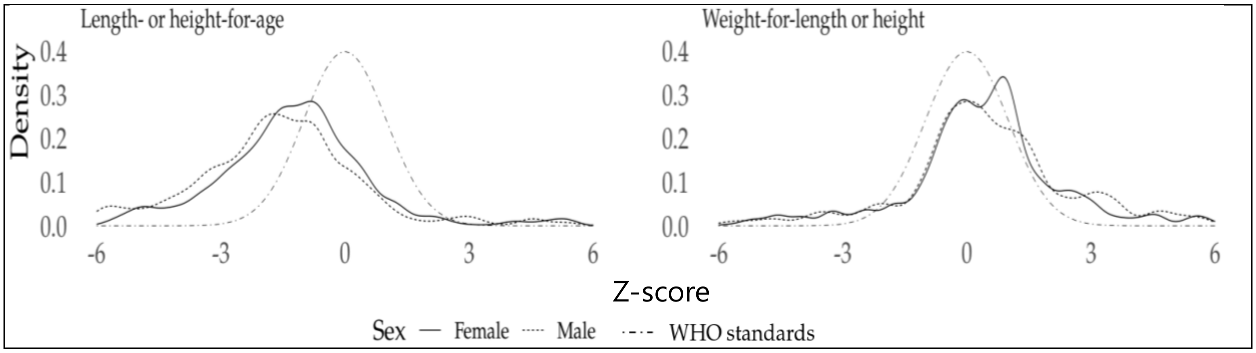
Example a) Length/height-for-age and Weight-for-length/height only



Example b) Length/height-for-age and weight-for-length/height only



Example c) Length/height-for-age and weight-for-length/height only

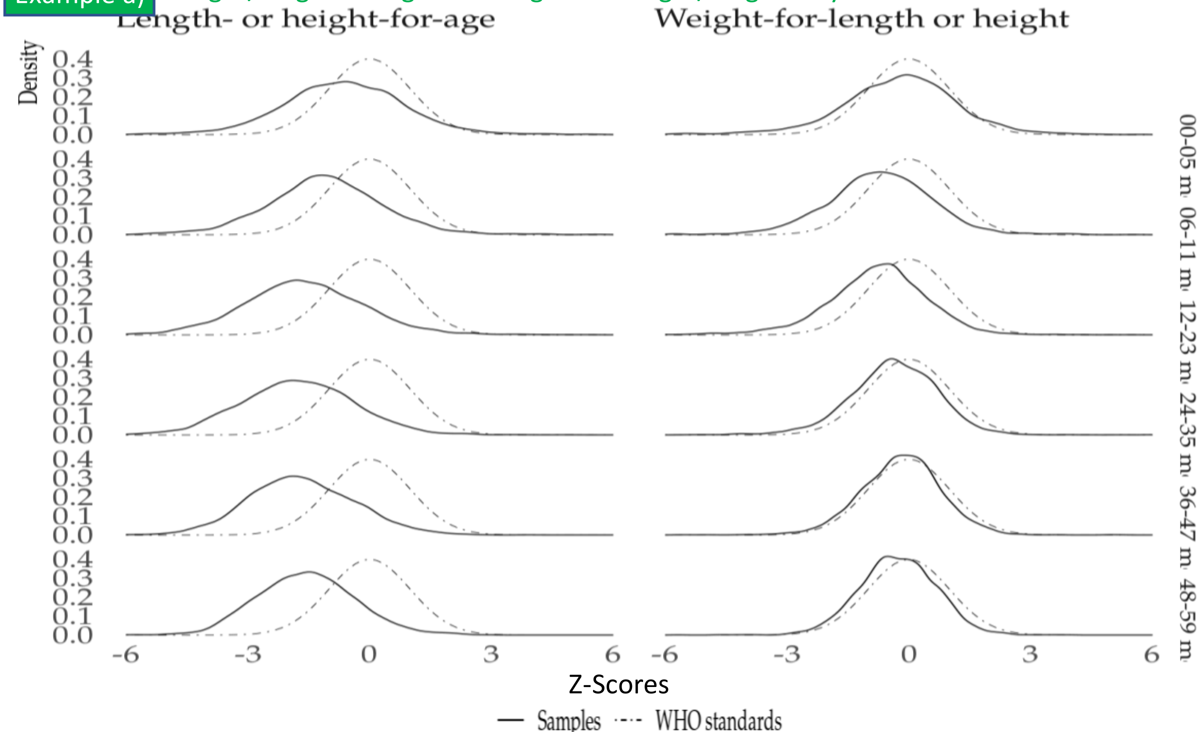


Interpretation notes: Similar to Section 5.a, these graphs provide information about the distribution of malnutrition in the survey population. However, in this subsection, the distribution for girls (darkest line) is shown separately from boys (medium light line), against the WHO standard (the light dotted line that has a normal distribution with a mean of 0 and a standard deviation of 1). The output in the standard Anthro Analyser report contains four graphs, one for each of: (i) length/height-for-age; (ii) weight-for-length/height; (iii) weight-for-age; and (iv) BMI-for-age. However, as only length/height-for-age and weight-for-length/height are required for the SDG indicators of stunting, wasting and overweight, the three scenario examples (a, b and c) only contain two graphs.

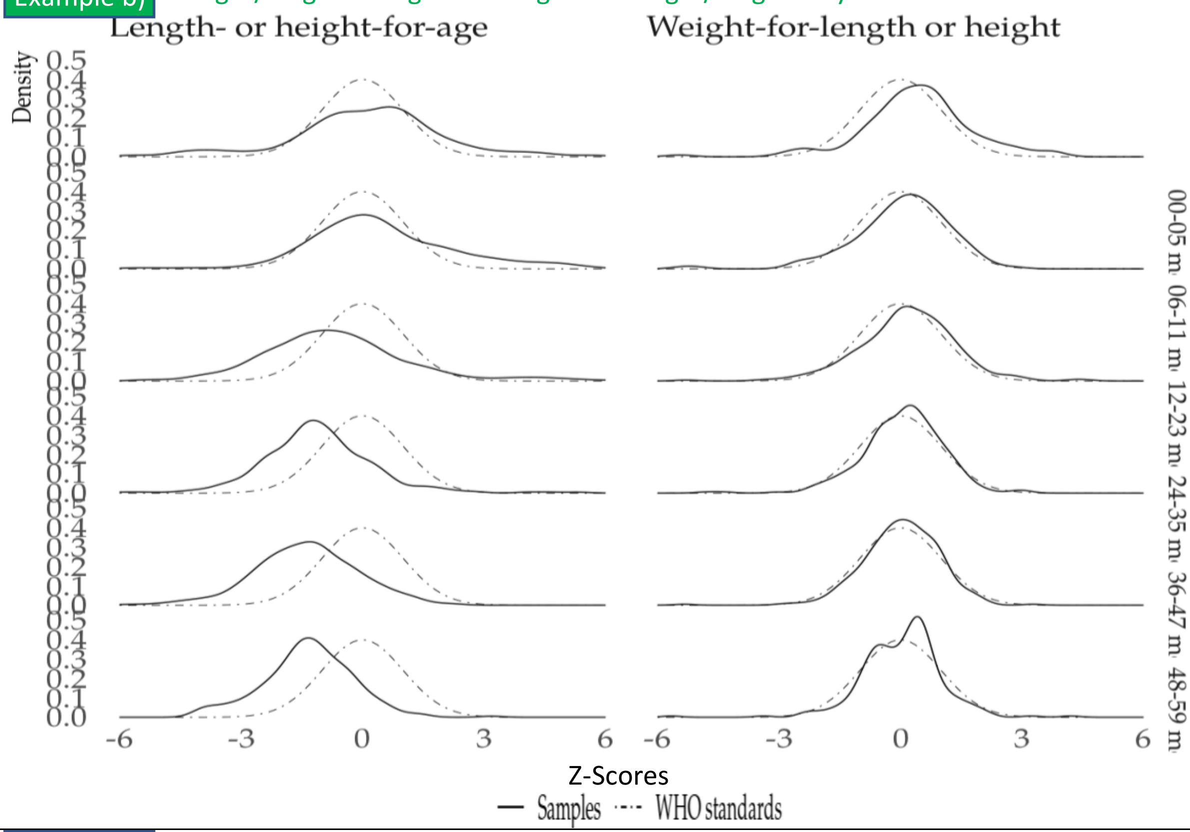
The length/height-for-age distribution shown in **Example a** represents the expected shape for a population that is stunted: the distribution is shifted to the right, where boys and girls have the same shape curves, but with the boys being shifted slightly more to the left than girls. The distributions are smooth and bell shaped, smoothly decreasing with no bumps at the right tail (as expected) with negligible area beyond +2 z-scores for height-for-age. The shape and smoothness of the distributions as well as the slight difference between boys and girls in Example a do not present any quality concerns. However, the bumpiness of the curves, along with the notable differences in shape seen between the curves for boys and girls in **Examples b and c**, would most likely be due to data quality issues.

iii) Z-score distribution by index and age group

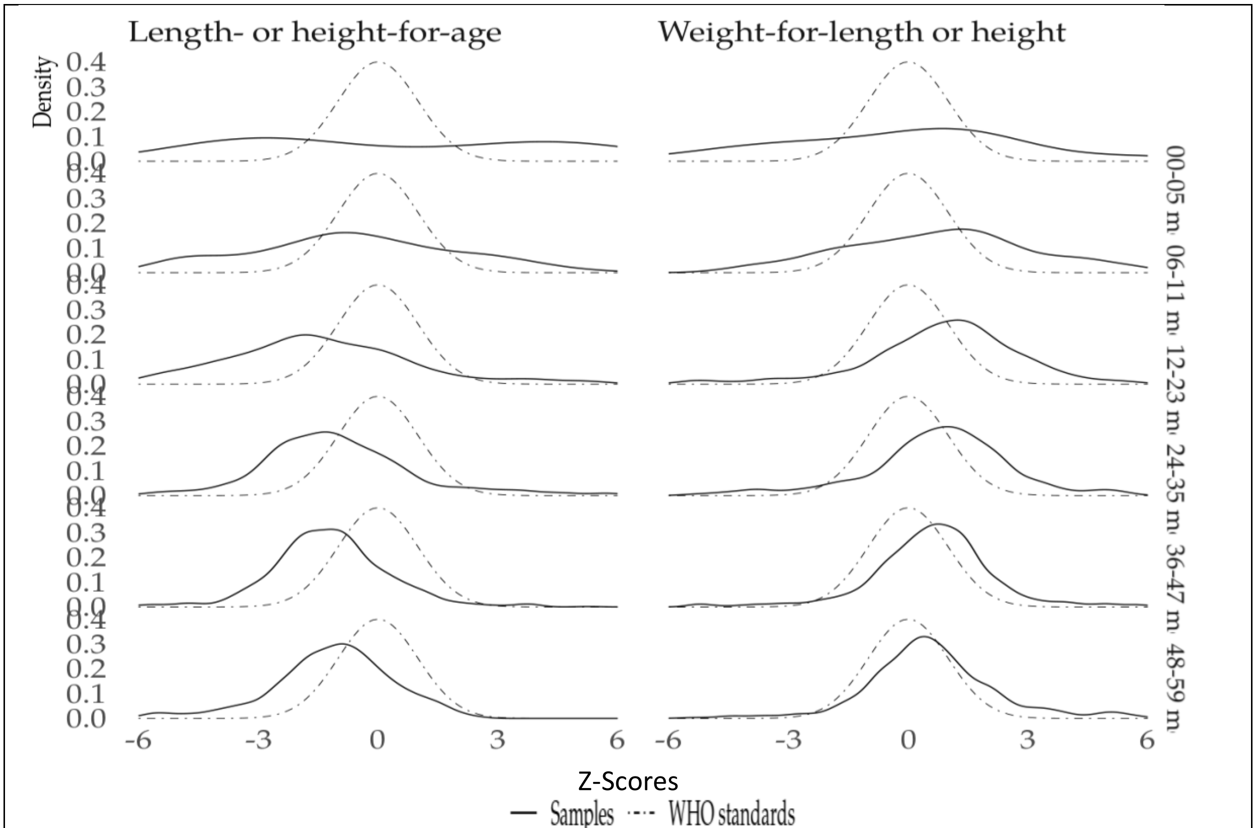
Example a) Length/height-for-age and weight-for-length/height only



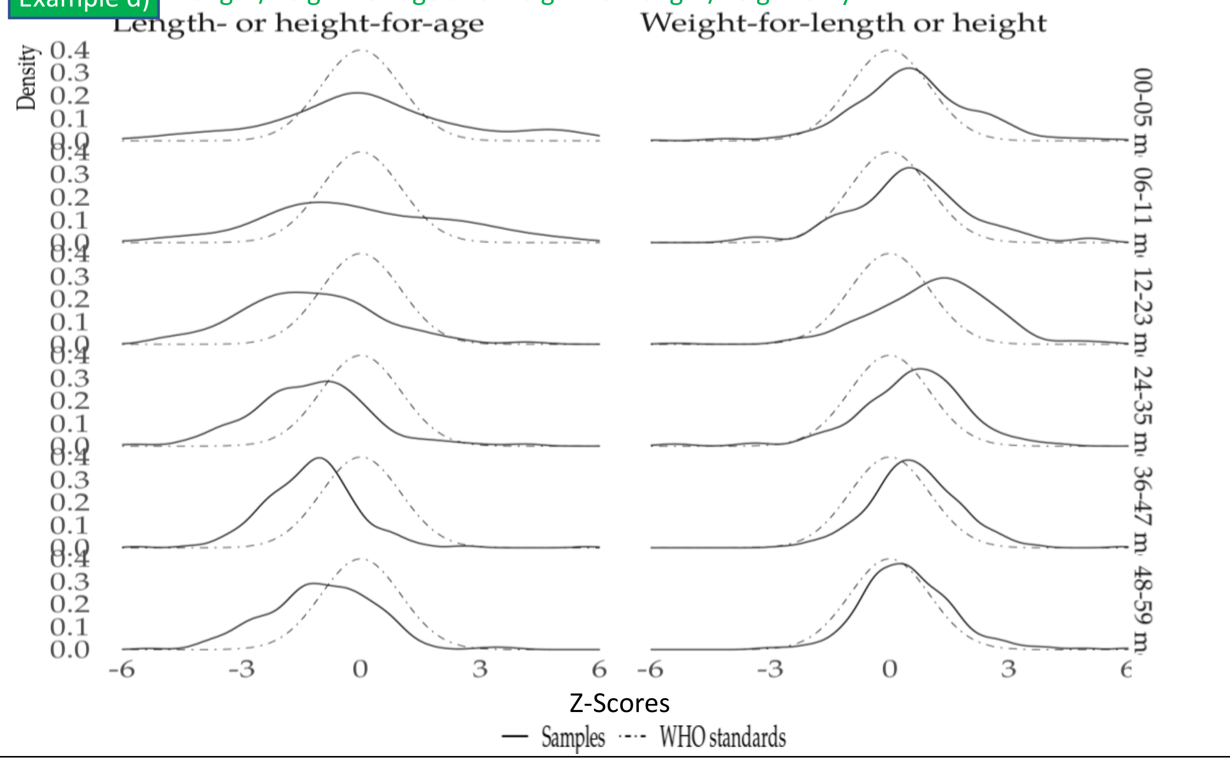
Example b) Length/height-for-age and weight-for-length/height only



Example c) Length/height-for-age and weight-for-length/height only



Example d) Length/height-for-age and weight-for-length/height only



Interpretation notes: Similar to Section 5.a, these graphs provide information about the distribution of malnutrition in the survey population. However, in this subsection, the distribution is shown for six separate sub-age groups (0–5 months, 6–11 months, 12–23 months, 24–35 months, 36–47 months and 48–59 months). The survey distribution is shown as a dark solid line against the WHO standard (the light dotted line that has a normal distribution with a mean of 0 and a standard deviation of 1). As in Sections 5.a and 5.b, the output in the standard Anthro Analyser report contains four graphs, one for each of: (i) length/height-for-age; (ii) weight-for-length/height; (iii) weight-for-age; and (iv) BMI-for-age. However, as only length/height-for-age and weight-for-length/height are required for the SDG indicators on stunting, wasting and overweight, the three scenario examples (a, b, c and d) in this section only contain two graphs.

The length/height-for-age distribution shown in **Example a** represents a population that is stunted, where the distribution for each age group is shifted to the right, and as expected, with the older children (24–59 months) being shifted more to the left than the younger age groups (0–11 months). The distributions are smooth and bell shaped, decreasing with no bumps (as expected) at the right tail and with negligible area beyond +2 z-scores. The similar shape of the distributions seen among the various age groups, as well as the smoothness of the distributions in Example a), do not present any quality concerns.

In contrast, **Examples b, c and d** would all raise concerns about data quality. In **Example b**, length/height-for-age has a larger than expected area above +2 for the age groups of 0–5 months, 6–11 months and 12–23 months, suggesting error in length and/or age data. The double peak (bimodal distribution) seen for 0–5 months as well as varying distribution shapes among age groups could also be due to data quality issues and warrants further investigation. **Example c** shows very flat distributions (i.e., no visible peak) for both the 0–5-month and 6–11-month age groups, showing substantial error on the right tail with a large portion of the entire area falling above +2. **Example d** also has severe inflation on the right tail for the youngest age groups.

The issues described for **Examples b, c and d** would most likely be due to data quality issues and should be a flag for further investigation of potential issues related to measurements or data entry. As mentioned before, concerns related to length/height distribution (e.g., many children with lengths/heights that seem excessive for their age), should also raise equal concerns about the weight-for-length/height distribution.

6. Percentage of flagged z-scores based on WHO flagging system by index

Interpretation notes: This graphic indicates the percentage of children that fall outside the range considered to be biologically plausible.[§§] The currently recommended flagging system to detect biologically implausible z-score values was defined based on the 2006 WHO Child Growth Standards. Cut-offs were defined on the basis of what is considered to be biologically implausible – in other words, incompatible with life. For example, a child with a weight-for-height z-score below -5 would be considered so dangerously thin that they would likely not have survived. These were defined as general cut-offs for systematically flagging z-scores in the sample that would generally indicate gross errors in measurement or data entry.

In some settings, however, there can be children beyond these cut-offs. In one national population, there was evidence of about 1 per cent of cases being beyond the right tail flag depicting overweight in this age group, although there was no evidence of biologically implausible heights in the 2– 5-year age group in this study.[***]

The flagged values are calculated using the z-scores for each child in the survey for whom indicators of malnutrition could be calculated. This means that children missing a key variable, such as height, weight, sex or age, are not included in the denominator of the percentages in the graph above, for indices that require the missing data. In other words, a child missing only age in months but who has a weight and a height will not be in the denominator of the percentage for zlen_flag (as the height-for-age index requires age in months). However, that child *can* be included in the denominator for the zwfl_flag (as calculation of weight-for length/height does not require age in months).

The implausible z-score value ranges are <-6 and >+6 for length/height-for-age (zlen_flag), <-5 and >+6 for weight-for-age (zwei_flag) and <-5 and >+5 for weight-for-length/height (zwfl_flag). While some children may truly have z-score values beyond the implausible value cut-offs recommended by the WHO (e.g., a child may be overweight to a degree that results in a weight-for-length/height z-score above +5), such cases would rarely occur in any sampled population. Using the study[‡] which found about 1 per cent of children had z scores beyond the flagged value for overweight, a threshold of 1 per cent was used to denote concern about the accuracy of the malnutrition estimates, with increasing values likely leading to larger inaccuracies in malnutrition estimates.

[§§] Note that the graphs in Section 5 only show the children within the ranges of -6 to +6 z-scores, closely aligned with the ranges considered to be biologically plausible for most malnutrition indices according to the 2006 WHO Child Growth Standards, but not any children beyond those ranges. Section 6 covers children that would be beyond the ranges in the graphs in Section 5.

[***] Freedman DS, Lawman HG, Skinner AC, McGuire LC, Allison DB, Ogden CL. 'Validity of the WHO cutoffs for biologically implausible values of weight, height, and BMI in children and adolescents in NHANES from 1999 through 2012', *Am J Clin Nutr*. 2015;102(5):1000-1006. doi:10.3945/ajcn.115.115576

7. Z-score summary table

Interpretation notes: The following tables contain information about the distribution of malnutrition in the survey population. The main indicator considered from this table with respect to data quality is the standard deviation (SD) for length/height-for-age z-score (SD (zlen)), weight-for-age z-score (SD (zwei)) and weight-for-length/height z-score (SD (zwfl)). The SD is a statistical measure that quantifies the amount of variability in a dataset. The smaller the SD, the closer the data points tend towards the mean. The higher the SD, the greater the spread of data points. When the distribution is skewed (i.e., asymmetric, as is usually seen for malnourished populations), the SD is a combination of the spread and the degree of skewness. Moreover, the spread of the distribution can also be affected by population heterogeneity.

Poor data quality, such as that which occurs with measurement errors, can lead to inflation of the SDs; however higher SDs are not necessarily linked to poor quality data. A study based on an analysis of the 474 nationally representative household surveys from 112 countries in the Joint Malnutrition Estimates (JME) dataset as of January 2019 was carried out aiming to find reasonable cut-offs for the indication of poor quality in anthropometric data. The median (and 5th and 95th percentiles) were 1.54 (1.21 and 2.03) for length/height-for-age z-score, 1.27 (1.04 and 1.72) for weight-for-length/height z-score and 1.22 (1.06 and 1.52) for weight-for-age z-score. The wide range of SDs derived from these surveys may be due to a combination of varying degrees of data quality and heterogeneity in the survey populations with regard to nutrition status and its determinants. Nevertheless, the 95th percentiles from the re-analysed surveys in the global database reflect very large SD values for both HAZ and WHZ; in fact, some of the SDs in the dataset are larger than would be reasonably explained by population heterogeneity or high degree of malnutrition, and are thus more likely to reflect poor data quality. It can be confidently stated that as SDs for anthropometric indices become larger, they can more reasonably be attributed to poor data quality rather than population heterogeneity or degree of malnutrition. While additional systematic investigations are needed to determine index-specific standard cut-offs, based on these findings using the JME dataset, SDs above the 95th percentile (the 95th percentile rounded to about 2.0 for length/height-for age and 1.7 for weight-for-length/height) are considered very concerning, and even those above the 80th percentile (the 80th percentile rounded to about ≥1.8 for length/height-for age and ≥1.5 for weight-for-length/height) should raise concerns. SDs for nationally representative populations that are <1 would also present potential data quality issues.

The data quality review also considers how SDs vary between population groups. Among the 473 surveys in the JME as of January 2019, the SDs between boys and girls were very similar, being on average 0.05 higher for boys than for girls for length/height-for-age and 0.06 higher for boys than girls for weight-for-length/height. Differences larger than 0.10 between boys and girls for height-for-age and greater than 0.12 for weight-for-length/height z-scores (be it for the entire age group or individual age groups) could indicate a data quality issue.

i) Z-score distribution of unweighted summary statistics by index

Group	Unweighted N	Mean (zlen)	Standard deviation (zlen)	Skewness (zlen)	Kurtosis (zlen)	Mean (zwei)	Standard deviation (zwei)	Skewness (zwei)	Kurtosis (zwei)
All	10175	-1.41	1.29	0.24	4.13	-0.86	1.08	-0.19	3.61

Group	Unweighted N	Mean (zlen)	Standard deviation (zlen)	Skewness (zlen)	Kurtosis (zlen)	Mean (zwei)	Standard deviation (zwei)	Skewness (zwei)	Kurtosis (zwei)
Age group: 00-05 mo	1099	-0.52	1.07	-0.13	1.33	-0.60	0.97	0.85	2.93
Age group: 06-11 mo	1187	-0.90	1.35	0.56	4.86	-0.83	1.20	-0.04	3.28
Age group: 12-23 mo	2253	-1.47	1.34	0.31	4.36	-0.93	1.18	-0.03	3.53
Age group: 24-35 mo	2155	-1.54	1.29	0.14	3.55	-0.84	1.14	-0.29	3.26
Age group: 36-47 mo	1941	-1.50	1.24	0.17	3.70	-0.79	0.93	-0.17	3.22
Age group: 48-59 mo	1512	-1.43	1.14	-0.20	3.47	-0.89	0.86	-0.60	4.22
Sex: Male	5040	-1.53	1.30	0.16	3.78	-0.93	1.10	-0.18	3.62
Sex: Female	5134	-1.30	1.28	0.32	4.48	-0.78	1.05	-0.17	3.57
Geographical region: 1	600	-1.44	1.25	0.01	4.24	-0.86	1.11	-0.34	3.37
Geographical region: 2	577	-1.44	1.23	0.22	3.15	-0.90	1.06	0.09	2.94
Geographical region: 3	637	-1.47	1.16	0.14	3.53	-0.86	1.00	-0.02	3.29

ii) Z-score distribution of unweighted summary statistics by index (continued)

Group	Unweighted N	Mean (zbmi)	Standard deviation (zbmi)	Skewness (zbmi)	Kurtosis (zbmi)	Mean (zwfl)	Standard deviation (zwfl)	Skewness (zwfl)	Kurtosis (zwfl)
All	10175	0.06	1.14	-0.19	3.71	-0.10	1.11	-0.23	3.59
Age group: 00-05 mo	1099	-0.41	1.12	0.34	2.55	-0.15	1.23	0.05	1.90
Age group: 06-11 mo	1187	-0.43	1.18	-0.24	3.24	-0.42	1.18	-0.28	3.37
Age group: 12-23 mo	2253	-0.04	1.17	-0.18	3.21	-0.29	1.15	-0.16	3.09
Age group: 24-35 mo	2155	0.22	1.16	-0.22	3.85	0.00	1.13	-0.21	3.72
Age group: 36-47 mo	1941	0.28	1.08	0.04	3.94	0.12	1.03	-0.02	3.68
Age group: 48-59 mo	1512	0.08	0.95	-0.28	4.49	0.00	0.96	-0.31	4.16
Sex: Male	5040	0.06	1.18	-0.19	3.71	-0.13	1.15	-0.24	3.54
Sex: Female	5134	0.05	1.09	-0.20	3.66	-0.07	1.06	-0.20	3.59
Geographical region: Bo	600	0.07	1.21	-0.06	3.74	-0.10	1.20	-0.17	3.48
Geographical region: 2	577	0.01	1.18	-0.05	3.61	-0.15	1.15	-0.06	3.51
Geographical region: 3	637	0.09	1.12	-0.15	3.50	-0.08	1.09	-0.19	3.32

8. Summary of recommended data quality checks

The Working Group on Anthropometry Data Quality recommends that data quality be assessed and reported based on assessment of the following seven parameters: (i) completeness; (ii) sex ratio; (iii) age distribution; (iv) digit preference of heights and weights; (v) implausible z-score values; (vi) standard deviation of z-scores; and (vii) normality of z-scores.

The Working Group recommends that (i) data quality checks should not be considered in isolation; (ii) formal tests or scoring should not be conducted; (iii) and the checks should be used to help users identify potential issues with the data quality to improve interpretation of the malnutrition estimates from the survey. Although not exhaustive, a summary of details on the various checks is provided below to facilitate their use. Full details and more comprehensive guidance, including on how to calculate, can be found at the full report on the Working Group's recommendations.[‡‡‡]

(i) Completeness: although not all statistics are included in the WHO Anthro Survey Analyser, the final survey report should include the following at minimum:

- Primary sampling units (PSUs): Percentage of selected PSUs that were visited.
- Households: Percentage of selected households in the PSUs interviewed or recorded as not interviewed (specifying why).
- Household members: Percentage of household roster members with interviews that were completed.
- Children: Percentage of all eligible children that were interviewed and measured, or recorded as not interviewed or measured (specifying why), with no duplicate cases.
- Dates of birth: Percentage of dates of birth for all eligible children that were complete.

(ii) Sex ratio:

- What – ratio of girls to boys in the survey and compare to expected for country. The observed ratios should be compared to the expected patterns based on reliable sources.
- Why – to identify potential selection biases.

(iii) Age distribution:

- What – age distributions by age in completed years (6 bars weighted), months (72 bars) and calendar month of birth (12 bars), as histograms.

[‡‡‡] Working Group on Anthropometric Data Quality, for the WHO-UNICEF Technical Expert Advisory Group on Nutrition Monitoring, *Recommendations for improving the quality of anthropometric data and its analysis and reporting*. Available at www.who.int/nutrition/team (under "Technical reports and papers").

- Why – to identify potential selection biases or misreporting.

(iv) Height and weight digit preference:

- What – terminal digits as well as whole number integer distributions through histograms.
- Why – digit preference may be a tell-tale sign of data fabrication or inadequate care and attention during data collection and recording. When possible, it should be presented by field team or other relevant disaggregation categories.

(v) Implausible z-score values:

- What – the percentage of cases outside of WHO flags[‡‡‡] for each HAZ, WHZ and WAZ.
- Why – a per cent above 1 per cent can be indicative of potential data quality issues in measurements or age determination. It should be presented by team or other relevant disaggregation categories.

(vi) Standard deviations:

- What – SD for each HAZ, WHZ and WAZ.
- Why – large SDs may be a sign of data quality problems and/or population heterogeneity. It is unclear what causes SDs size and more research is needed to determine appropriate interpretation. It should be noted that SDs are typically wider for HAZ than WHZ or WAZ, and that HAZ SD is typically widest in the youngest age groups (0–5 month) and increases as children age through to 5 years. No substantial difference should be observed between boys and girls. It should be presented by team or other relevant disaggregation categories.

(vii) Checks of normality:

- What – measures of asymmetry (skew) and tailedness (kurtosis) of HAZ, WHZ and WAZ, as well as density plots.
- Why – general assumption that three indices are normally distributed but unclear if applicable to populations with varying patterns of malnutrition. One can use the rule of thumb ranges of <-0.5 or >+0.5 for skewness to indicate asymmetry and <2 or >4 for kurtosis to indicate heavy or light tails. Further research is needed to understand patterns in different contexts. Comparisons among the distribution by disaggregation categories may help with the interpretation of results.

[‡‡‡] World Health Organization, *WHO Anthro Software for personal computers – Manual*, 2011. Available at www.who.int/childgrowth/software/anthro_pc_manual_v322.pdf?ua=1.

Annex 6: Sample JME data source review

This review form is a sample and does not pertain to any specific country. Text in green is explanatory and not part of a standard review form; it is intended to serve as a resource to provide the reader with additional information. This sample represents a survey of good quality that would be accepted into the JME dataset. For examples of some data quality outcomes that could be problematic, see Annex 4.

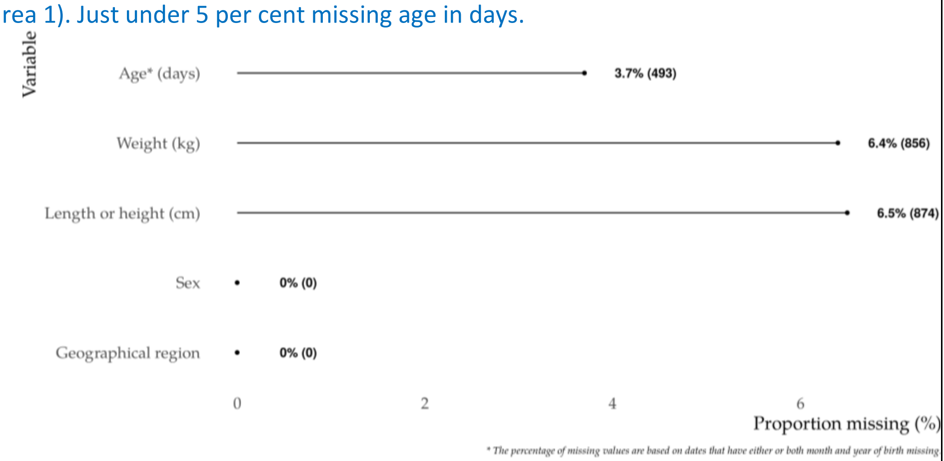
JME review form

COUNTRY	Country A
SURVEY NAME AND SOURCE	Survey name
DAY/MONTH/YEAR OF START AND END OF FIELDWORK	12 May – 15 August, 2018
METHODOLOGY *Key details such as those presented in this example are to be extracted from the survey report and any supporting documentation. A summary of the methodology is an important part of the review as it captures information on key aspects of data collection, such as equipment used and specifics about training of the measurers. When information on key topics is missing, incomplete or lacking detail (e.g., whether the scale was digital, whether the height measuring device was made from a rigid material like wood), it is not possible to conduct a thorough review.*	**Training**: Biomarker (including anthropometry) training was held from April 24 to May 2, 2018. Plenary lectures were held on the technical aspects; other training tools included video and hands-on demonstrations, instructions on how to fill out the questionnaire and transmittal sheets and instructions on data quality procedures. In addition, break-out sessions were held daily during which trainees had the opportunity for hands-on practice with both adults and children. A total of four anthropometry standardization exercises with 90 children (45 under age 2 and 45 over age 2) were undertaken. Following the standardization exercises, the results were presented. General observations on accuracy (difference between the reference value and the participant's value) and precision (difference between the first and second readings) were discussed. The field coordinators were trained on the use of the biomarker checklist. Random re-measurements were also implemented for quality assurance and households re-visits were undertaken for re-measurements for flagged cases involving children whose z-score values were less than -3 or greater than 3. A two-day field practice was also conducted. **Anthropometry equipment and measures**: Weight measurements were taken using lightweight scales with digital displays (model no. SECA 878U). Height/length measurements were taken using a standard measuring board (Shorr Board®). Recumbent length (lying down) was measured for children younger than age 24 months; standing height was measured for older children. **Data quality assessment during fieldwork:** The survey included quality assurance procedures, undertaken in real time during data collection, including remeasurement of all children with data outside of pre-specified flagged values on a subsequent day and remeasurement of the height and weight of a random selection of children (5 per cent) on a subsequent day from the initial measurement. **Supervision:** Fieldwork monitoring was an integral part of the survey, and several rounds of monitoring were carried out by the core team. The monitors were provided with guidelines for overseeing the fieldwork. Weekly field

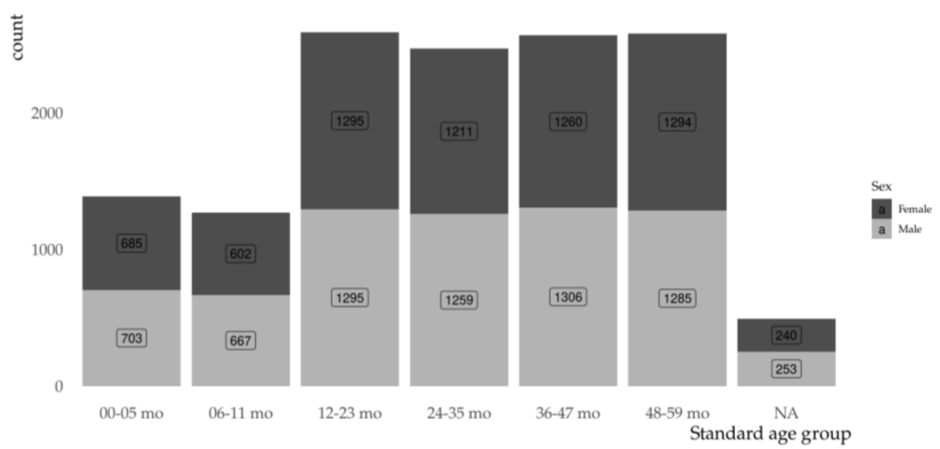
	check tables were generated from the completed interviews sent to the central office to monitor fieldwork progress, and regular feedback was sent to the teams. **Data collection and processing**: Computer-assisted personal interviewing tablets were used for data collection of the household and individual interviews, but data related to biomarkers (including anthropometry) were initially recorded on paper and subsequently entered into interviewers' tablet computers. The biomarker paper questionnaires were compared with electronic data files to check for any inconsistencies in data entry. Data entry and editing were carried out using the CSPro Systems software package.
COVERAGE *The intended coverage of the survey (e.g., nationally representative) is to be recorded here, as well as response rates at various levels (e.g., among enumeration areas, households and eligible children). In cases where response rates have been based on an inappropriate denominator (e.g., based on the number of children anticipated rather than the total number of eligible children taken from the household listing among interviewed households), the actual response rate would be unknown. When information on this topic is missing or based on the incorrect denominator, it is not possible to conduct a thorough review.*	The survey was sampled to be nationally representative with urban/rural stratification. Response rates: • **Enumeration area (EA) response rate**: 100 per cent (578/578) • **Household response rate**: 98.5 per cent • **Under-five year old response rate (completed interview)**: 98.5 per cent (13,157 out of 13,355 eligible children from the interviewed households). Note: individual items may be missing from some questionnaires among children with completed interviews (e.g., weights or heights may be missing among the completed interviews). This is noted below in the 'missing data' graphic of the Dataset Checking section, below.
TARGET POPULATION	Children aged 0–59 months who were de facto members of the household
ESTIMATES	Height-for-age (HAZ), Weight-for-height (WHZ) and Weight-for-age (WAZ) related indicators

SAMPLING DESIGN *The following are to be recorded here based on the survey report and other supporting documents: (i) source of first stage sampling frame (and year); whether the frame was updated, and if so how; how EAs were selected, etc.; (ii) how the second stage frame was developed and how households were selected; (iii) how children from within sampled households were selected. Some key points considered are the age of sampling frame (which should not be >10 years old), methods to develop the second stage frame need to be probabilistic (e.g., should not entail any non-recommended practices, such as: creation of the second stage sampling frame at the same time as implementation of survey fieldwork; selection of households by survey field teams, replacement of EAs, replacement of households, etc.). When information on this topic is missing, it is not possible to conduct a thorough review.*	The sampling frame was the Population and Housing Census from 2014, of which the list of EAs served as a foundation to estimate the number of households and distinguish EAs as urban or rural for the survey sample frame. The sample for the survey was a stratified sample selected in two stages. Stratification was achieved by separating each district into urban and rural areas. In total, 31 sampling strata were created. Samples were selected independently in every stratum via a two-stage selection process. Implicit stratifications were achieved at each of the lower administrative levels by sorting the sampling frame before sample selection according to administrative order, and by using a probability proportional-to-size selection during the first sampling stage. In the first stage, 578 EAs were selected with probability proportional-to-size. A household listing operation was carried out in all selected EAs, and the resulting lists of households were verified and served as a sampling frame for the selection of households in the second stage. In the second stage of selection, a fixed number of 24 households were selected in every cluster through equal probability systematic sampling, resulting in a total sample size of approximately 13,872 selected households. The household listing was carried out using tablets, while random selection of households was carried out through computer programming. The survey interviewers interviewed only the pre-selected households. To prevent bias, no replacements of households and no changes of the pre-selected households were allowed in the implementing stages. The Household Questionnaire listed all members of and visitors to selected households. All listed children 0–59 months of age who were household members or visitors who stayed the night before constituted the group that was eligible for anthropometry.
SAMPLE SIZE	578 EAs, 13,872 households (24 per EA) with a total of 13,355 children under 5 years eligible for anthropometry.
SAMPLE SIZE ESTIMATION METHOD	Based on multiple indicators in a multi-topic survey
SOFTWARE USED FOR ANALYSIS	CSPRO and STATA
FLAGS USED AND PERCENTAGE *This summarizes the percentage of children with a biologically implausible z-score (too tall for their age, too thin for their height, etc.) as defined by the 2006 WHO Child Growth Standards. While some children may truly have z-score values beyond the implausible value cut-offs recommended by WHO, they would rarely occur in any*	Standard WHO flags were applied and a very low percentage flagged as biologically implausible outliers was found in this survey (0.3 per cent for HAZ, 0.4 per cent for WHZ and 0.3 per cent for WAZ).

<table>
<tr>
<td>population. Therefore, a percentage flagged as biologically implausible of >1 per cent is of concern. See Section 6 of Annex 4 for more details.</td>
<td></td>
</tr>
<tr>
<td>DATASET CHECKING
This sample review represents a near optimal scenario. See Annex 4 for some examples of data quality outcomes that would represent scenarios that might negatively impact the accuracy of malnutrition estimates for the various outcomes presented here.</td>
<td>Missing Data: just over 5 per cent missing weight or length/height; only one geographic area had >10 per cent missing (11.1 per cent missing height and weight in area 1). Just under 5 per cent missing age in days.

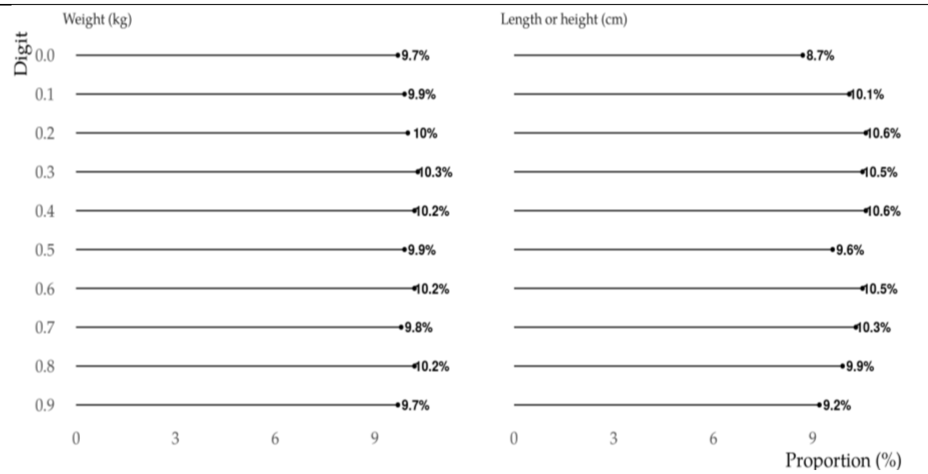


Age Distribution: age and sex distribution look good; very equal between all one-year age groups and proportion between the two age groups among <1 year olds.

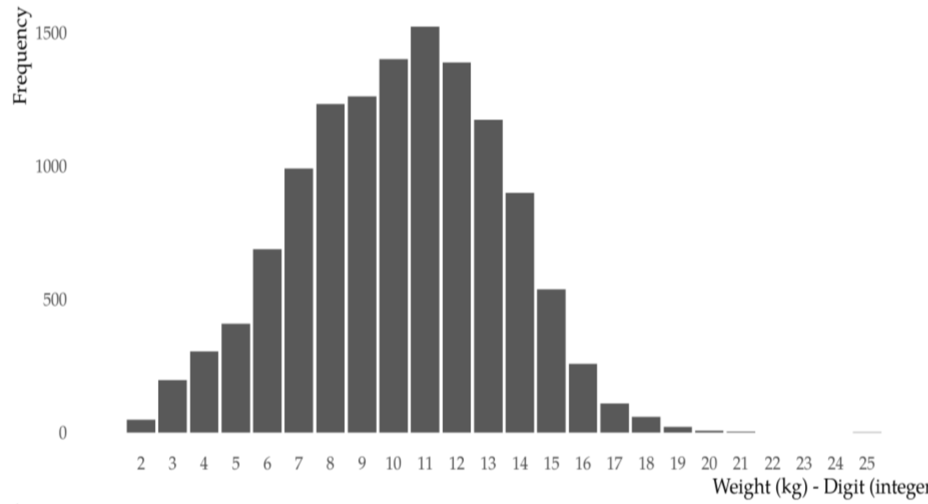


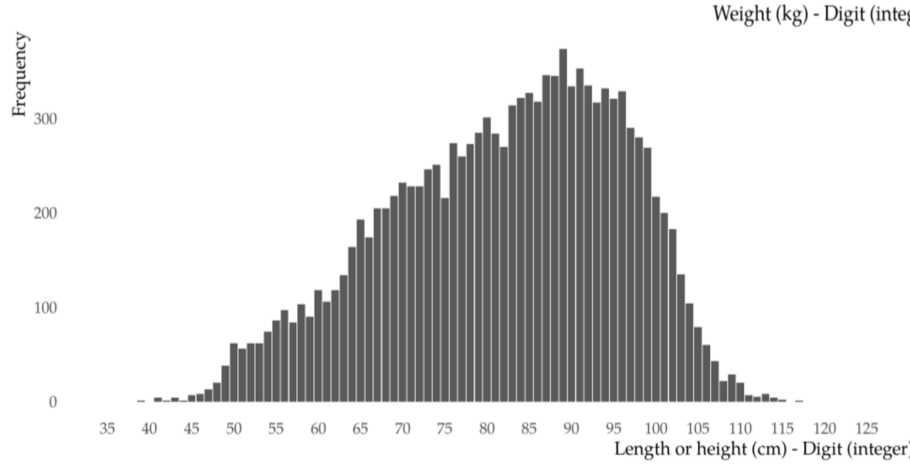
Infants <8 months of age reported to be measured in a standing position:
0 per cent of 0–8 month-olds were in the wrong (standing) position.

Decimal digit preference:
Terminal digit: looks good at national level and in all regions (regions not copied here).</td>
</tr>
</table>



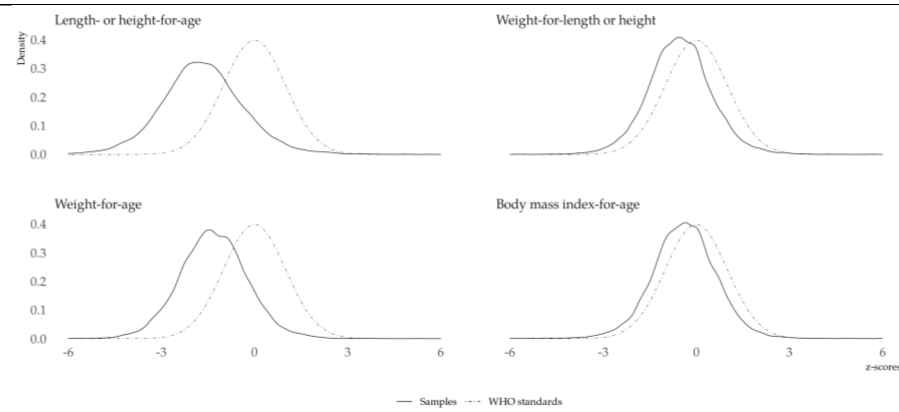
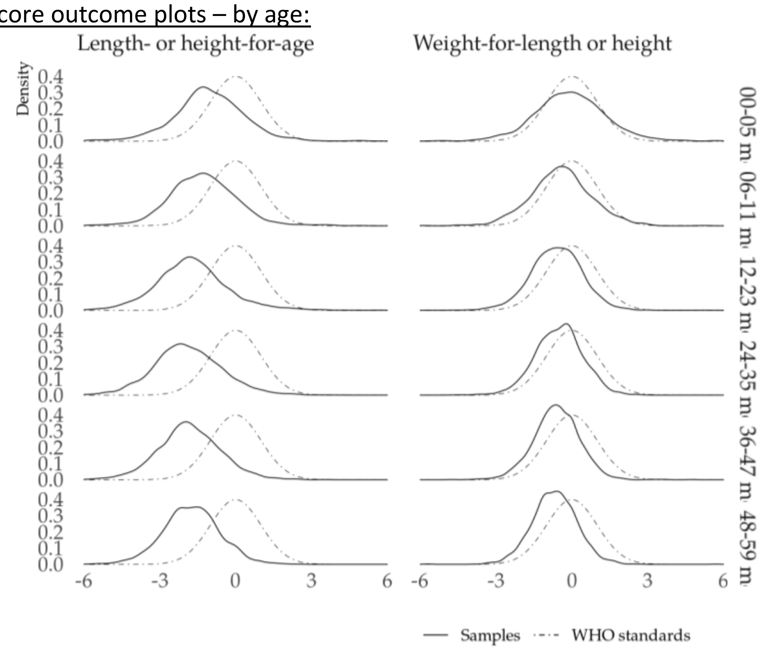
Whole number distributions: look very good for both weight and height.





Z-Score distribution: Distributions look very good for all indicators, even by age groups.

<u>Z-Score outcome plots – overall:</u>

	 <u>Z-Score outcome plots – by age:</u> 
GENERAL QUALITY ISSUES *This is a summary of the above data quality outputs – a space to highlight the key quality issues.*	No major issues, percentage of missing weights and heights is the main concern but not likely to affect estimates sufficiently to not include in the database.
REPORTED RESULTS *These are the reported results, and in cases where raw data are available, the re-analysed results would be included here.*	Estimate (95 per cent confidence interval) Severe wasting: 1.0 (0.8; 1.2) Wasting: 6.4 (5.9; 7.0) Overweight: 1.4 (1.2; 1.7) Stunting: 41.6 (40.1; 43.2) Underweight: 26.4 (25.2; 27.6)

TRENDS

Other points from the database are reviewed together with the new points under review (in yellow). In this case, the trends appear plausible, whereas large declines or increases in a short period of time (e.g., stunting reduced by one third, from 24–16 per cent in 2 years, would be considered implausible). In cases where changes between data points are large, UNICEF and WHO country teams are consulted to obtain information on national efforts to address malnutrition in the country.

Year	severe wasting	wasting	overweight	stunting	underweight
1992	1.2	6.6	1.7	61.2	36.2
1993-94		15.9		55.0	40.7
1995		8.7		55.8	31.2
1997	2.2	9.0	2.3	57.2	38.2
2003-04	5.3	14.9	4.3	53.1	36.5
2008-09				49.7	
2012-13	0.9	7.7	1.3	46.2	33.1
2018	1.0	6.4	1.4	41.6	26.4

Trends appear plausible

DISCUSSION

Some points to highlight from the review are listed here by the reviewer and a suggestion is made by the main reviewer who filled in this form to be discussed jointly by the JME team.

- No major issues based on data available to review.
- Largest issue is about 6 per cent of eligible children did not have weights or heights in the dataset, but this is not of a level to reject as many surveys with missing weights/heights beyond this level are included in the JME.
- Suggest accepting this survey for all indicators.

RECOMMENDATION

Final recommendation after the JME team has reviewed together.

Suggest accepting this survey for al indicators.