The previous two chapters laid out guidelines and checklists for the collection of high quality anthropometric data. For thorough and transparent reporting on survey quality, reporting on actions taken in its various stages including the planning, designing, field work, data entry and analysis are required as are thorough reports on data quality and estimates.

The present chapter provides guidance for best practices in data processing and reporting. It has four sections:

3.1. Data quality assessment;
3.2. Data analysis;
3.3. Data interpretation;
3.4. Harmonized reporting and recommended release of data.

A variety of software is available, some allowing a full range of activities from data entry to analyses and reports. For data analyses, the standard approach adopted for the WHO Global Database on Child Growth and Malnutrition (13) and the UNICEF-WHO-WB Joint Child Malnutrition Estimates (JME) (14) to ensure comparability across countries and years can be achieved using currently available Anthro software or macros (SAS, SPSS, STATA and R). WHO recently developed an online tool for anthropometric data analyses that updates Anthro methodology to provide more accurate estimates of standard errors and confidence intervals for prevalence and mean z-scores. The WHO Anthro Survey Analyser is a tool based on the R and R Shiny package that provides interactive graphics for data quality assessment and a summary report template offering key outputs, e.g. z-score distribution graphics in terms of various grouping factors and nutrition status tables with accompanying prevalence and z-score statistics.

### 3.1. Data Quality Assessment

It is recommended that data quality be assessed to determine whether there are any issues that might lead to biased estimates, have an impact on interpretability or limit the potential use of findings. In general, data quality assessment aims to pinpoint two main types of bias: selection bias and measurement bias. Selection bias is related to the representativeness of sampled households and children. Measurement bias is generally a consequence of inaccurate measurements of weight, height and date of birth. These biases may be due to either random or systematic errors.

High quality measurement of all information needed (length/height, weight and date of birth of children under five years of age) to generate the anthropometric indicators as well as sampling and field procedures are essential for generating accurate child malnutrition estimates. The following sections describe the checks recommended for assessing the quality of anthropometric survey data.

The checks described in this section address the following topics:

3.1.1. Completeness;
3.1.2. Sex ratio;
3.1.3. Age heaping;
3.1.4. Digit preferences of heights and weights;
3.1.5. Implausible z score values;
3.1.6. Standard deviation of z scores;
3.1.7. Normality (skewness and kurtosis) of z scores.

For each of these checks, text under bold headings explain what it is, why it is recommended to use it, how it should be used or calculated, and then describe how to interpret and report it.

Applied in combination, these checks can provide insight into the quality of the anthropometric data that serves to support the interpretation of malnutrition estimates. It is generally recommended that data quality be appraised not on the basis of isolated checks but by considering all of them conjointly. One limitation to current assessments of data quality is that no consensus exists about cut-offs for data quality checks that indicate a definitive problem. Further research is needed to determine appropriate cut-offs for data quality measures and whether other data quality checks might be helpful.

Data quality checks should be conducted for the entire sample population and separately for each main measurer or team. If potential data quality issues are detected at the national level, data quality checks may also be carried out for subpopulations within the sample, assuming the sample size is sufficient for the test/assessment for the specific disaggregation categories in question. Subpopulations should be disaggregated for sex and age, and if feasible
for region, mother’s education and wealth quintile. Disaggregation categories can provide valuable information to support the interpretation of data quality, although it is not always apparent whether differences are due to sample heterogeneity or quality issues.

Some data quality checks are made before excluding implausible values whereas other checks are made after exclusions: this is indicated below for each check. Prior exclusion of implausible values is only done for data quality checks relating to distribution.

Some data quality checks are made using unweighted and others using weighted samples, as indicated below for each check described. Weighted analyses are recommended where a comparison is being made to an external reference population. Conversely, unweighted analyses are recommended where measurement error is being evaluated, ensuring that each individual measurement has an equal sample weight.

It is recommended that data quality assessment findings be included in all survey reports that provide estimates for child anthropometric indicators. At the time of this publication, the WHO Anthro Survey Analyser includes most of the data quality checks described below, and follows the recommendations included in the present document (see sample report in Annex 9). Several other software packages are available that include some but not all the recommended checks. It is not known whether all these packages follow recommended calculation methods; some include formal tests, cut-offs or scoring systems not recommended in this report.

Errors in measurement and selection are important since they can lead to inaccuracies in prevalence estimates. Insight into data quality in every survey helps to interpret results especially when investigating trends over time. While it is not always possible to distinguish between them, there are two main types of measurement error, systematic and random error. In many other fields, random errors may have less of an influence on estimates and systematic error is the main concern, as reported indicators are based on the mean, median or estimated coverage. However, since the malnutrition indicators discussed in this report relate to prevalence at the tail ends of the distribution, both random and systematic measurement error are of concern and need to be minimized. In fact, for malnutrition estimates based on prevalence at the distribution tails, three major sources can threaten the accuracy of prevalence estimates: (a) selection errors (e.g. errors in identifying sampled households or eligible children in these households to be measured), (b) systematic measurement error, and (c) random measurement error. A number of different variables can be responsible for introducing systematic and/or random errors including date of birth (used to calculate age), length/height and weight. That is why transparent and thorough reporting on data quality and survey methodology is of such great importance for estimates related to malnutrition prevalence.

### 3.1.1 Completeness

**What**

When undertaking survey data collection in sampled households, it is necessary to ensure that the data collected are complete. In anthropometric surveys, this means ensuring not just that all eligible children are accounted for but includes checking the structural integrity of all aspects of the data. The following topics should be checked for structural integrity.

- **PSUs**: all selected PSUs are visited, although this may not be possible on some occasions, e.g. due to civil strife, flooding or other similar reason;
- **Households**: all selected households in the PSUs are interviewed or recorded as not interviewed (specifying why);
- **Household members**: all household rosters are complete, with all household members listed and information provided about their key characteristics, e.g. age, sex and residency;
- **Children**: all eligible children are interviewed and measured or recorded as not interviewed or measured (specifying why), with no duplicate cases;
- **Dates of birth**: dates of birth for all eligible children are complete.

**Why**

Assessing the completeness of collected data is an important aspect of verifying data quality. Errors during data collection and lack of data completeness in a survey can lead to non-representative or biased results. Being able to check data completeness provides confidence in the survey and how it has been implemented.
How to calculate

For each of the items listed above, the sampled proportion successfully interviewed should be reported, generally disaggregated by survey strata or sampling domain.

Usually it is possible to visit all PSUs, but in those surveys in which some PSUs are not visited the number of non-visited PSUs in each stratum should be stated. If selected PSUs are not visited, then it is generally necessary to make an adjustment in the analysis, e.g. sample weights to correct for under-sampling within a stratum. We recommend as a best practice that PSUs which cannot be visited, should not be replaced by other PSUs since this may introduce bias into the sample.

For **households**, the response rate (based on all contactable households) should be provided along with the completion rate (based on all selected households).

\[
\text{Household completion rate} = \frac{\text{Number of households with completed interviews}}{\text{Total number of households selected}}
\]

\[
\text{Household response rate} = \frac{\text{Number of households with completed interviews}}{\text{Total number of households contactable}}
\]

The total number of households contactable includes households completed (code 01), partially completed (code 02), with no household member at home or no competent respondent at home at the time of visit (code 03), refused (code 05), and dwelling not found (code 08), and excludes those where the entire household was absent for an extended period of time (code 04), dwelling vacant or address not a dwelling (code 06), dwelling destroyed (code 07) and other (code 96).

For **household members**, an assessment of the completeness of the household roster should be reported, comparing average household size and average number of children aged under five years by stratum or sampling domain with estimates of average household size and number of children from other sources.

\[
\text{Average household size} = \frac{\text{Number of household members}}{\text{Number of households completed}}
\]

\[
\text{Average number of children per household} = \frac{\text{Number of children under five}}{\text{Number of households completed}}
\]

Note: if the survey is using a de facto sample, the average number of children per household should be presented for de facto children, i.e. those who stayed in the household the previous night, rather than de jure children (usually resident).

For **eligible children**, typically all eligible children under 5 years should be reported unless a subsampling method is applied, showing the percentage of eligible children who had completed interviews.

\[
\text{Children completion rate} = \frac{\text{Number of children under five with completed interviews}}{\text{Number of eligible children under five}}
\]

Information on completeness of re-measurements should be presented for all eligible children, including random and flagged re-measurements.

\[
\text{Random remeasurement completion rate} = \frac{\text{Number of children with completed random remeasurements}}{\text{Total number of children selected for random remeasurements}}
\]
Flagged remeasurement completion rate = \( \frac{\text{Number of children with completed flagged remeasurements}}{\text{Total number of children with flagged measurements}} \)

In addition, present the percentage of children with a complete date of birth, including day of birth, and those with month and year of birth (but not day).

Children with complete date of birth = \( \frac{\text{Number of children with day, month and year of birth recorded}}{\text{Number of children with complete interviews}} \)

Children with partially complete date of birth = \( \frac{\text{Number of children with month and year of birth recorded, but not day}}{\text{Number of children with complete interviews}} \)

Children with incomplete date of birth = \( \frac{\text{Number of children with month or year of birth missing}}{\text{Number of children with complete interviews}} \)

Note: the second ratio related to children’s date of birth refers to the percentage of children with an imputed day of birth (i.e. where DD was imputed to 15 but MM and YYYY available to calculate age in months), and the third to the percentage of children with insufficient information to calculate an age in months (i.e. MM and/or YYYY are missing and these cannot be imputed for anthropometry z scores. The sum of the preceding three ratios—expressed as percentages—equals 100.

Completeness of measurement for length/height and weight should be presented by showing the proportion measured, the proportion absent, the proportion refused, and the proportion not measured owing to other reasons, for all eligible children.

For length/height, and for weight:

Children measured / absent / refused / other reason = \( \frac{\text{Number of children measured / absent / refused / other reason}}{\text{Number of children with complete interviews}} \)

Furthermore, the proportion of missing data for age, sex, residency based on the household questionnaire, whether measured lying or standing in the anthropometry module or for other variables used in the calculation of anthropometric z-scores should also be presented.

How to present

When presenting these results, numerators and denominators as well as resulting ratios should be displayed in the data quality survey report.

3.1.2 Sex ratio

What

The sex ratio is the proportion of males to females in a given population, usually expressed as the number of males per 100 females for a specific age group. It should be assessed for the survey dataset and compared to an expected sex ratio for the same age group. Because the sex ratio is generally not 100 boys to 100 girls in most countries, it is important to compare the survey sex ratio to a reference. A potential reference is the United Nations Population Division World Population Prospects (UNPD-WPP)\(^1\), which provides sex ratios estimated using smoothed distributions based

\(^1\) Check for the most recent version of the World Population Prospects at: [https://esa.un.org/unpd/wpp/Dataquery/](https://esa.un.org/unpd/wpp/Dataquery/)
on expected sex ratios at birth and mortality levels by country and year for different age groups\(^2\). The median sex ratio for all countries between 1995 and 2015 in the UNPD WPP is 104 boys per 100 girls with 5th and 95th percentiles of 101 and 108 respectively. It is therefore unlikely that a nationally representative survey would have a sex ratio outside this range. The only country that had a male-to-female ratio for children aged 0 to 4 years of 95 boys or lower per 100 girls was Rwanda in 1995, 1996 and 1997. At the other end of the spectrum, only a handful of countries such as Armenia, Azerbaijan and China have, over several consecutive years, shown sex ratios for children aged 0 to 4 years of 115 boys or more per 100 girls.

**Why**

The sex ratio of the survey population, when compared to an expected sex ratio, can be used to identify selection bias. This may result from problems of sampling (e.g., populations where members of one sex are more likely to be excluded from the household listing) or differences in response rates (e.g., higher absentee rates for one sex compared to the other).

**How to calculate**

The sex ratio should be calculated for all sampled children eligible for anthropometry from the household roster, whether or not measurements were made, or information is missing, and whether or not they are flagged as outliers on an anthropometry z-score. The sex ratio should be calculated using sampled weights in order to be comparable to the reference population. It is calculated as follows:

\[
\frac{\text{Weighted number of boys in survey under age 5 eligible for anthropometry}}{\text{Weighted number of girls in survey under age 5 eligible for anthropometry}} \times 100
\]

**How to present and interpret**

It is recommended that the sex ratio for the survey be compared against a country-specific expected sex ratio. Expected sex ratios can be obtained from UNPD-WPP\(^3\) or other national sources such as the latest censuses or other nationally representative survey reports for the same time period as the survey. If the survey sex ratio does not resemble the expected sex ratio for that country, sex ratio patterns should be examined by team and possibly by other disaggregation categories, but only if there is a sufficient sample size for this assessment for the disaggregation categories in question. The survey team should seek explanations for the unexpected sex ratio and include them in the survey report, including any problems with the reference used.

### 3.1.3 Age heaping

**What**

Age heaping refers to an unexpected distribution of observations for specific ages and/or months of birth. Three common age heaping patterns may occur, and various checks should be conducted to spot them as defined below:

- **Unequal distributions between single year age groups**

  The expected distribution for each single-year age group among children aged 0–4 years is about 20%. As confirmed by UNPD-WPP data between 1995 and 2015, the median ratio of children in each one-year age group between 0 and 4 years was 0.20 for all countries in the world. In other words, each single year interval (i.e., 0–11, 12–23, 24–35, 36–47 and 48–59 months old) contains 20% of all children aged 0–59 months. A survey team might see more 5 year olds than 4 year olds, for instance, if an unequal distribution existed between single-year age groups;

- **Distributions with peaks at single or multiple month age groups**

  In this scenario, peaks and troughs have a frequency distribution for age corresponding to specific months in the dataset. Common patterns include peaks at full years 0, 12, 24, 36 and 60 months or at half and full years (0, 6, 12, 18, 24, 30, 36, 42, 48, 54 and 60 months). However, in some surveys peaks may also occur at other ages; peaks for any single-month age group would be unexpected and a potential cause for concern. Unequal distributions may

---

\(^2\) Check for the most recent version of the World Population Prospects at: https://esa.un.org/unpd/wpp/Dataquery/

\(^3\) Check for the most recent version of the World Population Prospects at: https://esa.un.org/unpd/wpp/Dataquery/
also occur among multiple-month age groups, for example too few 0–3 month olds compared to 4–7 month olds in the dataset when a near equal distribution would be expected;

- **Distributions with peaks at specific months of birth**

  Estimates of age in months are calculated using the date of birth and date of interview and include the month of birth as obtained during data collection. Although there is evidence that the month of birth is not uniformly distributed in different countries, owing to various seasonal and climate factors (15), large peaks or troughs would not be expected.

**Why**

Unequal distributions between single-year age groups or too few children in a specific age group may be related to selection bias (e.g. aging out children who are just under five years) and/or measurement bias (e.g. misreporting of date of birth). It may be helpful, in the event of unequal distribution, to review the survey methodology and sampling design since they can shed some light on the type of bias. Distributions with peaks at a single month of age or peaks at specific months of birth are likely to indicate measurement bias.

Selection bias can occur if the household roster is filled incorrectly. Selection bias can also result from problems during interviews, e.g. where the interviewer records children close to the age of 5 as having already turned 5 in the household roster or there are different response rates, e.g. higher refusal rates for younger children or higher absenteeism for older children who may be attending school.

Measurement bias can occur where vital registration is not universal and information on the exact date of birth is not available. In this situation the interviewer is obliged to estimate the year and month of birth from incomplete records, maternal recall or by means of local events calendars. Misreporting can be due to the respondent genuinely not knowing a child’s date of birth, faulty date-of-birth records, incorrect handling of the local events calendar by the field team or data fabrication. Data heaping is more commonly observed for older (36–48 months) than younger children (12–24 months), most probably due to the fact that caretakers are less good at recalling the birthdates of older children. In some settings, children are issued with vital registration documents, but in populations where an application is not lodged for the child at the time of birth but instead sought some months or years following birth, a specific but arbitrary month—often “January”—is assigned as the month of birth on birth certificates, vaccination cards and other documentation when the true month of birth is unknown. Major data heaping on the month of birth can also be seen when local events calendars are not used according to the recommendations set out in this report (see Questionnaire development section in Chapter 1).

**How to calculate**

Age heaping should be investigated using histograms. These histograms should include all sampled children eligible for the child questionnaire for anthropometry from the household roster, whether or not measurements were made, or information is missing, and whether or not they are flagged as outliers on an anthropometry z-score. Information for children under 6 years old can be included as may show age displacement.

The following three histograms should be plotted.

- **Histogram 1**
  
  Calculated with sample weights and binned by year of age, i.e. six bars representing ages from 0 to 5 years (if data are collected only for children under 5 years, then five bars representing each year of age from 0 to 4);

- **Histogram 2**
  
  Calculated without sample weights and binned by month of age, i.e. 72 bars representing ages from 0 to 71 months (if data are collected only for children aged 0-4 years, there would be 60 bars representing each month of age from 0 to 59);

- **Histogram 3**
  
  Calculated without sample weights and binned by calendar month of birth, i.e. with 12 bars for the months January to December.

---

4 A histogram can also be plotted without sample weights when investigating misreporting but will probably not differ in a significant manner from the weighted histogram.
There are several methods to calculate age heaping numerically [e.g. index of dissimilarity (also known as Myers’ unblended index), Myers’ blended index, MONICA, Whipple’s index]. More research is needed to determine how different values for such indices influence estimates of malnutrition prevalence in order to develop cut-offs that might indicate poor data quality.

**How to present and interpret**

Plot all three histograms for the national sample as well as by team and examine them for unexpected distributions.

**Histogram 1 – by age in completed years:** Check to determine whether each of the 5 one-year age groups for children aged 0 to 4 years has an approximately 20% share of the entire under-five population (and/or each of the 6 one-year groups has an approximately 17% share of the entire under-six population if data for children aged under six years are available). If the proportion of any one-year age group for the 0 to 4 year olds deviates much from 20%, the country-specific expected age distribution from UNPD-WPP or other reliable sources (latest censuses or other nationally representative sources for the same time period as the survey) should be consulted. If an unexpected distribution is present at the national level, histograms should also be examined by other disaggregation categories. In some circumstances, for instance where child mortality is extremely high, or fertility rates have markedly changed in the previous five years, expected ratios may not follow a uniform distribution. A variation between the expected and survey age distributions may however also be due to a problem with the reference used. The team should seek explanations for unexpected age distributions and include them in the survey report. Examples of age distributions in years are shown in Figure 3.

![Figure 3. Examples of age distribution (in years) in different surveys](image)

**Histogram 2 – by age in completed months:** Check to see whether the length of each bar is approximately the same (see examples in Figure 4). Look for the common patterns noted above (e.g. peaks at 12, 24, 36, 48, and 60 months or 6, 12, 18, 24, 30, 36, 42, 48, 50 and 66 months). Obvious troughs or peaks at either end of the distribution should also be assessed to determine whether they represent a potential selection bias. If peaks and troughs are apparent in areas other than the beginning or end of the histogram, they may indicate a misreporting of the birth date, especially if they are observed at 6- or 12-month intervals. If an unexpected distribution is present at the national level, histograms should also be examined by other disaggregation categories but only if there is a sufficient sample size for this assessment for the disaggregation categories in question. Explanations should be sought for data heaping in months and included in the survey report.
Histogram 3 – by calendar month of birth: Check to see whether the length of each bar is approximately the same. It should be borne in mind that perfect distribution is not expected in any country since age distribution occurs as a function of monthly patterns of fertility (Figure 5). What is not expected is to find a large peak for a specific month, a finding which has typically been associated in some countries with the month of January. If an unexpected distribution is present at the national level, histograms should also be examined for other disaggregation categories but only if there is a sufficient sample size for this assessment for the disaggregation categories in question. Explanations should be sought for data heaping in the month of birth and included in the survey report.

3.1.4 Digit preferences for length/height and weight

What
Digit preference refers to an unexpected distribution of digits in weight and length/height measurements. Digit preference may affect the terminal digit or, less often, the integer part of the number. If survey teams use the equipment recommended in Chapter 1, each weight and length/height measurement in the survey should display one terminal digit:
this represents one tenth of a kilogram for weight and millimeters 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 expected distribution of each digit should be approximately 10%. Whole number preference refers to the process whereby data heaping occurs because the integer part of the number has been rounded off, e.g. 10 kg or 75 cm.

Common digit preference patterns include:

- a preference for the terminal digits 0 and 5;
- a preference for a terminal digit(s) other than 0 and 5;
- whole number digit preferences for height or weight (e.g. multiples of 5 or 10 cm for height or 2 or 5 kg for weight).

Why

Digit preference may be a tell-tale sign of data fabrication or inadequate care and attention during data collection and recording. Identifying which particular digits are overrepresented may provide insight into the type of error. For instance, if the frequency distribution indicates significant terminal digit entries for 0 and/or 5, this may indicate that measurers were rounding off. If there is a preference for digits other than 0 and 5, then it may be possible that data have been fictitiously constructed. A whole number digit preference is indicative of rounding of the integer part of the number or fictitious data.

If survey teams are using the currently recommended anthropometry equipment (digital weighing scale and height board with printed measuring tape) a digit preference for length/height is more likely to occur since the board needs to be read by counting lines and the 0 and 5 marks stand out from other markings on the board which represent the terminal digits. Since the recommended equipment for weight is a digital scale, the display provides easily readable numerical values. Rounding off is therefore less likely to occur for weights.

How to calculate

Digit preference in weight and length/height measurement should be investigated using histograms computed, without sample weights for all children measured and weighed in the entire sample, whether or not they are flagged as outliers on an anthropometry z-score.

- Histogram 1: binned by each terminal digit for weight (i.e. 10 bars from 0 to 9);
- Histogram 2: binned by each terminal digit for length/height (i.e. 10 bars from 0 to 9);
- Histogram 3: full range of weights in the dataset in whole numbers (i.e. approx. 25 bars from 0 to 25);
- Histogram 4: full range of lengths/heights in the dataset in whole numbers (i.e. approx. 90 bars from 35 to 125).

Terminal digit preference should also be calculated numerically, using the index of dissimilarity. computed, without sample weights for all children measured and weighed in the entire sample, whether or not they are flagged as outliers on an anthropometry z-score. See Annex 12 for the index of dissimilarity calculator for terminal digits. The index of dissimilarity is expressed by the following formula:

\[
\text{Index of dissimilarity} = \sum_{i=1}^{10} \left| \frac{\text{actual percentage}_{i} - \text{expected percentage}_{i}}{2} \right|
\]

where

\(\text{actual percentage}_{i}\) = percentages for terminal digits in the survey (e.g. number of height measurements with a terminal digit of zero/all height measurements), and

\(\text{expected percentage}_{i}\) = expected distribution percentages (i.e. 10% for each terminal digit).

\(^{5}\) Digit preference can only be assessed if length/height or weight values have not been rounded off at the data cleaning stage. For instance, in some DHS surveys, at the time of this publication, weight is recorded to the hundredth decimal place as either 0 or 5 and to the tenth place in the recode microdata (based on publicly available data).
How to present and interpret

Present the four histograms and examine them for unexpected distributions based on the overall sample as well as disaggregated by team. If an unexpected distribution is observed at the national level, histograms should also be examined by other disaggregation categories.

**Histogram 1:** Check to determine whether each of the 10 terminal digits for weight has an approximately 10% share of the entire sample as well as for each team or main measurer. If the proportion for any digit deviates much from 10% this suggests a terminal digit preference for weight.

**Histogram 2:** Check to determine whether each of the 10 terminal digits for length/height has an approximately 10% share of the entire sample as well as for each team or main measurer. If the proportion for any digit deviates much from 10% this suggests a terminal digit preference for length/height.

**Histogram 3:** Check to determine whether there are visible peaks for any particular weights. A roughly uniform distribution is not expected for weights at whole integers, but extreme peaks should not be strikingly visible.

**Histogram 4:** Check to determine whether there are visible peaks for any heights/lengths. A roughly uniform distribution is not expected for heights at whole integers, but extreme peaks should not be strikingly visible.

These histograms can provide information about the reasons for digit preference. For instance, in Figure 6, which presents terminal digit patterns, the digit preference for 0 and 5 suggests numbers have been rounded off; the preference for 3 and 7 in the adjoining histogram is more probably the result of fictitious data. Any noticeable peaks in whole number distributions (e.g. noticeable peaks at 70, 80 and 90 cm) would be indicative of severe problems with equipment or fictitious data. These are merely illustrative examples and scrutiny of actual survey data may bring to light more or less extreme cases. Prominent peaks on whole numbers give rise to inaccurate prevalence estimates since children in the peak value range are likely to have heights and/or weights which deviate substantially from their true values.

**Figure 6. Different possible patterns of terminal digit distribution (Histograms 1 and 2)**
It is recommended that the index of dissimilarity for terminal digits of weight and length/height be summarized for the entire sample as well as for individual teams. If an unexpected distribution is present at the national level, histograms should also be examined by other disaggregation categories. Index of dissimilarity outputs for terminal digits of height and weight range from 0 to 90 and represent the percentage of observations that would need to shift from overreported to underreported digits in order to achieve a uniform distribution. The ideal value of the index is 0 (0% needs to be redistributed) and the maximum score is 90 (all terminal digits are heaped on one value and 90% of all terminal digits in the dataset would need to be redistributed to achieve a uniform distribution).

A digit preference for terminal digits of weight will result in greater inaccuracies when estimating prevalence for WAZ and WHZ than a digit preference for terminal digits of length/height would have on HAZ or WHZ. Nevertheless, every digit preference is a data quality indicator and should be noted in the report.

### 3.1.5 Implausible z-score values

**What**

Implausible values are z-score values that fall outside a specified range. The currently recommended flagging system to detect implausible z-score values was defined in 2006 on the release of the WHO Child Growth Standards, replacing the NCHS/WHO reference for child growth\(^6\) (see discussion on flagging in section 3.2 on Data analysis). System cut-offs were defined on the basis of what is biologically implausible, in other words incompatible with life. These flagging cut-offs have been challenged based on observations of living children whose z-scores are beyond currently defined implausible values\(^{16}\), although true z-score values beyond the implausible value cut-offs recommended by the WHO rarely occur in any population. Nonetheless, this is a topic for future research.

---

\(^6\) The WHO macro excludes children if their length is outside of the ranges of 45 to 110 cm or if their height is outside of the ranges of 65 to 120 cm when calculating weight-for-height z-scores. This exclusion is done prior to flagging values outside of the plausible weight-for-height z-score ranges. Thus, when calculating the percentage of implausible weight-for-height the out-of-range length/height values need to be identified (not using the WHO macro) and added into the numerator and denominator.\(^6\) WHO Anthro Manual, https://www.who.int/childgrowth/software/anthro_pc_manual_v322.pdf?ua=1
Why
The percentage of implausible z-score values derived from the WHO Child Growth Standards is an important indication of data quality. Values outside the plausible range are usually due to poor measurement, inaccurate date of birth or data recording errors. As WHO flagging ranges are quite broad, they likely do not detect all values due to measurement errors because they fall within the plausible range.

How to calculate
The percentage of implausible z-score values should be calculated using unweighted sample weights for all children measured in the entire sample. For the currently recommended fixed exclusions approach, z-score values outside the following intervals are considered implausible: HAZ (-6, +6), WHZ (-5, +5), WAZ (-6, +5). These values should be flagged independently for each type of anthropometry z-score (HAZ, WAZ, WHZ), which means that some children may be flagged for one anthropometry z-score but not another. Statistical packages and software are available which calculate anthropometry z-scores and flag cases outside the predefined intervals for each anthropometry z-score¹ (17).

\[
\text{Percentage implausible HAZ} = \frac{\text{number children with } \text{HAZ} < -6 \text{ or } > 6}{\text{Total number children with height and DOB}}
\]

\[
\text{Percentage implausible WHZ} = \frac{\text{number children with } \text{WHZ} < -5 \text{ or } > 5}{\text{Total number children with height and weight}}
\]

\[
\text{Percentage implausible WAZ} = \frac{\text{number children with } \text{WAZ} < -6 \text{ or } > 5}{\text{Total number children with weight and DOB}}
\]

Note: DOB=date of birth, requiring at least the month and year of birth

How to present and interpret
Present the percentage of implausible values for each index separately, HAZ, WHZ and WAZ, for national sample as well as for each team. A percentage of implausible values exceeding 1% is indicative of poor data quality (18). This data quality threshold of 1% was based on the NCHS/WHO reference for child growth ranges for implausible values in use at that time as well as the deliberations of the WHO Expert Committee in 1995. The same threshold is expected to hold when based on the WHO Child Growth Standards ranges for implausible values, as the latter were developed to match the implications of the previously recommended. Examine the percentage of implausible values by other disaggregation categories if the percentage of implausible values is greater than 1%. While a high percentage of flagged values reliably indicates poor data quality, a low percentage does not necessarily imply adequate data quality since values that are inaccurate may still occur within the WHO flag range.

3.1.6 Standard deviation of z-scores

What
The standard deviation (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. Standard deviation cannot be negative; the lowest possible value for SD is zero, which would indicate that all data points are equal to the mean or that only one value exists in the entire dataset, e.g. if every child were to have exactly the same WHZ value.

The reference sample for the 2006 WHO Child Growth Standards displays, by definition, a standard normal distribution with zero mean and SD of 1 for each of the anthropometric indices including WAZ, WHZ and HAZ. The WHO Child

---

¹ The WHO macro excludes children if their length is outside of the ranges of 45 to 110 cm or if their height is outside of the ranges of 65 to 120 cm when calculating weight-for-height z-scores. This exclusion is done prior to flagging values outside of the plausible weight-for-height z-score ranges. Thus, when calculating the percentage of implausible weight-for-height the out-of-range length/height values need to be identified (not using the WHO macro) and added into the numerator and denominator.
Growth Standards are based on a sample of healthy children from six different countries on five different continents (Brazil, Ghana, India, Norway, Oman and the United States) with varying ethnic groups living in an environment that did not constrain optimal growth. The sample was purposively selected to be homogeneous with respect to variables that can affect optimal growth, e.g. economic status of the family, mother’s smoking behaviour, term delivery, feeding practices and absence of significant morbidity.

Less is known about expected SD in disadvantaged populations or those living in environments that do not support optimal growth. The 1995 WHO Technical Report on Anthropometry recommended using SD as a data quality criterion: it was stated that studies with a SD outside the following ranges would require closer examination for possible problems related to age assessment and anthropometric measurements: 1.1 to 1.3 for HAZ, 1.0 to 1.2 for WAZ and 0.85 to 1.1 for WHZ. These cut-offs for determining data quality need to be revised however, for various reasons:

- they were developed using a set of surveys not all of which were nationally representative and included several rapid nutrition surveys conducted in emergency situations, where the populations concerned were probably more homogeneous with respect to nutrition status and its determinants;
- they were based on the distribution of z-scores calculated using the NCHS/WHO reference for child growth which was replaced in 2006 with the WHO Child Growth Standards in use today; and
- the flagging system applied to exclude extreme values was more conservative (i.e., the ranges for exclusion were narrower) than the currently recommended flagging system (see Table 5), which would have resulted in narrower SD ranges.

### TABLE 5. EXCLUSION CRITERIA PREVIOUSLY AND CURRENTLY USED FOR DIFFERENT APPLICATIONS

<table>
<thead>
<tr>
<th>Purpose of ranges</th>
<th>Previously used for exclusion when generating SD ranges for data quality assessment but not recommended for any purpose at present</th>
<th>Previously used for exclusion before calculating prevalence estimates but not recommended for any purpose at present</th>
<th>Currently recommended for exclusion before calculating prevalence estimates and for generating SDs for data quality assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flag type</td>
<td>Fixed</td>
<td>Flexible*</td>
<td>Fixed</td>
</tr>
<tr>
<td>HAZ</td>
<td>&lt;-5 or &gt;3</td>
<td>&lt;-4 or &gt;4</td>
<td>&lt;-6 or &gt;6</td>
</tr>
<tr>
<td>WHZ</td>
<td>&lt;-4 or &gt;5</td>
<td>&lt;-4 or &gt;4</td>
<td>&lt;-4 or &gt;6</td>
</tr>
<tr>
<td>WAZ</td>
<td>&lt;-5 or &gt;5</td>
<td>&lt;-4 or &gt;4</td>
<td>&lt;-6 or &gt;6</td>
</tr>
</tbody>
</table>

*around the observed survey mean

The Joint Malnutrition Estimates (JME) country dataset lists, as of January 2019, estimates after re-analysis for 474 nationally representative household surveys from 112 countries. Standard deviations were estimated for HAZ, WHZ and WAZ for these 474 surveys after applying the exclusions discussed under “how to calculate” below. It contains surveys with a wide range of SD values for HAZ, WAZ and WHZ. Their median (and 5th and 95th percentiles) were 1.54 (1.21 and 2.03) for HAZ, 1.27 (1.04 and 1.72) for WHZ and 1.22 (1.06 and 1.52) for WAZ. 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

---

8 Even though the cut-offs are different in the last two columns, differences are small in terms of actual kg or cm values in the two international references (WHO Child Growth Standards and NCHS/WHO reference). This is because the WHO flags for exclusion criteria were identified to maintain inferences similar to those already in use with the NCHS/WHO reference.

re-analysed surveys in the global database reflect very large SD values for both HAZ and WHZ; some of the SDs in the dataset are larger than would be reasonably explained by population heterogeneity 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. Although data quality assessment based on z-score SDs is therefore fully justified, further research is required to develop recommended SD ranges for acceptable HAZ, WHZ and WAZ.

**Figure 9.** Box plots of Z scores for 474 nationally representative surveys included in the Joint Malnutrition Estimates database

![Box plots of Z scores](image)

Note: Middle line and value label represent the medians, edges of box represent first and third quartiles and whiskers represent minimum and maximum SD values by z score in the JME country dataset.

**Why**

Reporting on and identifying the causes of large SDs are important tasks in data quality assessment. Prevalence estimates of stunting, wasting and overweight are dichotomous variables which measure the percentage of children with z-score values beyond a specified cut-off (e.g. <-2 SD for wasting or stunting, >+2 SD for overweight). If the SD is artificially inflated as a result of poor quality data, prevalence estimates are therefore likely to be overestimated. The relative overestimation of prevalence will be even greater for estimates of severe categories of malnutrition (e.g. <-3 and >+3).

The higher the SD, the greater the likelihood that poor data quality is contributing to the wide SDs observed. Quantifying in definitive terms how much of the dispersion in z-scores can be attributed to heterogeneity in relation to environments prejudicial to optimal growth and how much to measurement error is a challenging research question. The following statements can be made about SDs for HAZ, WAZ and WHZ:

- The 1995 WHO Technical Report on Anthropometry suggested a set of SD ranges, beyond which data quality was of possible concern: these cut-offs need to be revised however, so that they reflect nationally representative surveys among populations with varying degrees of malnutrition and the currently used WHO Child Growth Standards. SDs are typically larger for HAZ than they are for WAZ or WHZ. Some of this difference in SDs is probably due to measurement error since height is more difficult to measure reliably than weight with currently available equipment; obtaining a reliable date of birth problematic in populations with substandard vital registration systems. Furthermore, the dispersion of z-scores which represent linear growth is likely to differ from that of z-scores which represent acute malnutrition, especially in malnourished populations. Also, length- or height-for-age z-scores can present wider
degrees of dispersion in malnourished than in well-nourished populations because deficits in length or height are cumulative; varying levels of substantial malnutrition in a country may be reflected in wider SDs for z-scores where data represent cumulative (i.e. HAZ) rather than non-cumulative deficits (i.e. WHZ);

- SDs for HAZ tend to decrease from youngest to the oldest group of children, as indicated in Figure 10 below, based on 571 surveys for which stratification by age is available and are included in the Joint Child Malnutrition Estimates database. Some of this SD spread is due to measurement error since length is more difficult to measure than height. It may, on the other hand, be easier to determine the date of birth in younger infants born just a few months before a survey, resulting in tighter SDs in this age group. However, smaller errors in age can impact z-scores for younger children more than for older children where there is the same degree of error (i.e. 15 days’ inaccuracy in age for a 1-month-old child would generally result in a different HAZ whereas 15 days’ inaccuracy in age for a 4-year-old would generally yield a similar HAZ). Furthermore, the WHO Child Growth Standards SDs for HAZ steadily increase with month of age from birth reflecting the divergent growth trajectories of full-term, well-nourished children when growth is viewed cross-sectionally at specific ages. In other samples of children that include pre-term births, the SDs may be larger at and just after birth.

Figure 10. SD for HAZ by age groups in 422 surveys included in the Joint Malnutrition Estimates database

Note: Middle line and value label represent the medians, edges of box represent first and third quartiles and whiskers represent minimum and maximum HAZ SD values by age group for surveys with SD estimates for each age group in the JME country dataset.

There is no expected substantial difference in SDs when comparing boys and girls, apart from slightly higher for boys which may be due to their higher rate of preterm births. Figure 11 shows these expected patterns between sexes based on 513 surveys for which stratification by sex is available and are included in the Joint Child Malnutrition Estimates database.
Figure 11. SD for HAZ, WFH and WFA by sex in 473 surveys included in the Joint Malnutrition Estimates database

Note: Middle line and value label represent the medians, edges of box represent first and third quartiles and whiskers represent minimum and maximum z score SD values by sex for surveys with SD estimates for boys and girls in the JME country dataset.

How to calculate
SD should be calculated using unweighted sample weights for all children measured and weighed in the entire sample and after WHO fixed flags (see section 3.2 on Data analysis for WHO fixed flag values) have been removed from the dataset. The formula is:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$$

where $n$=total number of data points, $\bar{Y} =$ mean of $Y_i$ and $Y_i =$ each value in the dataset.

How to present and interpret
It is recommended that the SD for each indicator (HAZ, WHZ, and WAZ) be presented separately, at the national level, and for teams and other disaggregation categories. Strata specific SDs that are greater than the SD for the national estimate, as well as large differences between groups where they would not be expected (e.g. large differences in SDs between girls and boys or large fluctuations between neighbouring age groups) should be examined and explained in the survey report.

Further investigations are needed to develop guidance on how to tease out the relative contribution of measurement error from expected population-associated spread in any given survey, and to establish cut-offs at which SDs for each anthropometric index might be more conclusively related to poor data quality.

3.1.7 Normality (skewness and kurtosis) of z-scores

What
Distribution of HAZ, WAZ and WHZ is a description of the relative number of times each z-score occurs in the survey population. A standard normal distribution is a symmetrical bell-shaped curve with a mean of zero and standard deviation of 1. Patterns of deviation from normal distribution include asymmetric, peaked or flat distribution curves. Skewness is
a measure of asymmetry: a normal distribution curve which is perfectly symmetrical will have a skewness value of zero with equal distribution on both right and left halves. When the skewness coefficient is positive, distribution is skewed to the right: this indicates that there are more cases on the right side of the distribution curve than on the left, usually an indication of extreme values in the right side or tail of the distribution curve. In turn, when the skewness coefficient is negative, the distribution curve is skewed to the left (see Figure 12). Like skewness, kurtosis is a description of deviation from the normal shape of a probability distribution (see Figure 13). Kurtosis is a measure of tailedness which also describes the sharpness or flatness of the frequency distribution peak: a kurtosis coefficient of 3 represents a population following normal distribution. When the kurtosis value is greater than 3, the curve is flat and peakedness reduced: this indicates that there are many extreme values in the tails than the expected normal distribution. Conversely, when kurtosis is less than 3, the peak is high, and tails are therefore relatively short.

**Figure 12. Different possible patterns of skewness**

- Positively skew population
- Normal distribution
- Negatively skew distribution

**Figure 13. Different possible patterns of kurtosis**

- Platykurtic distribution
- Normal distribution
- Leptokurtic distribution

**Why**

Understanding the shape of the frequency distribution can provide insights into the survey population and data quality. The WHO Child Growth Standards based on a sample of healthy children living in environment that did not constrain growth showed normal distribution for each of the anthropometry z-scores. It is sometimes assumed that survey populations will have a normal distribution and that distribution will shift depending on the degree of malnutrition affecting the population. However, probability distributions among malnourished populations can deviate from normal distribution especially when there are many inequities or severe forms of malnutrition are prevalent (e.g. severe stunting is high, or overweight is a significant problem in specific subpopulations) without necessarily indicating data quality issues.

Conclusions about data quality cannot be drawn solely on the grounds of skewness or kurtosis values. On the other hand, deviations from the normal, when combined with other issues identified by data quality checks, should raise concern. Further research is required to understand distribution patterns in populations with different forms of malnutrition and also how skewness and kurtosis values suggesting departure from normality might indicate problems with data quality.

**How to calculate**

The shape of the distribution curves for HAZ, WHZ and WAZ can be visualized using kernel density plots. These plots should be developed based on the entire sample of children that were measured and weighed, without sample weights, and after flagged z-scores (see section 3.2 on Data analysis for WHO fixed flag values) have been removed from the dataset.
Checks for normality of z-score distributions help to assess departures from a normal distribution, based on measures of skewness and kurtosis. Skewness and kurtosis coefficients should also be calculated for HAZ, WAZ and WHZ without sample weights after flagged z-scores have been removed.

There are various formulas for determining skewness and kurtosis coefficients: those proposed below are based on the Fisher-Pearson coefficient, although others may be applied.

Formula to assess skewness using the Fisher-Pearson coefficient:

\[
\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^3}{s^3} / n
\]

Formula to assess kurtosis using the Fisher-Pearson coefficient:

\[
\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^4}{s^4} / n
\]

where

\[ \bar{Y} = \text{mean}, s = \text{standard deviation} \text{ (calculated with } n \text{ in the denominator rather than } n-1) \text{ and } n = \text{sample size}. \]

**How to present and interpret**

Present the distribution curve graphics for the national sample as well as by team and examine them for unexpected distribution patterns. Deviations from the normal for HAZ, WHZ or WAZ are difficult to interpret, as it may genuinely represent malnourished populations with high levels of inequity and/or severe forms of malnutrition, as well as be due to poor data quality, or a combination of these. Further research needs to be carried out before practical guidance on interpreting the shape of the distribution curve in any given survey can be provided. Nevertheless, comparisons of distributions between groups can provide hints that can help interpretation.

Kernel density plots 1, 2 and 3: Check to see if the tails of the HAZ, WHZ, and WAZ distribution curve end smoothly rather than abruptly. If distribution ends abruptly, it may be indicative of data quality issues.

It is recommended that skewness and kurtosis coefficients for HAZ, WHZ and WAZ be summarized for the entire survey sample\(^1\). While there is no defined cut-off, it is an accepted rule of thumb that a coefficient of <-0.5 or >+0.5 indicates skewness. Similarly, while there is no defined cut-off, in general a coefficient of <2 or >4 indicates kurtosis. Moreover, since the kurtosis coefficient for standard normal distribution is 3, some formulas and most statistical software subtract 3 from the value obtained using the formula above to obtain a kurtosis value of 0 for standard normal distribution: these formulas therefore represent “excess kurtosis”. When these formulas are used, a kurtosis coefficient of <-1 or >1 indicates kurtosis. Should skewness or kurtosis coefficients fall outside these ranges, the coefficients should be examined by other disaggregation categories.

### 3.2. DATA ANALYSIS – THE STANDARD ANALYSIS APPROACH

The previous chapters and section described steps to enhance data quality during fieldwork and assess data quality. The present section sets out various considerations to bear in mind when conducting an analysis of anthropometric measurements, from early preparation of the dataset to actual calculation of the prevalence estimates.

---

\(^1\) A small adjustment to produce an unbiased estimate with respect to sample size is automatically included in standard software packages.
The calculation of child malnutrition prevalence involves two stages: comparing the anthropometric measurements of sampled children against reference data (z-score calculation) and then calculating the proportion of children whose z-scores are below or above the specified cut-offs for each of the nutrition indicators in question (e.g. prevalence estimates for stunting or wasting).

For these analyses, the recommended standard approach below adopts as its reference data the WHO Child Growth Standards (see Note 8) and can be performed using standard software such as the WHO Anthro Survey Analyser. Macros for the main statistical programs (complete procedures available in Stata, R, SAS and SPSS) are also available to analyse data based on the reference data for children’s z-scores and survey sampling design for prevalence estimates. Other tools such as the MICS6 and DHS7 syntax/programs which adhere closely to the recommended standard approach are available. Epi Info/ENA software can also be adapted to the standard approach if selecting the WHO flag system.

3.2.1 Why a “standard analysis” approach?

Data analysis is an important step and it deserves proper attention to ensure that country-specific results are accurate and comparable between countries or over time. Using different methods when preparing the data analysis file (e.g. imputing a missing date of birth, selecting children in the household for inclusion in the analysis) and applying different reference data to calculate individual z-scores or different exclusion criteria for implausible values may generate inconsistencies between estimates made at a different point in time. Such methodological differences are commonly observed even when surveys are carried out very close together in time, sometimes during overlapping time periods, consequently making it difficult to monitor country trends.

Since the launch of the WHO Child Growth Standards in 2006 (see Note 8), WHO, key partners and data collection programs (e.g. national nutrition surveys, DHS, MICS, SMART and others) have collaborated in an attempt to standardize analyses of anthropometric data emerging from national surveys or other non-emergency settings as fully as possible.

Based on these collaboration efforts, WHO and key partners have developed software and macros for survey analysis based on a standard approach, which is referred to hereafter as the “standard analysis”. The main steps of this standard analysis are shown in Table 6. Most surveys have a complex sampling design (e.g. two-level sampling) since use of the appropriate methodology enhances estimates of accuracy around prevalence and mean z-score. This is a significant improvement in data reporting and is a precondition for adhering to the Guidelines for Accurate and Transparent Health Estimates Reporting, which aim to define and promote best practices in reporting health estimates. Recently, WHO and UNICEF have updated the R and Stata macros to include methodology that takes into account complex survey sampling design, and syntax files from programs such as MICS and DHS have also adopted this approach.
NOTE 8: THE WHO GROWTH STANDARDS 2006

In 2006, WHO published its Child Growth Standards for children from birth to 5 years (21). They were constructed based on data from the WHO Multicentre Growth Reference Study (MGRS), which established the full-term breastfed infant with absence of significant morbidity born from a non-smoking mother as the normative growth model (21). The wealth of data collected made it possible to replace the international NCHS/WHO reference for attained growth (weight-for-age, length/height-for-age, and weight-for-length/height) and develop new standards for body mass index (BMI)-for-age, head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age. Detailed descriptions of how the MGRS was conducted and WHO Child Growth Standards were constructed are available elsewhere (21) (22). The standards are sex-specific, in the recognition that boys and girls have different growth patterns.

WHO Child Growth Standards were devised for children from birth to 60 completed months of age. For each anthropometric index, available standards cover the following ranges:

- weight-for-length: length from 45 to 110 cm;
- weight-for-height: height from 65 to 120 cm;
- weight-for-age: age from 0 to 60 completed months;
- length/height-for-age: age from 0 to 60 completed months;
- BMI-for-age: age from 0 to 60 completed months;
- arm circumference-for-age: age from 0 to 60 completed months.

WHA Resolution 63.23 recommended the implementation of WHO standards (23). Adherence to WHO Child Growth Standards ought to ensure optimal levels of health globally. At the time of publication, standards had been adopted by more than 160 countries and provide a tool for identifying countries where child malnutrition is a significant burden.
TABLE 6. SUGGESTED COMPONENTS AND KEY CONSIDERATIONS FOR STANDARDIZING THE ANALYSIS OF ANTHROPOMETRIC DATA

<table>
<thead>
<tr>
<th>COMPONENT</th>
<th>KEY CONSIDERATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reference data for z-score calculation</td>
<td>• Use the WHO Child Growth Standards for child malnutrition monitoring.</td>
</tr>
<tr>
<td>2. Missing data</td>
<td>• Any recode of missing values (depending on the software or code used for the analysis, e.g. 9998, 9999, 99 recode to blank cells) or imputation should be made by creating a new variable. The original variables should always be retained since their presence in the file guarantees data reproducibility and transparency; • It is important that all records, including those with missing measurements or sampling weights, are available for analysis, since they are important for data quality assessment (e.g. non-response); • Imputation of missing day of birth: if only the month and year of birth are provided, it is recommended that the missing information for the day of birth be imputed. This can be done in different ways but using the 15th of the month for all missing days of birth is recommended in standard analysis. The approach used for imputing the date of birth and the number or proportion of cases falling on the imputed day should be mentioned in the report for reasons of data quality assessment. • If the month or year of birth is missing, then the date of birth and consequently the child’s age should be considered as missing. In such cases, indicators related to age as stunting or underweight will not be calculated, while indicators not age-related as wasting will be calculated; • Some surveys use a code number for missing values such as 9999, 9998, 98, etc. Such numbers should always be treated as missing data and not as extreme values, since it is important to differentiate between implausible z-score values and missing measurements when assessing data quality.</td>
</tr>
<tr>
<td>3. Age calculation</td>
<td>• Age should be calculated based on the date of visit and date of birth and both variables kept in the analysis file; • If exact date of birth is unknown, the month and year of birth should be estimated using a local events calendar. In such cases, age should be calculated after imputing the day of birth as the 15th of the month.</td>
</tr>
<tr>
<td>4. Oedema (Although assessment of oedema is not recommended for systematic inclusion in all surveys except in settings where collecting this information is appropriate)</td>
<td>• Oedema measurement is only appropriate in surveys where local experts, specifically clinicians or individuals from the Ministry of Health working at a local level, can clearly indicate if they have seen recent cases where nutritional oedema was present (see Note 1 in Section 1.1 in Chapter 1 for more details); • If information on oedema is collected following the above recommendation, it should be included in each child’s dataset and used in the analysis. In this event: – all children, even those with oedema, should be weighed to reduce the likelihood of biased decisions in the field; – children with oedema should automatically be classified with “severe acute malnutrition” (&lt;-3 SD for weight-related indexes) when calculating prevalence estimates; – weight-related indices z-scores will not be calculated for children with oedema (i.e. set to missing); – the number of cases of oedema should be included in the survey report; – prevalence levels based on analyses both including and excluding oedema-related data should be included in the survey report.</td>
</tr>
<tr>
<td>COMPONENT</td>
<td>KEY CONSIDERATIONS</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------</td>
</tr>
</tbody>
</table>
| 5. Conversion of recumbent length to standing height or vice versa | Recumbent length or standing height  
  • Verify that the child’s measurement position (standing height or recumbent, i.e. supine or lying length) was recorded in the questionnaire during measurement to allow for age-linked adjustments in length/height depending on whether they were lying or standing;  
  • Based on the recorded measurement position, software performing the standard analysis will need to make automatic adjustments when calculating z-scores, adding 0.7 cm if the standing height was measured for children aged < 24 months and subtracting 0.7 cm if the recumbent (lying) length was measured for children aged ≥ 24 months;  
  • If data on the measurement position are missing, recumbent length is assumed to have been adopted for children aged <731 days (<24 months) and standing height for those with aged ≥ 731 days (≥ 24 months);  
  • For children under 9 months of age, data which suggests that the infant was “standing” rather than the expected “lying” should be disregarded in the analysis, i.e. set to missing, since this is deemed to be an error. This is done to avoid the wrong automatic adjustment in such cases (adding 0.7 cm), which would result in an overestimation of wasting and underestimation of stunting. |
| 6. Handling re-measurement data | • Re-measurements (height, weight, date of birth, and sex) of children randomly selected or flagged should be retained in the datafile (see section 2.4, where this operation is described). Use height, weight, date of birth and sex from the first measurement for children randomly selected for re-measurement when calculating z-scores. Use height, weight, date of birth and sex from the second measurement for children flagged for re-measurement when calculating z-scores. |
| 7. Exclusion of flagged z-scores (WHO flag system) | • The recommended flags for z-score values follow the WHO flag system \(^\text{16}\) (see section 3.2.1 below for a discussion of flagging systems):  
  – height-for-age: < -6 or > +6;  
  – weight-for-length/height: < -5 or > +5;  
  – weight-for-age: < -6 or > +5;  
  – body mass index-for-age: < -5 or > +5.  
  • The number and percentage of values excluded should be reported;  
  • Exclusions should be made based on the indicator (rather than child), e.g. measurements for a child with a HAZ of -6.5 and a WHZ of -4.5 would be included in an analysis of wasting (WHZ) but not of stunting (HAZ);  
  • All measurements should be retained in the dataset for transparency;  
  • Flagged z-scores are excluded before calculating prevalence estimates and other z-score summary statistics. |

### COMPONENT: SAMPLING DESIGN

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Key Considerations</th>
</tr>
</thead>
</table>
| 8.  | **Sampling design**<br>Strata and cluster/PSU<br>  
  • The purpose of stratification is to ensure that the sample is representative of the population of interest and divides the population into groups (typically geographic groups) before sampling. Stratification in the sampling design helps to reduce sampling errors when introduced at the initial stage of sampling (its effect on the sampling error is minor when introduced at the second or later stages);<br>  
  • Strata should not be confused with survey domains, i.e. a population subgroup for which separate survey estimates are desirable (e.g. urban/rural areas, see bullet point 9 below)\(^{17}\). Both categories may be the same, but do not need to be. A cluster/PSU is a group of neighbouring households which usually serves as the Primary Sampling Unit (PSU) for efficient field work;<br>  
  • Each child/household should be assigned to a cluster/PSU and strata and analyses should take that information into account in order to boost the stability of estimated variance. |
| 9.  | **Stratified analysis** for population sub-groups (when available)**<br>  
  • The most common population disaggregation categories are age (different age groups), sex (male or female), type of residence (urban or rural) and sub-national geographic areas (e.g. region, district). For age groups, standard analysis relies on the exact age in days (where available) to define age groups in months (e.g. <6, 6 to <12, 12 to <24, 24 to <36, 36 to <48 and 48 to <60). One month equals 30.4375 days;<br>  
  • Monitoring equity is of increasing importance for health and development. Disaggregated analysis is also recommended in order to derive estimates by wealth quintiles (1=lowest, 2, 3, 4, 5=highest) and mother’s education (no education, primary school and secondary school or higher), whenever this is possible. |

---

**3.2.2 Exclusion of extreme values before calculating malnutrition estimates**

It is recommended that extreme values be excluded from estimations of mean z-scores and malnutrition prevalence since exclusion makes it more likely that true population estimates will be accurately represented.

Before the WHO Child Growth Standards were developed, in 1995 the WHO Expert Committee provided two exclusion approaches (18).

- **Fixed exclusions**

  Fixed exclusions are centred on a reference mean z-score. Fixed exclusion values in 1995 were based on the NCHS/WHO reference for child growth but now refer to WHO Child Growth Standards. Fixed exclusions were intended to remove biologically implausible observations;

Flexible exclusions

Exclusion values are centred on the observed survey mean z-score. Flexible exclusions are based on statistical probability and premised on the statistical phenomenon that > 99.99% of values fall within ± 4 standard deviations of the mean in standard normal distribution.

A description of the fixed and flexible exclusion approaches recommended by the WHO Expert Committee in 1995 is provided in Table 7. Where there are few extreme values, e.g. in a population with very low prevalence for any form of malnutrition, the use of either fixed or flexible exclusion has very little or no impact on prevalence estimates. However, in surveys with a greater number of extreme values there can be significant differences in prevalence estimates depending on the exclusion approach adopted (24).

Excluding extreme z-score values requires a balance to be found between the risk of two events: excluding a child with a genuinely very extreme z-score and including a mistake that has given rise to an improbably extreme z-score. Extreme z-scores values defined by fixed exclusions are almost certainly the result of measurement error, but in surveys with substantial measurement error, fixed exclusions may not capture a large portion of incorrect values. In this situation, fixed exclusions will probably lead to an overestimation of malnutrition prevalence in surveys. This can be problematic because even a 1–3 percentage point increase in the prevalence of severe wasting can have profound programmatic implications. By contrast, the prevalence of moderate categories of stunting and wasting (e.g. z-scores between -2 and -3) may be less affected by the exclusion criteria. A flexible exclusion approach that has been applied in some surveys is to use standard deviation units (by assuming a normal distribution with a standard deviation of 1) and a narrower exclusion range. This is likely to give rise to lower prevalence estimates, and while they may address concerns about measurement error they may also truncate true values representing extremely undernourished children (in the negative tail of the distribution curve) or children who are extremely obese or very tall (in the positive tail of the distribution curve). Further research is needed on the use of existing and potentially new flexible exclusion approaches, especially in situations where anthropometry measurements are of poor quality.

Table 7. Characteristics of fixed and flexible exclusion approaches (18)

<table>
<thead>
<tr>
<th></th>
<th>FIXED</th>
<th>FLEXIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusions applied</td>
<td>Apply the same exclusion ranges to all surveys (without taking into consideration the survey-specific population distribution).</td>
<td>Exclusion ranges shift depending on distribution within the survey population, i.e. a positive shift in the population will result in a positive shift in the exclusion ranges and vice versa.</td>
</tr>
<tr>
<td>(range of z scores included and removed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source of reference statistical distribution</td>
<td>Statistical distribution aligned with international growth standards.</td>
<td>Individual survey population statistical distribution is used to obtain the mean while assuming a standard normal distribution with SD of 1.</td>
</tr>
</tbody>
</table>

The use of fixed exclusions based on the WHO Child Growth Standards is currently recommended (Table 8). Fixed exclusions, conventionally used in national surveys, are accepted worldwide and allow comparability between surveys and countries. Software packages have been developed to flag z-score values automatically based on the recommended fixed exclusion values (HAZ (-6, +6), WHZ (-5, +5), WAZ (-6, +5)18. Exclusions should be applied by indicator (rather than child), e.g. measurements for a child with a HAZ of -6.5 and WHZ of -4.5 would be included in the analysis of wasting (WHZ) but not of stunting (HAZ). All measurements should be retained in the dataset for transparency, although z-scores which are considered extreme on the basis of the fixed exclusion approach do not contribute to prevalence estimates. The number and percentage of values excluded should be reported.

TABLE 8. EXCLUSION CRITERIA PREVIOUSLY AND CURRENTLY USED

<table>
<thead>
<tr>
<th>Purpose of ranges</th>
<th>Previously used for exclusion before calculating prevalence estimates but not recommended for any purpose at present</th>
<th>Currently recommended for exclusion before calculating prevalence estimates and for generating SDs for data quality assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>NCHS/WHO reference (19)(20)</td>
<td>WHO Child Growth Standards (17)</td>
</tr>
<tr>
<td>Flag type</td>
<td>Fixed</td>
<td>Fixed</td>
</tr>
<tr>
<td>HAZ</td>
<td>&lt;-6 or &gt;6</td>
<td>&lt;-6 or &gt;6</td>
</tr>
<tr>
<td>WHZ</td>
<td>&lt;-4 or &gt;6</td>
<td>&lt;-6 or &gt;5</td>
</tr>
<tr>
<td>WAZ</td>
<td>&lt;-6 or &gt;6</td>
<td>&lt;-5 or &gt;5</td>
</tr>
</tbody>
</table>

3.2.3 Reporting the results of analysis

Once standard analysis has been performed, it is recommended that reporting of its results include measures of precision around prevalence estimates, as well for z-score means. At the very least, the report should include the following parameters.

1. Height-for-age: weighted and unweighted sample sizes, % < -3 SD (95% CI), % < -2 SD (95% CI), z-score mean (95% CI), z-score SD;
2. Weight-for-age: weighted and unweighted sample sizes, % < -3 SD (95% CI), % < -2 SD (95% CI), z-score mean (95% CI), z-score SD;
3. Weight-for-height: weighted and unweighted sample sizes, % < -3 SD (95% CI), % < -2 SD (95% CI), % > +2 SD (95% CI), % > +3 SD (95% CI), z-score mean (95% CI), z-score SD.

It should be noted that deriving estimates of standard errors from confidence intervals of prevalence estimates may not be straightforward since they are often not symmetrical; ideally, standard errors should also be reported. Standard errors are useful for feeding into modelling exercises to account for data variance.

As an example of the standard analysis application, Table 9 sets out the requisite variables and procedures for treating input variables when calculating individual z-scores for each child and prevalence estimates with tool developed by WHO, the WHO Anthro Survey Analyser. More detailed comments on the methodology for deriving the various anthropometric z-scores are described elsewhere (25).

---

19 Even though the cut-offs are different in the two columns, differences are small in terms of actual kg or cm values in the two international references (WHO Child Growth Standards and NCHS/WHO reference). This is because the WHO flags for exclusion criteria were identified to maintain inferences similar to those already in use with the NCHS/WHO reference.
<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>ACCEPTED VALUES</th>
<th>METHODOLOGICAL ASPECTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth and date of visit</td>
<td>Accepted date formats: DD/MM/YYYY or MM/DD/YYYY</td>
<td>Both variables, date of birth and date of visit, should be provided to calculate age in days (date of visit minus date of birth). If DAY is missing for the date of birth, a new variable should be created by imputing the missing day by 15 (e.g. ??/05/2014 should be set as 15/05/2014) in the analysis file before importing the dataset. In turn, if the month or year is missing, the date value should be set to missing/blank (see analysis data file preparation in the tool WHO Anthro Analyser Quick Guide). When date of birth and date of visit are missing, the variable age (in days or months) can be used (see below). When month or year of birth and age are missing, results will only be computed for weight-for-height. Invalid date of birth or date of visit or a negative value resulting from date of visit minus (-) date of birth entails a missing age.</td>
</tr>
<tr>
<td>Age</td>
<td>Accepted values for variable age:</td>
<td>It is recommended that age be calculated based on the date of visit and date of birth. Mapping of the variable age is available only when the user selects this option instead of the (default) recommended calculation based on date of visit and date of birth. This should be done only if the latter is not available. Age in months should be derived by dividing age in days by 30.4375 and not by rounding it off). Age in months should be provided with a precision of at least two decimal points for accurate estimation of age-related malnutrition.</td>
</tr>
<tr>
<td>Sex</td>
<td>Male (1, M or m) and Female (2, F or f)</td>
<td>If this variable (sex) is missing, z-scores will not be calculated for any index because the WHO Child Growth Standards are sex-specific.</td>
</tr>
<tr>
<td>Weight</td>
<td>Numerical, float value (in kilograms with one decimal point precision)</td>
<td>If missing, estimates for weight-related indices will not be calculated.</td>
</tr>
<tr>
<td>Length or height</td>
<td>Numerical, float value (in centimetres with at least one decimal point precision)</td>
<td>If missing, estimates for length- or height-related indices will not be calculated.</td>
</tr>
<tr>
<td>VARIABLE</td>
<td>ACCEPTED VALUES</td>
<td>METHODOLOGICAL ASPECTS</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Standing (height) or recumbent (length) position</td>
<td>Usually: recumbent length (L or l) or standing height (H or h).</td>
<td>Depending on information provided about the measurement position, standard analysis software should make adjustments automatically when calculating z-scores, adding 0.7 cm if standing height is measured for children aged &lt; 24 months, and subtracting 0.7 cm if recumbent (lying) length is measured for children aged ≥ 24 months. If this information is missing, the code will assume recumbent length for children aged &lt; 731 days (&lt; 24 months) and standing height for those aged ≥ 731 days (≥ 24 months). If this information is missing and the child's age is also missing, the code will assume that the measurement was recumbent length if the length/height value is below 87 cm (mean value from the Multicentre Growth Reference Study sample in boys and girls, for height-for-age and length-for-age at 24 months) and otherwise assume that the measurement was standing height. For children under 9 months of age, where the information indicates that standing height was measured, the code will assume this was an error and register the case as missing. This is done to avoid the wrong automatic adjustment in such cases (adding 0.7 cm) which can lead to overestimation of wasting and underestimation of stunting.</td>
</tr>
<tr>
<td>Oedema (assessment not recommended systematically except for settings where collecting this information is appropriate)</td>
<td>Usually: No (2, N or n) or Yes (1, Y or y)</td>
<td>If not provided as a variable, all values will be assumed to be missing. Missing values are treated as no oedema and z-score calculation is not affected. Z-scores for all weight-related indices will be set to “missing” when oedema is present. The report includes the number of children with bilateral oedema. For prevalence calculation purposes, children with oedema are classified as having severe malnutrition (i.e. weight-for-length/height &lt; -3 SD, weight-for-age &lt; -3 SD and BMI-for-age &lt; -3 SD). It is recommended that prevalence levels based on both analyses (including or excluding information on oedema) are included in the survey report. This should be performed by a separate analysis, one including the variable “oedema” and another excluding (or not mapping) it.</td>
</tr>
<tr>
<td>Sampling weight</td>
<td>Numeric, float value</td>
<td>If sampling weights are not provided, the sample will be assumed to be self-weighted, i.e. the sampling weight equals one (unweighted analyses will be carried out). If provided, all children with missing sampling weights will be excluded from the analysis.</td>
</tr>
<tr>
<td>Cluster</td>
<td>Numeric integer</td>
<td>If not provided, it will be assumed that all children belong to the same unique cluster/PSU. If provided, all children with missing cluster/PSU data will be excluded from the analysis.</td>
</tr>
</tbody>
</table>

---

20 The cut-off point of 87 cm reflects the standards’ median for boys and girls height-for-age z-score (HAZ) at 24 months. The WHO standards’ median height is 87.1 cm for boys and 85.7 cm for girls, and median length is 87.8 cm for boys and 86.4 cm for girls. The mean of these four values is 86.75 cm which was rounded to 87 cm in order to obtain the cut-off point for shifting from length to height in case age and the type of measurement are unknown ([https://www.who.int/childgrowth/mgrs/en/](https://www.who.int/childgrowth/mgrs/en/)).
### 3.3. DATA INTERPRETATION

#### 3.3.1 Reporting nutritional status

Prevalence-based data for children aged 0–59 months are commonly reported using cut-off points, usually < -2 SD and > +2 SD. The rationale for doing so is that statistically 95% of the international reference population can be found within this central range. The reference population recommended globally for calculating prevalence is the Multi Reference Growth Study (MRGS) Population (22).

The WHO Global Database on Child Growth and Malnutrition uses a z-score cut-off point of < -2 SD to classify low weight-for-age (underweight), low length/height-for-age (stunting) and low weight-for-length/height (wasting) as moderate and severe, and < -3 SD as severe undernutrition. The cut-off points of > +2 SD classifies high weight-for-height in children as moderate and severe overweight and of > +3 SD as severe overweight.

The assessment of nutritional indices among children at the population level is interpreted on the basis of the assumption that in a well-nourished population they normally follow the distribution of the bell curve shown below (Figure 14).

Use of -2 SD and +2 SD as cut-offs implies that 2.3% of the reference population at both tails or ends of the population curve will be classified as malnourished even if they are apparently “healthy” individuals with no growth impairment. Accordingly, 2.3% can be regarded as the baseline or expected prevalence at both ends of the spectrum of nutritional status calculations. Reported values in surveys would need to subtract this baseline value in order to calculate prevalence above normal if they seek to be precise. It is important to note however, that the 2.3% figure is customarily not subtracted from observed values.

**Figure 14. Standard normal distribution of a model population**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strata</td>
<td>Numeric integer</td>
<td>If not provided, it will be assumed that all children belong to the same unique stratum. If provided, all children with missing strata data will be excluded from the analysis.</td>
</tr>
</tbody>
</table>
3.3.2 Interpreting prevalence estimates

Prevalence ranges have conventionally been used since the early 1990s to classify levels of malnutrition in global monitoring.

In 2018, the WHO-UNICEF Technical Expert Advisory Group on Nutrition Monitoring (TEAM), an independent technical expert group created to provide advice on nutrition monitoring, revised prevalence ranges used to classify levels of stunting and wasting, and established prevalence ranges to classify levels of overweight (based on weight-for-length/height) (26).

Table 10 presents the new prevalence thresholds, labels and country groupings for wasting, overweight and stunting. Labels have been harmonized across indicators as “very low”, “low”, “medium”, “high”, and “very high”. TEAM described these classifications as “prevalence thresholds”, a term more in line with its intended population-based application, as opposed to “cut-offs”, which is a term mainly used for interpreting measurements of individual children. Prevalence levels were set depending on their degree of deviation from normality as defined by the WHO Child Growth Standards.

| TABLE 10. POPULATION LEVEL PREVALENCE THRESHOLDS, CORRESPONDING LABELS AND NUMBER OF COUNTRIES IN DIFFERENT PREVALENCE THRESHOLD CATEGORIES FOR WASTING, OVERWEIGHT AND STUNTING USING A “NOVEL APPROACH” |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| WASTING                        | OVERWEIGHT                      | STUNTING                        |
| Prevalence thresholds (%)      | Labels                          | # of countries                  | Prevalence thresholds (%)      | Labels                          | # of countries                  |
| < 2.5                          | Very low                        | 28                              | < 2.5                          | Very low                        | 16                              |
| 2.5 - < 5                      | Low                             | 41                              | 2.5 - < 5                      | Low                             | 35                              |
| 5 - < 10                       | Medium                          | 39                              | 5 - < 10                       | Medium                          | 50                              |
| 10 - < 15                      | High                            | 14                              | 10 - < 15                      | High                            | 18                              |
| ≥ 15                           | Very high                       | 10                              | ≥ 15                           | Very high                       | 9                               |

Wasting = weight-for-length/height <-2 SD; Overweight = weight-for-length/height >+2 SD; Stunting = Length/height-for-age < -2 SD

The revised prevalence thresholds presented here can be interpreted and exploited by the international nutrition community for various purposes: classifying and mapping countries according to levels of malnutrition severity (27); by donors and global actors to identify priority countries for action (28); and most importantly, by governments for monitoring and triggering action and target programmes aimed at achieving “low” or “very low” levels.

3.3.3 Interpreting mean z-scores

The mean z-score, though less commonly used, provides a direct description of the average nutritional status of the entire population without the need to refer to a subset of individuals below a given cut-off. A mean z-score significantly lower than zero, which is the value for distribution in the reference population (in this event the MGRS study population), means that overall distribution has shifted downwards, implying that most if not all individuals have been affected. More research is needed to understand the circumstances in which the mean z-score can most effectively be used.

3.3.4 Checking denominators

Attention needs to be paid to the denominators used when reporting on stunting, wasting, underweight and overweight in children aged under five years. While not recommended for global reporting, some surveys do not routinely include children aged 0–5 months in their anthropometry measurements. Using denominators that are not aligned with global indicators for children under 5 years of age may confuse the interpretation of estimations if reports are being compared across populations or examined for trends and even data quality. Denominators should always be clearly documented when reporting on the nutritional status of children, especially when observing trends.
3.3.5 Tracking trends

WHO, in collaboration with UNICEF and the EC, has developed a tracking tool to help countries set their national targets and monitor progress towards WHA targets, three of which are stunting, overweight and wasting. This tool allows users to explore different scenarios while taking into account different rates of progress to meet the targets and time left to 2025. Information and tools related to the tool can be accessed by users at the following link on the WHO website of the Global Targets Tracking Tool.

This tracking tool has been used to review trends in current data included and validated in the UNICEF-WHO-WB Joint Child Malnutrition review. Tools for estimating trends include stunting and overweight as indicators. While wasting is part of the tracking tool it is not included in the calculation of trends owing to high short-term variability. Countries are encouraged to use the Excel template provided with this guidance document (Annex 11), which follows a similar methodology to that of the tracking tool, to enter their own data for assessing trends.

3.4. HARMONIZED REPORTING AND RECOMMENDED RELEASE OF DATA

A harmonized method of reporting is essential if survey teams wish to develop a comprehensive set of indicators and ensure comparability between surveys. In addition, qualitative information about contextual factors, e.g. shocks and crises, can help survey managers and statisticians to gain a better understanding of anthropometric data from various types of survey and to use it more effectively. Providing systematic notes on seasonal and other relevant contextual factors and how to use these meta-data is recommended.

This section presents a harmonized scheme for reporting anthropometric data as well as contextual information gathered in various nutrition surveys.

Survey results must be reported at the national level. Data at the subnational level can also be presented where they exist. Results should be presented in a standardized manner, e.g. percentage of children with z-scores below or above standard cut-offs using WHO flags and age groups (< 6, 6 to < 12, 12 to < 24, 24 to < 36, 36 to < 48 and 48 to < 60 months).

The inclusion of the following information when presenting anthropometry data is recommended:

a) Cover page
Survey title, dates of the survey, author.

b) Executive summary

c) Introduction
- Survey title and details: geographic area surveyed (areas excluded if any and why), description of the population: total population, population surveyed, type of population surveyed (residents, immigrants, refugees, displaced, etc.).
- Contextual information: food security, nutrition, health situation or any other information likely to have an impact on the nutrition status of the population;
- Objectives: population including age group surveyed;

d) Methodology
- Sample size determination;
- Sample frame details including whether any region, district, PSU or other area or population has been excluded from the first stage sample (and why);
- Sampling design and procedure: full details about all sampling stages, especially the initial stage (i.e. selection criteria for PSUs), second stage (i.e. mapping and listing procedures) and last stage (i.e. selection of households and participants, etc.) and any additional step or stage applied in the survey (e.g. subsampling, etc). Include a definition of household and household member;
- Questionnaire: procedures for developing the questionnaire and interviewer instructions, development and instructions for using the local events calendar, pre-testing if any, procedures for translation and back-translation, etc;
- Measurement procedures;
- Case definitions and inclusion criteria;
- Training of field staff: content, number of days, number of trainees, description of standardization exercises implemented and results of the standardization exercise, pilot test in the field, etc;
- Field work procedures: data collection procedures, number and composition of teams, period of data collection, procedures for call-backs when children absent or for re-measuring children, etc;
- Equipment used and calibration procedures;
- Coordination and supervision process: checks for procedures in the field;
- Data entry procedure;
- Data analysis plan: software (name, version and link if available), data cleaning process, imputation factors (e.g. WHO Anthro Analyser imputes day 15 when day of birth is missing);
- Type of flags used.

e) Results
- Total number of PSUs sampled versus number of PSUs completed (and reason for non-completions);
- Total number of sampled households;
- Breakdown of survey outcomes for all sampled households: completed, refused, including random and flagged re-measurements;
- Total number of children under 5 years in sampled household (indicating if all children were eligible); if data are collected in a subsample, present the total number of eligible children in this subsample;
- Total number of eligible children under 5 years with weight measurement, length/height measurement and at least the month and year of birth;
- Total number of eligible children under 5 years selected for random re-measurements with weight measurement; length/height measurement; and at least month and year of birth;
- Prevalence of anthropometric indicators based on recommended cut-offs for each indicator together with confidence intervals (for stunting, wasting, overweight and underweight). Information should be presented as tables and/or plots;
- Design effects observed;
- Mean z-scores for each index;
- Z-score standard deviations;
- Standard errors (SE) for prevalence and mean z-score estimates;
- 95% confidence intervals for prevalence and mean z-score estimates;
- Frequency distribution plots versus reference population;
- Results presented by disaggregation categories for results where available: sex, age group, urban, rural and subnational regions, wealth quintiles and mother’s educational level;
- Weighted and unweighted total sample (n) for each indicator.

f) Reporting on indicators for data quality
The recommendation in section 3.1 is to include a data quality report similar to the model in Annex 9 using the WHO Anthro Analyser output. Since this may be too extensive for multitopic household surveys, it could be limited to a summary presentation following the bullet list below provided the raw data are publicly available.
- Number and percentage of cases excluded when applying the fixed exclusion criteria based on WHO Child Growth Standards for each anthropometric index: this should include the overall number and percentage of cases as well as for the best and worst performing teams;
- Missing data disaggregated by age group and other reporting categories. Number and percentage of children missing for height, weight and age expressed at least as month and year of birth;
- Digit heaping charts including for length, height, weight and age;
- Distribution issues: z-score distributions by age group, sex and geographical region;
- Percentage of date-of-birth information obtained from birth certificate, vaccination card, caretaker’s recall or other source out of the total number of eligible children. Children lying down/standing up for measurement by age: % of children below 9 months standing, % of children over 30 months lying down, % mismatches for position measured versus recommended position;
- Mean, SD, median, min, max, absolute difference between the first and second measurement for the random cases;
- Percentage of random remeasurements within the maximum acceptable difference;
- Indicate other eventual data quality pitfalls and survey limitations.

g) Discussion
Interpretation of the nutritional status of the children concerned including contextual factors which might have some bearing on the results. Limitations of the survey.

h) Conclusion
The conclusion should summarize the main findings of the survey, briefly mention any interpretative issues raised in the discussion and make recommendations (often as a list) that are logically related to points already made. It should not contain any new material and may merge with the executive summary.

i) Appendices
- Sample design details;
- Questionnaire;
- Local events calendars;
- Map of the area surveyed;
- Results of standardization exercises;
- Field check tables used.

• The online application WHO Anthro Survey Analyser generates a summary report template with the principal graphic outputs and tables of summary statistics as well as a data quality report (Annex 9). A completeness checklist is provided in Annex 10.

Public release of datasets from surveys collecting anthropometric data
As mentioned at the beginning of this document, it is recommended that an agreement be signed with central or local government in the very first stages of the survey for public release of its report and dataset once the data have been validated. Releasing survey datasets for public use ensures transparency and also allows for secondary analysis which can lead to a better understanding of the data and the context in which they were collected, thus enabling the data to be used for the benefit of the population from which they were collected. Raw datasets should be made available for public use including the quality assurance measures that were included in the dataset for both random and flagged re-measurement.

Datasets should still be made public even where a survey produces poor quality results, and problems with data quality explicitly addressed in the report, even when report does not include nutritional status findings.

In some cases, government is responsible for endorsing and releasing survey results and must be consulted for authorization to release datasets to specific individuals or to make them accessible on the internet.

Whatever the case, there is a need to strengthen commitment and advocacy to ensure public access to raw data and develop a database (e.g. registry or repository) containing survey datasets and protocols. Datasets should be released with minimally cleaned data, showing pre- and post-application of flags, so that researchers are subsequently able to apply uniform flags to datasets.

Datasets should include records for all sampled households even if interviews were not completed, and all children who ought to have been measured (even if they were not), actual measurements recorded (i.e. length/height and weight), their date of birth and date of measurement (date of visit), sampling weights, and all other variables. These datasets should be accompanied by clear documentation.

21 See an example of a Sampling annex of a MICS report here: [http://mics.unicef.org/surveys](http://mics.unicef.org/surveys)
The TEAM working group, which is composed of experts from different international organizations (CDC, DHS, SMART, UNICEF, WHO, etc.) recommends that raw datasets from national household surveys including anthropometry be released for public use as a way of strengthening the use of anthropometry data for public health purposes.

**TOOLS**

- On data anonymisation, consult a [Guide to data protection](#).
- [Archiving and dissemination tool](#).
- [USAID open data policy 2014](#).
RECOMMENDATIONS AND BEST PRACTICES

Section 3.1- DATA QUALITY ASSESSMENT

Recommendations (must)
- Report on the following using specifics outlined in the report regarding how to calculate and present:
  - Completeness;
  - Sex ratio;
  - Age distribution;
  - Digit preference for height and weight;
  - Implausible z score values;
  - Standard deviation;
  - Normality.

- Appraise data quality by considering the indicators conjointly and not in isolation;
- Do not undertake formal tests or scoring.

Good practices (optional):
- Use WHO Anthro Survey Analyser data quality report.

Section 3.2- DATA ANALYSIS

Recommendations (must):
- Use the standard approach as outlined in the report for analysis, including:
  - Use of WHO Child Growth Standards and WHO flags;
  - It is important that all records, including those with missing measurements or sampling weights, are available for analysis, since they are important for data quality assessment (e.g. non-response);
  - Oedema measurement is only appropriate in surveys where local experts, specifically clinicians or individuals from the Ministry of Health working at a local level, can clearly indicate if they have seen recent cases where nutritional oedema was present;
  - Calculate age using date of birth and date of visit and imputation of day 15 if no day available;
  - Imputation of missing day of birth: if only the month and year of birth are provided, it is recommended that the missing information for the day of birth be imputed. This can be done in different ways but using the 15th of the month for all missing days of birth is recommended in standard analysis;
  - Child’s measurement position (standing height or recumbent, i.e. supine or lying length) should be recorded in the questionnaire to allow for age-linked adjustments in length/height depending on whether they were lying or standing;
  - Ignore conversion of standing to lying position for children <9 months;
  - Re-measurements (height, weight, date of birth, and sex) of children randomly selected or flagged should be retained in the datafile;
  - Use height, weight, date of birth and sex from the first measurement for children randomly selected for re-measurement when calculating z-scores. Use height, weight, date of birth and sex from the second measurement for children flagged for re-measurement when calculating z-scores;
  - The number and percentage of values excluded should be reported;
  - All measurements should be retained in the dataset for transparency;
  - A sampling weight must be assigned to each individual in the sample to compensate for unequal probabilities of case selection in a sample, usually owing to the design.
**Good practices (optional):**
- Use WHO Anthro Survey Analyser or standard STATA and R syntax from JME;
- Monitoring equity is of increasing importance for health and development. Disaggregated analysis is also recommended in order to derive estimates by wealth quintiles (1=lowest, 2, 3, 4, 5=highest) and mother’s education (no education, primary school and secondary school or higher), whenever this is possible;

**Section 3.4– DATA INTERPRETATION AND REPORTING**

**Recommendations (must):**
- Include measures of precision around prevalence estimates, as well for z-score means in the report.
- Include the prevalence of moderate and severe forms of malnutrition as well as mean and SD for HAZ, WHZ and WAZ in the report.
- Include data quality assessment findings as per section 3.1 in all survey reports that provide estimates for child anthropometric indicators.
- Release complete and clearly labelled datasets to the public including initial measurements and re-measurements
- Include a detailed annex on sampling (to the level of detail in MICS and DHS reports)

**Good practices (optional):**
- Include the Anthro Analyzer data quality report in annex of the survey report