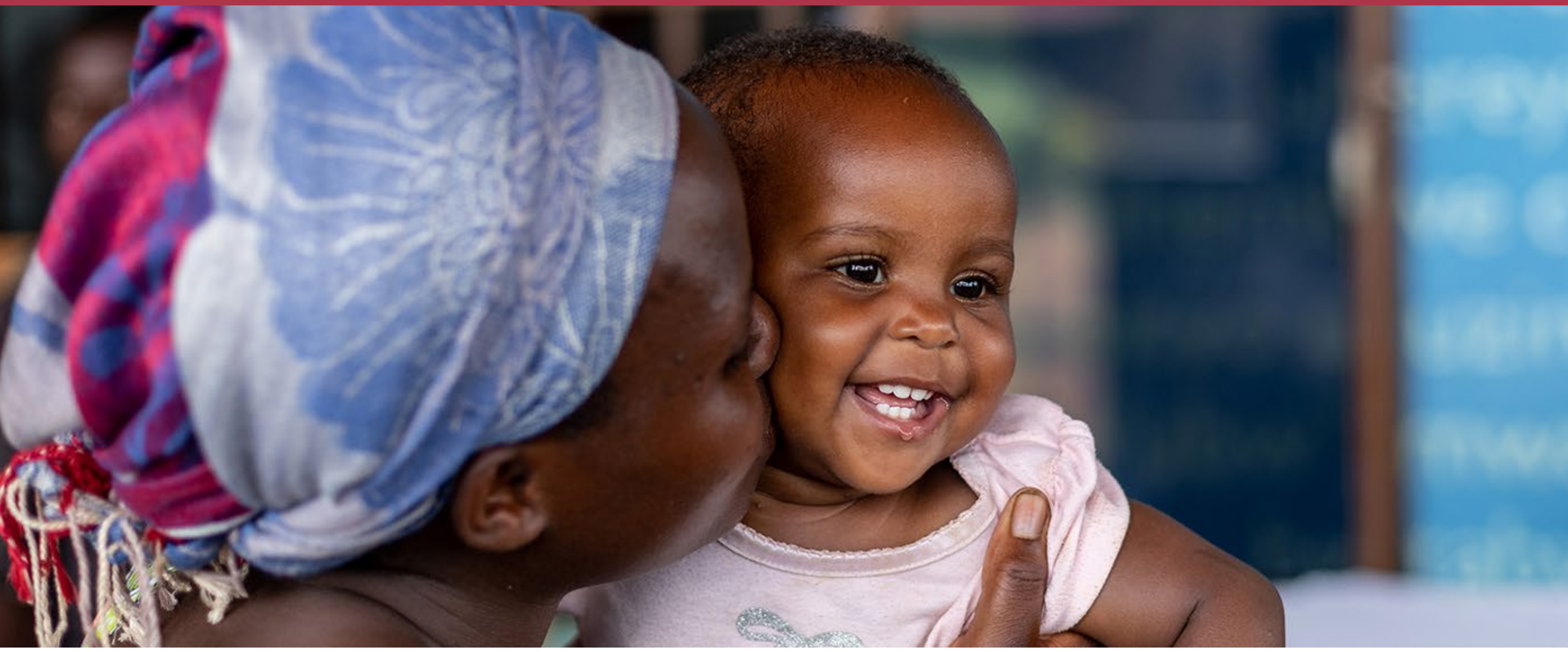
Research Agenda Brief

Research agenda to address gaps in data collection, analysis and reporting on anthropometric indicators in children under 5 years old



Assessment and reporting of anthropometric data quality in population-based surveys

Statement of problem

The accuracy of population-level nutritional status indicators depends on rigorous implementation of recommended practices for collection, management, cleaning and analysis of anthropometric survey data.[1] Data quality has improved over the past few decades as the result of closer supervision and quality control of field work, which can mitigate errors in real-time.[2,3] However, there remains substantial variability in data quality between surveys as well as across teams carrying out the same survey.[4-7] Survey methodology (e.g., household listing and sampling procedures, age eligibility) can also affect elements of quality scoring, such as the proportion of missing values.[6] Therefore, standard methods for assessing and reporting survey quality are required to contextualize the interpretation of survey results and support decisions regarding their inclusion in global databases used for monitoring trends. They are also required to enable differential weighting of data sources when aggregating surveys in global modelling exercises, and, when justified, to exclude unacceptably low-quality anthropometric datasets.

In the 2019 *Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old,*[1] the World Health Organization (WHO) and United Nations Children's Fund (UNICEF) described a comprehensive set of indicators for data quality assessment, elaborating on indicators originally proposed by the 1995 WHO Expert Committee.[8] Because anthropometric data collection and recording is affected by a variety of random and systematic errors,[9] a range of indicators are used to reflect the precision and accuracy of height, weight and date of birth data, as well as factors that affect the representativeness of the survey sample (i.e., completeness, age and sex distributions)[7]. Quality measures are not all correlated with one another,[7,10] which suggests the value of including complementary indicators that capture distinct dimensions of data collection and recording practices (**Table 1**). Multi-indicator scores or indices have also been used to provide an overall assessment of each survey.[3,6,7] Although such scores or indices have not been formally validated, surveys known to be of particularly low or high quality appear to be ranked accordingly by them.[3,6] The Standardized Monitoring and Assessment for Relief and Transitions (SMART) programme, which is widely used by international agencies and national nutrition surveys, includes a series of 'plausibility checks' that are aggregated into an overall quality score – a process that is automated using the SMART Emergency Nutrition Assessment software.[11]

The 2019 WHO/UNICEF *Recommendations* highlight numerous limitations of existing quality indicators.[1] A general concern is the lack of empirical evidence to define thresholds above or below which surveys should be flagged as low quality; as such, a specific cut-off was only endorsed for one indicator in the 2019 WHO/UNICEF *Recommendations*[1] (Table 1).

Table 1. Commonly used measures and indicators of anthropometric survey data quality

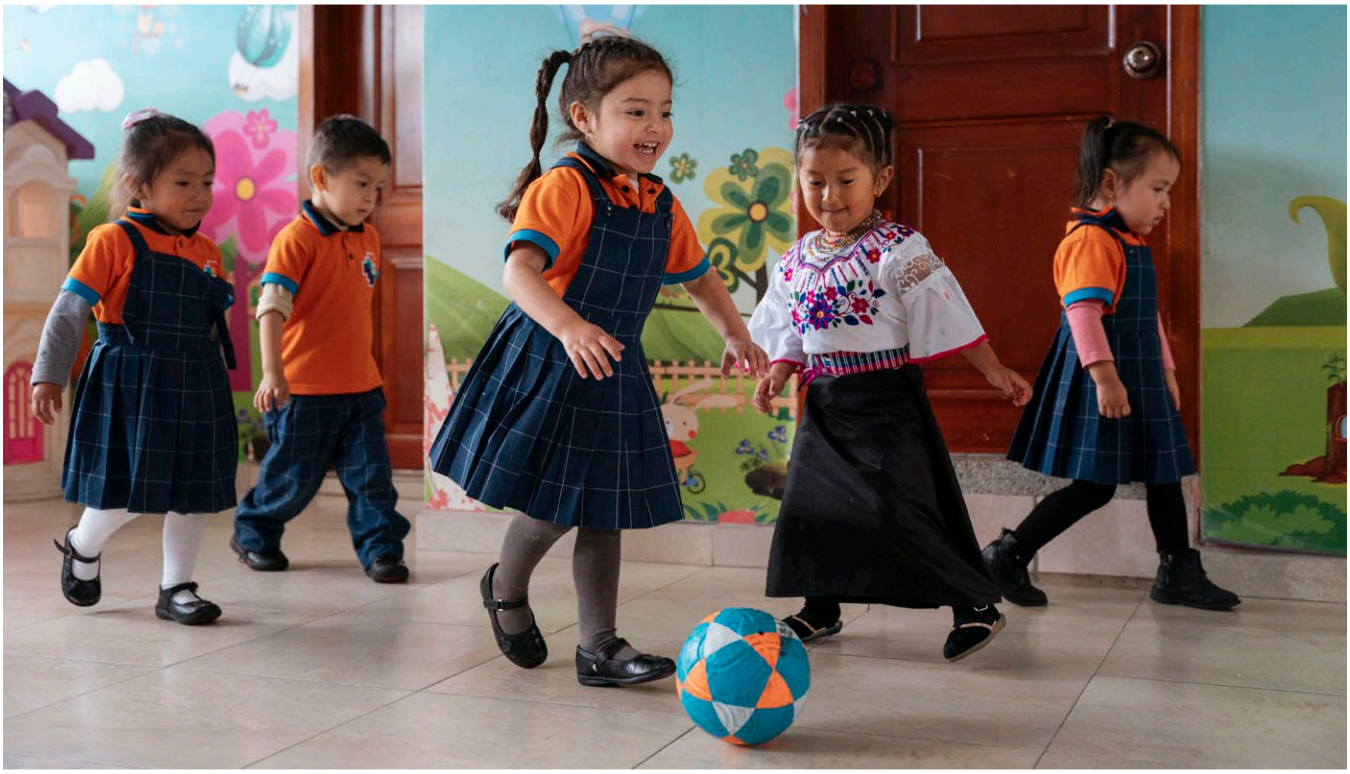
Indicator	Explanation	Reporting and thresholds according to the 2019 WHO/UNICEF Recommendations[1]	SMART quality scoring[11] *[Higher number of penalty points indicates lower quality]*
Completeness	Various indicators that reflect the proportions of potential survey respondents who were contacted or provided any data and/or valid data.	Report proportions of children or households with completed interviews or measurements. E.g., Child completion rate = Number of children under 5 years of age with completed interviews divided by the number of eligible children under 5 years of age. *No thresholds specified.*	Report proportion of children in the dataset without a complete birthdate. *Not scored.*
Age ratio	Ratio of the number of children in a defined 'young' age range versus the number in an older age range.	N/A	Chi-squared test of expected versus observed distributions. Up to 10 points if p ≤0.001.
Age heaping	Observed distributions of specific ages and/or months of birth, compared to the expected uniform distributions.	Examine unexpected distributions using histograms. *No thresholds specified.*	Visual plot of age distribution by month. *Not scored.*
Sex ratio	Ratio of the number of males to females.	Compare observed ratio to the expected country-specific ratio. *No thresholds specified.*	Expected versus observed distributions tested by Chi-squared test. Default setting assumes the expected ratio is 1:1. *Up to 10 points if p ≤0.001.*
Terminal digit preferences of heights, weights and mid-upper arm circumferences	Observed distribution of terminal digits, for a given variable, compared to the expected uniform distribution.	Report the index of dissimilarity and plot distributions of terminal digits and integers using histograms. *No thresholds specified.*	Chi-squared test of expected versus observed digit distributions and calculation of Digit Preference Score (DPS, range 0-100). *Up to 10 points if DPS >20.*
Flagged outliers or implausible z-score values[a]	Proportion of values for a particular anthropometric index that exceed specified lower and upper limits of acceptable ranges based on statistical criteria or biological plausibility.	Report the percentage of biologically implausible values for each index based on WHO criteria. *>1% is indicative of poor quality.*	Report the proportion of flagged values using SMART criteria for each index. *Up to 20 points if % >7.5.*
Standard deviation (SD) of z-scores[b]	Measure of the dispersion of anthropometric z-score values around the mean value.	Report the SD for each anthropometric index and examine patterns within disaggregated subgroups. *No thresholds specified.*	Report SD for each index, but only the SD for weight-for-height z-score (WHZ) is scored. *Up to 20 points if SD is ≤0.8 or ≥1.*
Skewness and kurtosis of z-scores[b]	Skewness is a measure of the symmetry and kurtosis is a measure of the 'tailedness' of the anthropometric z-score distribution.	Examine anthropometric z-score distributions using kernel density plots. Report and interpret coefficients based on typical ranges: Skewness: <-0.5 or >+0.5 Kurtosis: <2 or >4 (excess kurtosis of <-1 or >1) *No thresholds specified for quality.*	Report skewness and excess kurtosis coefficients for the WHZ distribution. *Up to 5 points if skewness is ≥0.6.* *Up to 5 points if excess kurtosis is ≥0.6.*

a Detection of individual outliers and biologically implausible values (BIVs) is addressed in detail in Brief 7. WHO-2006 BIV thresholds: height-for-age z-score (HAZ): -6, +6; WHZ: -5, +5; weight-for-age z-score (WAZ): -6, +5. SMART outlier thresholds (by default settings): +/-3 z-scores from the sample mean.

b Standard deviation, kurtosis and skewness are usually assessed in the cleaned dataset (following exclusion of outliers/BIVs) whereas other quality indicators are typically based on the raw dataset.

The inherent validity and generalizability of some indicators is uncertain given the diversity of populations in which anthropometric surveys are conducted. For example, sex and age distributions may not be uniform in all settings due to secular shifts in the demographic profile or mortality patterns.[6] Also, several measures of survey data quality are correlated with the estimated prevalence of stunting and wasting.[3] While these correlations can arise as direct effects of measurement error (e.g., excess dispersion of values at the lower tail of z-score distribution will increase the prevalence of stunting or wasting),[12] they may also represent confounding by contextual factors that interfere with rigorous data collection and compromise nutritional status, such as in conflict-affected areas or communities with high levels of within-population health inequities.[3]

There is substantial debate surrounding the use and interpretation of the statistical properties of anthropometric z-score distributions as measures of data quality.[1] A long-held assumption is that high-quality data produces a standard deviation (SD) for each anthropometric z-score distribution that is "close to the expected value of 1.0" based on reference distributions.[8] Support for the use of SD as a quality indicator has also been based on its lack of correlation with the mean



z-score.[5,13] However, in practice, SDs vary widely across surveys[4,6,7,14] and by different teams carrying out a given survey[7,15]. SDs are usually >1,[4,6,7,14] particularly for height-for-age z-scores (HAZ), even in high-quality surveys (e.g., the National Health and Nutrition Examination Survey).[6] There are considerable variations in SD by age[1,16] and other child and household characteristics,[4] but it is unclear whether these result from genuine differences in the magnitude of between-child variation or differential measurement errors across population subgroups. For example, irrespective of survey methodology, SDs for HAZ are usually wider at younger ages.[1,4,5,13] This may be due to the greater technical challenge of obtaining precise length measurements in very young infants,[4,16] not accounting for gestational age at birth,[17] the greater difficulty in measuring length versus standing height,[1] or that the same absolute error would have less effect on the SD at older ages. Wider dispersion of measurements at younger ages could also be related to the greater vulnerability of infants and toddlers to the effects of environmental conditions compared with 3 to 5 year olds.[18] However, there is a lack of evidence to directly support these explanations, and they do not fully account for the progressive decline in SD with increasing age from birth through to 5 years of age.[1]

Random measurement error causes excess dispersion of z-score distributions and can therefore spuriously increase the prevalence of undernutrition or overweight (since estimates are based on the proportions of values in the tails of the distributions).[14,19] Heterogeneity within and between population subgroups also contributes to the spread of z-scores;[1,4] however, this phenomenon has been more difficult to quantify.[4,19] Therefore, while there is consensus that SDs are useful for comparing data quality across otherwise similar surveys, and that very large SD values usually reflect poor data quality, there remains uncertainty about the range of acceptable SDs across diverse settings and the thresholds above which high SDs conclusively signal problematic datasets.[1] While z-score distributions in low- and middle-income countries are often symmetrical or bell-shaped, they may not be strictly normal (Gaussian)[5,7] and there is doubt about whether classical measures of normality, such as kurtosis and skewness, can be used as indicators of data quality.[1]

The WHO 2006 Child Growth Standards were based on generally healthy term-born children,[20] but in practice, they are applied to survey populations that include preterm-born children and children with chronic conditions or disabilities that affect growth, which could alter the shape and dispersion of observed z-score distributions. These considerations also pertain to the proportion of measurements considered outliers or biologically implausible values (BIVs) because such estimates are closely related to the characteristics of the z-score distribution, in addition to being influenced by the thresholds selected to define the acceptable range of values.[10*] The proportion of implausible z-scores was the only quality indicator for which a threshold (>1%) was specified in the 2019 WHO/UNICEF *Recommendations,*[1] drawing on the report of the 1995 WHO Expert Committee

* Methods for detection of outliers and biologically-implausible values (BIVs) are addressed in brief 7.

on anthropometry.[8] However, surveys frequently have >1% BIVs,[21] and the 2019 WHO/UNICEF *Recommendations* do not prescribe decisions or actions to be taken when the threshold is surpassed.[1]

Terminal digit preferences, which are unrelated to the location or dispersion of the underlying age or raw weight/height distributions, may reveal non-adherence to standard data collection and recording procedures; they could also indicate data fabrication.[1,11] Unexpected peaks in the distributions of specific ages, months of birth or other raw variables (referred to as 'heaping') may be caused by problems with survey administration or measurement errors. Visualization of various types of heaping (age in complete years, age in complete months and calendar month of birth) were included in the 2019 WHO/UNICEF *Recommendations;*[1] these can be quantified by similar methods as those used for terminal digit preferences (e.g., index of dissimilarity),[7] but there is a need to further develop methods for quantifying heaping and assessing its influence on malnutrition prevalence estimates.[1] Since month of birth misreporting was shown to cause an upward gradient in HAZ by month of birth,[22] the absolute difference between mean HAZ (or WAZ) for the first and last months of the local calendar (e.g., January versus December) provides an alternative measure for quantifying age misreporting that has shown promise in some recent quality assessments.[3,7]

Anthropometric data quality assessments are recommended and widely implemented. However, there is no standard method to support decisions regarding use of the data to generate malnutrition prevalence estimates or inclusion of surveys in global databases and models such as the UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates (JME).[23] The JME excludes some surveys due to quality concerns using an assessment framework based on the 2019 WHO-UNICEF *Recommendations;*[24] exclusions may be due to large amounts of missing data or high proportions of implausible z-scores (substantially higher than the 1% threshold), but qualitative appraisals are also used.[24]

In summary, there is insufficient evidence to establish the validity of specific measures and indicators, definitions of acceptable ranges or decision thresholds for each indicator (including cut-offs above/below which surveys are of unacceptably poor quality), or the role of narrative reviews in the interpretation of quality indicators. The advantages and disadvantages of composite quality scores or indices also remain unclear. There is interest in exploring new measures and indicators and methods of quality assessment, including techniques designed specifically to detect data manipulation or fabrication, which have been under-explored to date. This brief describes a research agenda aimed at the development and validation of standard methods and metrics for the assessment and reporting of anthropometric data quality in population-based surveys.*

Research questions

1	**What criteria and evidence support the validity of survey quality indicators and composite scores and indices currently used in the assessment of anthropometric data from population-based surveys across diverse settings?**
2	**What are the statistical properties of anthropometric z-score distributions (standard deviation, kurtosis, skewness) in high-quality population-representative surveys conducted in diverse settings?**
3	**What are the optimal ranges and threshold values for existing survey quality indicators to enable survey grading and classification of surveys with respect to data quality? What is the role of qualitative appraisal in the interpretation of anthropometric survey data quality indicators?**
4	**What are the advantages and disadvantages of composite quality scores or indices compared to individual indicators of anthropometric survey quality? Are there particular contexts in which composite quality scores/indices would be preferred over individual indicators to support decision-making?**
5	**What are potential new measures, indicators and methods of data quality assessment that may be applicable to anthropometric datasets, including specific techniques for detecting data manipulation or fabrication?**

* This brief focuses on the use of quality indicators and composite scores for the purposes of grading surveys, including the identification of low-quality surveys or fabricated data. However, decisions regarding the downstream use of low-quality survey data when estimating malnutrition indicators depend on additional considerations. For example, when aggregating multiple surveys in global modelling exercises, some low-quality datasets could be included but down-weighted rather than excluded entirely. In other situations, low-quality datasets may be excluded from primary analyses but included in sensitivity analyses. The topic of statistical modelling approaches to generate malnutrition indicators is addressed in Brief #9.

Research topic 1

Criteria and evidence to support the validity of anthropometric survey quality indicators.

APPROACH 1

Type of research	Outcomes	Data source(s)
Mixed methods review: Conduct a landscape review to identify criteria or evidence used to support the adoption of conventional anthropometric survey quality indicators and research methods used to evaluate and validate indicators of survey quality. Given the lack of a gold-standard global measure of survey quality, 'validity' refers primarily to the extent to which an indicator discriminates among surveys with respect to one or more elements of survey quality (i.e., completeness, population representativeness, precision and/or accuracy of malnutrition estimates). The suitability of an indicator may also depend on how readily it can be feasibly calculated from routinely collected survey data. Findings from this review will be used to inform the development of an assessment, validation and indicator selection framework (see approach 2).	**Primary (1):** Summary of available evidence that supports that adoption (or rejection) of specific quality indicators. **Primary (2)**: Summary of research methods and criteria that have been used to assess and establish the validity of anthropometric survey quality indicators (or composite scores/indices).	Peer-reviewed literature, including relevant validation studies, review papers and commentaries. Technical documents by international agencies (e.g., United States Agency for International Development, JME Group) that describe methods used to select or justify the implementation of specific anthropometric survey quality indicators.

APPROACH 2

Type of research	Outcomes	Data source(s)
Qualitative: Use the Delphi method or another consensus-oriented process to convene experts and stakeholders to determine: a) the quality indicators for which sufficient evidence already exists to support (or reject) their use; b) criteria and methods by which candidate quality indicators (or composite scores/indices) may be evaluated to establish their validity in distinguishing low- versus high-quality surveys, or identifying surveys with manipulated or fabricated data.	**Primary:** Framework to guide the generation and interpretation of evidence to support the validity and selection of anthropometric survey quality indicators. Hierarchical list of criteria and assessment methods that may be used to accept (or reject) candidate survey quality indicators.	Questionnaires administered to a panel of technical experts and other stakeholders, combined with virtual or in-person discussions.

APPROACH 3

Type of research	Outcomes	Data source(s)
Secondary analyses: Conduct multi-survey analysis of recent survey datasets from low- and middle-income countries, applying the assessment and validation framework (per approach 2). Prioritize indicators for which sufficient supportive evidence is not already available (per approach 1). Generate indicator values for each survey. In the multi-survey dataset, estimate correlations among quality indicators, between indicators and estimates of malnutrition prevalence, and between indicators and measures of the precision/accuracy of estimates of malnutrition prevalence. Calculate average indicator values in sub-groups of surveys categorized based on inclusion/exclusion from JME estimates. Compare indicator distribution between surveys known/suspected of having quality concerns versus acceptable surveys.	**Primary and secondary**: Outcome selection will be based on the framework developed per approach 2. Potential outcomes include: between- and within-survey variability of the indicator; correlations among quality indicators; correlations between each indicator and malnutrition prevalence; correlation between each indicator and metrics of precision/accuracy of malnutrition prevalence estimates; accuracy of indicator in identifying outliers in country-level secular trends in malnutrition prevalence estimates; differences in indicator values between surveys that were included versus excluded from the JME (or other global modelling activities); sensitivity and specificity of the indicator for the identification of surveys suspected or known to be fraudulent or manipulated; availability of data required to routinely generate the indicator; availability of relevant country-specific benchmark data (e.g., age distribution).	Large-scale population-representative anthropometric surveys; administrative datasets, particularly those that include neonatal measurements.

Research topic 2

Properties of anthropometric z-score distributions in high-quality population-representative surveys conducted in diverse settings

APPROACH 1

Type of research	Outcomes	Data source(s)
Secondary analyses: Conduct multi-survey analysis of data from selected recent surveys that adhered to exceptionally high standards of quality and rigour in anthropometric data collection. Estimate statistical properties of the anthropometric z-score distributions in the raw datasets (without exclusion of any outlier/BIVs). Examine changes in the distributional properties over time within countries, particularly where comparisons can be made between surveys conducted before versus after documented implementation of measures to improve quality of data collection. Compare z-score distributional properties across surveys sub-grouped by other measures of survey quality (e.g., data completeness). Examine within-survey differences in z-score properties across subgroups of children categorized by demographic factors (e.g., location of residence, socioeconomic factors) and survey team/personnel.	**Primary (1):** Mean, median, standard deviation, kurtosis and skewness of anthropometric z-scores in each survey, and the distributions/ranges of each property across all surveys. **Primary (2):** Properties of z-score distributions within sub-groups of children categorized by geography (e.g., rural, urban) and socioeconomic status (e.g., wealth quintile), and by survey team/personnel. **Secondary (1):** Within-country (between-survey) differences in standard deviation, kurtosis and skewness. **Secondary (2):** Differences in average standard deviation, kurtosis and skewness between sub-groups of surveys categorized by other measures of survey quality (e.g., completeness). **Secondary (3):** Within-survey differences in standard deviation, kurtosis and skewness between sub-groups of children categorized by clinical or demographic factors, and survey team/personnel.	Selected large-scale population-representative anthropometric surveys that are considered exemplary in terms of documented quality control methods used during anthropometric data collection. Prioritize countries with multiple surveys, particularly those for which there have been documented efforts and activities to improve survey data quality in recent compared to older surveys.

APPROACH 2

Type of research	Outcomes	Data source(s)
Primary data collection: Conduct high-quality anthropometric surveys in diverse settings, implementing gold-standard measurement methods (e.g., paired independent measures for each child, rigorous quality control). Estimate statistical properties of the anthropometric z-score distributions, as described for approach 1. Record observations and insights from anthropometrists regarding potential low-quality measurements due to practical challenges (e.g., child unable to be properly positioned).	**Primary (1):** Mean, median, standard deviation, kurtosis and skewness of anthropometric z-scores in each survey and the distributions/ranges of each property across all surveys. **Primary (2):** Properties of z-score distributions within sub-groups defined by geography (e.g., rural, urban) and demographic factors (e.g., wealth quintile), and by survey team/personnel. **Secondary:** Properties of z-score distributions upon exclusion of datapoints for which anthropometrists recorded practical challenges during measurement procedures.	Focused surveys with gold-standard measurement methods.

APPROACH 3

Type of research	Outcomes	Data source(s)
Qualitative: Use a Delphi method or other consensus-oriented process to review evidence generated in studies described above (approaches 1 and 2) to determine which of the statistical properties of anthropometric z-score distributions (standard deviation, kurtosis, skewness) are suitable for use as indicators of survey quality.	**Primary:** Summary of discussions and decisions regarding the suitability of each statistical property (standard deviation, kurtosis, skewness) for use as an indicator of anthropometric survey quality.	Questionnaires administered to a panel of technical experts and other stakeholders, combined with virtual or in-person discussions.

Research topic 3

Optimal ranges and threshold values for survey quality indicators, and the role of qualitative appraisals

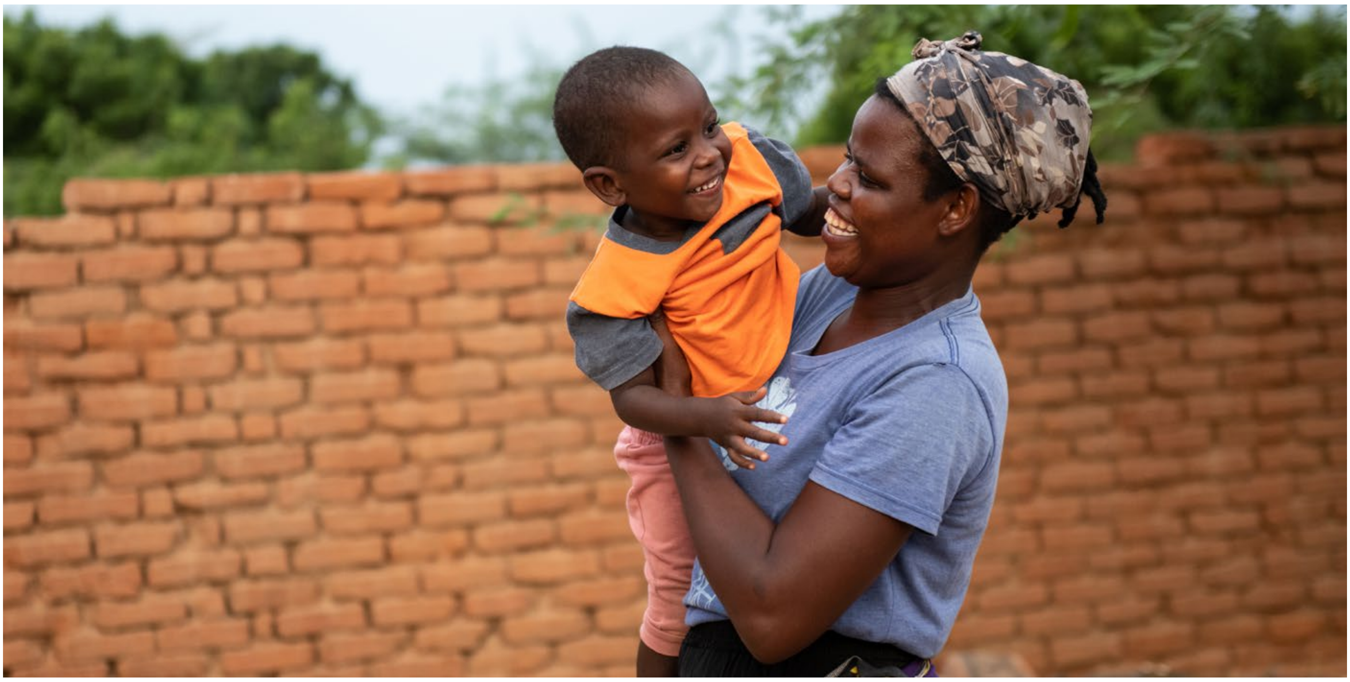
Type of research	Outcomes	Data source(s)
APPROACH 1		
Secondary analyses: Analyse selected surveys considered to have adhered to high standards of rigorous anthropometric data collection and quality control. Examine selected indicators considered suitable for use as quality metrics (based on evidence generated per research topics 1 and 2). Generate values for each indicator for each survey. In multi-survey analyses, define thresholds based on *a priori* defined percentiles indicating relatively low quality (i.e., <5th and/or >95th percentiles). Consider alternative data-driven thresholds based on the empirical distributions, across all surveys and in subgroups of surveys categorized by geographic region, year and other country characteristics. Setting thresholds based on normative reference ranges is proposed due to the lack of a global gold-standard measure of data quality against which indicator thresholds could be benchmarked.	**Primary (1):** Thresholds for categorizing values (e.g., low versus high) for each quality indicator, based on 5th and/or 95th percentiles. **Primary (2):** Alternative data-driven thresholds for low and/or high values for each indicator, based on shape and range of indicator distributions.	Selected large-scale population-representative anthropometric surveys that are exemplary in terms of rigorous anthropometric data collection and quality control methods
APPROACH 2		
Secondary analyses: Calculate indicators in available anthropometric surveys. Conduct multi-survey analyses; rank surveys and apply thresholds (per approach 1) to identify low-quality anthropometric datasets.	**Primary (1)**: Visualizations and descriptive statistical properties of each indicator distribution. **Primary (2):** Ranks of surveys and identification of low-quality surveys, by indicator.	Available large-scale population-representative anthropometric surveys.
APPROACH 3		
Qualitative: Present findings from studies conducted per approaches 1 and 2 to a panel of technical experts and other stakeholders. Assess perceived value and salience of the thresholds and rankings/ratings. Consider indicators for which qualitative/narrative assessments provide value in addition to (or in place of) rankings/classifications based on quantitative thresholds.	**Primary (1):** Summary of feedback from stakeholders regarding perceived value and salience of proposed thresholds. **Primary (2):** Identify indicators for which qualitative assessments may be used in addition to (or in place of) quantitative thresholds/ rankings.	Questionnaires administered to a panel of technical experts and other stakeholders, followed by a virtual or in-person discussion.

Research topic 4

Advantages and disadvantages of composite quality scores and indices		
APPROACH 1		
Type of research	**Outcomes**	**Data source(s)**
Scoping review: Conduct literature search to identify composite anthropometric quality scores and indices or other approaches that have been used to summarize information across multiple indicators.	**Primary:** List of composite scores/indices and summary of their prior use in published analyses.	Peer-reviewed and grey literature.
APPROACH 2		
Type of research	**Outcomes**	**Data source(s)**
Secondary analyses: Calculate quality indicators and composite scores/indices (per approach 1) in anthropometric survey datasets. Conduct analyses as described for research topic 1, approach 3, including descriptions of distributions of scores/indices, correlations and average scores/indices in sub-groups of surveys categorized based on inclusion/exclusion from JME estimates.	**Primary and secondary:** Selected outcomes similar to those proposed in research topic 1, approach 3, including: between- and within-survey variability of each score/index; correlations between scores/indices and individual quality indicators; sensitivity and specificity of the score/index for the identification of surveys excluded from JME analyses; etc.	Large-scale population-representative anthropometric survey datasets.
APPROACH 3		
Type of research	**Outcomes**	**Data source(s)**
Qualitative: Use a Delphi method or other consensus-oriented process to enable technical experts and stakeholders to review evidence generated per approaches 1 and 2; to assess advantages/disadvantages of composite scores/indices; and, to identify contexts in which composite scores/indices may be particularly useful for decision-making purposes.	**Primary:** Narrative summary and critical analysis of the advantages/disadvantages of candidate scores/indices or descriptions of specific use cases.	Questionnaires administered to a panel of technical experts and other stakeholders and virtual or in-person discussions.

Research topic 5

New indicators and methods, including techniques to detect data manipulation or fabrication		
APPROACH 1		
Type of research	**Outcomes**	**Data source(s)**
Mixed methods review: Conduct a landscape review of the literature and interviews with experts in data forensics and research fraud to identify candidate novel quality indicators and techniques for detection of data manipulation and fabrication in anthropometric surveys.	**Primary (1):** Inventory of candidate indicators and methods. **Primary (2):** Narrative summary and critical analysis of the advantages and disadvantages of candidate indicators/methods.	Peer-reviewed and grey literature. Key informant interviews including technical experts in methods of detecting research fraud.
APPROACH 2		
Type of research	**Outcomes**	**Data source(s)**
Secondary analyses: Implement novel indicators and candidate fraud detection techniques (per approach 1) in anthropometric survey datasets. Conduct analyses as described for research topic 1, approach 3. Identify surveys with potential data manipulation or fabrication. Compare results across different methods, and correspondence of results with conventional quality indicators. Examine characteristics of surveys with data found to be potentially manipulated or fabricated.	**Primary:** Outcomes used in research topic 1, approach 3. **Secondary (1):** Number (proportion) of surveys identified as containing potentially manipulated or fabricated data. **Secondary (2):** Characteristics of surveys identified as containing potentially manipulated or fabricated data.	Large-scale population-representative anthropometric surveys.



Research roadmap

Despite increasingly common use of measures and indicators of anthropometric survey quality, there is a lack of evidence to establish the validity of specific indicators or define thresholds for each indicator that enable identification of relatively poor-quality surveys. The proposed research agenda aims to address topics of uncertainty, including the use of statistical properties of anthropometric z-score distributions (e.g., standard deviation) as indicators of data quality, the value of composite data quality scores/indices, and the adoption of new methods specifically focused on detecting fraudulent data. A roadmap towards the establishment of global guidance on the assessment and reporting of anthropometric data quality is presented below, along with the input needed from key stakeholders to address different aspects of the research agenda.

1. CREATE AND APPLY
a framework for assessing, validating and selecting indicators of anthropometric survey quality
Who:
Researchers, anthropometry experts, survey implementers

2. DETERMINE
which statistical properties of anthropometric z-score distributions are suitable for use as indicators of survey quality
Who:
Researchers, anthropometry experts

3. DEFINE
the optimal ranges and threshold values for survey quality indicators
Who:
Researchers, anthropometry experts

4. ASSESS
the advantages and disadvantages of composite scores/indices of anthropometric survey quality
Who:
Researchers, anthropometry experts

5. IDENTIFY
novel indicators and methods to assess the quality of anthropometric surveys, including techniques to detect data manipulation or fabrication
Who:
Researchers, anthropometry experts, survey implementers

6. ESTABLISH
standards and develop global guidance for anthropometric survey quality assessment that can be disseminated for wide-scale use.
Who:
UNICEF, WHO and the Technical Expert Advisory Group on Nutrition Monitoring

References cited

1. World Health Organization and the United Nations Children's Fund (UNICEF). Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old. World Health Organization and the United Nations Children's Fund (UNICEF); 2019.
2. Grellety E, Golden MH. Change in quality of malnutrition surveys between 1986 and 2015. Emerg Themes Epidemiol 2018;15:8.
3. Perumal N, Namaste S, Qamar H, Aimone A, Bassani DG, Roth DE. Anthropometric data quality assessment in multisurvey studies of child growth. The American journal of clinical nutrition 2020;112:806S–15S.
4. Assaf S, Kothari MT, Pullum T. An assessment of the quality of DHS anthropometric data, 2005-2014. DHS Methodological Reports No. 16. Rockville, Maryland, USA: ICF; 2015.
5. Bilukha O, Couture A, McCain K, Leidman E. Comparison of anthropometric data quality in children aged 6-23 and 24-59 months: lessons from population-representative surveys from humanitarian settings. BMC nutrition 2020;6:60.
6. Corsi DJ, Perkins JM, Subramanian SV. Child anthropometry data quality from Demographic and Health Surveys, Multiple Indicator Cluster Surveys, and National Nutrition Surveys in the West Central Africa region: are we comparing apples and oranges? Glob Health Action 2017;10:1328185.
7. Allen CK, Croft TN, Pullum TW, Namaste SML. Evaluation of Indicators to Monitor Quality of Anthropometry Data during Fieldwork. DHS Working Paper No. 162. Rockville, Maryland, USA: ICF; 2019.
8. "Physical Status: The use and interpretation of anthropometry." Report of a WHO Expert Committee, 1995. World Health Organization; 1995.
9. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 1999;82:165-77.
10. Aimone A, Bassani DG, Qamar H, Perumal N, Namaste S, Roth DE. Alternative and Complementary Metrics of Linear Growth for Tracking Global Progress in Child Nutritional Status. DHS Working Paper No. 153. Rockville, Maryland, USA: ICF; 2021.
11. SMART: Standardized Monitoring and Assessment for Relief and Transitions. Manual 2.0.: SMART, Action Against Hunger Canada, and the Technical Advisory Group.; 2017.
12. Crowe S, Seal A, Grijalva-Eternod C, Kerac M. Effect of nutrition survey 'cleaning criteria' on estimates of malnutrition prevalence and disease burden: secondary data analysis. PeerJ 2014;2:e380.
13. Mei Z, Grummer-Strawn LM. Standard deviation of anthropometric Z-scores as a data quality assessment tool using the 2006 WHO growth standards: a cross country analysis. Bull World Health Organ 2007;85:441-8.
14. Grellety E, Golden MH. The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition. PLoS One 2016;11:e0168585.
15. Dwivedi LK, Banerjee K, Sharma R, et al. Quality of anthropometric data in India's National Family Health Survey: Disentangling interviewer and area effect using a cross-classified multilevel model. SSM Popul Health 2022;19:101253.
16. Roth DE, Krishna A, Leung M, Shi J, Bassani DG, Barros AJD. Early childhood linear growth faltering in low-income and middle-income countries as a whole-population condition: analysis of 179 Demographic and Health Surveys from 64 countries (1993–2015). The Lancet Global Health 2017;5:e1249-e57.
17. Perumal N, Roth DE, Perdrizet J, et al. Effect of correcting for gestational age at birth on population prevalence of early childhood undernutrition. Emerg Themes Epidemiol 2018;15:3.
18. Leroy JL, Ruel M, Habicht JP, Frongillo EA. Linear growth deficit continues to accumulate beyond the first 1000 days in low- and middle-income countries: global evidence from 51 national surveys. J Nutr 2014;144:1460-6.
19. Pullum TW, Allen C, Namaste SML, Croft T. Sensitivity of anthropometric estimates to errors in the measurement of height, weight, and age for children under five in population-based surveys. DHS Methodological Reports No. 28. Rockville, Maryland, USA: ICF; 2020.
20. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development. World Health Organization, 2006. Available at: https://www.who.int/tools/child-growth-standards.
21. Benedict R, Namaste SML, Croft T. Evaluation of Implausible Anthropometric Values by Data Collection Team in Demographic and Health Surveys 2010–2020. DHS Methodological Reports No. 33. Rockville, Maryland, USA: ICF; 2022.
22. Larsen AF, Headey D, Masters WA. Misreporting Month of Birth: Diagnosis and Implications for Research on Nutrition and Early Childhood in Developing Countries. Demography 2019;56:707-28.
23. United Nations Children's Fund (UNICEF), World Health Organization, International Bank for Reconstruction and Development/The World Bank. Levels and trends in child malnutrition: UNICEF / WHO / World Bank Group Joint Child Malnutrition Estimates. Key findings of the 2025 edition. Geneva: World Health Organization; 2025. Licence: CC BY-NC-SA 3.0 IGO.
24. World Health Organization, United Nations Children's Fund (UNICEF), The World Bank. Technical Notes from the background document for country consultations on the 2021 edition of the UNICEF-WHO-World Bank Joint Malnutrition Estimates. 2021. at https://data.unicef.org/resources/jme-2021-country-consultations/)

If interested in joining this effort

or if you have any questions or comments, please contact the TEAM Working Group on Anthropometric Data Quality at:

nfsdata@who.int | sdg2.2@unicef.org

unicef | World Health Organization