Outliers and biologically-implausible values in population-based anthropometric surveys

Statement of problem

Successful implementation of anthropometric surveys depends on skilled field personnel, standardized data collection equipment and methods, and rigorous quality control practices to recognize and remediate potentially problematic data in real-time, as detailed by the World Health Organization (WHO) and United Nations Children's Fund (UNICEF) in the 2019 *Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old.*[1] However, even strict adherence to best practices in the field is unlikely to mitigate all errors in the collection and recording of each participant's age, weight, height and other anthropometric measures. Therefore, as with all real-world datasets, it is recommended that anthropometric survey data are explored and cleaned before the data are used to estimate nutritional status indicators, such as the prevalence of stunting, wasting and overweight.[1]

A core component of anthropometric data cleaning is the identification of outliers,[*] which are datapoints sufficiently distant from the rest of the observed or expected distribution that they warrant consideration as potential errors.[2] Outliers in anthropometric z-score distributions are concerning because they can exert disproportionate influence on statistical summary measures, including mean and standard deviation (SD). They are particularly relevant to prevalence estimates of undernutrition and overweight because extreme values retained in the final dataset will be included when estimating the proportion of children above or below the z-score thresholds that define the nutritional status indicators.

Some extreme outliers can be readily attributed to data recording errors, particularly when values differ from expectations by orders of magnitude (e.g., length of "9.0 cm" rather than "90 cm"). However, the adoption of electronic data capture systems and real-time quality control mechanisms (e.g., range and consistency checks) can prevent many overt errors, such that most extreme values that remain in the dataset at the analysis stage tend to be arrayed near the fringes of the expected or observed range rather than separated from it by discontinuities in the tails of the distribution. The uncertain distinction between erroneous values and valid but extreme datapoints is a challenge in anthropometric data management, particularly in cross-sectional surveys where outliers cannot be identified based on deviations from child-specific growth trajectories.[3] The 2019 WHO/UNICEF *Recommendations* include guidance regarding outlier identification and exclusion, the goal of which is to generate a final dataset that produces more accurate estimates of malnutrition indicators than would be obtained with the uncleaned dataset.

[*] This brief focuses on the identification of outliers in the context of data cleaning prior to analysis and estimation of indicators of population nutritional status. Other considerations related to outliers are addressed in other briefs: outlier tracking during data collection to prompt repeat measurements is covered in brief 6 and the proportion of observed values defined as outliers or biologically-implausible values as an indicator of survey quality is discussed in brief 8.

However, when there are substantial numbers of extreme values, different methods to identify and exclude outliers can yield disparate prevalence estimates.[3,4] Given the debate and uncertainty about best practices in outlier identification, and the lack of an existing gold-standard method,[1] the development and validation of a standardized method to identify outliers remains a research priority.

Among the wide variety of candidate methods for outlier detection in cross-sectional quantitative datasets, the most common is fence-labelling, which imposes upper and lower bounds of an acceptable range of data.[5] Fence locations may be based on the statistical characteristics of the observed sample distribution or determined by external information (e.g., reference interval). The most widely-known fence-labelling based on statistical properties of a sample is Tukey's box-and-whisker plot (or 'boxplot'), which highlights outliers beyond 1.5-times the interquartile range from the median, although numerous variations on Tukey's method have been proposed.[5,6] For human physical characteristics, fences are often based on natural limits (in size or growth) or normative ranges beyond which observations are considered to be biologically-implausible values (BIVs).[3]

Two types of fence-labeling were defined by the 1995 WHO Expert Committee on anthropometry to identify and exclude z-score values "that are most likely to represent errors"[7], as elaborated in the 2019 WHO/UNICEF *Recommendations:*[1]

- *Fixed exclusion ranges*, defined by lower and upper bounds of z-scores that are universally applied to all surveys, and which are intended to remove BIVs.
- *Flexible exclusion ranges*, for which the fences are placed at defined distances (in z-score units) from the survey-specific mean z-score, and which are intended to remove values with very low probabilities of being true in the observed population, even if they would not be considered BIVs.

The thresholds of fixed and flexible exclusion ranges (and hence the range of included values) have been conventionally defined using the z-score scale of the WHO Child Growth Standards rather than the empirical SDs of observed sample distributions.[1] Fixed criteria are applied uniformly to every survey, irrespective of the location, shape or dispersion of its empirical distribution of z-scores relative to the WHO Child Growth Standards. Conversely, use of flexible criteria assumes that the survey-specific mean z-score is valid and relevant but does not inherently account for the empirical SD of z-scores, so the size of a standard z-score unit may not equal an empirical SD unit. For example, if the survey-specific SD is >1.0, a fence placed at +4 standard z-scores from the mean would be less than 4 survey-specific SD units above the mean. Anchoring the exclusion range on the survey-specific mean acknowledges that a population mean z-score lower than 0 generally reflects a downward shift in the entire z-score distribution, such that extremely low values are more likely to be true in the observed population than they would be in a healthy reference population with a mean z-score of 0.

Several sets of fixed and flexible criteria have been used to clean anthropometric datasets in the past several decades (**Table 1**).[4,7] These criteria are often described as 'flagging' and the thresholds referred to as 'flags' because software packages used to derive z-scores from raw anthropometric data automatically flag the z-score values that meet selected exclusion criteria.[8,9,*] The most widely used criteria apply thresholds to each anthropometric parameter separately (Table 1),[8,9] although within-child inconsistencies between indices of height and ponderal growth (e.g., height-for-age z-score [HAZ] >3.09 and weight-for-height z-score [WHZ] <-3.09) were also previously flagged in anthropometric survey analysis software developed jointly by WHO and the United States Centers for Disease Control and Prevention (CDC),[10] and have been used in data cleaning in longitudinal studies.[11,12] In 2019, WHO and UNICEF jointly recommended a fixed exclusion range based on the 2006 WHO Child Growth Standards (WHO 2006) as an essential data cleaning procedure prior to estimating nutritional status indicators and some data quality metrics.[1**] However, since 2002, many international agencies and national governments have adopted a flexible exclusion range that was developed and supported by the Standardized Monitoring and Assessment of Relief and Transitions (SMART) programme (Table 1).[8] The SMART method locates fences at +/-3 standard z-score units from the survey-specific mean, but permits increasing the range to +/-3.5 z-scores when data quality is high.[8]

Concerns have been raised about the validity of both fixed and flexible exclusion ranges. The WHO-2006 criteria were adapted from the National Center for Health Statistics (NCHS)/WHO growth references based on expert opinion, rather than validation studies that established empirical limits of biological plausibility. Furthermore, the WHO-2006 growth standards were based on healthy term-born children,[13] but in practice, are applied to diverse survey populations that include children born preterm and children with chronic health conditions and disabilities. Studies in the United States using longitudinal outlier detection methods[11,14] and high-quality survey datasets

* In this brief in which data cleaning procedures are addressed, we use the terms outliers and BIVs according to their common usage in the epidemiological literature. The term 'flagging' is preferentially used to refer to the real-time identification of potential errors during data collection, which may prompt re-measurement (addressed in Brief 6).

** In conventional data quality assessments, outliers/BIVs are excluded prior to estimating standard deviation, kurtosis and skewness; however, outliers/BIVs (and the children contributing such values) are retained in the dataset used for other data quality indicators (see Brief 8 for more details regarding quality indicators).

(in which extreme values were verified in the field)[15] have demonstrated that even with permissive (wide) ranges of acceptable values, a substantial proportion of values labelled as BIVs by cross-sectional fixed-fence methods are likely valid, especially in the upper tails of the weight-for-age (WAZ) and BMI-for-age z-score distributions.[3,15] Such findings prompted the CDC to increase the upper thresholds in 2016 (**Table 1**).[16]

The SMART flexible method is comparatively stringent, leading to higher proportions of values identified as outliers (and thus conventionally excluded from analyses) compared to fixed methods, such as those using the WHO-2006 criteria.[4,17,18] The proportion of values considered outliers by either method is correlated with the empirical SD.[17] Therefore, the extent to which SMART and WHO-2006 criteria yield disparate malnutrition prevalence estimates is related to the magnitude of over-dispersion of the survey dataset compared to the standard SD of 1 (based on the WHO growth standards).[4] The rationale for using the SMART +/-3 z-score cut-offs to identify outliers is that given a normal distribution with an empirical SD of 1, a very low proportion of true values would be expected to be excluded (~0.3%), thereby sacrificing a negligible number of true values while removing errors that are extreme and which may distort measures of the z-score distribution (e.g., mean, SD).[8] However, over-dispersion (SD>1), excess kurtosis and skewness are commonly observed in real-world survey datasets,[19,20] such that a non-negligible proportion of true values are likely excluded when SMART criteria are applied. It is unknown if the SMART method leads to a more accurate estimation of nutritional status indicators compared to other methods, but because trimming at +/-3 z-scores from the survey-specific mean generates cleaned datasets with distributions that are more symmetrical and less dispersed than the raw distributions or those resulting from conventional fixed exclusion ranges, users of the method may conclude that the SMART method yields higher-quality data.[21] However, this is uncertain, because the extent to which such features of the z-score distribution (e.g., skewness, kurtosis) indicate poor data quality remains unknown.[1,19] *

Table 1. Fixed and flexible outlier identification thresholds recommended for anthropometric survey data cleaning

	WHO 1995[a]		**WHO 2006**[b]	**CDC 2016**[c]	**SMART**
Method type	Fixed	Flexible[d]	Fixed	Fixed	Flexible[d,e]
HAZ	< -5 and >+3	<-4 and >+4[f]	<-6 and >+6	<-5 and >+4	<-3 and >+3
WHZ	< -4 and >+5	<-4 and >+4	<-5 and >+5	<-4 and >+8	<-3 and >+3
WAZ	<-5 and >+5	<-4 and >+4	<-6 and >+5	<-5 and >+8	<-3 and >+3
BMIZ	NA	<-4 and >+4	<-5 and >+5	<-4 and >+8	NA

HAZ, height-for-age z-score; WHZ, weight-for-height z-score; WAZ, weight-for-age z-score; BMIz, body-mass-index-for-age z-score.

a From the 1995 WHO Expert Committee;7 provided for historical context, but not currently recommended for use.[1]

b Approach endorsed by the 2019 WHO/UNICEF Recommendations.[1]

c From: https://www.cdc.gov/growth-chart-training/hcp/computer-programs/sas.html?CDC_AAref_Val=https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm

d Thresholds represent distances in standard z-score units from the survey-specific mean.

e According to the SMART manual, thresholds may be increased to +/-3.5 if there are high-quality data.8

f Up to a maximum HAZ of +37.

* The utility of the statistical properties of z-score distributions as indicators of survey data quality is addressed in Brief 8.

Outlier/BIV detection ultimately entails a trade-off between error mitigation and exclusion of valid data. Simulations have shown that both fixed and flexible methods tend to have low sensitivity for errors in general, meaning that most errors are not outliers.[12] Outlier/BIV exclusion is therefore unlikely to offer a satisfactory solution for mitigation of random errors due to measurement imprecision throughout the entire distribution.[22]

Moreover, outlier/BIV exclusion could lead to biased estimates of malnutrition indicators if extreme values are disproportionately contributed by children at high risk of undernutrition (e.g., younger infants, children born preterm or small-for-gestational age, or those with disabilities) and children with severe overweight. Use of thresholds applied symmetrically around the sample mean (e.g., SMART criteria) may be particularly problematic in settings with a double burden of malnutrition; for example, the WHZ distribution may be shifted down to the extent that the upper fence (e.g., +3 z-scores above the survey-specific mean) would be close to, or even below, the WHO/UNICEF overweight cut-off of WHZ=2, resulting in an artifactually low prevalence of overweight in the cleaned dataset.

Standardized and harmonized anthropometric data cleaning practices are a priority because nutritional status indicators from multiple survey types are routinely compared within and across countries and regions, and compiled in global databases such as the UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates.[23] However, while different methods for outlier/BIV exclusion may yield incongruent malnutrition prevalence estimates, it is unknown which method yields the most accurate estimates. There is also concern about the extent to which some methods may lead higher-risk groups of children to be inequitably excluded from cleaned datasets, even if such exclusions have a relatively minor effect on overall population malnutrition prevalence estimates. A relatively limited set of fence-labelling methods have been used to date, and the variations among them relate primarily to the specific location of the fences, whereas methods of outlier detection adapted from other scientific disciplines, including machine learning methods,[24] have been virtually unexplored. Conventional fixed and flexible methods consider each anthropometric metric separately, whereas alternative methods could use multivariable modelling to incorporate other child-level health or demographic data available in the survey.

This brief proposes a research agenda for the assessment and comparison of candidate methods to identify outliers and BIVs in population-based anthropometric surveys.* The goal of this research is to identify an outlier/BIV detection and exclusion method that optimizes the accuracy of estimates of the prevalence of various forms of malnutrition but does not lead to the inequitable exclusion of data from medically or socially identifiable sub-groups of children.

Research questions

1	**What characteristics of children or data collection procedures are associated with higher risks of extremely low and high values in anthropometric datasets across diverse settings?**
2	**What outlier detection methods used for other types of population health data or in other scientific disciplines may be applicable to anthropometric survey data cleaning?**
3	**What are the effects of conventional (fixed and flexible) and other candidate outlier/BIV detection methods on estimates of the prevalence of undernutrition and overweight and other statistical summary measures of z-score distributions (e.g., mean, median, standard deviation) in population-based anthropometric surveys, across diverse settings and with variable levels of survey data quality? Where gold-standard estimates are available or can be generated in simulation studies, which outlier/BIV methods result in the most accurate estimates of the prevalence of stunting, wasting and overweight?**
4	**What proportion of anthropometric measurements are identified as outliers or BIVs by each candidate approach, overall and in medically or socially identifiable sub-groups of children (e.g., children born preterm, children with disabilities), across diverse settings?**

* Consideration of outliers/BIVs in this brief assumes two steps involving data cleaning (including outlier/BIV identification) followed by data analysis. Conventional practice is to exclude outliers/BIVs from analyses to estimate malnutrition indicators but to retain them when generating estimates of most data quality indicators (see Brief 8). When making comparisons among different outlier/BIV methods, it is assumed that a consistent analytical method is applied to the cleaned datasets from which outliers/BIVs have been excluded. However, the effect of outlier/BIV exclusion on malnutrition prevalence estimates may be modified by the choice of statistical method. Some methods may be relatively robust even when outliers are retained. The topic of statistical modelling to generate malnutrition indicators will be addressed in Brief 9.

Research topic 1

Characteristics of children or data collection procedures associated with higher risks of extremely low and high values across diverse settings.

APPROACH 1

Type of research	Outcomes	Data source(s)
Secondary analyses: Conduct a multi-survey analysis of survey datasets to identify the characteristics of children (including parental and household factors) and/or data collection procedures that are associated with extremely low and/or high positions in the anthropometric z-score distributions. Using regression models (or analogous approaches), estimates (e.g., relative risks) will reflect the magnitude/direction of associations of factors (alone or in combinations) with low or high percentile rankings (on a continuous scale) or with categorization of values as extremely low/high values or outliers/BIVs. Conduct analyses among surveys stratified by country characteristics (e.g., world region, economic classification), nutritional status of the survey population (e.g., prevalence of stunting), and indicators of survey quality.	**Primary (1):** Percentile ranking of individual children's z-scores for each index, within each survey population. **Primary (2):** Categorization of individual children's values as "extremely low" (<5th percentile) or "extremely high" (>95th percentile) within each survey population. **Secondary (1):** Extremely low or high individual child values defined more stringently than in primary analyses, (e.g., <1st and >99th percentiles of each survey population). **Secondary (2):** Individual child z-score values classified as low or high BIVs based on WHO-2006 definitions or conventional SMART criteria.	Large-scale population-representative anthropometric surveys from low- and middle-income countries. Highest-priority surveys are those that include raw data (e.g., observations for children who were eligible but not measured) and child-level clinical and physical characteristics, such as preterm/term status and disabilities. Surveys may also be prioritized if there are metadata related to anthropometric measurement procedures, personnel or team identifiers linked to each measurement, and indicators of personnel skill level based on training and standardization exercises (e.g., technical error of measurement). Longitudinal studies with data related to gestational age at birth, small-for-gestational age, and clinical information relevant to the determination of presence/absence of physical disabilities.

APPROACH 2

Type of research	Outcomes	Data source(s)
Primary data collection: Conduct high-quality anthropometric surveys in diverse settings, targeting representation of children who may disproportionately contribute outliers/BIVs (e.g., preterm-born children). Implement gold-standard measurement methods (e.g., paired independent measures for each child, rigorous field-level quality control) and include indicators of data collection procedures and experiences of anthropometrists (e.g., concerns about discordance between reported age and child measurements), and performance characteristics of each measurer (e.g., technical error of measurement). As described above for approach 1, effect estimates (e.g., relative risks) will reflect the magnitude/direction of associations of child-, encounter- or anthropometrist-related factors (alone or in combinations) with low or high percentile rankings (on a continuous scale) or with categorization of values as extremely low/high values or outliers/BIVs.	**Primary:** Percentile ranking of individual children's z-scores for each index, within each survey population. Other primary and secondary outcomes as listed for approach 1.	Small-scale surveys with adequate representation of at-risk children, collection of measurement-related metadata and gold-standard anthropometric measurement methods.

Research topic 2

Outlier detection methods used in other fields.

APPROACH 1

Type of research	Outcomes	Data source(s)
Secondary analyses: Conduct multi-survey statistical analyses of anthropometric survey and/or administrative datasets. Generate estimates of the prevalence of malnutrition and other statistical summary measures of z-score distributions in raw datasets (no outlier exclusion), and following application of several candidate methods to exclude outliers/BIVs, including: a) WHO and SMART approaches as currently implemented; b) fixed and flexible approaches that incorporate z-score cut-offs other than those currently used by WHO and SMART; and, c) alternative or novel methods identified in activities related to research topic 2 (see above). Consider approaches that incorporate values from >1 anthropometric variable for the same child to identify BIVs. Conduct analyses in groups of surveys stratified by world region, economic classification and data quality (e.g., completeness of data).	**Primary:** Prevalence of stunting, wasting, overweight and obesity. **Secondary:** Summary measures of the z-score distributions (HAZ, WAZ, WHZ and BMIZ), including mean, median, standard deviation, kurtosis, skewness.	Large-scale population-representative anthropometric surveys from low- and middle-income countries. Administrative databases, particularly those that include newborn measurements.

APPROACH 2

Type of research	Outcomes	Data source(s)
Qualitative: Use the Delphi method to undertake a consensus-oriented discussion and decision-making process to identify key characteristics of promising outlier detection methods. Generate a list of specific high-priority methods that may be applied in analyses to address research topic 3.	**Primary (1):** Key characteristics of outlier detection methods that indicate their potential suitability for application to anthropometric datasets. **Primary (2):** List of selected outlier identification methods that should be prioritized for application to anthropometric survey data cleaning.	Questionnaires administered to a panel of technical experts and other stakeholders in two or more rounds, followed by a virtual or in-person discussion.

Research topic 3

Effects of outlier/BIV detection methods on estimates of the prevalence of undernutrition and overweight and other statistical summary measures of z-score distributions; and, determination of the outlier/BIV methods that generate the most accurate estimates of the prevalence of stunting, wasting and overweight.

Type of research	Outcomes	Data source(s)
APPROACH 1		
Mixed method review: Conduct a landscape assessment to identify all possible mitigation or adjustment approaches to address interference on height/length and/or weight measurements.	**Primary:** Identification of mitigation and adjustment approaches.	A systematic search of peer review and grey literature. An online survey and/or key informant interviews with stakeholders.
APPROACH 2		
Simulation: Create simulated anthropometric datasets without BIVs/outliers (gold-standards) that reflect z-score distributions commonly observed in low- and middle-income countries. Introduce outliers/BIVs in each dataset by applying random and/or systematic errors. Apply candidate outliers/BIV detection methods to original and modified datasets. Compare estimates of the prevalence of malnutrition and other statistical summary measures of the z-score distributions among the original (outlier/BIV-free) datasets, simulated/modified datasets prior to outlier/BIV exclusion, and datasets following application of each candidate outlier/BIV exclusion method (per approach 1).	**Primary:** Differences between estimates of prevalence of malnutrition for each outlier/BIV detection method versus gold-standard. **Secondary:** Differences between estimates of summary measures of the z-score distributions for each outlier/BIV detection method versus the unmodified (gold-standard) dataset.	Simulated anthropometric datasets.
APPROACH 3		
Primary data collection: Conduct high-quality anthropometric surveys across diverse settings, using gold-standard measurement methods (e.g., paired independent measures for each child), as described for research topic 1, approach 2. Simulate outliers and BIVs by introducing random and/or systematic errors that modify a subset of real values. Conduct analyses as described above for approach 2.	**Primary:** Differences between estimates of prevalence of malnutrition for each outlier/BIV detection method versus unmodified dataset. **Secondary:** Differences between estimates of summary measures of the z-score distributions for each outlier/BIV detection method versus unmodified dataset.	Small-scale surveys with gold-standard anthropometric measurement methods (e.g., paired independent measures for each child).

Research topic 4

Proportion of anthropometric measurements identified as outliers or BIVs by each candidate method, overall and in medically or socially identifiable sub-groups of children

Type of research

Secondary analyses: Conduct multi-survey statistical analyses of existing large-scale anthropometric survey and/or administrative datasets from low- and middle-income countries. Stratify survey participants using identifiers of medical or social vulnerability compared to other children in the same survey population, including preterm birth, physical disabilities, developmental delays, low socioeconomic status, geographic location of residence (e.g., remote or rural). Apply selected candidate outlier/BIV detection methods, based on results of studies per research topic 3. Estimate associations between medical or social factors and risk of outlier/BIVs, by outlier/BIV detection method.

Outcomes

Primary (1):
Number/proportion of values labelled as low or high outliers/BIVs by each outlier/BIV method, overall and in medically or socially identifiable sub-groups of children.

Primary (2):
Relative risks of low or high outliers/BIVs by each outlier/BIV method in medically or socially identifiable sub-groups of children, compared to children without medical or social risk factors.

Secondary (1):
Associated characteristics of low or high outliers/BIVs (e.g., extreme length, implausible age, reported difficulty in completing measurement procedure) among children with low or high outliers/BIVs by each method, overall and in medically or socially identifiable sub-groups of children.

Secondary (2):
Number/proportion of children with missing or incomplete anthropometric z-score data, overall and in medically or socially identifiable sub-groups of children.

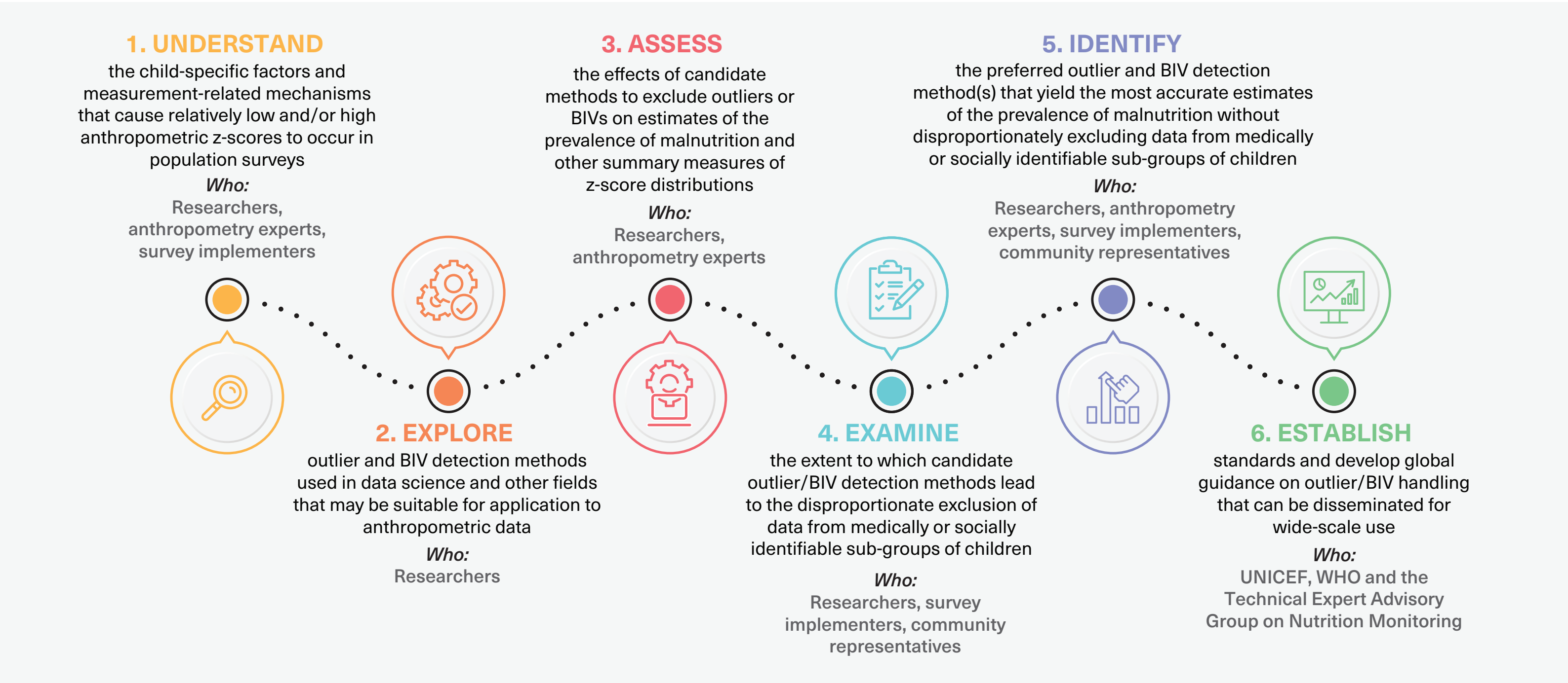
Data source(s)

Large-scale population-representative anthropometric surveys from low- and middle-income countries that include child-level clinical and physical characteristics, such as preterm/term status and disabilities (as described for research topic 1).

Administrative databases, particularly those that include newborn measurements.

Research roadmap

There is debate and uncertainty about the optimal outlier/BIV detection methods that most efficiently enable exclusion of extreme anthropometric z-score values that are likely to be errors, without disproportionately leading to the exclusion of data from medically or socially identifiable subgroups of children. The proposed research agenda aims to establish a new evidence base for the selection of preferred outlier/BIV detection methods to be routinely applied to large-scale anthropometric survey datasets from low- and middle-income countries. A roadmap towards the establishment of global guidance on this topic is presented below, along with the input needed from key stakeholders to address different aspects of the research agenda.



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). (https://www.who.int/publications/i/item/9789241515559).
2. Van den Broeck J, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. PLoS Med 2005;2(10):e267. DOI: 10.1371/journal.pmed.0020267.
3. Lawman HG, Ogden CL, Hassink S, Mallya G, Vander Veur S, Foster GD. Comparing Methods for Identifying Biologically Implausible Values in Height, Weight, and Body Mass Index Among Youth. Am J Epidemiol 2015;182(4):359-65. DOI: 10.1093/aje/kwv057.
4. 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. DOI: 10.7717/peerj.380.
5. Dastjerdy B, Saeidi A, Heidarzadeh S. Review of Applicable Outlier Detection Methods to Treat Geomechanical Data. Geotechnics 2023;3(2):375-396. (https://www.mdpi.com/2673-7094/3/2/22).
6. Jones PR. A note on detecting statistical outliers in psychophysical data. Atten Percept Psychophys 2019;81(5):1189-1196. DOI: 10.3758/s13414-019-01726-3.
7. "Physical Status: The use and interpretation of anthropometry." Report of a WHO Expert Committee, 1995. World Health Organization. (https://apps.who.int/iris/handle/10665/37003).
8. SMART: Standardized Monitoring and Assessment for Relief and Transitions. Manual 2.0. SMART, Action Against Hunger Canada, and the Technical Advisory Group. (https://smartmethodology.org/survey-planning-tools/smart-methodology/).
9. WHO Anthro for personal computers, version 3.2.2, 2011: Software for assessing growth and development of the world's children. WHO. (http://www.who.int/childgrowth/software/en/).
10. Sullivan KM, Gorstein J. Anthro: Software of Calculating Anthropometry. Version 1.02. 29 June 1999. Center for Disease Control and Prevention & World Health Organization.
11. Boone-Heinonen J, Tillotson CJ, O'Malley JP, et al. Not so implausible: impact of longitudinal assessment of implausible anthropometric measures on obesity prevalence and weight change in children and adolescents. Ann Epidemiol 2019;31:69-74 e5. DOI: 10.1016/j.annepidem.2019.01.006.
12. Shi J, Korsiak J, Roth DE. New approach for the identification of implausible values and outliers in longitudinal childhood anthropometric data. Annals of Epidemiology 2018;28(3):204-211.e3.
13. World Health Organization. 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. Geneva: World Health Organization, 2006.
14. Freedman DS, Lawman HG, Pan L, et al. The prevalence and validity of high, biologically implausible values of weight, height, and BMI among 8.8 million children. Obesity (Silver Spring) 2016;24(5):1132-9. DOI: 10.1002/oby.21446.
15. Freedman DS, Lawman HG, Skinner AC, McGuire LC, Allison DB, Ogden CL. Validity of the WHO cutoffs for biologically implausible values of weight, height, and BMI in children and adolescents in NHANES from 1999 through 2012. Am J Clin Nutr 2015;102(5):1000-6. DOI: 10.3945/ajcn.115.115576.
16. The SAS program for CDC growth charts that includes the extended BMI calculations (updated January 9, 2023). Centers for Disease Control and Prevention (CDC) (https://www.cdc.gov/growth-chart-training/hcp/computer-programs/sas.html).
17. 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.
18. Harkare HV, Corsi DJ, Kim R, Vollmer S, Subramanian SV. The impact of improved data quality on the prevalence estimates of anthropometric measures using DHS datasets in India. Sci Rep 2021;11(1):10671. DOI: 10.1038/s41598-021-89319-9.
19. 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.
20. 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(1):1328185. DOI: 10.1080/16549716.2017.1328185.
21. Grellety E, Golden MH. Change in quality of malnutrition surveys between 1986 and 2015. Emerg Themes Epidemiol 2018;15:8. DOI: 10.1186/s12982-018-0075-9.
22. Ulijaszek SJ, Kerr DA. Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 1999;82(3):165-77. DOI: 10.1017/s0007114599001348.
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. (https://data.unicef.org/resources/jme/).
24. Estiri H, Klann JG, Murphy SN. A clustering approach for detecting implausible observation values in electronic health records data. BMC Medical Informatics and Decision Making 2019;19(1):142. DOI: 10.1186/s12911-019-0852-6.

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.