

Considerations for assessment of measurement quality of mid-upper arm circumference data in anthropometric surveys and mass nutritional screenings conducted in humanitarian and refugee settings

Oleg Bilukha  | Behzad Kianian

Emergency Response and Recovery Branch,
Division of Global Health Protection, Center
for Global Health, Centers for Disease
Control, Atlanta, Georgia, USA

Correspondence

Oleg Bilukha, Emergency Response and
Recovery Branch, Division of Global Health
Protection, Center for Global Health, Centers
for Disease Control, 1600 Clifton Rd, Atlanta,
GA 30329, USA.
Email: obb0@cdc.gov

Abstract

Despite frequent use of mid-upper arm circumference (MUAC) to assess populations in humanitarian settings, no guidance exists about the ranges for excluding implausible extreme outliers (flags) from MUAC data and about the quality assessment of collected MUAC data. We analysed 701 population-representative anthropometric surveys in children aged 6–59 months from 40 countries conducted between 2011 and 2019. We explored characteristics of flags as well as changes in survey-level MUAC-for-age z-score (MUACZ) and MUAC means, SD and percentage of flags based on three flagging approaches: ± 3 and ± 4 MUACZ z-scores from observed MUACZ survey mean and a fixed interval 100–200 mm of MUAC. Both ± 4 and 100–200 flagging approaches identified as flags approximately 0.15% of records; about 60% of all surveys had no flags and less than 1% of surveys had >2% of flags. The ± 3 approach flagged 0.6% records in the data set and 3% of surveys had >2% of flags. Plausible ranges (defined as 2.5th and 97.5th percentiles) for SD of MUACZ and MUAC were 0.8–1.2 and 10.5–14.4 mm, respectively. Survey-level SDs of MUAC and MUACZ were highly correlated ($r = 0.68$). The average SD of MUACZ was 0.96 using the ± 4 flagging approach and 0.94 with ± 3 approach. Defining outliers in MUAC data based on the MUACZ approach is feasible and adjusts for different probability of extreme values based on age and nutrition status of surveyed population. In assessments where age is not recorded and therefore MUACZ cannot be generated, using 100–200 mm range for flag exclusion could be a feasible solution.

KEYWORDS

anthropometry, assessment, humanitarian, malnutrition, measurement, mid-upper arm circumference, quality

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Published 2023. This article is a U.S. Government work and is in the public domain in the USA. *Maternal & Child Nutrition* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Each year, hundreds of population-representative anthropometric surveys are conducted in humanitarian and refugee settings worldwide with the primary objective of estimating the prevalence of wasting in children aged 6–59 months. To achieve this objective, these surveys typically measure two separate anthropometric indicators: weight-for-height z-score (WHZ) and mid-upper arm circumference (MUAC), as both of these indicators have been endorsed by the United Nations Children's Fund (UNICEF) and the World Health Organization (WHO) as independent criteria for admitting children in selective feeding programmes (World Health Organization, United Nations Children's Fund, & Standing Committee on Nutrition, 2006; World Health Organization, World Food Programme, United Nations System Standing Committee on Nutrition, & United Nations Children's Fund (UNICEF) 2007; World Health Organization & United Nations Children's Fund, 2009). In addition, MUAC is the primary indicator routinely measured in blanket mass community screenings to identify and refer children eligible for nutrition treatment.

However, currently, no clear guidance exists either about the ranges for excluding implausible extreme outliers (flags) from MUAC data or about the quality assessment of collected MUAC data based on two key parameters of measurement quality, standard deviation (SD) of the distribution and percentage of flags in the data (World Health Organization, 1995; World Health Organization & The United Nations Children's Fund [UNICEF], 2019). SD is a key metric of measurement error because the introduction of random errors results in widening the observed distribution. A wider distribution, in turn, results in an artificial inflation of prevalence when a continuous variable (e.g., WHZ or MUAC value) is converted into a categorical variable using defined cutoffs (Biehl et al., 2013; Grellety & Golden, 2018).

The primary difficulty is the fact that the weight-for-height indicator is based on standardised z-scores comparing measured population to World Health Organization (2006) growth standards, whereas MUAC is a raw measure not adjusted for age. Therefore, defining the fixed cutoffs for MUAC flag exclusion is suboptimal for several reasons: it takes into account neither the age of the child nor the severity of malnutrition in the measured population. For example, some very low MUAC measurements may still be plausible in the youngest child aged 6 months, but clearly implausible in the oldest child close to his 5th birthday and vice versa. Similarly, very low MUAC measurements may be plausible in a population with very high malnutrition rates, but implausible in a population where children are on average overweight rather than wasted and vice versa.

In this manuscript, we endeavour to provide considerations that fill this gap and provide simple and practical solutions for field epidemiologists and nutritionists on how to identify flags and evaluate measurement quality in MUAC data. We are basing our analysis on a large set of recent anthropometric field surveys conducted during 2011–2019 in humanitarian and refugee settings, and use z-scores based on the MUAC-for-age indicator (MUACZ) for

Key messages

- Mid-upper arm circumference (MUAC) is routinely used in surveys and screenings to assess nutrition status of pre-school children in humanitarian settings.
- To date, there is no guidance on cleaning (detecting implausibly extreme outliers) and assessing quality of MUAC data.
- Defining flags in MUAC data based on the MUAC-for-age z-score (MUACZ) distribution is feasible and in line with flagging approaches used for other anthropometric indicators.
- In assessments where age is not recorded and therefore MUACZ cannot be generated, using a fixed range 100–200 mm range for flag exclusion in children aged 6–59 months could be a feasible solution.

which WHO growth standards exist, taking the same approach as is used for other z-score based indicators, such as WHZ. For assessments such as mass screenings where this approach is not possible because the age of the child is not collected, we use empirical evidence from our large survey data set based on the observed statistical relationship between MUAC and MUACZ indicators.

2 | METHODS

Data for these analyses were single stratum, cross-sectional, population-representative anthropometric surveys generally conducted at the district, subdistrict or refugee camp level (referred to as 'small-scale surveys') provided by the United Nations High Commissioner for Refugees (UNHCR) and by Action Contre le Faim (ACF). Surveys were conducted between 2011 and 2019 and measured age, sex and MUAC in children aged 6–59 months following standard procedures (Standardized Monitoring and Assessment of Relief and Transitions [SMART], 2012; UNHCR, 2018). Survey sampling designs followed the Standardized Monitoring and Assessment (SMART) methodology guidelines (SMART, 2012) and were either two-stage cluster, exhaustive or simple random. MUACZ z-scores were calculated for all children using the WHO SAS macro, which applies the WHO 2006 growth standards (World Health Organization, 2006, 2007).

Surveys were included if they measured a minimum set of anthropometric indicators: sex, age and MUAC, and had a sample size between 200 and 1200 children. Cluster surveys with fewer than 25 clusters or those missing cluster identification variables for more than 20% of sampled children were excluded from the analysis. Within surveys retained for analysis, children were excluded if the child's record was an exact duplicate of another for all variables including

identifier variables. Children with missing or invalid data for sex, age or MUAC were excluded from analysis.

To describe characteristics of flags excluded by MUACZ, we explored two flag exclusion ranges: z-score outside of ± 4 z-scores from the observed survey sample mean, as described by the World Health Organization (1995) and z-score outside of ± 3 z-scores from the observed survey sample mean used in the SMART standard data quality evaluation procedure (so called 'plausibility check') (SMART, 2012). In addition, to address the problem of identifying flags in assessments where age is not collected and thus MUACZ cannot be generated, we explored characteristics of flags based on the fixed range of 100–200 mm, which is mnemonically easy to remember for field practitioners and takes into account the fact that most populations where mass MUAC screenings are conducted have high levels of malnutrition. The lower boundary of this range (100 mm) corresponds to MUACZ of < -5 in 59-month-old children and to approximately MUACZ of -4 in 6-month-old children (with minor variations by sex). The upper boundary (200 mm) corresponds to MUACZ of approximately $+2$ in 59-month-old children and to MUACZ of approximately $+4.5$ in 6-month-old children. Previous analysis reported that the distribution of MUAC showed no departure from a normal distribution in 319 (37.7%) of 852 surveys using the Shapiro–Wilk test (Frison et al., 2016). The characteristics we explored included the percentage of flags at the lower versus upper end of the distribution and distribution of flags by age compared to the overall age distribution in the survey database.

We calculated the mean and SD of MUAC and MUACZ in each survey after excluding flags with MUACZ outside of ± 4 z-scores from the observed survey sample mean and described distributions of these parameters and Pearson correlations between parameters based on MUAC and those based on MUACZ. We empirically defined plausible ranges for quality assessment as 2.5th and 97.5th percentiles of the distribution of survey-level SD seen in our data set. Finally, we also calculated the mean and SD of MUAC and MUACZ as well as the prevalence of low MUAC (based on < 125 mm cutoff) by survey using alternative flagging approaches (± 3 and 100–200) to explore how much effect these different approaches would have on the mean, SD and low MUAC prevalence.

This study was determined as nonresearch by the institutional review board of the Centers for Disease Control and Prevention as it entailed secondary analysis of routinely collected programmatic data. No individual identifiers were included in the data set used for analysis. Data were aggregated, analysed and visualised using SAS Version 9.4 and R 4.0.3 (R Core Team, 2020; SAS Institute, 2014).

2.1 | Ethics statement

Not applicable. Analysis based on secondary analysis of anonymized data where no individual data could be identified so formal ethical clearance was not required. Permission to use the data were obtained from the organisations collecting the datasets.

3 | RESULTS

The data set for this analysis included 701 surveys conducted by ACF or UNHCR between 2011 and 2019 in 40 countries and included over 378,000 child records. The number of surveys and prevalence of low MUAC by region are shown in Supporting Information. The populations surveyed were malnourished on average, with a median survey-level prevalence of low MUAC of 4.64% and an average MUACZ mean of -0.81 ; less than 1% of surveys had mean MUACZ above 0. Both MUACZ and MUAC means by survey were normally distributed with both skewness and excess kurtosis within ± 0.2 from zero (Table 1, Figure 1). MUAC and MUACZ survey means were almost perfectly correlated ($r = 0.99$) (Figure 2a).

Characteristics of flags generated using three approaches (± 4 , ± 3 and 100–200 mm) described in the Methods section are presented in Table 2. Based on the ± 4 approach, 541 flags were identified. About 1% of all surveys had more than 1% of flags and 61% of surveys had no flags. Based on the more restrictive ± 3 approach, 2387 flags were identified. Twenty percent of surveys had no flags, in 17% of surveys percentage of flags was between 1% and 2% and 3% of surveys had more than 2% of flags. Based on the fixed 100–200 mm approach, 609 flags were identified, of those 576 (94.6%) were also flagged using the ± 3 approach and 381 (62.6%) were flagged based on the ± 4 approach.

The mean percentage of flags per survey was 0.13% based on the ± 4 approach, 0.58% based on the ± 3 approach and 0.14% based on the 100–200 mm approach. Similarly, the percentage of flagged children among all child records in the combined data set was 0.14%, 0.63% and 0.16% for the ± 4 , ± 3 and 100–200 mm approaches, respectively.

TABLE 1 Summary statistics of MUACZ and MUAC survey-level means and standard deviations using ± 4 MUACZ flagging approach.

Measure	MUACZ		MUAC	
	Mean	SD	Mean	SD
Mean	-0.81	0.96	144.38	12.31
Median	-0.81	0.95	144.22	12.29
SD	0.36	0.10	4.27	1.06
Skewness	0.01	0.92	0.13	0.21
Kurtosis	-0.18	1.07	-0.11	-0.21
P1	-1.58	0.79	135.54	10.24
P2.5	-1.50	0.81	136.50	10.48
P25	-1.07	0.89	141.23	11.55
P75	-0.56	1.01	147.26	13.04
P97.5	-0.15	1.20	152.25	14.42
P99	-0.01	1.25	155.07	14.89

Abbreviations: MUAC, mid-upper arm circumference; MUACZ, mid-upper arm circumference-for-age z-score; PX, X percentile of distribution; SD, standard deviation.

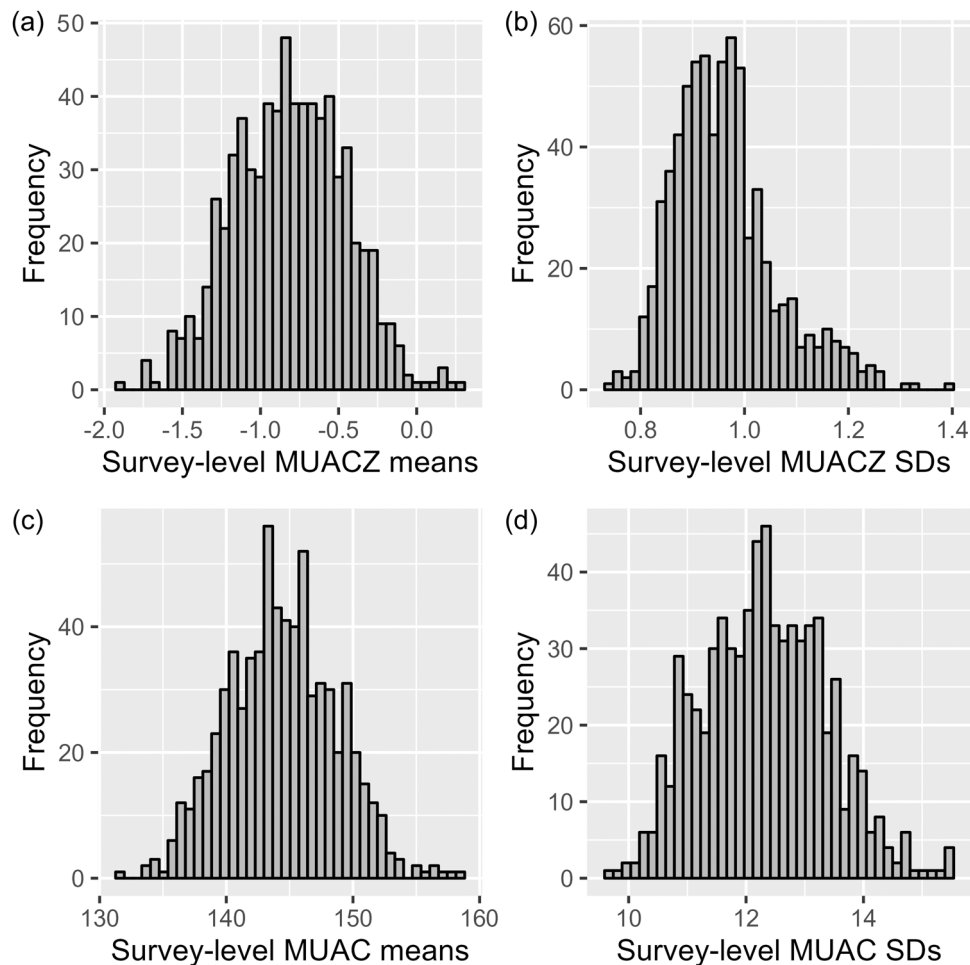


FIGURE 1 Distributions of MUACZ and MUAC means and standard deviations across 701 surveys. Distributions of (a) MUACZ means, (b) MUACZ standard deviations, (c) MUAC means (mm) and (d) MUAC standard deviations across 701 surveys using the ± 4 MUACZ flagging approach. MUAC, mid-upper arm circumference; MUACZ, mid-upper arm circumference-for-age z-score; SD, standard deviation.

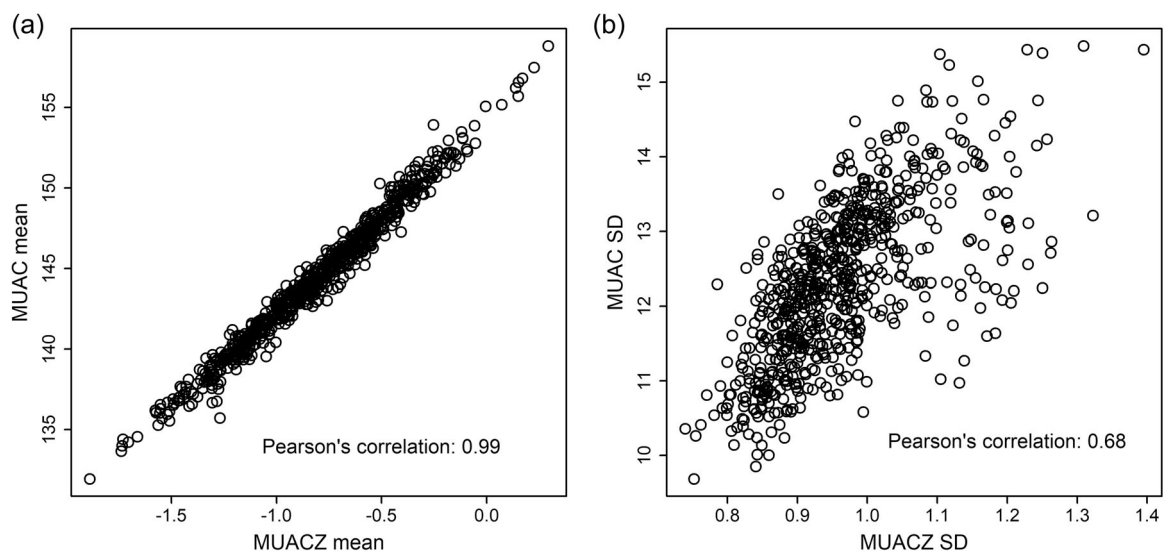


FIGURE 2 Comparison of survey-level MUAC and MUACZ means and standard deviations. Comparison of (a) survey-level MUAC and MUACZ means and (b) survey-level MUAC and MUACZ standard deviations under ± 4 MUACZ flagging criteria. MUAC, mid-upper arm circumference; MUACZ, mid-upper arm circumference-for-age z-score; SD, standard deviation.

Approximately 82% of flags were at the lower end of the distribution using the ± 4 and 100–200 mm approaches and 77% were at the lower end using the ± 3 approach. Children younger than 2 years of age had disproportionately more flags than older children: 45%, 48% and 63% of all flags using ± 4 , ± 3 and 100–200 mm approaches, respectively, were in the under-2 age group although this group accounted for only 36% of children in the combined data set.

Mean SDs for MUACZ and MUAC by survey with ± 4 flag exclusion were 0.96 and 12.3, respectively (Table 2). MUAC SDs were approximately normally distributed, with skewness and excess kurtosis within ± 0.25 from zero. The MUACZ SD distribution was skewed to the right with heavy tails, with a skewness of 0.9 and

excess kurtosis of 1.1 (Figure 1, Table 1). The plausible range of SDs defined by 2.5th and 97.5th percentiles was 0.8–1.2 mm for MUACZ and 10.5–14.4 mm for MUAC, respectively. SDs of MUAC and MUACZ were highly correlated ($r = 0.68$) (Figure 2b).

Using alternative approaches for flag exclusion (± 3 and 100–200 mm) did not change the averages of MUAC or MUACZ survey-level means (Table 3). Average SDs of both MUACZ and MUAC decreased slightly (from 0.96 to 0.94 and from 12.3 to 12.1, respectively) when the ± 3 approach was used.

4 | DISCUSSION

Our analysis, based on a large number of recent field surveys conducted in 2011–2019 by reputable international agencies and following standard methodology guidelines, produced several interesting observations important for interpreting results based on MUAC versus MUACZ indicators and/or instrumental in defining criteria for data cleaning and quality assessment of MUAC measurements.

First, as expected, survey-level means of MUAC and MUACZ were almost perfectly correlated ($r = 0.99$). This suggests that MUAC and MUACZ means are fully consistent with each other in describing anthropometric status of a population.

Second, SDs of MUAC and MUACZ, which are the key parameters of data quality, were also highly correlated, albeit less so than the means ($r = 0.68$). This is also as expected, as MUAC SD is only influenced by the quality of MUAC measurements, whereas the SD of MUACZ is also dependent on the quality of age ascertainment. Accurate age ascertainment is challenging in resource poor settings with substandard vital registration and low literacy (Assaf et al., 2015), such as many of the populations surveyed in this analysis.

Third, the populations included in this analysis were substantially malnourished on average, as evidenced from the overall mean MUACZ of -0.8 (nonmalnourished populations are expected to have mean MUACZ scores close to 0). However, the overall quality of the survey data was very good, with an average SD of 0.96, which is very close to the WHO Growth Standards' SD of 1 z-score. The plausible range for MUACZ SD quality assessment using an empirical approach based on the available surveys was 0.8–1.2, in line with SMART quality assessment guidelines for other anthropometric indicators.

Fourth, as expected mathematically, the more rigorous/restrictive flagging approach (± 3) produces several (4.4) times more flags than the less restrictive one (± 4). But even with the most rigorous flags exclusion, the average proportion of flags identified in this data set was 0.6% and only 3% of all surveys had more than 2% of flags, which, again, speaks to the overall good quality of these survey data. This also suggests that $<2\%$ of flags may be a sensible plausibility criterion for quality assessment if the most rigorous ± 3 flagging approach is used, as in the SMART plausibility check (SMART, 2012). The less restrictive ± 4 approach identified only 0.1% flags on average in our data set.

TABLE 2 Characteristics of flagged children using ± 4 MUACZ, ± 3 MUACZ and MUAC 100–200 mm as flagging approaches.

Flagging approach	± 4 MUACZ	± 3 MUACZ	MUAC: 10–20 cm
Percentage of children with flags	0.14%	0.63%	0.16%
Number of flags	541	2387	609
Mean percentage of observations flagged across surveys	0.13%	0.58%	0.14%
Surveys by range of flagged observations			
0%	61.2%	20.3%	61.3%
>0% to 0.5%	32.8%	37.5%	31.0%
$\geq 0.5\%$ to 1%	4.85%	22.4%	5.99%
$\geq 1\%$ to 2%	0.86%	17.1%	1.28%
$\geq 2\%$	0.29%	2.71%	0.43%
Percentage of flags in lower tail	81.7%	76.7%	82.6%
Percentage of flags in age group < 2 years	44.9%	48.4%	63.1%

Abbreviations: MUAC, mid-upper arm circumference; MUACZ, mid-upper arm circumference-for-age z-score.

TABLE 3 Averages of survey-level means and standard deviations of MUAC-for-age and MUAC and mean prevalence of low MUAC (<125 mm) using different flagging approaches.

Measure	MUACZ ± 4	MUACZ ± 3	MUAC ± 4	MUAC ± 3	MUAC 10–20 cm
Mean	-0.81	-0.81	144.4	144.5	144.4
SD	0.96	0.94	12.3	12.1	12.3
Mean prevalence of low MUAC	-	-	6.11%	5.81%	6.11%

Abbreviations: MUAC, mid-upper arm circumference; MUACZ, mid-upper arm circumference-for-age z-score; SD, standard deviation.

Fifth, we note some interesting features of identified flags: 82% were at the lower end of the distribution and children <2 years of age were overrepresented: they accounted for almost half of all flags, although the proportion of this age group in the overall sample was only about 36%.

Sixth, when the age of the child is not collected and MUACZ cannot be generated, we explored the possibility of using 100–200 mm fixed interval for MUAC flags exclusion. This fixed interval identified a percentage of flags that is similar to that identified by the ± 4 approach and 95% of these flags were the same as those identified by the ± 3 approach. As before, most of the flags were at the lower end of the distribution and children aged < 2 years were overrepresented. The plausible quality assessment range of MUAC SD was 10.5–14.4 mm.

Finally, flagging approaches had virtually no effect on the means of MUAC or MUACZ. The most restrictive ± 3 flagging approach slightly reduced both mean prevalence of low MUAC (from 6.11% to 5.81%) and SDs of MUACZ and MUAC (from 0.96 to 0.94 and from 12.3 to 12.1, respectively) compared with the ± 4 and 100–200 mm approaches. This is mathematically expected, as this approach excluded 0.6% flags from the data set, about four times more than approximately 0.15% excluded by the other two approaches. Removing additional outliers from the data is expected to decrease the resulting SD of the distribution.

It should be noted that we are describing different flagging approaches, including those more (± 3) or less (± 4) restrictive, but make no argument as to which of those is preferable or more valid. As noted in the 2019 WHO survey guidelines, defining the optimal flag exclusion ranges (not only for MUACZ but for any z-score-based anthropometric indicators, such as WHZ, HAZ, etc.) requires further investigation (World Health Organization & The UNICEF, 2019). Different flagging approaches may be considered depending on the levels of malnutrition in the surveyed population (World Health Organization, 1995) or on the overall quality of survey data.

Further, although we suggest using 100–200 mm flagging ranges when analysing data from mass MUAC screenings, this flagging approach will be applied posthoc and will not affect in the field referral to feeding programmes of children whose MUAC may be measured as below 100 mm.

A major strength of this study is the high number and ascertained quality of the cross-sectional surveys it builds upon as well as the fact that all these surveys were conducted in the last 10 years and therefore best reflect current field practices. The 701 surveys contributing to this analysis were conducted in 40 countries around the world. During these surveys, planning, data collection and analysis followed standardised methods embedding rigorous quality controls and were supervised and validated by highly qualified and trained staff (SMART, 2012; UNHCR, 2018). This study, however, has several notable limitations. First, only surveys from humanitarian or refugee settings were included. Further, the surveys were small scale, generally with the objective of providing anthropometric estimates at the district or refugee settlement level. Thus, the results we obtained may not be representative of the overall countries or regions. Second,

the surveys we analysed were conducted predominantly in populations with elevated levels of wasting. Including more surveys with normal or overweight populations could produce slightly different results. However, arguably including undernourished populations as we did is more relevant as MUAC data are collected predominantly in malnourished populations where nutrition treatment programmes for acute malnutrition are implemented because MUAC is used as one of admission criteria for these programmes. Lastly, this analysis did not include children aged 0–6 months, as this age group is not routinely included in small scale anthropometric surveys in humanitarian settings.

5 | CONCLUSION

In summary, we suggest that the MUACZ approach to flag identification in raw MUAC data collected in field surveys is feasible, in line with flagging approaches used for other anthropometric indicators such as WHZ and has advantages over the fixed MUAC ranges in that it adjusts for different probabilities of outliers based on age and nutrition status of the surveyed population. To simplify the analysis of survey data in the field, automatic flag generation and analysis can be programmed into user friendly software universally used by field practitioners, such as ENA software for SMART (Standardized Monitoring and Assessment of Relief and Transitions [SMART], 2020). Based on this analysis, the plausible range for quality assessment for MUACZ SD can be defined as 0.8–1.2, fully in line with SMART guidance for WHZ SD plausible range. The maximum plausible limit for MUACZ flag percentage in survey data can be set at 2% when the most rigorous ± 3 flagging approach (that generates the most flags out of the three approaches explored here) is used.

In mass MUAC screenings and other assessments where the age of children is not collected and therefore MUACZ cannot be generated, the use of the fixed MUAC range to define flags remains the only option. According to our preliminary exploration, 100–200 mm range seems a sensible solution. It is easy to remember and generated approximately the same percentage of flags as the ± 4 MUACZ approach. Almost all records flagged using 100–200 mm range were also flagged by ± 3 MUACZ approach, suggesting that the 100–200 range does not tend to exclude children with MUAC values that are clearly plausible in a given population. Further exploratory work to determine if some other fixed MUAC range would work better than 100–200 for flagging purposes in these generally malnourished populations may be indicated. Of course, this analysis of flags and SDs in mass screenings would only be possible if MUAC measurements are collected. In mass screenings where actual MUAC values are not recorded and only the total numbers of children are tallied based on MUAC category (red <115 mm, yellow 115 to <125 mm and green ≥ 125 mm), none of the measurement quality analysis described in this paper would be possible.

AUTHOR CONTRIBUTIONS

Oleg Bilukha designed the study. Behzad Kianian analysed the data. Behzad Kianian produced the tables and figures. Oleg Bilukha drafted the manuscript. Oleg Bilukha and Behzad Kianian critically revised the manuscript for important intellectual content.

ACKNOWLEDGEMENTS

The authors thank Action Contre la Faim (ACF); United Nations High Commissioner for Human Rights (UNHCR).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Relevant survey-level data are provided within the paper and in the Supporting File. The underlying data that support the findings of this study are available from Action Contre la Faim and the United Nations High Commissioner for Human Rights but restrictions apply to the availability of these data, which were used under license for the current study and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Action Contre la Faim and the United Nations High Commissioner for Human Rights.

ORCID

Oleg Bilukha  <http://orcid.org/0000-0003-2175-8104>

REFERENCES

- Assaf, S., Kothari, M. T., & Pullum, T. W. (2015). *An assessment of the quality of DHS anthropometric data, 2005-2014*. ICF International.
- Biehl, A., Hovengen, R., Meyer, H. E., Hjelmæsæth, J., Meisfjord, J., Grøholt, E. K., Roelants, M., & Strand, B. H. (2013). Impact of instrument error on the estimated prevalence of overweight and obesity in population-based surveys. *BMC Public Health*, 13(1), 146. <https://doi.org/10.1186/1471-2458-13-146>
- Frison, S., Checchi, F., Kerac, M., & Nicholas, J. (2016). Is middle-upper arm circumference "normally" distributed? Secondary data analysis of 852 nutrition surveys. *Emerging Themes in Epidemiology*, 13, 1–8. <https://doi.org/10.1186/s12982-016-0048-9>
- Grellety, E., & Golden, M. H. (2018). Change in quality of malnutrition surveys between 1986 and 2015. *Emerging Themes in Epidemiology*, 15, 1–13. <https://doi.org/10.1186/s12982-018-0075-9>
- R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.3) [software]. R Foundation for Statistical Computing. Available from: <https://www.R-project.org/>
- SAS Institute. (2014). SAS software version 9.4 for windows. SAS Institute Inc.

- Standardized Monitoring and Assessment of Relief and Transitions (SMART). (2012). *Sampling methods and sample size calculation for the SMART methodology*. Retrieved April 6, 2021 from <http://smartmethodology.org/survey-planning-tools/smart-methodology/>
- Standardized Monitoring and Assessment of Relief and Transitions (SMART). (2020). *ENA software for SMART*. Retrieved April 6, 2021 from <https://smartmethodology.org/survey-planning-tools/smart-emergency-nutrition-assessment/>
- UNHCR. (2018). *SENS: Pre-module: Survey steps and sampling*. Retrieved April 6, 2021 from <http://sens.unhcr.org/introduction/pre-module-survey-steps-and-sampling/>
- World Health Organization. (1995). *Physical status: The use and interpretation of anthropometry, report of a WHO Expert Committee* (World Health Organization Technical Report Series, No. 854). World Health Organization.
- World Health Organization. (2006). *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. (2007). *WHO child growth standards: Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development*.
- World Health Organization, & The United Nations Children's Fund (UNICEF). (2019). *Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old*.
- World Health Organization, & United Nations Children's Fund. (2009). *WHO child growth standards and the identification of severe acute malnutrition in infants and children: A joint statement by the World Health Organization and the United Nations Children's Fund*.
- World Health Organization, United Nations Children's Fund, & Standing Committee on Nutrition. (2006). *Informal consultation on community-based management of severe malnutrition in children*. SCN Nutrition Policy.
- World Health Organization, World Food Programme, United Nations System Standing Committee on Nutrition, & United Nations Children's Fund (UNICEF). (2007). *Community-based management of severe acute malnutrition: A joint statement by the World Health Organization, the World Food Programme, the United Nations System Standing Committee on Nutrition and the United Nations Children's Fund*.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bilukha, O., & Kianian, B. (2023). Considerations for assessment of measurement quality of mid-upper arm circumference data in anthropometric surveys and mass nutritional screenings conducted in humanitarian and refugee settings. *Maternal & Child Nutrition*, 19, e13478. <https://doi.org/10.1111/mcn.13478>