Population Health Metrics

RESEARCH

Open Access

Composite anthropometric data quality index for children under the age of 5 on the Brazilian National Food and Nutrition Surveillance System, 2019–2021

Rafaela Oliveira-Santos[1,2], Priscila Ribas de Farias Costa[1,2], Natanael de Jesus Silva[2,3], Juliana Freitas de Mello e Silva[2], Laís Silva Sacramento[2], Gilberto Kac[4], Rita de Cássia Ribeiro-Silva[1,2] and Mauricio Lima Barreto[2*]

Abstract

Background A composite evaluation that merges various data quality indicators separately enabled the researchers to score the overall data quality of the research. In this context, the objective of the present study is to develop composite anthropometric data quality indices for children under 5 registered on the Brazilian National Food and Nutrition Surveillance System (SISVAN) from 2019 to 2021.

Methods Anthropometric data quality indicators were generated for 5,210 Brazilian municipalities: coverage, completeness, the ratio between the sexes, age difference index, preference for height and weight digits, biologically implausible z-score values, and standard deviation. Principal component analysis [PCA] was used to generate a composite anthropometric data quality index for standardized height-for-age (HAZ) and weight-for-height z score (WHZ) indices. The municipalities were ranked in descending order, following their anthropometric quality index values: lowest [worst quality] and highest values [best quality].

Results In total, 29,367,435 records and 8,930,881 children with anthropometric measurement information were identified. The dispersion indicators, the percentage of biologically implausible values [BIV] and the digit preference had the highest factor loadings. We observed that the worst index values were found in municipalities in the country's poorest and most vulnerable regions [North, Northeast, and Central-West]. The correlation between the HAZ and WHZ quality indices was 0.74.

Conclusion The proposed index provides a coherent measurement to discriminate municipal anthropometric data quality.

Keywords Anthropometric evaluation, Anthropometric data quality indicators, Food and nutrition surveillance system, Children

*Correspondence:
Mauricio Lima Barreto
barreto.mauricio@gmail.com
[1]School of Nutrition, Federal University of Bahia (UFBA), Salvador, Bahia, Brazil
[2]Centre for Data and Knowledge Integration for Health (CIDACS), Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador, Bahia, Brazil
[3]ISGlobal, Hospital Clínic, Universitat de Barcelona, Barcelona, Spain
[4]Josué de Castro Institute of Nutrition, Nutritional Epidemiology Observatory, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

Introduction

Anthropometric data are pivotal for estimating nutritional status across and within populations, tracking secular trends, and evaluating the effectiveness of public health interventions [1]. However, the validity and reliability of metrics derived from this data are contingent upon the quality of the anthropometric measurements [2]. Analyzing anthropometric data in large-scale research presents significant complexities and challenges, particularly in children. Inaccuracies in the quality of anthropometric data have substantial implications for assessing the burden of malnutrition within populations. Several stages of the research can compromise data quality, including the questionnaire development to staff training, fieldwork design, data entry into information systems, and finally in subsequent data analysis [3]. From the outset, there are challenges related to the validity of anthropometric equipment, as well as errors in data collection and recording [4]. Healthcare professionals often round numbers to simplify data reporting, but this tendency can introduce subtle biases, particularly when multiple field professionals are involved. Such rounding errors can skew measurements, leading to inaccuracies in data collection and recording, which is particularly critical in anthropometric studies where precision is essential for accurate nutritional assessments and effective public health planning [3, 5].

Various indicators have been employed to assess the quality of anthropometric data, examining aspects such as coverage, completeness of date of birth and anthropometric measurements, digit preference in reporting age, length/height, and weight, percentage of biologically implausible z-score values (BIV), and the distribution of standardized weight and height measurements [6–11]. While individual indicators are crucial for evaluating different dimensions of data quality, their interpretation can be complex due to potential conflicting signals. For instance, one indicator might suggest issues with data completeness, while another might highlight problems with data accuracy. To address these challenges, a more integrated approach that combines multiple data quality indicators may offer a comprehensive evaluation. This aggregated method allows researchers and health managers to assess the overall quality of the data more effectively [2]. Despite its importance, the methodological development of assessing the quality of anthropometric data through an aggregated index or composite score has been limited [3, 12, 13]. One challenge is that a weighted summative score of numerous individual indicators may include elements that are not informative regarding anthropometric quality in each survey. Additionally, the weights assigned to each parameter in the global score are often not empirically derived and can be influenced by indicators with greater variance [3]. Another approach proposed to assess data quality is the use of algorithms, which, despite offering systematic and comparable criteria, have limitations. This method is resource-intensive and relies heavily on the quality of the information available in reports, which often only partially reflect the integrity of the data. Furthermore, algorithms often assume that ideal practices have not been implemented if they are not explicitly mentioned, which can result in biased assessments, often based on worst-case scenarios [12]. As a consequence, some studies may be wrongly classified as of lower quality, even when they have adopted appropriate methodologies. In contrast, Perumal et al. (2020) addressed this issue by assessing the quality of anthropometric data from 145 Demographic and Health Surveys (DHS), predominantly conducted in Sub-Saharan Africa, using Principal Component Analysis (PCA) [2]. This method evaluated the weights of the indicators based on factor loadings derived from the correlation matrix, thereby avoiding arbitrary weight assignments and reducing the dominance of any single indicator in the final score. Moreover, this approach incorporated only those indicators recommended in the recent WHO/UNICEF Anthropometric Data Quality Working Group report (2019) [5].

The fact is that the use of a general data quality index is easier to understand and interpret, particularly for the policy manager, and can be useful for planning and monitoring food and nutritional surveillance actions at various levels of aggregation. Furthermore, the proposed anthropometric data quality indices can be easily derived from the data set. Therefore, we will use PCA to generate a composite index to evaluate the quality of data from Brazil's Food and Nutrition Surveillance System (SISVAN) from 2019 to 2021.

Materials and methods

Study design and population

This descriptive cross-sectional study evaluates anthropometric data quality for children under 5 years old, assisted at Unified Health System primary healthcare (SUS) services between 2019 and 2021. The data was obtained from the Food and Nutrition Surveillance System (SISVAN) individualized database. SISVAN uses as a reference the Protocols of the Food and Nutrition Surveillance System and the "Guidelines for the collection and analysis of anthropometric data in health services: SISVAN Technical Standard", which determine specific procedures for measuring weight and height in different age groups [6].

The SISVAN anonymized data was made available for use in this project by the Centre for Data and Knowledge Integration for Health (CIDACS), Oswaldo Cruz Foundation (FIOCRUZ). All the CIDACS procedures include technical and physical measures to manage, document, store, and protect data, as required by the 2018 Brazilian General Data Protection Law, in force since August 2021, and National Health Council National Research Ethics Commission Resolution Nº 466/2012.

The project was submitted and approved as an amendment to the "100 Million Brazilian Cohort" study by the Federal University of Bahia, Institute of Collective Health Ethics Committee (CAAE: 41695415.0.0000.5030).

Data source

The SISVAN database includes food and nutrition follow-up records from the e-SUS AB, the Bolsa Familia Programme Management System (SIGPBF) and data placed on the SISVAN-Web [14]. The data, which monitor the Bolsa Familia program (PBF) beneficiaries' nutritional status, collected at least twice per year, are incorporated into SISVAN at the end of each period (the first period from January to June and the second from July to December). The primary care teams gradually integrate the e-SUS AB records into the Basic Healthcare Information System (SISAB), following the schedule for sending data and then exporting to SISVAN following validation. The Information Technology Centre (NTI/DAB/SAS/MS) conducts the system integration process with the Unified Health System IT Department (DATASUS). The result of this research is data consolidation on the SISVAN database [14].

Anthropometric data quality indicators

The anthropometric data quality was evaluated using multiple indicators. For children under the age of 5, the indicators recommended in a recent WHO/UNICEF team Working Group on Anthropometric Data Quality (2019) publication and prior experiences were used to qualify data of this type [5]. Indicators related to the following aspects were analysed in this study: (1) completeness, (2) sex ratio, (3) age dissimilarity index, (4) digit preference for length/height and weight, (5) implausible z-score values, and (6) z-score standard deviation. We also included coverage, calculated as the percentage of children under 5 years of age with a SISVAN nutritional status record, divided by the reference population (based on data from the 2000 and 2010 censuses) of children under 5 using the SUS, multiplied by 100 [15].

The dissimilarity index of age evaluates the unexpected concentration of observations in specific ities or months of birth. This index measures the percentage of records that deviate from a uniform distribution of amounts in months (0–59), with values that vary from 0 (ideal) to 90. Digit preference for height/length and weight identifies unexpected patterns. An even distribution of final digits is expected, with about 10% occurrence for each digit (0 to 9). However, preference for digits such as 0 or 5 may occur due to the environment, especially in height/length measurements. To quantify these discrepancies, the dissimilarity index is used. Implausible Z-score values correspond to anthropometric measurements falling within two biologically plausible ranges defined by the WHO child growth standards. For example, HAZ (< -6 or > +6) and WHZ (< -5 or > +5) values are considered implausible [5] [Chart 1].

The following were used for indicator analysis: length/height (cm) and weight (kg) measurements, sex, date of birth, age (months and years), and the anthropometric index z-scores commonly used to evaluate the nutritional status of children: height-for-age (HAZ) and weight-for-height (WHZ). According to SISVAN guidelines, pediatric scales should have a minimum accuracy of 10 g, while platform scales should have a minimum accuracy of 100 g [6]. The z-scores were calculated using the WHO R package *anthro* and child growth reference curves (2006) [16, 17].

Data processing and analysis

The HAZ and WHZ quality indices were generated using PCA. Before conducting the analyses, all duplicates were excluded based on person identification criteria, date of birth, weight, and length/height. The Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity were then computed to verify the sample suitability [18]. The KMO value is considered more reliable for large samples, while Bartlett's test of sphericity is sensitive to sample size. Bartlett's test, especially in studies with large samples (such as the present case), tends to generate significant results even when correlations between variables are not substantial, which can lead to misinterpretations. Therefore, we prioritized KMO as the main criterion in this study. The KMO results (0.53 for HAZ quality index and 0.55 for WHZ quality index) indicate that the factor analysis is adequate for the data. We considered the following indicators to generate the composite index: coverage, age dissimilarity index, digit preference for height and weight, implausible z-score values, and z-score standard deviation. We excluded the completeness indicators due to the low variability using the scree plot, where the points on the largest gradient indicate the appropriate number of components to be retained (the factors with eigenvalues > 1.0 were retained in this analysis). We used the first principal component to produce the composite index, which explained the greatest variation based on the individual variable's correlation matrix (62% for HAZ quality index and 66% for WHZ quality index). The principal component factor loadings were multiplied by the standardized data. The scores represent the original data predictions on the new axes defined by the principal components [18] This strategy was employed since it is the most adequate interpretation of an index expressed in a single aspect [2]. The composite quality index

Chart 1 Anthropometric data quality indicators

Quality indicator	Indicator details	Population (children < 5 years old)
Completeness		
Target population coverage (%)	Percentage of the number of children with a SISVAN nutritional status record, divided by the SUS under 5 user population, multiplied by 100. The SUS user population was obtained by subtracting the total population (estimated by the IBGE) from the population covered by private health insurance (obtained from the ANS).	All children with a SISVAN nutritional status record.
Completeness of date of birth (%)	Percentage of the number of records with the full month and year of birth, divided by the total number of SISVAN child records, multiplied by 100.	All SISVAN nutritional status follow-up records.
Completeness of anthropometric measurements (%)	Percentage of the number of records with completed weight and lengtht/height measurements, divided by the total number of SISVAN child records, multiplied by 100.	All SISVAN nutritional status follow-up records.
Sex ratio		
Sex ratio	Ratio of the number of boys divided by the number of girls. The ratio expected for the Brazilian population under the age of 5, estimated by the IBGE, was used as a reference for comparison purposes.	All SISVAN nutritional status follow-up records.
Age distribution		
Age dissimilarity index (in months)	The Myers' unblended dissimilarity index was analysed to identify the percentage of records deviating from an equal distribution of age in months (0–59). It was calculated by the sum of the observed percentages subtracted by the expected percentages, divided by 2. The dissimilarity index varies between 0 and 90, with 0 being the ideal value.	All records with a complete month and year of birth.
Digit preference for lengtht/height and weight		
Dissimilarity index for the last digits for lengtht/height (cm) and weight (kg)	The Myers' unblended dissimilarity index was analysed to identify the percentage of records deviating from an uniform distribution of last lengtht/height and weight digits. It was calculated by the sum of the observed percentages, subtracted by the expected percentages, divided by 2. The dissimilarity index varies between 0 and 90, with 0 being the ideal value.	All records with weight and/or height information.
Implausible z-score values		
Percentage of implausible z-scores for each anthropometric index (%)	Percentage of the number of records with implausible z-score values, divided by the total number of SISVAN child records, multiplied by 100. Using the WHO "anthro" flagging system, implausible Z-score values were detected according to plausibility criteria (WHO, 1995): HAZ (-6, +6), WHZ (-5, +5), WAZ (-6, +5) and B/AZ (-5, +5).	All records with date of birth, weight, and/or lengtht/height information.
Dispersion of z-score values		
Standard deviation of plausible z-score values for each anthropometric index	Standard deviation was calculated using the formula below: $\sqrt{\frac{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}{n-1}}$ Where n is the total number of observations, Yi is each database value, and $\bar{y}$ overall mean of the variable.	All records with biologically plausible Z-score values for the anthropometric index of interest.

IBGE: Brazilian Institute of Geography and Statistics; ANS: National Agency of Supplementary Health; HAZ: height-for-age; WHZ: weight-for-height; WAZ: weight-for-age; B/AZ: Body mass index-for-age

Source: Silva, N. J., et al. (2023). "Quality of child anthropometric data from SISVAN, Brazil, 2008–2017." Rev Saude Publica 57: 62

values evaluated from the 5,570 municipalities, 5,210 were ranked in descending order, with the lowest [worst quality] and highest values [best quality]. Three hundred and sixty municipalities [360] were excluded from the analysis due to a lack of information on at least one of the indicators used to construct the index. Due to the absence of data for these indicators, it was impossible to generate the index for these municipalities since the creation of it depends on a linear combination of all indicators. Also, we performed stratified analyses by age group for the L/HAZ indicator.

All the analyses were conducted using R software version 4.3.2 and the RStudio interface.

Results

In total, 29,367,435 records and 8,930,881 children with anthropometric measurement data were identified on the SISVAN database between 2019 and 2021, following the exclusion of duplicated records. In general, 100%

Table 1 Results of individual indicators for 5,570 municipalities. Brazilian National food and nutrition surveillance system (SISVAN), 2019–2021

Data Quality Indicator	Summarised statistics		
	Median	1st and 3rd quartiles	Range (minimum and maximum
Sex ratio	1.1	1–1.1	0.1–3.8
SUS population coverage	0.5	0.3–0.7	0–1
Digit preference for weight, dissimilarity index, %	31.8	20.2–46	2.9–89
Digit preference for length/height, dissimilarity index, %	59.4	49.9–73.9	38.5–90
Digit preference for age, dissimilarity index, %	14.2	10.1 – 20.3	2.6–55.2
Biologically implausible HAZ values ("flagged"), %	1.6	0.9–2.5	0.1–34.8
Biologically implausible WHZ values ("flagged"), %	2	1.3–2.9	0.1–33
HAZ standard deviation	1.5	1.4–1.7	0.8–3.1
WHZ standard deviation	1.5	1.4–1.6	0.5–2.4

Table 2 Factor loadings of individual anthropometric quality indicators in principal component analyses for 5,210 municipalities [HAZ/WHZ]. Brazilian National food and nutrition surveillance system (SISVAN), 2019–2021

Data Quality Indicator	HAZ index factor loadings	WHZ index factor loadings
Target population coverage	0.16	0.06
Length/height dissimilarity index	0.34	0.03
Age dissimilarity index	0.60	-
Biologically implausible HAZ values	0.81	-
HAZ standard deviation	0.87	-
Weight dissimilarity index	-	0.77
Biologically implausible WHZ values	-	0.83
WHZ standard deviation	-	0.90

completeness of the date of birth (i.e., a record of the month and year of birth, as a minimum) and weight and lengh/height measurements were found. The target population coverage varied between 31.8% and 42.8% over the period studied [Table 1].

Across the municipalities, the median ratio between the number of records of boys and girls on SISVAN was 1.1 (interquartile range: 1-1.1). The dissimilarity index showed that, in general, 14.0% of the records (interquartile range: 10.1–20.3%) would need to be redistributed to attain an uniform distribution of age in months, 31.8% (interquartile range: 20.2–46%) of last digits for weight (kg) and 59.4% (interquartile range: 49.9–73.9%) for length/height (cm). We observed that the median standard deviation values for the plausible z-score measurements (i.e., the median value across the standard deviations of all municipalities) were 1.5 for HAZ (interquartile range: 1.4–1.7) and 1.5 for WHZ (interquartile range: 1.4–1.6). The percentage of biologically implausible HAZ values (based on the WHO definition) varied from 0.1 to 34.8% (interquartile range: 0.9–2.5%) and from 0.1 to 33% for WHZ (interquartile range: 1.3–2.9%) (Table 1).

We observed weak correlations between the individual data quality indicators, suggesting that each indicator used in the PCA reflects a different aspect of anthropometric data quality. However, the strongest correlation was identified between the index WHZ and the z-score standard deviation [corr: 0.91] [Supplementary Table 1]. The PCA results revealed that the standard deviation indicators and percentage of BIV had the highest factor loadings, followed by digit preference for age in the HAZ and WHZ data quality indices [Table 2].

The composite quality index varied from – 15.12 to 5.12 for HAZ quality index and from – 17.17 to 5.51 for WHZ quality index. The municipalities were ranked in descending order according to the composite quality index values evaluated. To facilitate the visualization of the results on the map, we selected 600 municipalities for each composite index, composed of the 300 with the highest and 300 with the lowest values. We observed that the worst index values were found in municipalities located in the North and Northeast regions (Figs. 1, 2, 3 and 4). The correlation between the HAZ and WHZ quality indices was 0.74.

Stratified analyses by age group for HAZ (under 24 months and over 24 months) are available in the supplementary material (Supplementary Figs. 3 and 4). We did not find significant differences in the overall data quality index calculated for each municipality when we stratified the analyses by age [0–24 months; 25–59 months].

Discussion

This study examines the anthropometric data quality of children followed up by SISVAN between 2019 and 2021. This is the first study to evaluate SISVAN data quality in

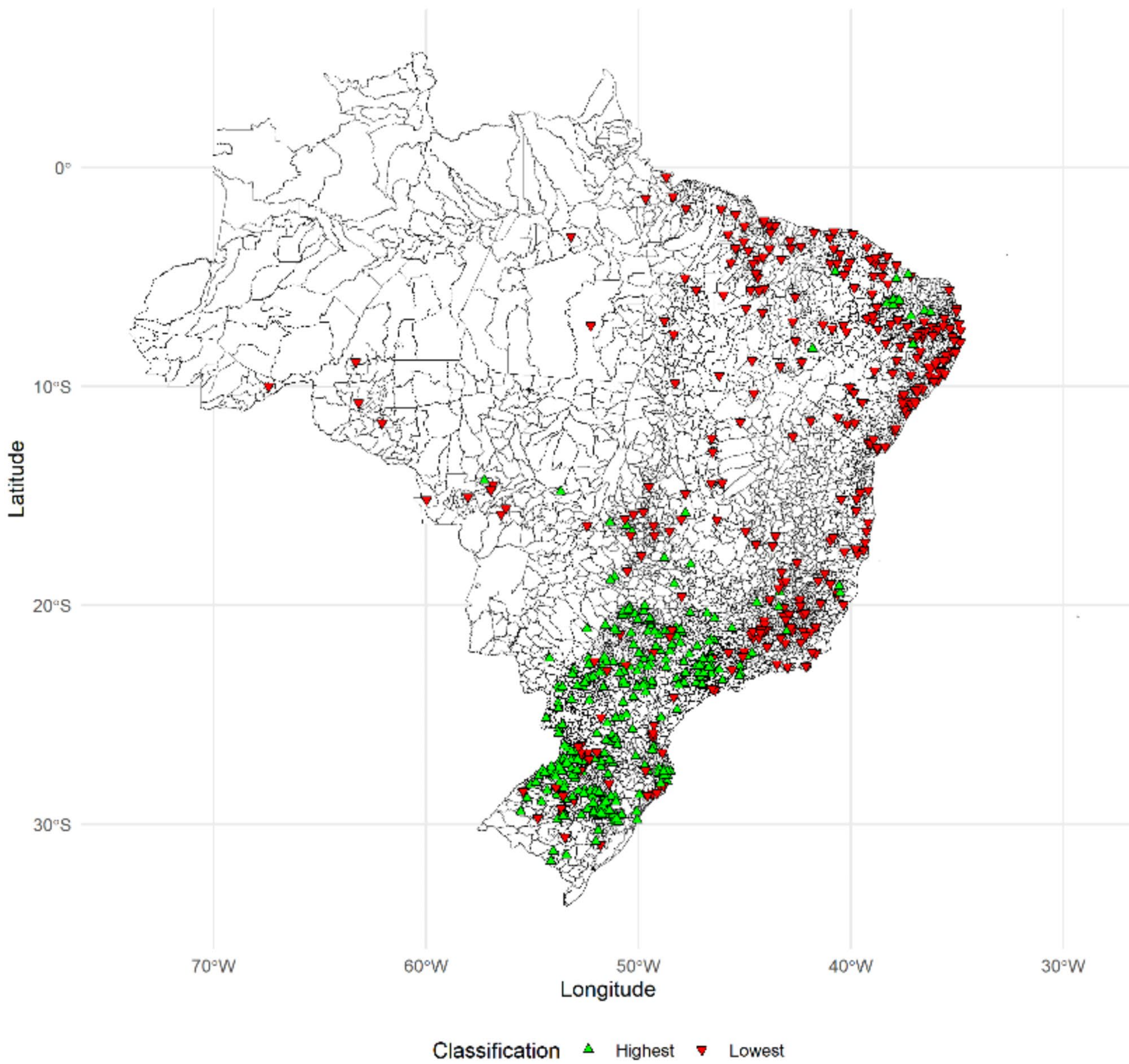

Fig. 1 300 municipalities with the highest and lowest HAZ indices. Global HAZ quality indicator by municipality based on data from the Brazilian National Food and Nutrition Surveillance System (SISVAN), 2019–2021

Brazil, covering multiple aggregated data quality markers in a single index. The results indicate that the composite anthropometric data quality index using a set of individual markers may effectively discriminate the municipalities according to data quality; the worst values were found in municipalities located in the country's poorest and most vulnerable regions. Similar findings were identified when the indicators were evaluated separately [7, 8]. The high correlation between the HAZ and WHZ quality indices also suggests that a composite index, which includes individual indicators for HAZ quality alone, also captures quality issues for other nutritional status metrics.

An evaluation of these SISVAN data individual quality markers was investigated extensively [7] and updated by Conde et al. (2024) [8]. According to these authors, the anthropometric data quality for children under 5 improved substantially between 2008 and 2020; however, as this study confirms, a substantial regional heterogeneity in the anthropometric data quality remains throughout the country. In other words, we advanced in anthropometric data quality during the period analysed. However, it is still insufficient to attain the measurement quality standard recommended by the WHO [5, 7, 8], particularly in the poorest regions of the country.

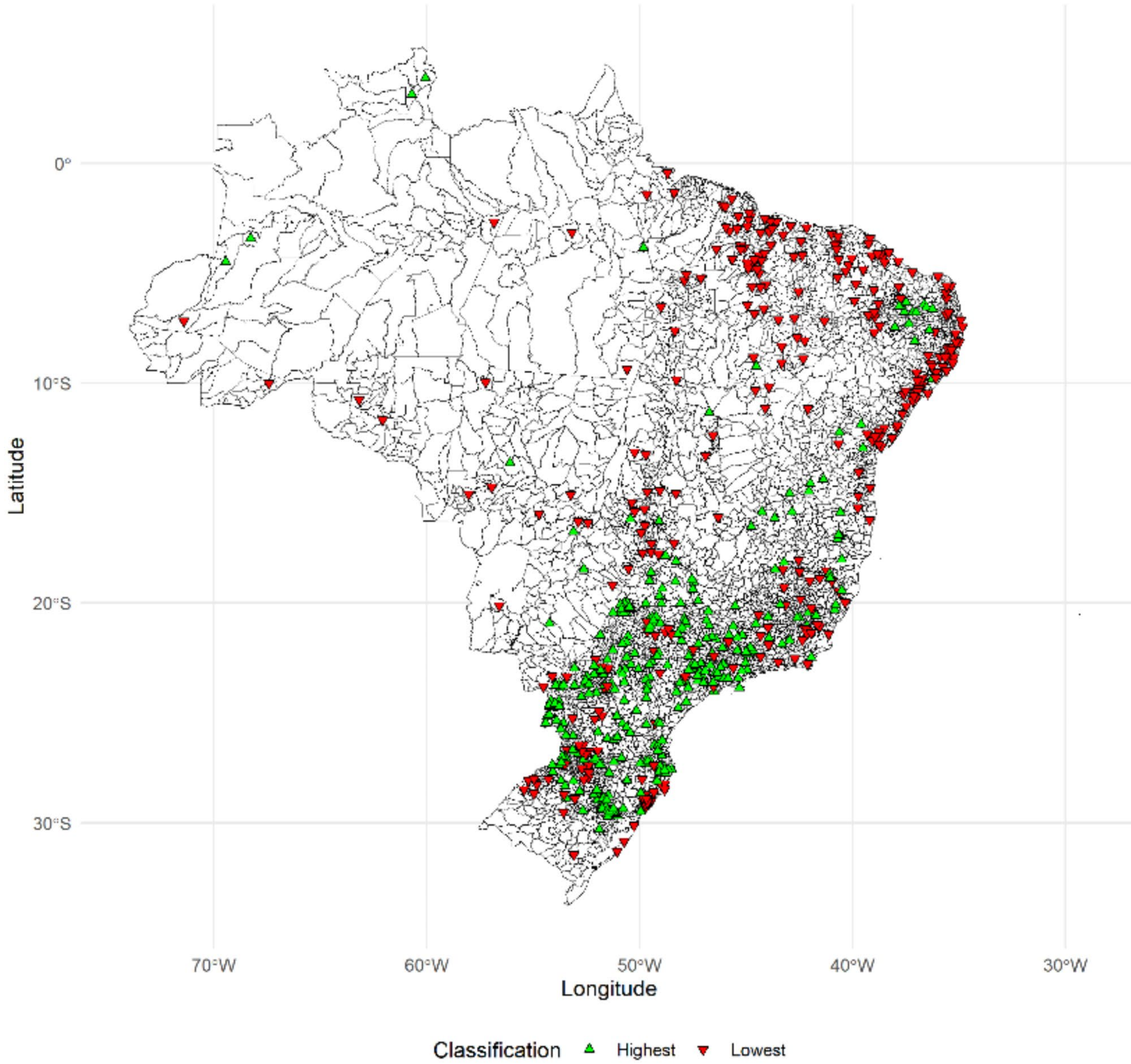

Fig. 2 300 Municipalities with the highest and lowest WHZ indices. Global WHZ quality indicator by municipality based on data from the Brazilian National Food and Nutrition Surveillance System (SISVAN), 2019–2021

We highlight several individual quality markers we evaluated in our research; the z-score standard deviations exceeded 1 for the anthropometric indices [HAZ: 1.5 and WHZ: 1.5]. It is known that each of these tests is an independent quality marker, providing information on different types of possible measurement errors. Large standard deviations indicate measurement errors [19, 20]. Silva et al. (2023) demonstrated that the SD is often larger for children under 2 comparing to the SD of those between 2 and 5. Younger children are usually more challenging to measure, given differences in children's behaviour and cooperation while being measured [7]. On the other hand, for the most part, the [BIV] are the results of typing errors instead of measurement (for example, recording 58 cm instead of 85 cm for length/height, or 1st November instead of 11th January). Our study observed that the BIV frequency was higher for length/height than weight, and both were greater than 1%. Minor measurement errors are also evaluated by rounding them up. The digit preference for weight is unusual in research using digital scales but remains an issue for length/height. Since the centimetre marks are much larger on length/height signs and easier to read than those in millimetres, rounded-up values are often recorded by less diligent teams and may indicate other unobserved issues related to training and supervision. In accordance with our study, almost 59.4%

Fig. 3 HAZ global distribution indicator by state. HAZ global quality indicator by state. Brazilian National Food and Nutrition Surveillance System (SISVAN), 2019–2021

of the length/height records required redistribution to obtain an equal distribution of the last digits, while for the end digits for weight, this value was approximately 31.8%; these values are far from WHO recommendations. Digit preference higher than 20 and a BIV frequency higher than 1% on nutrition index dispersion indicate that the data lacks quality and, possibly, estimating nutritional diagnoses at the population level for children under 5 may be problematic if no prior adjustment is conducted.

The composite anthropometric data quality index may also be an effective tool to defend the reinforcement of rigorous anthropometric data collection [2]. It provides a snapshot of current anthropometric quality, and over time, within the country, and its ranking in an aggregated form, which may be used to supplement recommendations on improving data quality. When producing this global indicator, we used the first principal component to construct our HAZ and WHZ data quality indices [which explained the higher variance in a single anthropometric data quality dimension (62% for HAZ quality index and 66% for WHZ quality index)]. However, the total variance explained by our indicators is relatively low since the individual anthropometric data quality indicators were not strongly correlated. Nevertheless, we present a more straightforward interpretation of an index expressed in a single aspect. These global indicators reinforce our previous disaggregated results that the registered anthropometric data quality for residents in the North, Northeast, and Central-West regions is lower than others in the country [7, 21]. It is important to highlight that the correlation between the HAZ and WHZ data quality indices was 0.74, indicating that research with higher-quality HAZ data also has high-quality WHZ data. These results indicate well-known errors and limitations when collecting and recording anthropometric measurements. The

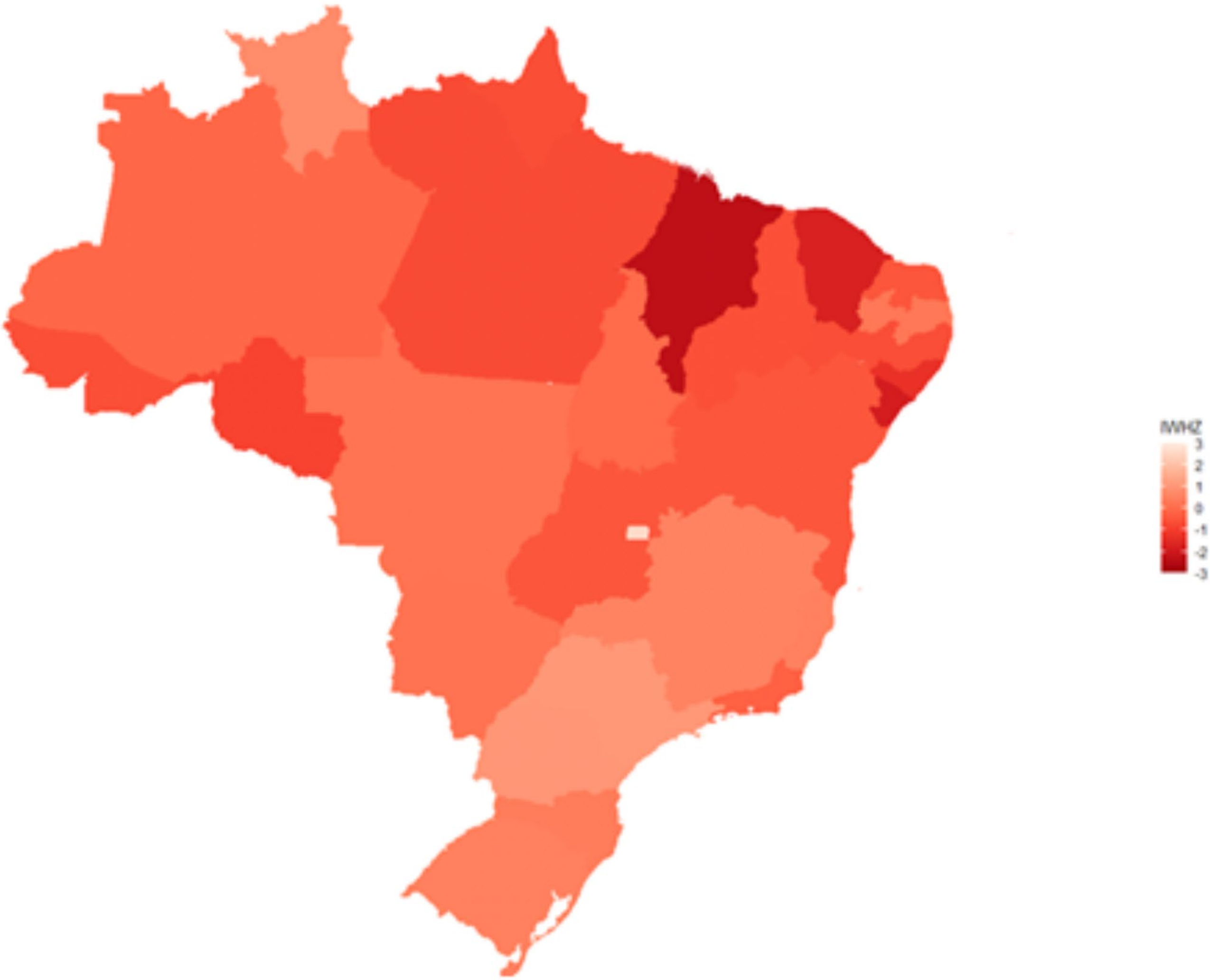

Fig. 4 WHZ global distribution indicator by state. WHZ global quality indicator by state. Brazilian National Food and Nutrition Surveillance System (SISVAN), 2019–2021

anthropometric measurement location may represent a source for such errors. In a study by Lima et al. [2010] at public healthcare services (PHS) of Alagoas, 60% of the healthcare establishments did not have a place and appropriate conditions to conduct anthropometry (18). Similar conditions were also found by Dos Santos (2021) for other establishments in the North and Northeast regions. Therefore, the low data quality observed in these regions can be attributed to the combination of these factors, which are related to greater socioeconomic vulnerability and structural limitations that compromise the accuracy and consistency of anthropometric measurements [22, 23].

Conclusion

This quality index may be constructed in any population database in which anthropometric data is collected. We also consider the number of children under 5 in all Brazilian municipalities contributing anthropometric information for this analysis as a strength. Utilizing a parsimonious and informative set of individual data quality indicators recommended for use in the recent WHO/UNICEF Working Group on Data Quality report (2019) is also a strength.

However, we highlight that we did not use z-score value normality indicators [asymmetry and kurtosis], which had been used previously in in another assessment of anthropometric data quality [7]. It is argued that there is a lack of evidence to suggest that the deviation of a normal distribution is only due to measurement errors. The WHO reference population used to derive the z-scores was restricted to a healthy population that lived in environmental conditions favourable to healthy growth [5]. Thus, unusual distributions may occur in more heterogeneous populations, such as countries with significant social inequalities. The SISVAN population represents users of primary care under the Unified Health System in Brazil, the majority of whom are beneficiaries of the conditional income transfer program-Bolsa Família Program; in other words, it is a more socio-economically vulnerable population. Therefore, further research on indicator development is required, which may better differentiate between data quality problems and population heterogeneity. We used PCA for factor extraction, but the total variance explained by the indicator was relatively low. Additional principal components could be incorporated to increase the variability explained by the data quality indices. Although the variance explained was limited, the method proved to be effective for the proposed objective. It allows the classification of surveys based on the relative quality of anthropometric data and presents itself as a valuable tool for comparative analyses.

We conclude that the proposed anthropometric data quality index may be easily derived from the data set and provide a continuous, coherent measurement for a regional breakdown. Despite advances, the worst values were found in municipalities in the country's North, Northeast, and Central-West regions. Therefore, permanent qualification and education actions and maintaining and extending financial support to the municipalities for food and nutrition surveillance structuring in Unified Health System basic healthcare through the acquisition and periodic calibration of anthropometric equipment should be intensified so that measure of nutritional deterioration is more accurate.

Glossary

BIV	Biologically implausible values.
CIDACS	Centre for Data and Knowledge Integration for Health.
DAB	Department of Primary Care.
DATASUS	Unified Health System IT Department.
FIOCRUZ	Oswaldo Cruz Foundation.
HAZ	Height-for-age z score.
KMO	Kaiser-Meyer-Olkin.
MS	Ministry of Health.
NTI	Information Technology Center.
PBF	Bolsa Familia program.
PCA	Principal Component Analysis.
SAS	Health Care System.
SIGPBF	Bolsa Familia Programme Management System.
SISAB	Basic Healthcare Information System.
SISVAN	Food and Nutrition Surveillance System.
SUS	Unified Health System primary healthcare.
UNICEF	United Nations Children's Fund.
WHO	World Health Organization.
WHZ	Weight-for-height z score.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12963-025-00371-3.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Acknowledgments: We thank the CIDACS/FIOCRUZ data production team's staff for their work on linking the data used in this study, and providing information on data quality. We would also like to thank the IT team for their efforts in helping to facilitate our access to data. We also thank Nandita Perumal for the contributions to the critical revision of the manuscript.

Author contributions

Authors' Contributions: Study concept and planning: ROS, PRFC, NJS, LSS, MLB, and RCRS. Data analysis or interpretation: ROS, PRFC, NJS, LSS, JFMS and RCRS. Manuscript preparation or review: ROS, PRFC, NJS, LSS, JFMS, MLB, GK and RCRS. Approval of the final version: ROS, PRFC, NJS, LSS, JFMS, MLB, GK and RCRS. Public responsibility for the content of the article: ROS, PRFC, NJS, LSS, MLB, and RCRS.

Funding

MS-SCTIE-Decit/ CNPq [442948/2019-0], Decit/SECTICS/MS [25000.148278/2022-10] funded this study. CIDACS received financial support from MCTI / CNPq / MS / SCTIE / Decit / Bill & Melinda Gates Foundation's Grandes Desafios Brasil – Desenvolvimento Saudável para Todas as Crianças (call number 47/2014) (grant number OPP1142172). Research Support Foundation of the State of Bahia (FAPESB) (grant number BOL0719/2023).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

The project was submitted and approved as an amendment to the "100 Million Brazilian Cohort" study by the Federal University of Bahia, Institute of Collective Health Ethics Committee (CAAE: 41695415.0.0000.5030).

Consent for publication

Not applicable. The data were anonymized and de-identified.

Competing interests

The authors declare no competing interests.

Received: 17 September 2024 / Accepted: 22 February 2025
Published online: 13 May 2025

References

1. Corsi DJ, Neuman M, Finlay JE, Subramanian SV. Demographic and health surveys: a profile. Int J Epidemiol. 2012;41:1602–13.
2. Perumal N, Namaste S, Qamar H, Aimone A, Bassani DG, Roth DE. Anthropometric data quality assessment in multisurvey studies of child growth. Am J Clin Nutr. 2020;112:S806–15.
3. 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.

4. Fujimura MS, Conkle J, Van Wyk M, Jimba M. Nutritional status of children under 5 years old in Namibia: adjusting for poor quality child anthropometry. J Nutr Sci. 2022;11:e66.
5. WHO. Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old. Geneva: World Health Organization, UNICEF; 2019.
6. Brasil. Orientações Para a Coleta e análise de Dados antropométricos Em serviços de Saúde: norma técnica do sistema de vigilância alimentar e Nutricional - SISVAN. Ministério da Saúde Brasília, DF; 2011.
7. Silva NJ, Silva J, Carrilho TRB, Pinto EJ, Andrade R, Silva SA, Pedroso J, Spaniol AM, Bortolini GA, Fagundes A, et al. Quality of child anthropometric data from SISVAN, Brazil, 2008–2017. Rev Saude Publica. 2023;57:62.
8. Santos I, Conde WL. [Quality of anthropometric data of children under 5 years in the Brazilian National food and nutrition surveillance system, 2008–2020]. Cad Saude Publica. 2024;40:e00070523.
9. Leidman E, Mwirigi LM, Maina-Gathigi L, Wamae A, Imbwaga AA, Bilukha OO. Assessment of anthropometric data following investments to ensure quality: Kenya demographic health surveys case study, 2008 to 2009 and 2014. Food Nutr Bull. 2018;39:406–19.
10. Dwivedi LK, Banerjee K, Sharma R, Mishra R, Ramesh S, Sahu D, Mohanty SK, James KS. 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.
11. de Mello e Silva JF, de Jesus Silva N, Carrilho TRB, Jesus Pinto Ed, Rocha AS, Pedroso J, Silva SA, Spaniol AM, da Costa Santin, de Andrade R, Bortolini GA, et al. Identifying biologically implausible values in big longitudinal data: an example applied to child growth data from the Brazilian food and nutrition surveillance system. BMC Med Res Methodol. 2024;24:38.
12. Prudhon C, de Radiguès X, Dale N, Checchi F. An algorithm to assess methodological quality of nutrition and mortality cross-sectional surveys: development and application to surveys conducted in Darfur, Sudan. Popul Health Metr. 2011;9:57.
13. Hunger AA. The SMART plausibility check for anthropometry. Toronto: Action Against Hunger; 2015.
14. Brasil. Manual operacional Para Uso do sistema de vigilância alimentar e nutricional: SISVAN VERSÃO 3.0. Ministério da Saúde Brasília, DF; 2017.
15. Freire FHMA, Gonzaga MR, Gomes MMF. Sex and Age-Specific population projections for small areas in Brazil. Rev Latinoam Poblac. 2019;14:124–49.
16. Organization WH. WHO anthro survey analyser and other tools. Geneva: World Health Organization; 2020.
17. WHO Child Growth. Standards based on length/height, weight and age. Acta Paediatr Suppl. 2006;450:76–85.
18. Artes R, Barroso LP. Métodos multivariados de análise estatística. 2023.
19. Grellety E, Golden MH. The effect of random error on diagnostic accuracy illustrated with the anthropometric diagnosis of malnutrition. PLoS ONE. 2016;11:e0168585.
20. 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;11:10671.
21. Lima MA, Oliveira MA, Ferreira Hda S. [Reliability of anthropometric data obtained in children seen at the primary public healthcare service network in Alagoas, Brazil]. Rev Bras Epidemiol. 2010;13:69–82.
22. Santos D, Alves CMC, Rocha TAH, Queiroz RCS, Silva NCD, Thomaz E. Structure and work process regarding child care in primary health care in Brazil: an ecological study with data from the program for primary health care access and quality improvement 2012–2018. Epidemiol Serv Saude. 2021;30:e2020425.
23. Machado PMO, Lacerda JT, Colussi CF, Calvo MCM. Estrutura e processo de Trabalho Para as Ações de alimentação e Nutrição Na Atenção primária à Saúde no Brasil, 2014. Epidemiol Serv Saude. 2021;30:e2020635.