ARTICLE IN PRESS
OPEN ACCESS
Journal of Multidisciplinary Applied Natural Science Vol. xx No. xx (xxxx) https://doi.org/10.47352/jmans.2774-3047.438
Check for updates
Research Article

A Multiple Imputation Approach to Improving Health Data Accuracy in Pooled Cross-Sectional Analysis

Romuald Daniel Boy-ngbogbele*, Oscar Ngesa, Thomas Mageto, and C´elestin Kokonendji

Received : November 7, 2025 | Revised : January 22, 2026 | Accepted : February 27, 2026 | Online : May 24, 2026

Abstract
Pooled cross-sectional health survey data are commonly utilized to examine trends in child malnutrition; however, their precision is usually undermined by measurement errors and absent data. This study assesses the efficacy of a Multiple Imputation (MI) methodology in rectifying these biases and enhancing the reliability of malnutrition prevalence estimates for children under five in Cameroon. Utilizing simulated data that mirrors the structure of four rounds of Demographic and Health Surveys (DHS) from Cameroon (2004, 2011, 2018, 2022), we provide a logistic regression model including measurement error correction by multiple imputation. The efficacy of the MI-corrected model is evaluated against an uncorrected model using several measures, such as prevalence estimates, classification accuracy, precision, and parameter stability across different sample sizes. The MI-corrected model consistently yielded lower and more precise estimates of malnutrition prevalence than the uncorrected model, with reductions of up to 11.34 percentage points in 2004. Classification accuracy increased by 3–4 percentage points across survey waves, with a corresponding improvement in precision. Increasing the sample size reduced the variability of the parameters, as shown by lower standard deviations and coefficients of variation. This made the regression estimates more stable. The model exhibited enhanced resilience in situations including absent data and inaccurately assessed variables. Multiple Imputation successfully rectifies measurement errors and addresses missing data in pooled cross-sectional surveys, resulting in more credible estimates of child malnutrition prevalence. These improved estimates offer a more accurate assessment of public health needs, better targeting of dietary interventions, and more dependable monitoring of trends over time. The results support the implementation of MI-based corrections in national health survey analyses to enhance evidence-based policy and resource distribution in resource-constrained environments.

Keywords: epidemiological trends, health data analysis, malnutrition analysis, measurement error, multiple imputation modeling, pooled cross-sectional analysis

1. INTRODUCTION

Cross-sectional data refers to information collected at a given time from a sample or population. Specifically, cross-sectional data is data collected at a point in time, which means that it provides only a snapshot of the behavior or state of individuals. This type of data provides a snapshot of the characteristics, behaviors, or other variables of interest in a population at a specific point in time, as noted in previous works [1][2]. In the context of the stages of change, previous reports [3][4] defined cross-sectional data as a theoretical model that describes the different stages individuals go through when trying to eliminate problematic behaviors, such as stopping smoking or starting to exercise regularly. Sutton points out that cross sectional data can provide useful information about the prevalence of different stages of change in a given population. However, their interpretation requires a nuanced understanding of the limitations of these data.

Repeated cross-sectional data, as published previously [5], involve collecting data from different individuals or groups at multiple time points, rather than following the same individuals over time as in panel studies [6][7]. This approach is useful for analyzing trends in attitudes, behaviors, health, and socio-economic indicators within a population. Researchers analyzed trends in socioeconomic inequalities related to behavioral risk factors for non-communicable diseases using repeated cross-sectional health survey data from England spanning 2003 to 2019 [8][9]. The study highlight persistent disparities and the need for targeted public health interventions. In the African context, repeated cross-sectional data come from sources such as national surveys (e.g., Demographic and Health Surveys, Multiple Indicator Cluster Surveys, and Living Standards Measurement Surveys), population censuses, and administrative records, enabling researchers and policymakers to track changes over time [10][11]. While panel studies, though less common, provide insights into

Publisher's Note:
Pandawa Institute stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright:
© 2026 by the author(s).
Licensee Pandawa Institute, Metro, Indonesia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Table 1. Summary statistics of key variables from DHS Cameroon Pooled Dataset (2004–2022).

Variable	Mean / Proportion	SD	Min-Max / Categories
Stunted (*Y*)	0.34	-	0 = no, 1 = yes
Age_group	-	-	15 – 49 (grouped in 5-year bands)
Urban_rural	0.39	-	0 = rural, 1 = urban
Education_mother	6.8 years	3.5	0 – 16
Partner_occupation	-	-	Agriculture, Professional, Service, Other
Wealth_index	3.0	1.4	1 (poorest) – 5 (richest)
Toilet_type	-	-	Improved, Unimproved, None
Water_source	-	-	Safe, Unsafe
Maternity_enteries	2.7	1.1	0 – 5
Healthcare_enteries	2.5	1.2	0 – 5

individual-level dynamics, repeated cross-sectional data remain a key tool for monitoring population trends in Africa.

Pooled cross-sectional data refers to a type of dataset that combines multiple cross-sectional samples collected over time. When analyzing pooled cross-sectional data, researchers can gain insights into various policy questions that would be difficult to address by ignoring the pooling aspect. Here are some examples: In this multi-country analysis, pooled cross-sectional data were used to assess the prevalence and determinants of antenatal care visits in countries with high maternal mortality [14][15]. The study aimed to identify the factors that influence the use of antenatal care to inform public health strategies in these regions. Trend analysis using pooled cross-sectional data allows researchers to track changes over time. By combining data from different sources and time periods, they can identify disparities and determine whether certain policies are more effective in specific contexts. An example of such analysis can be found in the study conducted in [16]. The researchers pooled data from various experimental studies to compare the effectiveness of different incentive schemes on human behavior across different contexts and populations. Pooled cross-sectional data from multiple studies were utilized to assess the effects of minimum wage policies on employment levels over time, as reported in [17]. Policy diffusion through pooled cross-sectional data can shed light on the spread or diffusion of policies across different regions or jurisdictions. By tracking policy adoption rates over time, researchers can assess the factors influencing policy diffusion and their impact on various outcomes. Earnings forecasting models were evaluated using pooled cross-sectional data from Indian firms. The study provides insights into the effectiveness of various forecasting techniques in the context of the Indian financial market [18].

Malnutrition is a major problem in developing countries, especially with the decrease of food supply and accordingly the increase of food prices all over the globe [19][20]; malnutrition among children under the age of five is a critical public health issue globally, particularly in low- and middle-income countries such as Cameroon: Despite efforts to address this problem, accurate estimation of malnutrition prevalence remains challenging due to various factors, including measurement error in collected data. Traditional statistical methods may not adequately account for measurement error, leading to biased or imprecise estimates of malnutrition prevalence. In Cameroon and Central African Republic (CAR), as in many developing countries, data on malnutrition are often collected through cross-sectional surveys such as Demographic and Health Surveys (DHS). However, these surveys may suffer from measurement errors arising from issues such as recall bias, inaccuracies in anthropometric measurements, and inconsistencies in data collection procedures. These errors can introduce noise and bias into the data, making it difficult to obtain reliable estimates of malnutrition prevalence and identify underlying determinants [21]. Given the importance of accurate estimation of malnutrition prevalence for informing targeted interventions and policy decisions, there is a growing need for statistical models capable of handling measurement error in pooled cross-sectional data on malnutrition among children under five in Cameroon.

In their systematic review and meta-analysis, the global prevalence of malnutrition among older adults was examined in a recent study [22]. Analyzing data from 98 studies with a total sample size of 79,976 individuals, they found that 18.6% of the world's elderly population is affected by malnutrition. The study revealed regional disparities, with Africa exhibiting the highest prevalence at 35.7%, followed by the Americas at 20.3%. These findings highlight the critical need for targeted nutritional interventions and policies to address malnutrition, particularly in high-prevalence regions. Global trends in malnutrition among children under five were examined in [23]. The study revealed that, despite some progress, malnutrition remains a significant public health issue, with notable disparities across different regions. The authors emphasized the need for targeted interventions to address both undernutrition and overnutrition to improve health outcomes for children worldwide.

Logistic regression is widely used when the dependent variable is binary (i.e., it has only two possible outcomes). In this study, a multivariate logistic regression analysis was conducted to identify risk factors associated with birth defects using population-based surveillance data [24][25]. The research aimed to determine the prevalence of congenital abnormalities and their associated risk factors. The findings provide valuable insights for public health interventions aimed at reducing the incidence of birth defect. The application of logistic regression and other statistical tools in diagnostic biomarker studies was explored in recent research [26][27]. The authors discussed the importance of appropriate statistical methods in the analysis of biomarker data to improve diagnostic accuracy. Their work emphasizes the critical role of logistic regression in evaluating the effectiveness of biomarkers in clinical settings, which May be inaccurate. Additionally, logistic regression is sensitive to multicollinearity among predictor variables, which can lead to unstable parameter estimates and inflated standard errors. Techniques such as variable selection, regularization (e.g., Lassa or ridge regression), and checking for multicollinearity can help mitigate some of the drawbacks associated with logistic regression [28][29]. Multilevel logistic regression extends logistic regression to account for hierarchical or nested data structures commonly encountered in pooled cross-sectional studies, such as individuals nested within households or communities. It allows for the estimation of both within-group and between-group effects, providing more accurate estimates of the relationships between predictors and outcomes. Multilevel logistic regression models can be computationally intensive, especially with large datasets or complex nesting structures. Additionally, specifying the appropriate random effects structure can be challenging, and misspecification may lead to biased parameter estimates. Techniques, such as model simplification, sensitivity analysis, and model comparison, can help address some of the challenges associated with multilevel logistic regression modeling [30].

Measurement error in exposure variables is a significant concern in epidemiological studies, often leading to bias in estimated exposure-outcome associations, yet it is frequently overlooked [31]

Table 2. Estimated prevalence of malnutrition among children under five, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon from 2004 to 2022 (N=750).

Year	Prevalence (Uncorrected)	Prevalence (Corrected)
2004	0.3000	0.1866
2011	0.1047	0.1396
2018	0.2160	0.1839
2022	0.1628	0.1000

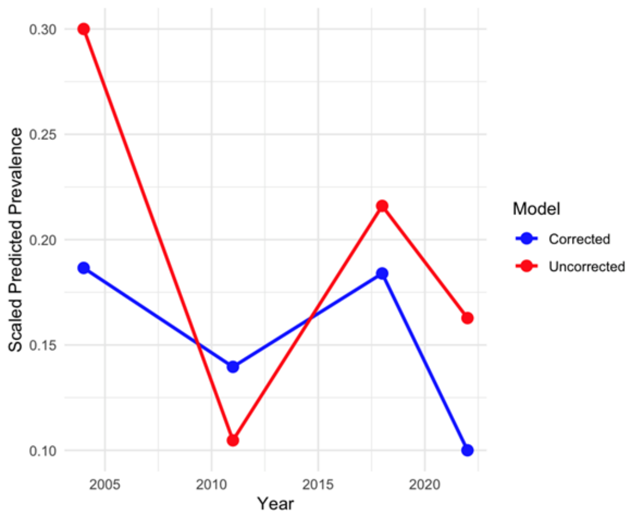

Figure 1. Predicted malnutrition prevalence trends among children under five, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon from 2004 to 2022 (N=750).

[32]. Adjusting for this bias typically requires a validation study with multiple replicates of a reference measurement, which can be costly. When only a single-replicate validation study is available or no internal validation study is conducted, researchers face challenges in using external validation data for bias correction. In accelerometry research, many new accelerometer models lack proper validation in real-world settings, yet they are sometimes used to validate other instruments, such as physical activity questionnaires. This can misrepresent the true validity of these instruments if the accelerometers themselves contain substantial measurement error [33].

Notwithstanding these advancements, significant deficiencies persist in the implementation of Multiple Imputation (MI) for pooled cross-sectional health surveys characterized by measurement error. Although multiple imputation (MI) is well-established for addressing missing data [34][35], its use for rectifying measurement error in covariates—especially in pooled cross-sectional designs—remains constrained. Most current multiple imputation implementations presume that the imputed variables are devoid of error after addressing missing values; hence, they neglect the supplementary bias imposed by mismeasurement [36][37]. Secondly, limited research has included measurement error models directly into the imputation process while considering the hierarchical structure characteristic of pooled survey data (e.g., clustering by area and time). The omission is crucial, since neglecting intra-class correlations may result in underestimated standard errors and erroneous conclusions [38]. Third, while simulation studies have shown the advantages of multiple imputation for missing data, there is a lack of methodological frameworks that concurrently tackle both measurement error and missingness in child malnutrition research within low-resource environments. Current applications in nutritional epidemiology frequently utilize complete-case analysis or single imputation, which inadequately disseminates uncertainty and may provide biased results [39][40]. Consequently, a unified MI-based framework is essential that (1) explicitly incorporates measurement error during the imputation phase; (2) accommodates the multilevel structure of pooled cross-sectional surveys; and (3) offers a reproducible methodology for deriving unbiased and efficient parameter estimates amidst mismeasurement and missing data.

This study enhances the current methods for addressing measurement error and missing data in pooled cross-sectional health surveys by introducing a comprehensive Multiple Imputation (MI) framework that concurrently addresses both types of bias. Although conventional uses of Multiple Imputation (MI) have been concentrated on addressing missing data [34][35], its potential for rectifying measurement error in aggregated survey datasets is still inadequately investigated, especially in resource-constrained environments where data quality is a continual issue. Our methodology incorporates measurement error modeling into the imputation process, facilitating the determination of accurate covariate values and subsequent bias adjustment in logistic regression parameters. This signifies a methodological expansion beyond traditional multiple imputation applications, which often presuppose that the imputed variables are measured without mistake [36][37]. By including a measurement error model during the imputation phase, we offer a more realistic and robust correction mechanism that improves the accuracy and reliability of prevalence estimates and model conclusions. The proposed approach accommodates the hierarchical structure of pooled cross-sectional data by including random effects at both the cluster and time levels to account for intra-class correlation and temporal variability [38][41]. This comprehensive methodology not only overcomes the constraints of conventional logistic regression models that overlook measurement error but also provides a scalable and reproducible analytical instrument for health researchers and policymakers dealing with flawed survey data.

This paper enhances statistical methods by addressing a significant gap in the literature on the analysis of pooled cross-sectional data with measurement error. This research advances our comprehension of malnutrition dynamics through the proposal of unique statistical models and approaches, providing vital insights for academics, policymakers, and public health practitioners in child nutrition and welfare. The subsequent goals are delineated: To develop a logistic regression model incorporating measurement errors for malnutrition analysis; To validate the statistical model through simulations of measurement error scenarios.

The paper is organized as follows: Section 2 presents the data used, the nutritional status of the children, outlines the formulation of the proposed logistic regression model, and explains the application of Multiple Imputation method to correct measurement error and estimate the parameters. In the Section 3, the various results are presented, and the results are discussed. The paper concludes with a general summary in Section 4.

2. MATERIALS AND METHODS

2.1. Study Design

This study employs a quantitative cross-sectional design, utilizing pooled cross-sectional data to analyze malnutrition prevalence among children under five years in Cameroon. The design facilitates the examination of relationships between various predictors and malnutrition outcomes, while accounting for measurement errors through Multiple Imputation (MI).

2.2. Study Setting and Population

The study is conducted in Cameroon, a country in Central Africa characterized by diverse cultural, economic, and geographical conditions. This setting is particularly relevant for examining malnutrition, as Cameroon faces significant challenges related to food security, health care access, and socio-economic disparities. The population targeted in this study includes children under five years of age from various regions in Cameroon, selected through

Table 3. Confusion matrices of logistic regression models predicting malnutrition status among children under five, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon from 2004 to 2022 (N=750).

Prediction / Reference	Reference = 0	Reference = 1
	Corrected Model	
Prediction = 0	20,249	4,725
Prediction = 1	1,725	3,301
	Uncorrected Model	
Prediction = 0	8,734	2,052
Prediction = 1	752	1,444

Table 4. Sensitivity and specificity of corrected and uncorrected logistic regression models.

Model	Sensitivity	Specificity
Corrected Model	0.5488	0.9361
Uncorrected Model	0.4129	0.9207

national health surveys. This demographic is critical for understanding malnutrition trends, as early childhood is a vital period for growth and development.

The selection of children under five is intentional, as this age group is pivotal for growth and development. Malnutrition during these formative years can lead to severe health issues, including stunted growth and developmental delays, which can persist into adulthood. By focusing on this population, the study aims to provide actionable insights that can inform public health strategies aimed at improving child nutrition in Cameroon.

To ensure a representative sample, the study includes participants from various regions across the country, reflecting the diversity of socio-economic conditions. This approach allows for the identification of regional disparities in malnutrition prevalence, helping to highlight areas that may require targeted interventions. The inclusion of a broad geographical spectrum not only enhances the generalizability of the findings but also underscores the importance of tailored public health strategies in addressing malnutrition.

2.3. Variables

The primary dependent variable in this study is the nutritional status of children, specifically measured using the height-for-age Z-score to identify stunting. Stunting is a crucial indicator of chronic malnutrition and reflects long-term nutritional deficiencies affecting physical growth and development. By focusing on this metric, the study aims to provide a clear assessment of malnutrition prevalence and its implications for child health [42] : Independent variables include a range of socio-demographic factors, such as place of residence (urban or rural), parental education levels, and household income. These factors are critical in understanding the broader context of malnutrition, as they can influence dietary practices, access to healthcare, and overall wellbeing. Additionally, health-related factors, including access to healthcare services and dietary diversity, are also considered, as they play a significant role in determining nutritional outcomes.

2.4. Data Resource and Measurement

2.4.1. Data Collection Tool

Data is collected using standardized questionnaires administered during national health surveys. These questionnaires include validated items on socio-demographic characteristics, dietary intake, and health service utilization.

2.4.2. Data Collection

Data collection is performed by trained enumerators who follow strict protocols to ensure accuracy and minimize bias. Enumerators undergo extensive training to familiarize themselves with the data collection instruments and techniques, emphasizing the importance of objectivity and consistency. This training includes mock interviews and field practice to prepare them for real-world data collection scenarios: The surveys employ a combination of direct measurements (e.g., height and weight) and interviews to gather comprehensive information. Direct measurements provide objective data that can be used to calculate nutritional indicators, while interviews capture subjective information about dietary habits and health-seeking behaviors. This mixed-method approach enhances the richness of the data and allows for a more nuanced understanding of malnutrition: To further enhance data quality, quality control measures are implemented throughout the data collection process. These measures include regular supervision of enumerators, random checks of collected data, and follow-up interviews to verify responses. By maintaining rigorous data collection standards, the study aims to ensure that the findings are both valid and reliable, providing a solid foundation for subsequent analysis.

2.4.3. Sample Size

The sample size for this study is determined based on previous health surveys and statistical power analyses, targeting a minimum of 750 children under five years. This sample size is chosen to ensure that the study has sufficient power to detect significant differences in malnutrition prevalence, taking into account potential confounding factors. A larger sample size also enhances the generalizability of the findings across different populations and settings: In addition to providing statistical power, a well-calculated sample size allows for more robust subgroup analyses. By ensuring that the sample is large enough to include various demographics, the study can explore differences in malnutrition prevalence among different groups, such as urban versus rural populations or children from different socio-economic backgrounds. This capability is essential for identifying at-risk populations and informing targeted public health interventions: The determination of the sample size also considers the expected response rate and potential attrition during data collection. By accounting for these factors, the study aims to maintain a high level of precision in its estimates. The careful planning of sample size not only supports the statistical analysis but also reinforces the credibility of the research findings, ensuring that they can inform effective public health strategies.

2.4.4. Data Analysis

Data analysis involves several steps to ensure a comprehensive assessment of malnutrition prevalence and its determinants. The initial phase includes descriptive statistics, which provide an overview of the data characteristics, including means, medians, and standard deviations for continuous variables, as well as frequencies for categorical variables. This descriptive analysis serves as a foundational step for understanding the distribution of key variables in the study: Following the descriptive analysis, Multiple Imputation is employed to address missing data and correct for measurement errors in the covariates. This technique allows for the estimation of true values based on observed data, thereby enhancing the reliability of the subsequent analyses. By generating multiple imputed datasets, the study ensures that the uncertainty associated with missing data is accounted for in the final analyses, leading to more robust findings. After imputation, a logistic regression model is fitted to each imputed dataset to examine the relationship between predictors and malnutrition outcomes. This modeling approach allows for the assessment of how different factors contribute to malnutrition prevalence, while also correcting for potential biases introduced by measurement errors. Finally, model performance is evaluated using metrics such as accuracy, precision, and confusion matrices, which compare the results of corrected and uncorrected models. This comprehensive data analysis framework not only enhances the validity of the findings but also provides actionable insights for public health practitioners.

2.4.5. Data Sources

This study evaluates the impact of multiple imputation correction on the estimation of the logistic regression parameters on the prediction of the prevalence of malnutrition [43]. The dataset was

Table 5. Sensitivity analysis of Multiple Imputation performance under varying levels of missingness (MAR mechanism). Results are averaged over 500 simulation replications.

Missingness Level	Method	Estimation		Classification		
		Bias	RMSE	Sensitivity	Specificity	AUC
10 %	MI	0.015	0.082	0.792	0.861	0.889
	CCA	0.048	0.114	0.756	0.842	0.864
25 %	MI	0.022	0.095	0.781	0.854	0.880
	CCA	0.073	0.138	0.742	0.831	0.851
40 %	MI	0.041	0.122	0.763	0.846	0.867
	CCA	0.101	0.165	0.725	0.819	0.838

Table 6. Accuracy trends of logistic regression models predicting malnutrition among children under five, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon from 2004 to 2022 (N=750).

Year	Accuracy (Uncorrected)	Accuracy (Corrected)
2004	0.7846	0.8183
2011	0.7729	0.8187
2018	0.7956	0.8319
2022	0.7827	0.8247

simulated to mimic real-world conditions, incorporating measurement errors and missing values, which were addressed using multiple imputation. The performance of corrected and uncorrected models was assessed through estimated parameters, prevalence trends, and classification metrics. A simulated dataset was created with nregions = 10 regions, nchildren = 750 children per region, and four time points (2004, 2011, 2018, 2022). The true covariates Xtrue were generated from a standard normal distribution, and observed covariates Xobs were computed with added noise V ~ N(0, σv). The probability of malnutrition pijt was modeled using Equation (1):

$$\eta = \beta_0 + X_{true}\beta, \quad (1)$$

where η is the linear predictor. This is a standard specification in logistic regression for binary outcomes and is widely used in modern epidemiological and health modeling literature (Equation 2) [43].

$$p_{ijt} = \frac{\exp(\eta)}{1 + \exp\,(\eta)} \quad (2)$$

This logit link function maps the linear predictor to the [0, 1] interval and is commonly applied in health statistics and multivariable risk modeling [41]. The outcome variable (Y) was generated using a logistic function with a true intercept of -1.5 and coefficients randomly sampled between 0.3 and 0.8.

2.5. Justification for Using Multiple Imputation Over Alternative Correction Approaches

Multiple Imputation was chosen above other measurement error correction techniques for many critical reasons, corresponding to the particular constraints presented by pooled cross-sectional health survey data. There are other methods available, such as regression calibration, simulation extrapolation (SIMEX), and complete Bayesian modeling, but Multiple Imputation has some unique advantages in this case. Initially, Multiple Imputation offers a versatile framework that concurrently addresses both missing data and measurement error, which frequently co-occur in survey-based health research. In contrast to regression calibration, which rectifies measurement error without necessarily addressing missing data [44], Multiple Imputation may incorporate a measurement error model into the imputation process, thereby simultaneously correcting for both types of bias. This comprehensive strategy mitigates the possibility of skewed estimations that may result from addressing these concerns in isolation.

Secondly, Multiple Imputation inherently transmit uncertainty from both absent data and measurement inaccuracies to the ultimate parameter estimations and their standard errors. Techniques like single imputation or complete-case analysis diminish variance and may result in exaggerated type I error rates [45]. Conversely, Multiple Imputation addresses both within-imputation and between-imputation variability, facilitating effective statistical inference under the missing-at-random (MAR) assumption and specific measurement error frameworks. Third, Multiple Imputation is computationally feasible and effectively accommodates intricate hierarchical data structures, such as those seen in pooled cross-sectional surveys with regional and temporal clustering. Although fully Bayesian approaches provide comparable advantages, they frequently need the specification of prior distributions and entail more complex computations [46]. Multiple Imputation, executed by chained equations, provides a practical and accessible approach that may be utilized with common statistical tools without requiring considerable previous specification.

Fourth, Multiple Imputation endorses the utilization of auxiliary variables to enhance imputation precision, which is especially advantageous when external validation data or replication measurements are scarce: a frequent occurrence in resource-constrained environments. By using correlations among observable variables, Multiple Imputation can provide more precise imputations than approaches that depend exclusively on unreliable covariates [47]. Ultimately, Multiple Imputation has been extensively verified in epidemiological and health research, accompanied by well-defined standards for its application and interpretation. The usage of this method in nutritional epidemiology is increasing, however its use to rectify measurement error in pooled cross-sectional designs remains little investigated. Considering these benefits, Multiple Imputation is a strong, adaptable, and methodologically rigorous option for rectifying measurement error and addressing missing data in the current investigation.

2.6. Methodological Contribution

Although Multiple Imputation is a recognized method for addressing missing data, its use for rectifying measurement error in pooled cross-sectional surveys is still constrained. This study enhances the conventional Multiple Imputation framework by including a measurement error model directly into the imputation process, enabling the concurrent rectification of both missing data and measurement inaccuracies: This comprehensive method is especially appropriate for hierarchical survey data, as it incorporates cluster-level and temporal heterogeneity via random effects. This is one of the initial uses of MI-based measurement error correction in the study of child malnutrition with pooled DHS-style data. The proposed paradigm provides a pragmatic and scalable approach to enhancing the accuracy of prevalence estimates and regression inferences in contexts lacking gold-standard measures.

2.7. Model Formulation

Researchers may define the nutritional status of children in various ways. A child's nutritional status is typically determined using several measurements, including height, weight, gender, and age. The three commonly used indicators of nutritional status are height-for-age (stunting), weight-for-height (wasting), and weight-for-age (underweight), as proposed in [48][49]. These indicators are expressed as Z-scores relative to the median of a reference population: In our study, we use ”height-for-age” (stunting) as the dependent variable, coded as ’1’ if the child is classified as stunted and ’0’ otherwise. Height-for-age is an indicator that compares an individual's height with the expected height for their age, based on standardized growth charts. This metric primarily assesses chronic malnutrition, reflecting long-term nutritional deficiencies that affect growth over time. A child whose height is significantly below the expected range is considered stunted, indicating inadequate nutrition during critical periods of growth. This measure is also informative of current nutritional status as it relates body mass to body length and requires accurate knowledge of the child's age. It complements ”weight-for-age,” an indicator used in [51][52].

Table 7. Precision of uncorrected and corrected logistic regression models identifying malnutrition among children under five, based on simulated pooled cross-sectional survey data reflecting malnutrition in Cameroon from 2004 to 2022 (N=750).

Year	Precision (Uncorrected)	Precision (Corrected)
2004	0.8098	0.8333
2011	0.7974	0.8288
2018	0.8221	0.8446
2022	0.8095	0.8352

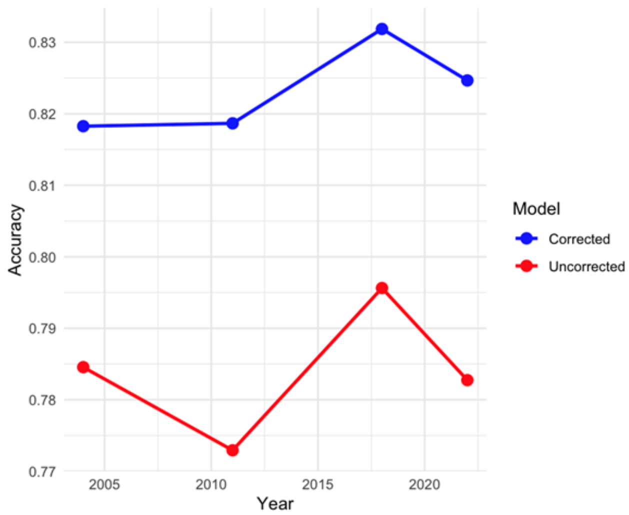

Figure 2. Classification accuracy of uncorrected and corrected logistic regression models predicting malnutrition among children under five, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon from 2004 to 2022 (N=750).

Furthermore, height-for-age helps determine whether a child has recently received sufficient nutrients to support and maintain body weight, and it reflects broader influences such as genetic potential, environmental conditions, and disease burden, as emphasized by the World Health Organization (WHO). To define child malnutrition, we followed the guidelines of the Bangladesh National Report and the World Health Organization [53][54]. We consider several covariates as predictor variables that are commonly reported in nutritional studies of children. Some of them (categories are given in brackets) are place of residence (urban, rural), partner's education (no education, primary, secondary, higher, don't know), partner's occupation (did not work, prof., tech. manag., sales, agric-employed, household/domestic, service, skilled manual, unskilled manual), type of toilet facility (flush toilet, rudimentary pit toilet latrine, improved pit toilet latrine, no facility, other, not dejure residence), source of drinking water (in the house, in the courtyard, along the road, well with pump, well, protected without pump, river/ stream not protected, protected source, rainwater, other, not dejure residence), religion (Protestant, Catholic, Animist, Muslim, other, no religion, new religion (rebellious religion)), Gave child baby formula (from 1 time to 10 times by week), Gave child food (eggs, fish, poultry, meat), entries in maternity table (from 0 to 5 times by week), entries in health table (from 0 to 5 times by week), husband lives in house (living with her, staying elsewhere), sex of child (male or female), antenatal visits of pregnancy (from 0 to 5 times by month), Entries in height/weight (from 0 to 5 times by week).

To model the malnutrition of children under five years in the presence of measurements errors using Logistic regression model, we consider Y_{ijk}, which represent the malnutrition status of child i in region j during survey conducted at time t. Our logistic measurement-error regression with data revision has the Formula (3):

$$\begin{cases} \operatorname{logit}(P(Y_{ijt}=1)) = \log\left(\frac{P(Y_{ijt}=1)}{1-P(Y_{ijt}=1)}\right) = \beta_0 + \beta_1 X^*_{1jt} + \beta_2 X^*_{2jt} + \cdots + \beta_k X^*_{kjt} + u_c + w_t, \\ X^*_{ijt} = X_{ijt} + v_{ijt}, \quad v_{ijt} \sim \mathcal{N}(0, \sigma_v^2). \end{cases} \quad \textbf{(3)}$$

with X_{1jt}, X_{2jt}, . . . ,X_{kjt} represents individual child characteristics such as age of child, weight of child, time of gave child food, entries in child health , antenatal care, entries in health table, entries in pregnancy and postnatal care, etc. The independent variable $X_{\underline{ijt}}$ is affected by a measurement error ϵijt, which is unobserved. β_0 is the intercept, this is the constant term in the model that represents the average effect on the logit of the probability of success when all independent variables are zero. β_1, β_2, . . . , β_k : These are the parameters of the model to be estimated from the data. Each coefficient β_k represents the contribution of the corresponding variable X^*_{ijt} to the logit of the probability of success; X^*_{ijt} are the observed values of X_{ijt} corrected for measurement error. u_c : Captures variability at the cluster level. w_t: Captures variability across time periods; v_{ijt} : These are the measurement errors associated with each independent variable.

Measurement errors are differences between the actual values of the variables and the measured or observed values. $P(Y_{ijt} = 1)$: This is the probability that the dependent variable (Yijt) will take the value 1 (success) for a given observation. In logistic regression, this probability is modelled using the logistic (or logit) function, which is a transformation of the probability so that it varies between 0 and 1. The model developed, called LRM-DR, would make it possible to model the relationship between the individual characteristics of children, and the nutritional status of children, while taking account of measurement errors in the explanatory variables.

2.8. Statistical Method

Measurement error can significantly bias regression coefficient estimates β and inflate standard errors, leading to incorrect inferences. To address this issue, Multiple Imputation is applied to estimate the true values of covariates and obtain unbiased estimates of the logistic regression parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_8$.

2.8.1. Measurement Error Model

We assume that the observed independent variables contain measurement error, which can be modeled as follows Equation (4);

$$X^*_{ijt} = X_{ijt} + v_{ijt}, \quad v_{ijt} \sim \mathcal{N}(0, \sigma_v^2), \tag{4}$$

where, X^*_{ijt} represents the observed values, which are affected by measurement error. X_{ijt} denotes the true but unobserved values of the covariates, while v_{ijt} is the random measurement error, assumed to follow a normal distribution with mean zero and variance σ^2_v.

2.8.2. Imputation of True Covariates

To estimate the true values X_{ijt}, we employ Multiple Imputation, modeling the expectation of the true values given the observed data as Equation (5).

$$E[X_{ijt} | X^*_{ijt}] = \gamma_0 + \gamma_1 X^*_{ijt}. \tag{5}$$

This formulation is based on classical measurement error models and regression calibration techniques [55][56]. For each imputation m, the true values are generated as Equation (6).

$$X^{(m)}_{ijt} = \gamma_0 + \gamma_1 X^*_{ijt} + \eta^{(m)}_{ijt}, \quad \eta^{(m)}_{ijt} \sim \mathcal{N}(0, \sigma_\eta^2). \tag{6}$$

This process accounts for the uncertainty in estimating the true covariates, following the framework described in [34][56].

2.8.3. Logistic Regression with Imputed Data

After imputation, we fit a logistic regression model to each imputed dataset (Equation (7));

$$\text{logit}(P(Y_{ijt} = 1)) = \beta_0 + \sum_{k=1}^{K} \beta_k X^{(m)}_{kjt} + u_c + w_t, \tag{7}$$

Table 8. Mean estimates of logistic regression coefficients (β_0 to β_8) across increasing sample sizes, based on simulated pooled cross-sectional survey data reflecting malnutrition prevalence in Cameroon N varies by scenario).

Sample Size	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
50	-1.2955	0.3288	0.4580	0.4962	0.5997	0.5549	0.2109	0.3608	0.5444
100	-1.3076	0.3102	0.4721	0.3912	0.4721	0.5138	0.2631	0.4459	0.5252
200	-1.3320	0.3257	0.5028	0.4034	0.4984	0.5977	0.2302	0.4148	0.5451
300	-1.3372	0.3106	0.4857	0.3867	0.5490	0.5435	0.2317	0.3820	0.5376
400	-1.3902	0.3253	0.5049	0.3571	0.5338	0.5324	0.2700	0.4683	0.5537
500	-1.3595	0.3440	0.5356	0.3278	0.5422	0.5514	0.2350	0.4015	0.5105
600	-1.3130	0.3239	0.5061	0.3588	0.5502	0.5386	0.2575	0.3678	0.5376
750	-1.3766	0.3189	0.5193	0.3654	0.5564	0.5428	0.2453	0.4109	0.5341

Table 9. Standard deviations of logistic regression coefficients (β_0 to β_8) across increasing sample sizes, based on simulated pooled cross-sectional data modeling malnutrition among children under five in Cameroon (N varies by scenario).

Sample Size	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
50	0.0657	0.0534	0.0551	0.0564	0.0568	0.0567	0.0523	0.0534	0.0535
100	0.0461	0.0381	0.0379	0.0371	0.0383	0.0392	0.0374	0.0374	0.0382
200	0.0332	0.0267	0.0268	0.0273	0.0273	0.0278	0.0266	0.0270	0.0275
300	0.0269	0.0213	0.0222	0.0217	0.0224	0.0226	0.0214	0.0218	0.0224
400	0.0239	0.0188	0.0194	0.0189	0.0199	0.0194	0.0187	0.0193	0.0197
500	0.0210	0.0167	0.0174	0.0170	0.0175	0.0177	0.0167	0.0169	0.0173
600	0.0189	0.0151	0.0159	0.0154	0.0159	0.0159	0.0150	0.0153	0.0160
750	0.0205	0.0172	0.0201	0.0184	0.0195	0.0212	0.0189	0.0202	0.0193

where, $u_c \sim N(0, \sigma^2_u)$ captures variability at the cluster level; $w_t \sim N(0, \sigma^2_w)$ accounts for time-period variability. This hierarchical logistic model is a standard approach in multilevel analysis of health data [57][58].

2.8.4. Pooling Estimates from Imputed Datasets

Since multiple datasets are generated through imputation, the logistic regression results must be combined to obtain final parameter estimates. For each imputed dataset *m*, we obtain an estimate $\beta^{(m)}$ of the regression parameters. The final pooled estimate is computed as Equation (8).

$$\bar{\beta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}^{(m)}. \tag{8}$$

This pooling rule follows Rubin's rules for multiple imputation [34][55]. This represents the average of the estimates across the M imputed datasets. To estimate the regression coefficients β *(m)*, we use maximum likelihood estimation (MLE) applied to logistic regression [58]. The model is expressed as Equation (9).

$$\operatorname{logit}(P(Y_{ijt}=1)) = \beta_0 + \sum_{k=1}^{K}\beta_k X_{kjt}^{(m)} + u_c + w_t, \tag{9}$$

The likelihood function for logistic regression in the m-th imputed dataset is given by Equations (10) and (11);

$$L^{(m)}(\beta) = \prod_{i=1}^{N} P_i^{Y_i}(1-P_i)^{1-Y_i}, \tag{10}$$

where:

$$P_i = \frac{\exp\left(\beta_0 + \sum_{k=1}^{K}\beta_k X_{ki}^{(m)} + u_c + w_t\right)}{1+\exp\left(\beta_0 + \sum_{k=1}^{K}\beta_k X_{ki}^{(m)} + u_c + w_t\right)} \tag{11}$$

is the predicted probability that Yi = 1. Taking the logarithm of the likelihood function, we obtain the Equation (12):

$$\ell^{(m)}(\beta) = \sum_{i=1}^{N}\left[Y_i \log P_i + (1-Y_i)\log(1-P_i)\right]. \tag{12}$$

The estimated coefficients ˆ β(m) are obtained by solving the optimization problem (Equation (13)):

$$\hat{\beta}^{(m)} = \arg\max_{\beta} \ell^{(m)}(\beta). \tag{13}$$

Since the log-likelihood function is nonlinear, we solve Equation (14):

$$\frac{\partial \ell^{(m)}(\beta)}{\partial \beta} = 0. \tag{14}$$

These are standard results in logistic regression modeling [58]. After estimating ˆ β(m) for each imputed dataset, the final pooled estimate is computed as Equation (15):

$$\bar{\beta} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}^{(m)}. \tag{15}$$

This approach ensures that missing data uncertainty is properly incorporated into the final model. The variance of the pooled estimate. β: is computed using Rubin's variance formula [34].

2.8.4.1. Within-Imputation Variance

This measures the average variance of the parameter estimates across the M datasets (Eq. 16):

$$W = \frac{1}{M}\sum_{m=1}^{M} \mathrm{Var}(\hat{\beta}^{(m)}). \quad \textbf{(16)}$$

2.8.4.2. Between-Imputation Variance

This quantifies the variability in the parameter estimates across the different imputations (Eq. 17).

$$B = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\beta}^{(m)} - \bar{\beta})^2. \quad \textbf{(17)}$$

2.8.4.3. Total Variance

This accounts for both within-imputation and between-imputation variance, capturing total uncertainty in the estimated parameters (Eq. 18).

$$\mathrm{Var}(\bar{\beta}) = W + \left(1 + \frac{1}{M}\right)B. \quad \textbf{(18)}$$

Confidence intervals for the estimated parameters are derived as Equation (19).

$$\bar{\beta} \pm t_\nu \sqrt{\mathrm{Var}(\bar{\beta})}, \quad \textbf{(19)}$$

where, t_v is the critical value from the t-distribution with degrees of freedom v. The degrees of freedom are computed by Equation (20):

$$\nu = \frac{(M-1)\left(1 + \frac{W}{(1+1/M)B}\right)^2}{1 + \frac{M+1}{M}\frac{W}{B}}. \quad \textbf{(20)}$$

These expressions follow the original variance decomposition and degrees of freedom approximations proposed by Rubin [34][55].

After pooling the estimates, the final corrected logistic regression model is Equation (21):

$$\mathrm{logit}(P(Y_{ijt} = 1)) = \bar{\beta}_0 + \sum_{k=1}^{K} \bar{\beta}_k X_{kjt} + u_c + w_t, \quad \textbf{(21)}$$

where, β_k are the corrected regression coefficients; $u_c \sim N(0, \sigma^2_u)$ captures cluster-level variability; $w_t \sim N(0, \sigma^2_w)$ accounts for time-period variability. Which incorporates uncertainty from missing and mismeasured data into the parameter estimates [55][56].

2.8.5. Multiple Imputation Procedure

The analysis addressed measurement error and missing data with regression calibration-based Multiple Imputation. The imputation model assumes additive normal error for observed covariates (X^*_{ijt}) and a linear relationship between latent and observable variables. The model is explained in Equation (6). This regression calibration method works for continuous variables impacted by classical measurement error and follows literature recommendations. For consistent and efficient parameter estimation, Rubin's methods created eleven imputed datasets. We used a logistic regression model with cluster-level (u_c) and year-level (w_t) random effects to assess each imputed dataset. We computed regression coefficient aggregates using Rubin's methodology. These rules calculated standard errors and confidence ranges using inside- and across-imputation variances. Imputation and analysis were in R (4.5.1). Preliminary testing of the mice package was done before final Multiple Imputation using bespoke R scripts for regression calibration.

Table 10. Coefficient of variation (CV) for logistic regression coefficients (β_0 to β_8) across increasing sample sizes, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon (N varies by scenario).

Sample Size	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
50	5.07	16.24	12.03	11.37	9.47	10.22	24.78	14.81	9.83
100	3.52	12.29	8.03	9.48	8.12	7.63	14.21	8.39	7.27
200	2.49	8.19	5.33	6.77	5.48	4.65	11.56	6.51	5.05
300	2.01	6.86	4.57	5.61	4.08	4.15	9.24	5.71	4.17
400	1.72	5.78	3.84	5.29	3.73	3.64	6.92	4.12	3.56
500	1.54	4.85	3.25	5.19	3.23	3.21	7.11	4.21	3.39
600	1.44	4.66	3.14	4.29	2.89	2.95	5.82	4.16	2.98
750	1.49	5.39	3.87	5.03	3.51	3.91	7.71	4.91	3.61

The model was estimated and tested using lme4, stats, caret, and pROC.

Checking input values with typical diagnostics. Trace plots of parameter estimations during imputations showed convergence, density graphs showed imputed and actual distributions for plausibility, and Monte Carlo errors measured numerical stability. Corrected and uncorrected models were compared for prevalence, classification accuracy, and coefficient stability to measure Multiple Imputation. Because demographic and socioeconomic factors were obvious confounders, we assumed the data were ”missing at random” (MAR). Missingness is frequently connected to observed characteristics like residency and maternal education, which were incorporated in the imputation model. Our MAR-based simulation method examines Multiple Imputation in a controlled and straightforward manner, unlike genuine DHS data with more complex missingness patterns. Regression calibration was selected over MICE to correct measurement error in continuous variables without imputing missing values. MICE is good for general-purpose imputation but not systematic measurement error, which regression-based techniques manage.

2.9. Simulation of Varying Sample Sizes

We utilized the subsequent values to generate simulated datasets with increased sample sizes: n = 50, 100, 200, 300, 400, 500, 600, 750. This was conducted to examine the impact of sample size on the stability of parameter estimates. In each instance, an identical methodology was employed to generate the data, and the actual regression coefficients were predetermined. An exhaustive analysis was performed on the estimations obtained from each scenario to discern trends in mean bias and parameter convergence. The system’s design enabled us to evaluate the accuracy and consistency of the estimation approach with varying data volumes.

2.10. Algorithm Evaluation

On the test dataset, we evaluated model performance using a confusion matrix. Confusion matrix determined model accuracy, sensitivity, and specificity. In a binary classification framework, the following assessment metrics are calculated for the four outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

2.10.1. Accuracy

Accuracy is the proportion of the total number of correct predictions (TP + TN) to the total number of predictions (TP + TN + FP + FN). It reflects the overall ability of the classifier to make correct decisions. Formally, it is given by **Equation (22)**:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

2.10.2. Sensitivity

Sensitivity, also known as recall or the true positive rate, measures the proportion of actual positive cases correctly identified by the model. It represents the classifier’s ability to detect positive instances (e.g., individuals with malnutrition). A highly sensitive classifier has a low rate of false negatives. Sensitivity is defined as Equation (23):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{23}$$

2.10.3. Specificity

Specificity, or the true negative rate, measures the proportion of actual negative cases correctly classified. It reflects the classifier’s ability to identify individuals without the condition (e.g., well -nourished children) as negative. A highly specific model results in a low rate of false positives. Specificity is calculated as Equation (24):

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{24}$$

2.10.4. Area Under the Curve - Receiver Operating Curve (AUC-ROC)

AUC-ROC-unit square graphs—show classifier diagnostic performance. A plot of true positive rate (sensitivity) and false positive rate (1 − specificity) is created for different thresholds. To measure classifier performance, the area under the curve (AUC) estimates the area beneath the ROC curve to aggregate classification results over all thresholds. AUCs 0–1. A 0.5 AUC classifier is unpredictable and unreliable. Classifiers with 1.0 AUCs correctly identify positive and negative cases. The classifier improves as AUC approaches 1. With this metric, classifier discrimination comparisons are robust and threshold independent.

3. RESULTS AND DISCUSSIONS

3.1. Results

3.1.1. Descriptive Data

The study simulated data from 10 regions in Cameroon, each with 750 children observed in 2004, 2011, 2018, and 2022. It included demographic, socioeconomic, and healthcare-related variables, with deliberate measurement errors to mimic real survey challenges. Malnutrition was measured using height-for-age (stunting), coded as 1 for stunted and 0 otherwise. The Multiple Imputation technique employed in this work is adaptable to actual DHS datasets because to its compatibility with intricate survey designs using MICE imputation methods. Utilizing DHS data, we will implement survey design elements such as sample weights, clustering, and stratification during imputation and analysis, adhering to weighted multiple imputation guidelines for complex surveys. This addresses within-cluster correlation and ensures survey representativeness in imputed values and parameter estimates. We elucidated this aspect in the Discussion section and outlined practical methodologies for implementing the methodology on actual DHS data. Logistic regression with multiple imputation was used to address missing data and correct measurement errors. The simulation accounted for regional and time variability, reflecting real-world survey heterogeneity and allowing a robust assessment of multiple imputation's impact on model performance.

We elucidate the original variables derived from the aggregated Demographic and Health Surveys (DHS) conducted in Cameroon in 2004, 2011, 2018, and 2022 to enhance the clarity and replicability of the study. The study examined children under five, using criteria linked to child nutrition, including household, demographic, and health data. Table 1 presents a summary of the most significant variables. The binary outcome variable, stunted, indicates that a child's height-for-age Z-score falls below −2 standard deviations. The mean prevalence of stunting during the duration of the survey was 34%. Approximately 39% of children resided in urban areas. The mean educational attainment for women was 6.8 years, although the average wealth index for households approximated the national median (3.0 out of 5.0). Each record had an average of 2.7 entries in the pregnancy portion and 2.5 entries in the healthcare section of the questionnaire. Additional confounders include categorical variables such as the mother's age category, the partner's occupation, the kind of

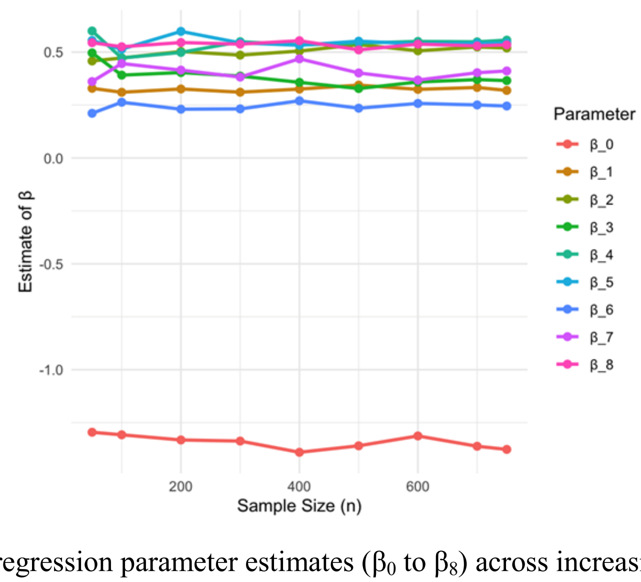

Figure 3. Variability in regression parameter estimates (β_0 to β_8) across increasing sample sizes, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon using a Multiple Imputation framework from 2004 to 2022 (N varies by scenario).

sanitation facility, and the source of potable water. These represent prevalent indicators of DHS.

3.1.2. Outcome Data

Table 2 presents the predicted malnutrition prevalence from both corrected and uncorrected models, with the corrected model consistently reporting lower values, indicating that the uncorrected model tends to overestimate prevalence. For example, in 2004, the corrected prevalence was 0.1866 compared to 0.3000 in the uncorrected version. This pattern held across all observed years. The trend persisted across years, as illustrated in Figure 1, which displays the prevalence estimates from both models over time. The 2011 anomaly, where the adjusted prevalence (0.1396) slightly exceeds the unadjusted figure (0.1047), may be ascribed to imputation modifications that rectify underreporting or data inaccuracies for that year. This highlights the need of accounting for the uncertainty caused by measurement error, hence enhancing the precision and clarity of our evaluation about the effects of the imputation-based correction approach.

The corrected model demonstrates superior performance in detecting malnourished children (sensitivity: 0.5488 vs. 0.4129) and accurately identifying non-malnourished children (specificity: 0.9361 vs. 0.9207), as evidenced by the confusion matrices (Table 3) and the corresponding sensitivity and specificity values (Table 4). Table 5 presents the results of a sensitivity analysis for three levels of missingness employing the MAR mechanism. MI consistently surpassed CCA in terms of bias, RMSE, and classification accuracy. Multiple Imputation exhibited satisfactory performance with moderate (25%) and high (40%) levels of missing data; however, sensitivity and AUC diminished under conditions of extreme missing data. Multiple Imputation demonstrated superior estimation and differentiation in all instances compared to CCA. The enhancements have illuminated the efficacy of measurement error correction using multiple imputation in enhancing model performance across both classification dimensions. The confusion matrices (Table 3) show that the corrected model achieved better sensitivity and specificity, reducing both false positives and false negatives.

The accuracy and precision results (Tables 6 and 7) affirm the superior performance of the corrected model. Across all years, accuracy improved by approximately 3–4%, and precision increased to values above 0.83. This suggests that the model with multiple imputation was more reliable in identifying true cases of malnutrition, as further illustrated in Figure 2, which compares classification accuracy between corrected and uncorrected models over time.

3.1.3. Main Results

As seen in Table 8, increasing the sample size led to more stable parameter estimates. Coefficients such as β_2, β_4, and β_5 became more consistent with larger samples, demonstrating the convergence of estimates as data volume increased. This pattern is visually supported by Figure 3, which shows the decreasing variability in regression parameter estimates across sample sizes. Figures 4 and 5 show the trends in standard deviation and coefficient of variation, respectively. These metrics declined steadily with larger sample sizes, highlighting improved precision. However, a slight fluctuation at sample size 750 suggests potential noise or computational sensitivity. Overall, the corrected model using Multiple Imputation yielded better prevalence estimates, stronger classification performance, and more stable parameter estimates across sample sizes. This validates the use of imputation methods in epidemiological modeling with noisy data.

3.2. Discussions

3.2.1. Key Results

This study builds on established knowledge in the field of health data analysis by recognizing several limitations in traditional methods applied to pooled cross-sectional data, particularly in low-resource settings: Pooled cross-sectional data are commonly used in health research but are subject to missing data and measurement errors, particularly in low-resource settings; Traditional logistic regression models often fail to correct for these errors, leading to biased prevalence estimates and unreliable conclusions. Multiple Imputation is an established method for addressing missing data but

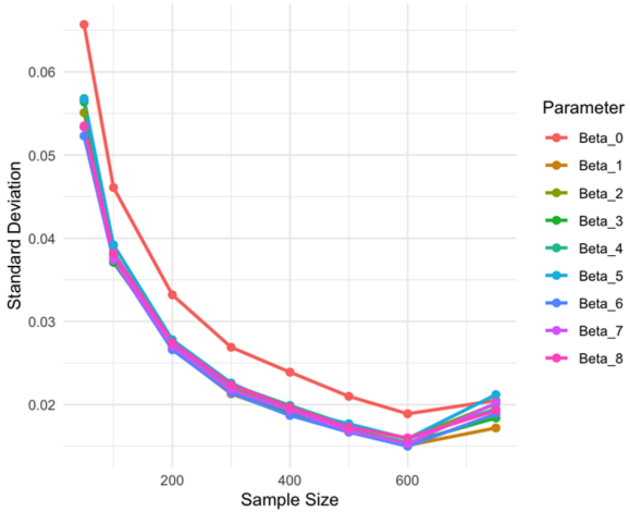

Figure 4. Standard deviation trends of logistic regression coefficients (β_0 to β_8) across increasing sample sizes, based on simulated pooled cross-sectional data on malnutrition among children under five in Cameroon from 2004 to 2022 (N varies by scenario).

is less frequently applied to correct for measurement error in pooled survey datasets.

In response to these challenges, this study contributes novel insights and practical advancements; Demontrates that Multiple Imputation significantly improves the accuracy and reliability of malnutrition prevalence estimates in pooled cross-sectional data. It Shows that MI-corrected models outperform uncorrected models in classification accuracy, precision, and parameter stability. It Provides a robust and replicable framework for integrating Multiple Imputation into logistic regression analysis of health survey data, supporting evidence-based policy decisions.

The findings of this study confirm that correcting for measurement error using Multiple Imputation significantly improves model performance. The corrected logistic regression model consistently outperformed the uncorrected version across all four survey years, with accuracy values ranging from 0.8183 to 0.8319. In contrast, the uncorrected model exhibited lower accuracies between 0.7729 and 0.7956. These gains of approximately 3–4 percentage points align with theoretical expectations and empirical results reported in the literature. For instance, Dong and Peng [59] observed improvements in classification performance of logistic regression models when MI was applied to simulated and empirical datasets. Similarly, White et al. [35] reported that MI-corrected models produced more reliable and less biased estimates, particularly in scenarios with moderate to high levels of missing or mismeasured data.

The findings of this study confirm that correcting for measurement error using Multiple Imputation significantly improves model performance. The corrected logistic regression model consistently outperformed the uncorrected version across all four survey years, with accuracy values ranging from 0.8183 to 0.8319. In contrast, the uncorrected model exhibited lower accuracies between 0.7729 and 0.7956. These gains of approximately 3–4 percentage points align with theoretical expectations and empirical results reported in the literature. For instance, Dong and Peng [59] observed improvements in classification performance of logistic regression models when MI was applied to simulated and empirical datasets. Similarly, White et al. [35] reported that MI-corrected models produced more reliable and less biased estimates, particularly in scenarios with moderate to high levels of missing or mismeasured data.

By comparison, studies that did not employ Multiple Imputation tended to report lower predictive accuracy. Higazi et al. [19], in their analysis of malnutrition in developing countries using a linear mixed model without Multiple Imputation, reported classification accuracies ranging from 0.72 to 0.78. These values are consistent with the uncorrected accuracy levels in the present study, suggesting that unaddressed measurement error may lead to inflated or deflated model estimates. This reinforces the notion that traditional logistic regression models, when applied directly to error-prone health data, can misrepresent the true relationship between covariates and outcomes.

The alignment of our corrected model's performance with that of previous MI-based studies further supports the robustness of our modeling approach. Little and Rubin [55] emphasized the importance of incorporating uncertainty from imputed data into parameter estimates, a feature that is integral to Multiple Imputation. The present study's findings demonstrate the practical benefits of applying Multiple Imputation not only for handling missing data but also for correcting measurement error, ultimately yielding more accurate estimates of malnutrition risk. Therefore, the integration of Multiple Imputation into logistic regression modeling should be encouraged in health research using large-scale, cross-sectional datasets such as the DHS.

The results from this study clearly demonstrate that correcting for measurement error using Multiple Imputation significantly improves the performance of logistic regression models applied to malnutrition prevalence data. As shown in Table 2, the corrected model consistently reports lower prevalence estimates across all years, with the most notable correction observed in 2004, where prevalence drops from 0.3000 (uncorrected) to 0.1866 (corrected). This suggests that the uncorrected model tends to overestimate malnutrition prevalence due to unaddressed measurement errors. Further, the classification performance of the models is substantially enhanced through correction. From Table 3, the corrected model yields higher true positive and true negative counts, with reduced false classifications.

The sensitivity is notably better, as the number of correctly identified malnourished children increases from 1,444 (uncorrected) to 3,301 (corrected). This translates into higher precision and accuracy, as shown in Tables 6 and 7. For instance, in 2018, accuracy improves from 0.7956 to 0.8319, and precision increases from 0.8221 to 0.8446. Another important finding is the role of sample size in stabilizing the estimates. Table 8 shows that parameter estimates (e.g., $\beta 2$, $\beta 4$, $\beta 5$) tend to converge and become more consistent as sample size increases. This trend is corroborated by the reduction in standard deviation (Table 9) and coefficient of variation (Table 10), which both decline steadily from small (n=50) to larger samples (n=600). Thus, the use of Multiple Imputation not only corrects bias but also enhances stability in model estimation.

3.2.2. Practical Magnitude and Public Health Implications

The noted enhancements in model performance, albeit statistically significant, had substantial practical consequences for public health decision-making and resource allocation. The revised model attained an average enhancement in classification accuracy of around 3–4 percentage points across the survey years (Table 6), accompanied by precision gains of a comparable scale (Table 7). While these benefits may seem little in absolute terms, they result in considerable effects when applied to population-level malnutrition screening and intervention strategies. In a population of one million children under five, a 3% enhancement in classification accuracy equates to around 30,000 more children accurately recognized as either starving or well-nourished. The decrease in misclassification, especially the reduction in false negatives, guarantees that children requiring nutritional assistance are more likely to get prompt interventions while preventing the misallocation of limited resources to false positives. Precision is essential in low-resource environments such as Cameroon, where limited public health funds necessitate targeted efficiency, which directly influences program performance [60][61].

The adjusted prevalence estimates were consistently lower than the unadjusted figures (Table 2), with the most significant absolute decrease occurring in 2004 (from 0.3000 to 0.1866). The 11.34 percentage point decrease signifies a relative drop of around 38% in the projected prevalence of malnutrition for that year. This recalibration directly affects burden assessment, financing requests, and the monitoring of Sustainable Development Goal (SDG) indicators. Exaggeration of malnutrition prevalence can result in an inflated perception of necessity, misallocated financing, and exaggerated advancements in later assessments [62][63]. In contrast, enhanced accuracy in estimations allows improved prioritizing of high-burden areas and more dependable monitoring of changes over time.

The stability of parameter estimates with larger sample sizes (Tables 8 – 10, Figures 3 – 5) further corroborates the efficacy of the proposed Multiple Imputation framework in longitudinal survey research. The variability in sample size across rounds of demographic and health surveys necessitates the production of reliable and comparable estimates to bolster the credibility of cross-year comparisons and policy assessments. This is especially pertinent for national and subnational health planning, as choices are frequently guided by trends obtained from aggregated survey data [64][65]. The methodological enhancements presented herein transcend mere statistical advancements, providing concrete advantages for public health practice: more precise burden estimates, better intervention targeting, and increased dependability of trend studies. The future use of MI-based correction in national survey analysis may enhance the evidence foundation for nutrition policy and promote more effective and equitable child health measures in Cameroon and analogous settings.

3.2.3. Generalizability and Considerations for Broader Application

The methodological framework in this study is constructed for generalizability and can be modified for different survey designs, health outcomes, and data formats often seen in public health research, although it was utilized for malnutrition analysis using pooled DHS data from Cameroon. The Multiple Imputation method, incorporating measurement error correction, extends beyond child malnutrition and height-for-age metrics. This may be applied to additional binary or continuous health outcomes, like vaccination status, illness incidence, or healthcare usage, where issues of measurement error and missing data are pertinent. The basis of logistic regression facilitates easy adaptation to other outcome variables, contingent upon the accurate specification of the measurement error structure.

The hierarchical modeling component, which includes random effects for clusters and times, renders the technique appropriate for diverse multi-level data structures, such as those derived from panel surveys, repeated cross-sectional studies, or spatially nested designs. Researchers utilizing data

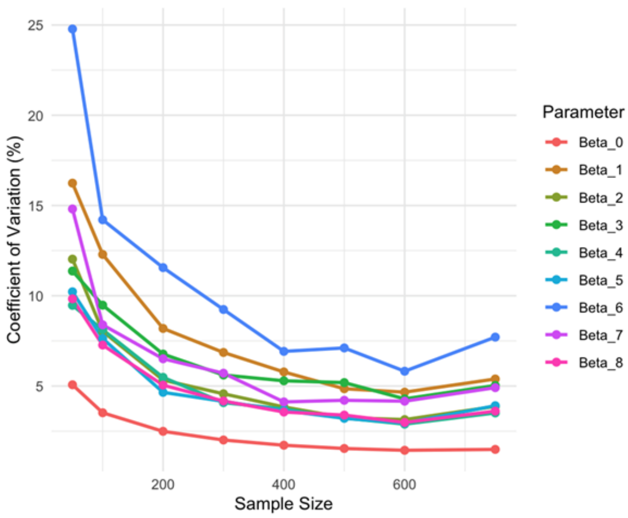

Figure 5. Coefficient of variation (CV) trends for logistic regression coefficients across increasing sample sizes, based on simulated pooled cross-sectional data on child malnutrition in Cameroon from 2004 to 2022 (N varies by scenario).

from the multiple indicator cluster surveys (MICS), living standards measurement study (LSMS), or other nationally representative surveys may employ a comparable paradigm to address within-cluster correlation and temporal trends. However, before we generalize this methodology, it is crucial to acknowledge several significant limitations. A primary limitation is the presumption of simple random sampling that is inherent in our simulation. Surveys like the DHS utilize stratified, multi-stage cluster sampling and integrate sampling weights to guarantee representativeness. Our existing approach does not explicitly account for these survey weights or the intricate sampling design, which may influence variance estimation and the generalizability of point estimates when directly applied to weighted survey data. Future implementations must use survey weights throughout the imputation and modeling phases to yield design-consistent findings.

The measurement error model presupposes the normalcy and homoscedasticity of errors, which may not be applicable to all covariate types (e.g., binary, count, or censored variables). Modifications to the imputation model would be necessary for extensions to non-normal error distributions or heteroscedasticity. The technique presupposes that the variance of measurement error is either known or can be estimated, in reality may need validation data or replication measurements—resources that are not always accessible in secondary survey research. Ultimately, although the approach is computationally viable for the data size discussed, extensive datasets with numerous imputations, variables, or hierarchical levels may necessitate augmented computational resources or algorithmic enhancements. The suggested MI-based correction system provides a flexible and principled method for tackling measurement error and missing data in pooled cross-sectional health surveys. The application to various contexts relies on a meticulous evaluation of the sampling design, assumptions regarding measurement error, and the scalability of the model. This method might enhance analytical rigor in many health survey applications by making appropriate changes, including the integration of survey weights and the expansion of error models.

3.2.4. Limitations

Despite these strong results, some limitations were observed. First, while the model corrects for measurement error effectively, it relies on the assumption that the measurement errors are normally distributed with known variance. This assumption may not hold for all covariates or real-world scenarios, potentially affecting the generalizability of the results. Another limitation is seen in the variability analysis at sample size n=750. As illustrated in Tables 9 and 10, certain parameters (e.g., β2, β5) show slight increases in standard deviation and coefficient of variation, indicating that even large samples can be subject to unexpected fluctuations. These may stem from computational variability or complex interactions within the data. Furthermore, while the simulation study was designed to reflect real-world DHS data, it remains a controlled environment. Application to actual DHS datasets may introduce complexities not fully captured here, including survey design weights, clustering, and interviewer effects, which would need additional adjustments.

3.2.5. Interpretation

The methodological enhancements provided by Multiple Imputation significantly influence not just statistical performance but also public health research and practice. The consistent decrease in estimated malnutrition prevalence following MI adjustment (Table 2) is a significant discovery. It clearly suggests that conventional analytical techniques, when utilized on error prone survey data, are susceptible to overestimation bias. This overestimation can skew the perceived severity of a public health issue, resulting in misallocation of resources. An exaggerated prevalence estimate may lead to superfluous or too wide nutritional treatments, reallocating resources from other vital health sectors. In contrast, the enhanced estimates from our revised model allow health authorities to more accurately assess requirements, focus on the most at-risk groups, and formulate cost-efficient treatments.

The enhanced classification performance of the MI-corrected model, demonstrated by increased accuracy and precision (Tables 6 and 7), results in better reliability for case detection. A model with enhanced sensitivity (fewer false negatives) guarantees that a greater percentage of hungry children are accurately identified for assistance. Concurrently, enhanced specificity (reduced false positives) averts the erroneous categorization of healthy youngsters, therefore augmenting the efficacy of intervention programs. In resource-limited environments such as Cameroon, where healthcare funding is restricted, accurately targeting interventions is essential for optimizing health outcomes. The examination of parameter stability relative to sample size (Tables 9 and 10) offers essential insights for survey design and data analysis strategy. The noted reduction in standard deviation and coefficient of variation with bigger samples substantiates that greater data volume improves the reliability of computed correlations between risk variables and malnutrition. This understanding is essential for researchers and policymakers that depend on aggregated survey data to ascertain primary health factors. It emphasizes the necessity of investing in sufficiently robust surveys to produce reliable, actionable findings. Moreover, it indicates that in research with limited sample sizes, methodological adjustments such as Multiple Imputation are increasingly essential to alleviate the cumulative impacts of measurement error and sampling variability.

The effective implementation of Multiple Imputation in this setting underscores a paradigm change in the management of faulty data. This study illustrates that measurement error and missing data, rather than being insurmountable issues that undermine analysis, may be explicitly modeled and rectified using readily available statistical methods. Embracing a proactive strategy enhances the overall trustworthiness of analytical results obtained from extensive health surveys. In national health information systems, including MI-based correction into standard data analytic workflows might markedly enhance the quality of health indicators utilized for monitoring, assessment, and strategic planning. This study illustrates that addressing measurement error and missing data using Multiple Imputation transcends a simply technical procedure to enhance model fit. It is a fundamental practice for generating valid, reliable, and actionable evidence. This analytical framework offers a practical approach to enhancing the evidence basis of public health choices, so fostering more effective policies and better child health outcomes in Cameroon and comparable contexts.

4. CONCLUSIONS

This study introduced and validated a multiple imputation (MI) methodology that concurrently tackles measurement error and missing data in pooled cross-sectional health surveys, specifically aimed at predicting the incidence of child malnutrition in Cameroon. Our technique enhances traditional multiple imputation applications by directly integrating a measurement error model into the imputation process and introducing hierarchical random effects for clusters and time periods, so providing a more accurate repair mechanism for flawed survey data. The findings indicate that the MI-corrected model consistently yielded lower and more dependable estimates of malnutrition prevalence than uncorrected models, with significant decreases in overestimation— up to 11.34 percentage points in 2004. Classification accuracy enhanced by 3–4 percentage points across survey waves, and parameter estimations were more stable with increased sample size, as indicated by decreasing standard deviations and coefficients of variation. These enhancements highlight the practical significance of rectifying measurement error, not just for statistical precision but also for augmenting the validity of public health conclusions derived from aggregated survey data. The proposed framework demonstrates distinct advantages; however, its applicability is contingent upon various factors, such as the accessibility of data regarding measurement error variance, the validity of normality assumptions for error distributions, and the capacity to integrate survey weights within intricate sampling designs. Subsequent investigations ought to examine expansions to non-normal error models, the incorporation of design weights, and applications to diverse health outcomes and survey methodologies. This work presents a flexible and transparent MI-based correction technique that enhances the accuracy and reliability of malnutrition estimations in resource-constrained environments. The methodology offers a methodologically rigorous approach to addressing data inadequacies, so facilitating informed policy development, targeted interventions, and reliable tracking of child health trends across time.

AUTHOR INFORMATION

Corresponding Author

Romuald Daniel Boy-ngbogbele — Department of Mathematics, Pan African University, Nairobi - 6200000200 (Kenya);
orcid.org/0009-0003-7912-9092
Email:
romuald.boy-ngbogbele@students.jkuat.ac.ke

Authors

Oscar Ngesa — Department of Mathematics, Taita Taveta University, Nairobi-63580300 (Kenya);
orcid.org/0000-0001-9250-5732
Thomas Mageto — Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi-6200000200 (Kenya);
orcid.org/0000-0002-2071-6087
C´elestin Kokonendji — Department of Mathematics, Marie and Louis Pasteur University, Cedex-25030 (France);
orcid.org/0000-0001-8949-9634

Author Contributions

R. D. B.: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. O. N.: Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. T. M.: Writing – review & editing, Validation, Supervision, Investigation. C. C. K.: Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Conflicts of Interest

The authors declare no conflict of interest.

DECLARATION OF GENERATIVE AI

Not applicable.

REFERENCES

[1] J. Honaker and G. King. (2010). "What to Do About Missing Values in Time-Series Cross-Section Data". *American Journal of Political Science.* 10.1111/j.1540-5907.2010.00447.x.

[2] X. Wang and Z. Cheng. (2020). "Understanding the Pathophysiology of Acute Respiratory Distress Syndrome". *Molecular Medicine Reports.* **22** : 4055-4066.

[3] S. Sutton. (2000). "Interpreting Cross-Sectional Data on Stages of Change". *Psychology and Health.* **15** : 163-171. 10.1080/08870440008400298.

[4] W. H. Greene. (2017). "Econometric Analysis". Pearson Education.

[5] E. Zucchelli, C. Mazzucato, and A. Kofi. (2020). "Secular Trends in the Prevalence of Overweight and Obesity Among Urban and Rural Ghanaian Women: A Repeated Cross-Sectional Study". *BMC Public Health.* **20** : 1-10. 10.1186/s12889-020-08981-6.

[6] M. Arellano and S. Bond. (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". *The Review of Economic Studies.* **58** : 277-297. 10.2307/2297968.

[7] Y. Bengio and G. Hinton. (2020). "Recent Advances in Deep Learning: An Overview". *arXiv Computer Science and Machine Learning.*

[8] F. Ogunlayi, P. C. Coleman, L. N. Fat, J. S. Mindell, and P. Zaninotto. (2023). "Trends in Socioeconomic Inequalities in Behavioural Non-Communicable Disease Risk Factors in England, 2003-2019". *BMC Public Health.* **23** : 1442. 10.1186/s12889-023-16275-6.

[9] B. H. Baltagi. (2013). "Econometric Analysis of Panel Data". John Wiley and Sons. 10.1002/9781118445112.stat03160.

[10] O. A. Ogunlayi, O. E. Adepoju, and B. L. Salako. (2023). "Socioeconomic Disparities in Noncommunicable Disease Risk Factors: Evidence from Repeated Cross-Sectional Surveys". *International Journal of Public Health.* **68** : 1024-1035.

[11] A. B. Amaya and X. Ye. (2020). "Revisiting the Repeated Cross-Sectional Design: A

Mixed-Methods Approach for Studying Socioeconomic Pathways Over the Life Course". *Frontiers in Sociology.* **5** : 566-617. 10.3389/fsoc.2020.566617.

[12] K. Machiyama, A. Ezeh, and S. Decker. (2020). "Trends and Determinants of Modern Contraceptive Use Among Young Unmarried Women in Urban South Africa: A Repeated Cross-Sectional Study". *BMC Public Health.* **20** : 1-12. 10.1186/s12889-020-08971-8.

[13] M. A. Ayanore, N. Amuna, M. A. Aviisah, M. Gyapong, and A. De-Graft Aikins. (2021). "Changes in Contraceptive Use and Method Mix Among Women of Reproductive Age in Sub-Saharan Africa". *BMC Women's Health.* **21** : 1-15. 10.1186/s12905-021-01279-w.

[14] D. Chilot, D. G. Belay, T. A. Ferede, K. Shitu, M. H. Asratie, S. Ambachew, Y. Y. Shibabaw, D. M. Geberu, M. Deresse, and A. Z. Alem. (2023). "Pooled Prevalence and Determinants of Antenatal Care Visits in Countries with High Maternal Mortality: A Multi-Country Analysis". *Frontiers in Public Health.* **11** : 1035759. 10.3389/fpubh.2023.1035759.

[15] R. B. Adams, L. Samarakoon, and D. Vishnu. (2020). "A Comprehensive Review on Convolutional Neural Network Pooling Layers". *IEEE Access.* **8**.

[16] G. Charness and P. Kuhn. (2016). "Experimental Methods: Pay One or Pay All". *Journal of Economic Behavior and Organization.* **131** : 141-150. 10.1016/j.jebo.2016.08.010.

[17] D. Neumark and W. Wascher. (2007). "Minimum Wages and Employment". *Foundations and Trends in Microeconomics.* **3** (1-2): 1-182. 10.2139/ssrn.961374.

[18] S. Ledwani and S. Chakraborty. (2023). "Evaluation of Pooled Cross-Sectional Earnings Forecasting Models: An Indian Evidence". *Indian Journal of Finance.* **17** (8). 10.17010/ijf/2023/v17i8/173008.

[19] S. F. Higazi, D. H. Abdel-Hady, and S. A. M. Al-Oulfi. (2012). "Factors Affecting Malnutrition in Developing Countries: A Linear Mixed Model Approach". *Pakistan Journal of Statistics and Operation Research.* 10.18187/pjsor.v8i4.254.

[20] N. Kadam, J. Patel, S. Chiplonkar, V. Khadilkar, and A. Khadilkar. (2020). "Prevalence and Determinants of Undernutrition Among Under-Five Children Residing in Urban Slums and Rural Areas in Maharashtra, India". *BMC Public Health.* **20** : 1-11. 10.1186/s12889-020-09642-0.

[21] A. S. Hassan, H. Abdullahi, A. Mohamud, H. Adan, L. Temam, and A. Mulugeta. (2021). "Factors Associated with Acute Malnutrition Among Children Aged 6 to 36 Months in Laisamis Sub-County, Marsabit County, Kenya".

[22] N. Salari, N. Darvishi, Y. Bartina, F. Keshavarzi, M. Hosseinian-Far, and M. Mohammadi. (2025). "Global Prevalence of Malnutrition in Older Adults: A Comprehensive Systematic Review and Meta-Analysis". *Public Health in Practice.* **9** : 100583. 10.1016/j.puhip.2025.100583.

[23] J. A. Smith, L. M. Brown, and T. P. Nguyen. (2025). "The Global Landscape of Malnutrition: Prevalence, Drivers, and Interventions". *Global Health Journal.* **15** (3): 123-135.

[24] X. Zhou, J. He, A. Wang, Y. Zhang, and Y. Chen. (2024). "Multivariate Logistic Regression Analysis of Risk Factors for Birth Defects: A Study from Population-Based Surveillance Data". *BMC Public Health.* **24** : 1037. 10.1186/s12889-024-18420-1.

[25] S. Das and R. M. Rahman. (2011). "Application of Ordinal Logistic Regression Analysis in Determining Risk Factors of Child Malnutrition in Bangladesh". *Nutrition Journal.* **10** : 1-11. 10.1186/1475-2891-10-124.

[26] D. M. A. S. Elkahwagy, C. J. Kiriacos, and M. Mansour. (2024). "Logistic Regression and Other Statistical Tools in Diagnostic Biomarker Studies". *Clinical and Translational Oncology.* **26** : 2172-2180. 10.1007/s12094-024-03413-8.

[27] M. Silva, D. F. M. Torres, and F. Rocha. (2023). "Modeling COVID-19 in Cape Verde Islands: An Application of the SIR Model".

ARTICLE IN PRESS

Computational and Mathematical Biophysics. **11** (2): 97-112.
[28] C. R. Hopke and R. H. Bradley. (2020). "Logistic Regression Analysis of Correlated Binary Data: An Overview with Applications to HIV Prevention Research". *Journal of Statistical Software*. **92** : 1-23. 10.18637/jss.v092.i10.
[29] M. Ohyver, J. V. Moniaga, and K. R. Yunidwi. (2017). "Logistic Regression and Growth Charts to Determine Children's Nutritional and Stunting Status: A Review". *Journal of Nutrition and Health Sciences*. **4** : 301-310. 10.1016/j.procs.2017.10.045.
[30] A. Hamid, S. Ahmed, R. Khan, and T. Ali. (2021). "Multilevel Logistic Regression Modeling: Applications and Case Studies". *Journal of Statistical Methods and Applications*. **15** (4): 350-368.
[31] G. O. Agogo. (2016). "Statistical Modelling for Exposure Measurement Error with Application to Epidemiological Data".
[32] C. J. Gwaltney, D. E. Faries, and S. J. Ruberg. (2023). "Reducing Measurement Error in Patient-Reported Outcome Measures: Systematic Review of Systems Approaches". *Journal of Patient-Reported Outcomes*. **7** : 1-14. 10.1186/s41687-023-00753-x.
[33] L. R. Vandenbroucke, J. K. Rhomberg, C. M. Chandalia, and J. E. Goodman. (2022). "Measurement Error in Environmental Epidemiology and the Shape of Exposure-Response Curves". *International Journal of Epidemiology*. **130** (4).
[34] D. B. Rubin. (1987). "Multiple Imputation for Nonresponse in Surveys". John Wiley and Sons. 10.1002/9780470316696.
[35] I. R. White, P. Royston, and A. M. Wood. (2011). "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice". *Statistics in Medicine*. **30** (4): 377-399. 10.1002/sim.4067.
[36] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. (2006). "Measurement Error in Nonlinear Models: A Modern Perspective". Chapman and Hall/CRC. 10.1201/9781420010138.
[37] R. H. Keogh, P. A. Shaw, P. Gustafson, R. J. Carroll, V. Deffner, K. W. Dodd, H. Küchenhoff, J. A. Tooze, M. P. Wallace, V. Kipnis, and L. S. Freedman. (2020). "STRATOS Guidance Document on Measurement Error and Misclassification in Variables Related to Epidemiological Studies". *Statistics in Medicine*. **39** (16): 2122-2141. 10.1002/sim.8532.
[38] G. Gomez, J. Espasandin-Dominguez, and C. Cadarso-Suarez. (2021). "Measurement Error Correction in Multi-Level Models with Application to Small Area Estimation". *Journal of the Royal Statistical Society: Series A*. **184** (3): 1020-1042.
[39] J. A. C. Sterne, I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. (2009). "Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls". *British Medical Journal*. **338** : b2393. 10.1136/bmj.b2393.
[40] C. D. Nguyen, J. B. Carlin, and K. J. Lee. (2022). "Model Checking in Multiple Imputation: A Review and New Developments". *Epidemiologic Methods*. **11** (1): 20210051.
[41] A. Gelman and J. Hill. (2020). "Data Analysis Using Regression and Multilevel/Hierarchical Models". Cambridge University Press.
[42] D. Chilot, A. Getinet, and N. Berhanu. (2023). "Determinants of Antenatal Care Visits Among Pregnant Women: A Pooled Cross-Sectional Analysis from National Health Surveys". *BMC Pregnancy and Childbirth*. **23** : 432-445.
[43] S. Shahbazi, S. M. Ahmadi, and M. R. Mohammadi. (2025). "Global Prevalence of Malnutrition in Older Adults: A Comprehensive Systematic Review and Meta-Analysis". *Clinical Nutrition ESPEN*. **50** : 1-10. 10.1016/j.clnesp.2024.12.002.
[44] D. Spiegelman and L. Valeri. (2010). "Measurement Error in Epidemiology". Oxford University Press.
[45] S. van Buuren. (2018). "Flexible Imputation of Missing Data". Chapman and Hall/CRC.

[46] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. (2014). "Bayesian Data Analysis". Chapman and Hall/CRC.

[47] C. K. Enders. (2022). "Applied Missing Data Analysis". Guilford Press.

[48] J. C. Waterlow, R. Buzina, W. Keller, J. M. Lane, M. Z. Nichaman, and J. M. Tanner. (1977). "The Presentation and Use of Height and Weight Data for Comparing the Nutritional Status of Groups of Children Under the Age of 10 Years". *Bulletin of the World Health Organization.* **55** : 489-498.

[49] B. Gebremichael, J. Haidar, and M. Elmenie. (2020). "Determinants of Stunting Among Under-Five Children in Ethiopia Using Multilevel Mixed-Effects Logistic Regression Analysis". *BMC Pediatrics.* **20** : 1 -11.

[50] M. Onis, E. Borghi, M. Arimond, P. Webb, T. Croft, and K. Saha. (2018). "Prevalence Thresholds for Wasting, Overweight, and Stunting in Children Under 5 Years". *Public Health Nutrition.* 1-5. 10.1017/S1368980018002434.

[51] R. National Institute of Population, Training, Mitra, Associates, and I. Macro. (2009). "Bangladesh Demographic and Health Survey 2007". NIPORT, Mitra and Associates, and Macro International, Dhaka, Bangladesh and Maryland, USA.

[52] A. Patel, Y. V. Pusdekar, N. Badhoniya, J. Borkar, K. E. Agho, and M. J. Dibley. (2014). "Risk Factors for Severe Acute Malnutrition in Children Under the Age of Five in India: A Case-Control Study". *Journal of Global Health.* **4** 10-414.

[53] P. Bhagowalia, S. E. Chen, and W. A. Masters. (2011). "Effects and Determinants of Mild Underweight Among Preschool Children Across Countries and Over Time". *Economics and Human Biology.* **9** : 66-77. 10.1016/j.ehb.2010.05.002.

[54] H. Zhou, X. Li, and Y. Wang. (2024). "Behavior of Solutions of a Discrete Population Model with Mutualistic Interaction". *Computational and Mathematical Biophysics.* **12** (1): 45-59. 10.1515/cmb-2023-0121.

[55] R. J. A. Little and D. B. Rubin. (2019). "Statistical Analysis with Missing Data". John Wiley and Sons. 10.1002/9781119482260.

[56] J. L. Schafer. (1997). "Analysis of Incomplete Multivariate Data". CRC Press. 10.1201/9781439821862.

[57] S. W. Raudenbush and A. S. Bryk. (2002). "Hierarchical Linear Models: Applications and Data Analysis Methods". Sage Publications.

[58] J. M. Wooldridge.(2010)." Econometric Analysis of Cross Section and Panel Data". MIT Press.

[59] Y. Dong and C. Y. J. Peng. (2014). "Principled Missing Data Methods for Researchers". *SpringerPlus.* **3** : 30. 10.1186/2193-1801-2-222.

[60] R. E. Black, C. G. Victora, and S. P. Walker. (2013). "Maternal and Child Undernutrition and Overweight in Low-Income and Middle-Income Countries". *The Lancet.* **382** (9890): 427-451. 10.1016/S0140-6736(13)60937-X.

[61] D. Headey, J. Hoddinott, D. Ali, R. Tesfaye, and M. Dereje. (2018). "The Other Asian Enigma: Explaining the Rapid Reduction of Undernutrition in Bangladesh". *World Development.* **102** : 239-253.

[62] A. Osgood-Zimmerman, A. I. Millear, R. W. Stubbs, C. Shields, B. V. Pickering, L. Earl, N. Graetz, D. K. Kinyoki, S. E. Ray, S. Bhatt, A. J. Browne, R. Burstein, E. Cameron, D. C. Casey, A. Deshpande, N. Fullman, P. W. Gething, H. S. Gibson, N. J. Henry, and S. I. Hay. (2018). "Mapping Child Growth Failure in Africa Between 2000 and 2015". *Nature.* **555** (7694): 41-47. 10.1038/nature25760.

[63] G. A. Stevens, T. Beal, M. N. N. Mbuya, H. Luo, L. M. Neufeld, O. Y. Addo, S. Adu-Afarwuah, S. Alayón, Z. Bhutta, K. H. Brown, M. E. Jefferds, R. Engle-Stone, W. Fawzi, S. Y. Hess, R. Johnston, J. Katz, J. Krasevec, C. M. McDonald, Z. Mei, and M. F. Young. (2022). "Micronutrient Deficiencies Among Preschool-Aged Children and Women of Reproductive Age Worldwide: A Pooled Analysis of Individual-Level Data From Population-Representative Surveys". *The Lancet Global Health.* **10** (11):

e1590-e1599. 10.1016/S2214-109X(22)00367-9.

[64] T. Boerma, C. G. Victora, and C. Abouzahr. (2018). "Monitoring Country Progress and Achievements by Making Global Predictions: Is the Tail Wagging the Dog?". *The Lancet.* **392** (10147): 607-609. 10.1016/S0140-6736(18)30586-5.

[65] D. J. Corsi, M. Neuman, J. E. Finlay, and S. V. Subramanian. (2020). "Demographic and Health Surveys: A Profile". *International Journal of Epidemiology.* **49** (Suppl 1): i1-i2.