

# A Generalized Small Area Estimation Model: Validated with the Children Anemia Study

## Abstract

Small area estimation (SAE) is a well-known method to produce reliable estimates for target variables associated with small areas and small sample sizes, and it is remarkably growing for public health applications. Direct survey estimates (DESvy) typically produce imprecise and unreliable estimates since they are obtained from the target variable of interest under the sampling model. To address this problem, this paper proposes generalized direct survey estimates (GDESvy) by incorporating the survey independent variables into the current DESvy. This approach improves the SAE model estimates by including survey independent variables in the sampling model through the proposed generalized small area estimation (GSAE). To validate and assess the performance of the proposed GSAE model, we first utilized independent variables from the Ethiopian Demographic and Health Survey (EDHS) to produce GDESvy estimates. Subsequently, we employed auxiliary variables from the population and housing census at the local level of Ethiopian administrative zones to provide precise GSAE estimates under the Fay-Herriot model. The results demonstrate that the GDESvy and GSAE estimates outperform the corresponding DESvy and SAE estimates, respectively, for anemia status among children aged 6–59 months, by producing lower standard errors. These findings are crucial for informing policy formulation and budget allocation at lower levels of government administration.

**Keywords:** Small area estimation, Survey sampling, Children’s anemia, Generalized direct survey estimation, Generalized small area estimation.

# 1 Introduction

The need for small-area statistics has grown significantly in recent years. Standard survey estimation methods (such as design-based direct survey estimators) for small areas (or small domains), which only use small area samples, are unreliable since these domains are subsets of the population with small sample sizes [1]. Although increasing the survey sample size can solve the issue, this approach is typically rarely taken because it requires time and financial resources, particularly for low- and middle-income countries. Health planners and policymakers rely on population health data at detailed geographic levels to describe health needs and to establish and assess health programs. However, estimates for local authorities are typically not possible due to the lack of geographic precision in population-based health survey data [2, 3].

Small area estimation (SAE) solves the problems of producing reliable estimates of a variable of interest for which sample sizes are too small for adequate precision [3]. The SAE provides a methodical approach to integrating direct survey estimates and auxiliary information, exploring the link between the outcome of interest and auxiliary data, and taking into consideration the sources of uncertainty in the model components [3]. By offering accurate estimates for areas or domains with small sample sizes, SAE approaches have emerged as essential tools in many scientific fields, supporting well-informed decisions. The SAE under the area-level model has sampling model and linking model components [1, 3]. The sampling model is used to obtain direct survey estimates (DES<sub>vy</sub>), and the linking model is employed to link the survey target variables (response variables) to the auxiliary variables obtained from the housing and population census datasets.

Thus, the effectiveness of SAE methods to produce acceptable small-area estimates has attracted a lot of attention [3]. In this context, reliable small-area estimates are often generated using model-based SAE methods that "borrow strength" via statistical models and auxiliary variables [1, 3, 4]. Compared to conventional direct survey estimates, these estimates are typically more efficient [5]. Model-based small area estimates are typically generated to assess the efficiency advantages over the survey-based estimates [3, 6, 7]. The most important thing in the SAE is reducing the variabilities of area-level estimates for small sample sizes and enhancing the reliability of the direct survey estimates [3, 8, 9].

Several scholars have extended the concept of the SAE to various statistical models

since it was introduced, and its value has been established. These include Bayesian SAE, multivariate SAE, spatial SAE, spatiotemporal SAE, and others [4, 10, 11]. Extensions of the Fay–Herriot model that draw strength from time have been proposed by several authors [4, 12, 13] since better small area estimators can be obtained by using the relevant information provided by historical data. A model with several temporal instants and autocorrelated structure for sampling errors was presented by Choudry and Rao [14]. A model that borrows data over time and across regions was put forth by Rao and Yu [8]. To determine the median income of four-person families for American states, Ghosh [15] suggested a time correlated area-level model [1, 3, 10].

Some extensions of the Rao–Yu model with applications to the estimation of labor or poverty indicators were provided by many authors [10, 16–20]. For spatiotemporal models with time-varying random slopes that follow an autoregressive process was introduced by Singh [4]. A more complex model with area-by-time random effects and regression coefficients that vary by area and time was put out [21]. Other studies used data from many sources, particularly survey datasets, to increase the reliability of the SAE [22–28].

To the best of our knowledge, the sampling model in SAE applications did not explore the independent variables of the survey, which are already present in the survey dataset to compute the direct survey estimates of an outcome variable [3, 9, 9, 20, 29–31]. This produces large variabilities in the direct survey estimates and the small area estimates. In research literature, the direct survey estimates of the variables of interest are obtained directly from the survey without including the survey’s independent variables [9, 31], and then lead to producing large standard errors. Suitable direct survey estimates are required since these estimates serve as the foundation for model-based SAE computations. Thus, to fill this knowledge gap in this field, we propose to include the survey independent variables to compute the direct survey estimates using weighted survey logistic regressions, which we subsequently name as the generalized direct survey estimates (GDES<sub>vy</sub>). To provide more accurate small area estimates, this research aims to validate the GDES<sub>vy</sub> to combine the DES<sub>vy</sub> with housing and population census auxiliary information under the Fay–Herriot model, and then subsequently named generalized small area estimation (GSAE).

Anemia in children is used to verify the improvements of new concepts for GSAE and GDES<sub>vy</sub> with their related SAE and DES<sub>vy</sub>. Anemia is a condition in which the concen-

tration of blood hemoglobin drops below predetermined thresholds. Anemia in mothers and children is a serious public health problem worldwide, such as low birth weight, impaired neurocognitive and motor development in children, as well as an increased risk of postpartum hemorrhage, infection, and maternal mortality for women [32–34].

To achieve the ambitious plan of decreasing the risk of anemia in children and reproductive age women by 2030, the UN Sustainable Development Goals (SDGs) of 2012 include prevention and prompt treatment of anemia [13, 35]. The risk of anemia influences all SDGs, including ending poverty in all its manifestations worldwide (Goal 1), promoting good health and well-being (Goal 3), improving educational quality (Goal 4), and gender equality (Goal 5) [33, 36]. In addition, Goal 2 aims to end hunger, attain food security and better nutrition, and advance sustainable agriculture by 2030, which also has an impact on anemia risk [36]. All these efforts did not sufficiently reduce the risk of anemia. According to estimates from 2019, anemia affects 30% (571 million) of women of reproductive age (15–49 years), 37% (32 million) of pregnant women, and 40% (269 million) of children aged 6–59 months. The WHO African and South-East Asia regions are the areas most affected [37, 38].

Anemia is a severe public health problem in Ethiopia, particularly for children. The 2016 Ethiopian Demographic Health Survey (EDHS) data showed that 57% of children aged 6–59 months were anemic, a 13% increase from the 44% reported in the 2011 survey [39]. Furthermore, the severity varies geographically and is more severe in different regions and administrative zones. As a result, we used the anemia status of children aged 6–59 months for GSAE model validations to see the improvements over the usual SAE approaches.

The rest of the paper is structured as follows. Section 2 presents the methods and materials, including data sources and the survey-based DESvy and GDESvy, and model-based SAE and GSAE methods; Section 3 presents the results; and Section 4 includes the discussion and conclusions.

## 2 Methods and Materials

This study considers the estimates of design-based and model-based (Fay-Herriot model) estimations for 6–59 months anemia status in Ethiopian administrative zones

with (i.e., GSAE) and without (i.e., SAE) survey explanatory variables.

## 2.1 Data Sources

The data used in this study were taken from the housing and population census as well as the EDHS. The target (i.e., response ) variable of anemia status for children under the age of 6 to 59 months, together with associated independent variables, has been taken from the 2016 EDHS data. The independent variables are utilized to enhance the design-based GDESvy and then the model-based GSAE estimates. The independent variables are provided from both the DHS survey and the population and housing census datasets for different purposes. The survey independent variables are used for improving the design-based estimations (Table 1), and the census auxiliary variables are used for model-based estimation under the Fay-Herriot model [29, 40] (Table 1). The auxiliary variables reviewed based on literature that are associated with 6-59 months anemia status [29, 29] (Table 1).

## 2.2 Advancing from DESvy to GDESvy in Survey Estimates

### 2.2.1 DESvy: Direct Survey Estimates

The sample survey design is taken into consideration while evaluating the estimator's characteristics of the survey estimations [41]. This is the usual and mostly explored method in SAE history [42]. The direct survey estimates (DESvy) only use the observed values for the variable of interest (i.e., the target variable) and the survey sample weights for the estimation of means and variances. For DESvy estimations of proportion and variance without considering covariates. For zonal level DESvy estimates, we use the primary sampling units (PSU) weight to adjust the zonal survey weights [3, 9]. The new zonal level weight  $w_z$  (or  $w_i$ , if we align with the notation for zones) is the aggregated value of the PSU weights (  $w_{ik}$ ) for a specific zone  $i$ . The new weight for zone  $i$  is given by:

$$w_i = \frac{1}{n_k} \sum_{k \in i}^{n_k} w_{ik},$$

where  $n_k$  is the number of PSU in zone  $i$  and  $w_{ik}$  denotes the  $k^{th}$  PSU weight in zone  $i$ . Therefore, the weighted zonal level estimates with the new weights are computed.

Table 1: Study variables

Survey variables	Categories
Anemia status (target variable)	0 = no, 1 = yes
Children's sex	1 = male, 2 = female
Children's age in months	continuous
Place of residence	1 = urban, 2 = rural
Age of mother	1 = 15–24, 2 = 25–34, 3 = 35–49
Educational level of mothers	0 = no education, 1 = primary, 2 = secondary, 3 = higher
Source of drinking water	0 = unimproved, 2 = improved
Household head	1 = male, 2 = female
Literacy	1 = illiterate, 2 = literate
Number of children	continuous
Daughters who have died	0 = no, 1 = yes
Currently pregnant	0 = no, 1 = yes
Sons who have died	0 = no, 1 = yes
Type of toilet facility	0 = no toilet, 1 = have toilet
Wealth index combined	0 = poor, 1 = middle, 2 = rich
Current marital status	0 = single, 1 = married, 2 = other
<b>Census auxiliary variables</b>	
Sex of children	% male, % female
Age of children	% below 2, % 2–4, % above 4
Parents' sex	% male parents, % female parents
Parents' age	% 15–25, % 25–35, % 35–49
Place of residence	% rural, % urban
Source of drinking water	% improved, % unimproved
Educational levels	% non-educated, % primary, % secondary and above
literacy	% literate, % illiterate
Marital status	% married, % single, % others
Type of toilet	% has toilet, % no toilet
Number of family members	% less than five, % more than five
Number of died sons	% no, % one, % more than two
Number of died daughters	% no, % one, % more than two
Disability	% disabled, % not disabled
Employment status	% government, % private, % self-employed, % unemployed, % others

Let  $Z$  be the set of zones,  $w_{ij}$  be a weight (new\_weight) for individual  $j$  in zone  $i$  and  $y_{ij}$  be a binary indicator for children anemia (1 if a child is anemic status, 0 otherwise), then the weighted proportion of anemia status in children for zone  $i$ , denoted as  $\hat{p}_i$ , is given by:

$$\hat{p}_i = \frac{\sum_{j \in i} w_{ij} y_{ij}}{\sum_{j \in i} w_{ij}},$$

where  $i \in Z$  denotes  $i$ -th zone and  $j \in i$  indicates  $j$ -th individual within zone  $i$ .

Note that the variance of children's anemia status for zone  $i$ , considering the new zonal weights, is obtained as:

$$\widehat{\text{Var}}(y_i) = \frac{\sum_{j \in i} w_{ij} (y_{ij} - \hat{p}_i)^2}{\sum_{j \in i} w_{ij}}.$$

### 2.2.2 GDESvy: Generalized Direct Survey Estimates

Standard design-based DESvy approaches rely on only weighted survey data to directly estimate characteristics of interest for small areas [3]. While effective, these estimates have large variabilities and suffer from a lack of precision. To resolve this problem and fill this knowledge gap, we propose a new framework that includes independent variables of a survey into the DESvy approach to reduce the variability and then improve the precision and reliability. The proposed GDESvy includes the survey independent variables in the design-based DESvy approach to compute the direct survey estimates. This concept could robust the standard DESvy, which only computes the direct survey estimates from the response variables of interest. The weighted proportions with covariates were employed in the design-based GDESvy approach to estimate the variables of interest. Typically, to determine the estimates, a model for the distribution of characteristics of interest is often assumed, given the relevant explanatory variables. Survey-weighted logistic regression is used to include covariates in the survey design in order to improve the precision of the design-based estimates and further enhance model-based estimates.

The likelihood function for survey weighted logistic regression is as follows:

$$L(\beta) = \prod_{j \in i} [\hat{p}_{ij}^{y_{ij}} (1 - \hat{p}_{ij})^{1-y_{ij}}]^{w_{ij}}.$$

Then, the log-likelihood becomes as:

$$\ell(\beta) = \sum_{j \in i} w_{ij} [y_{ij} \log(\hat{p}_{ij}) + (1 - y_{ij}) \log(1 - \hat{p}_{ij})].$$

The survey regression error function in the survey design inherently adjusts for weights and covariates to obtain coefficient estimates as follows:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i \in Z} \sum_{j \in i} w_{ij} [y_{ij} \log(\hat{p}_{ij}) + (1 - y_{ij}) \log(1 - \hat{p}_{ij})].$$

The predicted probability of anemia status for individual  $j$  in zone  $i$ ,  $\hat{p}_{ij}$ , is obtained

using the fitted model as

$$\hat{p}_{ij} = \frac{\exp(X_{ij}^\top \hat{\beta})}{1 + \exp(X_{ij}^\top \hat{\beta})}.$$

The proposed design-based estimated prevalence of childhood anemia status in zone  $i$ , is denoted by  $\hat{p}_i^*$  to differentiate the DESvy zonal level estimates ( $p_i$ ) at subsection 2.2.1, and given as:

$$\hat{p}_i^* = \frac{\sum_{j \in i} w_{ij} \hat{p}_{ij}}{\sum_{j \in i} w_{ij}}.$$

The variance of children's anemia status in zone  $i$ , considering covariates, is based on the variability of the predicted probabilities  $\hat{p}_{ij}$  across the zone. It accounts for design-based uncertainty and individual-level variation. We compute the weighted variance as follows:

$$\widehat{\text{Var}}(y_i) = \frac{\sum_{j \in i} w_{ij} (\hat{p}_{ij} - \hat{p}_i^*)^2}{\sum_{j \in i} w_{ij}},$$

where  $\hat{p}_{ij}$  is the predicted probability for an individual  $j$ , derived from the fitted survey-weighted logistic regression model using the `svyglm` function in R, and  $\hat{p}_i^*$  is a zone-specific weighted mean of predicted probabilities for children's anemia status.

## 2.3 Advancing from SAE to GSAE in Area-Level Estimation

Area-level SAE is a statistical method that generates more accurate estimates for small sample sizes by using a statistical model that combines auxiliary data from various sources to "borrow strength" with more data. This produced accurate estimates for areas where direct survey data alone might not be enough for estimation [3, 8].

The basic area-level Fay-Herriot model combines the direct aggregate zone-level survey estimates with the available auxiliary variables obtained from various secondary sources, mainly from census or records. Thus, the model has two components: (i) the sampling model for the direct survey estimates and (ii) the linking model consisting of the area-specific auxiliary variables.

### 2.3.1 Standard Small Area Estimation

Small area estimates are most frequently used in official statistics, using the sampling and linking model approach and combining data from various sources. An area-level model introduced by Fay and Herriot [43] was developed to generate small area estimators



for median income in small localities across the United States. There is an extensive body of literature in the field of SAE that explores and builds upon this model, highlighting its significance and widespread application in various statistical contexts [1, 3, 8, 13, 14]. In this study, we use the DESvy for both estimated means and variances for sampling and linking model parameters. The sampling and linking model for  $p_i$ , with Gaussian errors in the logit scale and a binomial distribution for  $y_i$  with success probability  $p_i$  (DESvy estimated value), is presented as follows,

$$\text{Sampling model: } y_i \mid p_i, n_i \sim \text{Bin}(n_i, p_i), \quad i = 1, 2, \dots, m$$

$$\text{Linking model: } \text{logit}(p_i) = \mathbf{x}_i^\top \boldsymbol{\beta} + \nu_i,$$

where  $\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$ ,  $\nu_i$  is distributed as  $\mathcal{N}(0, \sigma_\nu^2)$ ,  $n_i$  is the sample size of zone  $i$ ,  $\mathbf{x}_i$  is the vector of the area-level aggregated census auxiliary variables in this study, specifically, zone-level aggregated values.

### 2.3.2 Generalized Small Area Estimation

The design-based survey regression methodologies are used to calculate the estimated prevalence of childhood anemia for each zone. We now apply the proposed design-based GDESvy to introduce the GSAE model, which takes into account both survey and census covariates. The proposed childhood anemia prevalence estimates (GDESvy) under subsection 2.2.2 ( $p_i^*$ ) are used in the Fay-Herriot model, which is defined as:

$$\text{Sampling model : } y_i \mid p_i^*, n_i \sim \text{Bin}(n_i, p_i^*), \quad i = 1, 2, \dots, m$$

$$\text{Linking model : } \text{logit}(p_i^*) = \mathbf{x}_i^\top \boldsymbol{\beta} + \nu_i$$

$$\text{logit}(p_i^*) = \log\left(\frac{p_i^*}{1-p_i^*}\right), \quad \nu_i \sim \mathcal{N}(0, \sigma_\nu^2).$$

## 2.4 Model Comparison

We compare the performance of DESvy, GDESvy, SAE, and the GSAE for children aged 6-59 months with anemia status using the root mean square error (root MSE) and the coefficient of variation (CV). The CV and standard error are used as measures of variability associated with the estimate.

## Ethical considerations

Procedures and questionnaires for standard DHS surveys have been reviewed and approved by ICF Institutional Review Board (IRB). Additionally, country-specific DHS survey protocols are reviewed by the ICF IRB and typically by an IRB in the host country. ICF IRB ensures that the survey complies with the U.S. Department of Health and Human Services regulations for the protection of human subjects (45 CFR 46), while the host country IRB ensures that the survey complies with laws and norms of the nation. Therefore, Central Statistical Agency (CSA) is the national statistical agency of Ethiopia, with a national mandate to produce timely, accurate official statistics to support democracy and economic growth and development in Ethiopia with aid of international stakeholders. Therefore, CSA ethics council authorized all DHS data. Before taking part in the survey, all participants provided written informed permission. All the data were fully waived to the requirement for informed consent. There were no medical records used in the research since it was a DHS dataset. we did get formal permission from the DHS program to utilize the data for research purposes. The data is available at website <https://www.dhsprogram.com>.

## 3 Results

### 3.1 The analysis of DESvy vs GDESvy

In the DESvy method, the estimations of the prevalence of childhood anemia (zonal-wise estimations) use only the observed values for the variable of interest without considering the survey independent variables. However, as stated in subsection 2.2.2, the GDESvy method incorporates the survey independent variables to determine the prevalence of anemia status and subsequently to enhance the DESvy estimates. Table 2 displays the results of the survey weighted logistic regression for complex sampling data analysis. GDESvy is the sampling model for GSAE, which takes into consideration the survey independent variables listed in Table 2.

Table 2: Survey weighted logistic regression coefficients (GDESvy)

Independent variables	Estimate	Std. Error	t-statistic	Pr(> t )
Intercept	0.84950	0.44217	1.921	0.059383 .
<b>Age of child</b>	-0.41068	0.02282	-17.998	< 2e-16 ***
<b>Sex of child</b>				
Male (ref.)				
Female	-0.05344	0.08514	-0.628	0.532516
<b>Literacy</b>				
Illiterate (ref.)				
Literate	-0.40885	0.10549	-3.876	0.000262 ***
<b>Educational level</b>				
None educated (ref.)				
Primary	0.15864	0.11905	1.333	0.187637
Secondary	0.16100	0.19066	0.844	0.401725
Higher	0.37872	0.25429	1.489	0.141551
<b>Age of mothers</b>				
15-24 years (ref.)				
25-34 years	-0.19904	0.10606	-1.877	0.065336 .
35-49 years	-0.58925	0.13370	-4.407	4.31e-05 ***
<b>Place of residence</b>				
Urban (ref.)				
Rural	0.16253	0.16521	0.984	0.329109
<b>Head of household</b>				
Male (ref.)				
Female	0.29698	0.09642	3.080	0.003101 **
<b>Number of children</b>	0.08577	0.02018	4.251	7.41e-05 ***
<b>Wealth index</b>				
Poor (ref.)				
Medium	-0.37405	0.11761	-3.180	0.002312 **
Rich	-0.30105	0.11257	-2.674	0.009598 **
<b>Daughters died</b>				
No (ref.)				
Yes	0.08499	0.12378	0.687	0.494901
<b>Current pregnancy</b>				
No (ref.)				
Yes	0.21951	0.11387	1.928	0.058553 .
<b>Sons died</b>				
No (ref.)				
Yes	-0.19477	0.11259	-1.730	0.088708 .
<b>Current marital status</b>				
Single (ref.)				
married	0.26809	0.42976	0.624	0.535084
<b>Source of drinking water</b>				
Unimproved (ref.)				
Improved	0.08597	0.08510	1.010	0.316390

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### 3.2 Analysis of SAE vs GSAE

In this study, the survey independent variables are considered under the SAE children’s anemia status applications. The independent survey variables are incorporated in GDESvy for the survey-based direct estimates to further improve the SAE at zonal-level estimation. Therefore, for this analysis, the SAE and GSAE methods are associated with the DESvy and GDESvy, respectively.

In small area statistics, the auxiliary variables borrowed from the population and housing census dataset are linked to the survey target variables. The auxiliary variables from population and housing census data are aggregated at the zonal level for model-based small area applications. Consequently, an explanatory data analysis was conducted for the selection of appropriate auxiliary variables using LASSO techniques prior to the identification of appropriate auxiliary variables for the SAE modeling approach. Thus, Table 3 displays the appropriate auxiliary variables both in the SAE and GSAE analyses. The area level random effect residual variances for the GSAE model are lower than the SAE model estimates (Table 3).

Table 3: Regression coefficients and residual variance for SAE and GSAE models

SAE method					
Variables	Coefficient	Mean	SD	2.5%	97.5%
Intercept	$\beta_1$	-16.31704	13.92627	-42.29657	8.434
Male for child	$\beta_2$	20.26810	19.72906	-15.26438	60.152
Child age 2-4 years	$\beta_3$	14.05058	12.21944	-7.84759	36.433
Below 2 years child age	$\beta_4$	-4.48501	11.06860	-22.19256	17.673
Urban	$\beta_5$	1.08156	1.41940	-1.53411	3.860
Female for parents	$\beta_6$	2.25077	1.99036	-1.20908	5.983
Private employment	$\beta_7$	-1.75765	4.02210	-9.59369	5.089
Unemployed	$\beta_8$	0.23919	2.79859	-4.85546	5.641
Other employment	$\beta_9$	-0.08746	2.16394	-4.08551	4.457
Parents age of 35-49 years	$\beta_{10}$	3.19297	5.01742	-5.62235	13.676
Residual Variance		0.118655	0.140515	0.007083	0.511
GSAE methods					
Intercept	$\beta_1$	-8.3497	5.7373	-20.0037	2.876
Male for child	$\beta_2$	13.0368	9.6365	-5.8663	32.585
Child age 2-4 years	$\beta_3$	7.5268	4.4489	-1.2070	16.049
Below 2 years child age	$\beta_4$	-5.8993	3.2838	-12.5978	0.429
Urban	$\beta_5$	0.9945	0.6071	-0.2064	2.214
Female for parents	$\beta_6$	0.8742	0.7658	-0.5086	2.440
Private employment	$\beta_7$	-2.6343	1.1598	-4.7859	-0.184
Unemployed	$\beta_8$	-0.3799	1.1704	-2.6777	1.968
Other employment	$\beta_9$	0.1742	0.7606	-1.2650	1.680
Parents age of 35-49 years	$\beta_{10}$	0.9746	1.8080	-2.6645	4.546
Residual Variance		0.07820	0.05554	0.01146	0.215

### 3.3 Model Comparison

The estimated values for the DESvy and the standard SAE estimates are presented in Figure 1 (left). On the other hand, the survey-based regression estimates, GDESvy, and the corresponding model-based GSAE small area estimates are shown in the same figure (right). In both figures, the standard SAE estimates are nearly consistent with the direct survey estimates across almost all zones; however, the proposed model-based estimates GSAE are nearly perfectly consistent with the corresponding survey-based estimates

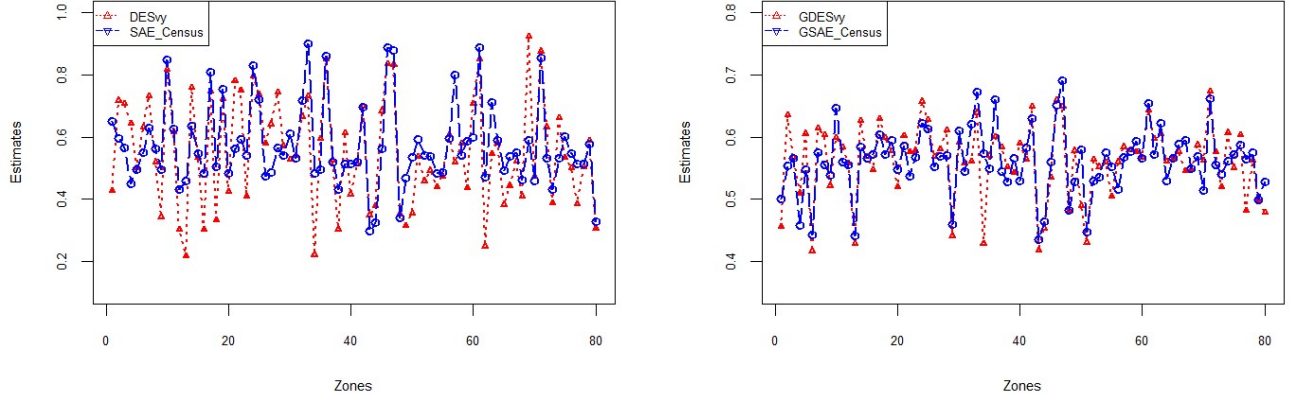


Figure 1: Direct survey and model-based estimates. The standard SAE estimates are on the left, and the proposed method estimates are on the right.

The root MSE and CV are used to compare the DESvy, GDESvy, SAE, and GSAE models. While the SAE and GSAE are model-based approaches that fall under the Fay-Herriot model, the DESvy and GDESvy are design-based survey methodologies under the survey weighted logistic regression. Model-based (SAE and GSAE) and design-based (DESvy and GDESvy) methods were applied for finding the estimates of children's anemia in the Ethiopian zone, together with their CVs and root MSE. The sampling variability is displayed as the estimates in CV and root MSE, and in general, larger CVs or root MSE estimates are regarded as unreliable. A 20% CV was maintained as a threshold by the United Kingdom's Office for National Statistics, however there is no exact guideline for CV cut-off values [44].

The zonal wise CV values for GSAE, SAE, GDESvy, and DESvy are displayed in Figure 2 for the purpose of comparing and identifying the most reliable estimates. In this study, the generalized model-based estimates produced by the GSAE approach are compared with the conventional SAE methods, as measured by the CV. It is also compared with the typical DESvy, which is determined solely by response variables, with the GDESvy.

The CVs for DESvy vs GDESvy are shown in Figure 2 (left) for purpose of comparison. The DESvy estimates of childhood anemia prevalence have CVs over 20 for all local administrative zones based on the findings presented in Figure 2 (left). But, the

GDESvy estimates for some zones have CVs above, while the other zones have CVs below this threshold. These results clearly demonstrate that the GDESvy estimates are more credible than the DESvy since they have smaller CV values (see supplementary file 5).

The SAE vs GSAE CV are presented in Figure 2 (right) for model-based comparison. The CV for GSAE estimates are below the thresholds for all zones, while the CV for SAE estimates are greater than the expected thresholds. In addition, the GSAE estimates are smaller than the corresponding SAE estimates across all the local administrative zones. Thus, the generalized model-based estimates produced by the GSAE approach are more precise and reliable than the corresponding SAE estimates. This is due to the inclusion of survey independent variables in the survey weighted regression for direct survey estimation procedures. Therefore, we can conclude that the GSAE improved the accuracy of standard SAE model estimates by including survey-independent variables in the sampling model.

Zonal wise root MSE for survey-based (left) and model-based (right) are displayed in Figure 3. Figure 3 (left) shows that the root MSE of the survey-based estimates, DESvy, are larger than the corresponding GDESvy estimates. Furthermore, Figure 3 (right) shows that the root MSE of GSAE is smaller than the comparable SAE estimates. This demonstrates how the new method improved the accuracy and reliability of estimates of childhood anemia throughout the Ethiopian administrative zones. In general, small area estimates produced by the GSAE are more efficient than the corresponding SAE model estimates, and the GDESvy is also better than the corresponding DESvy values.

Table 4 displays the summary statistics of the CV for both survey-based and model-based estimates. The results include the minimum, first quartile (1st Q), median, mean, third quartile (3rd Q), and maximum for the design-based estimates (DESvy and GDESvy) and the model-based estimates (SAE and GSAE). In general, the summary statistics for GDESvy are lower than those for the corresponding DESvy estimates, and similarly, the GSAE values are lower than those for SAE. The results show that the generalized models provide more precise estimates by reducing the variability of estimates for 6-59 months anemia status across Ethiopian local administration zones.

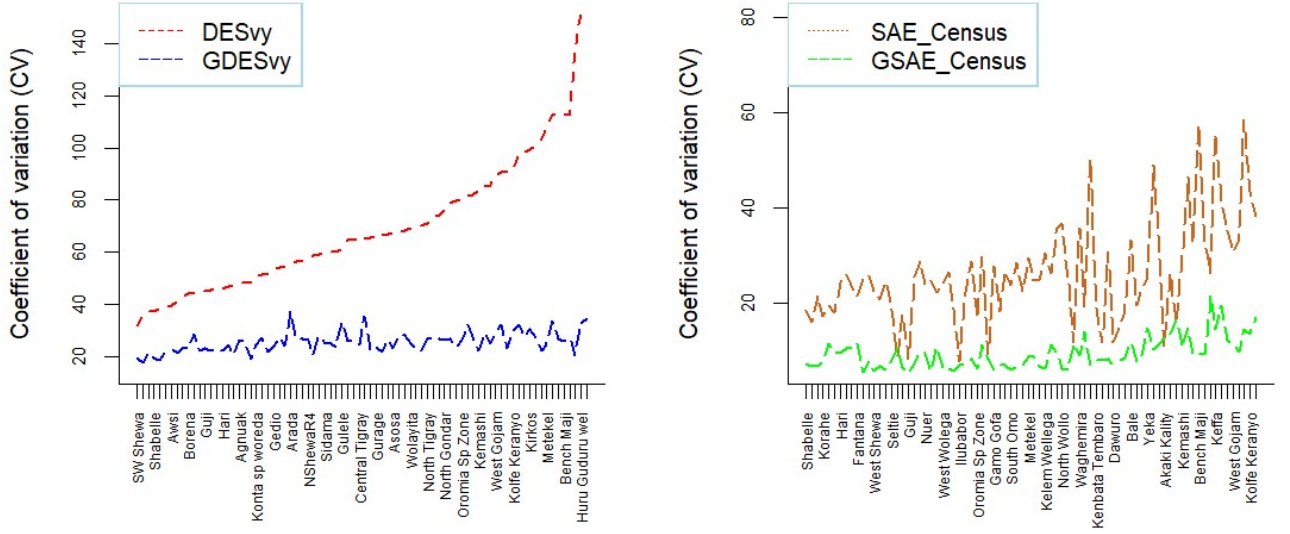


Figure 2: Zonal wise percentage coefficient of variation (CV, %) for survey-based (left) and model-based (right).

Table 4: Summary statistics of CVs for the design-based and model-based estimates.

	DESvy	GDESvy	SAE_Census	GSAE_Census
Min	31.83	17.97	7.695	5.530
1st Qu.	48.75	22.84	22.474	7.672
Median	65.36	25.13	27.853	9.179
Mean	69.48	25.73	29.075	10.079
3rd Qu.	82.51	27.63	33.273	11.522
Max	152.98	37.26	70.407	19.864

## 4 Discussion and Conclusion

The study characteristics of interest for children aged 6-59 months and their anemia status were derived from the 2016 EDHS data [39]. To integrate the survey data for the SAE application, the auxiliary variables were extracted from the 2007 housing and population census dataset [45]. The survey-weighted logistic regression analysis was applied to include the survey-independent variables in the generalized model. We used the target variable of the survey, the independent variables of the survey, and the auxiliary variables



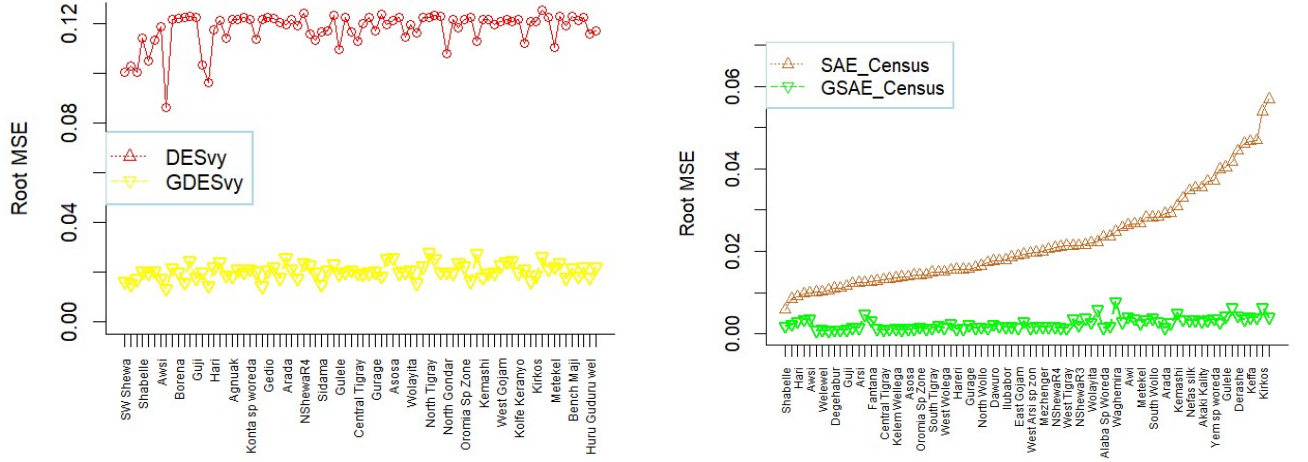


Figure 3: Zonal-wise root MSE for survey-based (left) and model-based (right) estimations of childhood anemia.

of the census for the status of children’s anemia to validate the GSAE model. This study used the well-known Fay-Herriot model for GSAE to produce more precise and reliable estimates by including independent variables from the survey. The prevalence of anemia within 6-59 months in children was estimated using an area-level model that linked data from the 2007 housing and population census with the 2016 survey datasets.

In many research findings, the precise estimates of disaggregated-level are enhanced by the SAE approach in comparison to survey estimates [3, 19, 20, 46]. Furthermore, other researchers extended the SAE method to integrate different statistical methods to enhance the reliability of the disaggregate level estimations [4, 7, 47]. The Behavioral Risk Factor Surveillance System (BRFSS) of the U.S. Centers for Disease Control and Prevention implemented a transformation to the survey data to reduce extremely unstable survey variances using a hierarchical Bayes small area estimate model [48], which is consistent with the current findings.

Singh proposed a spatiotemporal regression model for the SAE problem, which has large variabilities, in the general mixed effects model framework [4]. The small area estimates have improved due to a common autocorrelation parameter among the small areas. Additionally, the inclusion of fixed spatial auto-correlation across the small areas gives spatial-temporal SAE models an advantage over SAE models without spatial consideration [4]. Using data from the Spanish living conditions survey, multivariate Fay-Herriot

models for estimating small area indicators were introduced and further explored. As a result, survey estimates using a new method are better than those using the standard SAE method [12]. Our findings are consistent with the those study findings in producing reliable estimates.

In 2018, Anjoy and his colleagues carried out a study to estimate the disaggregated-level poverty incidence in Odisha using an area-level hierarchical Bayes SAE model. The results demonstrated that the model-based estimates produced by the hierarchical Bayes SAE method perform more accurately than the SAE and direct survey estimates [49]. Similar research findings, which are in line with our findings, are multivariate, spatial, and spatiotemporal SAE approaches for Ethiopian undernutrition survey data for children under five, demonstrated that the estimates are more accurate and reliable than the comparable conventional SAE estimates [11, 50, 51].

Some SAE methods, cited in this discussion, improved the survey estimates and the standard SAE of the estimated values for the characteristics of interest: poverty, health, and agriculture [4, 12, 48]. However, these studies did not use the survey’s independent variables for the sampling model and then in the linking model. Our study considered the survey independent variables in the GDESvy model to introduce the GSAE model and then improve the standard SAE estimates of the children’s anemia study.

The GSAE is used in this study to fill in the current research gaps and improve the SAE estimates while taking sample survey independent variables into account. In order to lower the variability of the small area estimates of the anemia status of children aged 6 to 59 months in the Ethiopian administrative zones, the survey’s independent variables were incorporated into the design-based models. The GSAE approach outperforms the survey estimates and the traditional SAE, which is consistent with the other research findings. Figure 2 (left) shows the comparison of the design-based direct survey estimates with and without survey independent variables. The GDESvy estimates are more precise than the comparable DESvy estimates. This is due to the fact that GDESvy estimations produced accurate survey estimates by including the survey independent variables into the design-based survey model. Direct survey-based variance estimates, calculated only using the target variables of interest, are frequently applied to compute the SAE under the Fay-Herriot model. In order to compute the design-based direct survey estimates (both the mean and variance estimates), we incorporated the survey independent variables into this

investigation. Consequently, there is consistency in the estimated means when comparing these two model estimations.

The numerical improvements of the new approach compared to existing methods were assessed based on the CVs. In terms of design-based survey estimates, the new approach showed improvements ranging from a minimum of 31.83% for DESvy to 17.97% for GDESvy, a mean improvement of 69.48% for DESvy to 25.73% for GDESvy, and a maximum improvement of 152.98% for DESvy to 37.26% for GDESvy. Comparably, switching from the SAE to the recently suggested GSAE model significantly lowers the range of estimates for child anemia. Specifically, the minimum CV reduced from 7.695% for SAE to 5.53% for GSAE, the mean decreased from 29.07% for SAE to 10.08% for GSAE, and the maximum dropped from 70.41% for SAE to 19.86% for GSAE. The first and third quartiles also indicated significant improvements in variability reduction under the GSAE model. This is a significant improvement in reducing the variability of the survey estimates and the standard SAE estimates for children aged 6 – 59 months of anemia status.

In these findings, we found additional insights beyond the existing Fay-Herriot model estimates to further improve the estimates of anemia status for the 6-59 months children by reducing the variabilities and by improving the precision of survey estimates. These improvements are due to including survey independent variables into the direct survey estimation process and then to the Fay-Herriot models. The performance measures of these findings are in line with other research findings [4, 11, 31]. As demonstrated by the findings from CV and the root MSE, GDESvy estimates are significantly more precise than comparable DESvy estimates in reducing the variabilities. Similarly, GSAE estimates reduce the variability of children's anemia status better than traditional SAE estimations. In general, the GSAE model significantly increases the efficiency of obtaining zonal-level estimations of anemia status in children aged 6-59 months. In summary, incorporating independent variables from the survey improved the performance of the newly developed GSAE model, leading to more accurate estimates of childhood anemia status across Ethiopian administrative zones. This research leads us to the conclusion that survey independent variables are crucial for small area statistics because they lower the variability of estimates, which improves precision. These findings offer valuable insights for policymakers, decision-makers, legislative bodies, and non-governmental organizations.

## **Statements and declarations**

Not applicable

## **Consent to participate**

Not applicable

## **Consent for publication**

Not applicable

## **Declaration of conflicting interest**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## **Funding Statement**

Although there was no specific external funding for this study, the University of Pretoria would waive the article processing charges (APCs) as part of a member of open access publishing arrangement.

## **Ethical considerations**

Procedures and questionnaires for standard DHS surveys have been reviewed and approved by ICF Institutional Review Board (IRB). Additionally, country-specific DHS survey protocols are reviewed by the ICF IRB and typically by an IRB in the host country. ICF IRB ensures that the survey complies with the U.S. Department of Health and Human Services regulations for the protection of human subjects (45 CFR 46), while the host country IRB ensures that the survey complies with laws and norms of the nation. Therefore, Central Statistical Agency (CSA) is the national statistical agency of Ethiopia, with a national mandate to produce timely, accurate official statistics to support democracy and economic growth and development in Ethiopia with aid of international stakeholders. Therefore, CSA ethics council authorized all DHS data. Before taking part in the survey, all participants provided written informed permission. All the data were fully waived to the requirement for informed consent. There were no medical records

used in the research since it was a DHS dataset. we did get formal permission from the DHS program to utilize the data for research purposes. The data is available at website <https://www.dhsprogram.com>.

## Data availability

The 2016 EDHS childhood anemia data was obtained from the DHS program website <https://www.dhsprogram.com>. Additionally, the zonal shapefiles were accessed without restrictions from the <https://africaopendata.org> website. The IPUMS International database, which offers census and survey data from all over the world, provided the census data used in the present study. The URL where the population and housing census dataset is available in <https://international.ipums.org/>

## Acknowledgments

This work was based upon research supported in part by the National Research Foundation (NRF) of South Africa, Ref.: RA210106581084, grant No. 150170, the South African DST-NRF-MRC SARChI Research Chair in Biostatistics (Grant No. 114613), and the RDP grant at the University of Pretoria, South Africa. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

## References

- [1] Danny Pfeffermann. New Important Developments in Small Area Estimation. *Statistical Science*, 28(1):40 – 68, 2013.
- [2] Barry Portnoy, Simon J Craddock Lee, Jennifer Kincheloe, Nancy Breen, Jean L Olson, Judy McCormally, and E Richard Brown. Independent state health surveys: responding to the need for local population health data. *Journal of Public Health Management and Practice*, 20(5):E21–E33, 2014.
- [3] J N K Rao and I Molina. *Small area estimation*. John Wiley & Sons, 2015.
- [4] Bharat Bhushan Singh, Girja Kant Shukla, and Debasis Kundu. Spatio-temporal models in small area estimation. *Survey Methodology*, 31(2):183, 2005.

- [5] Yong You and Mike Hidirolou. Application of sampling variance smoothing methods for small area proportion estimation. *Journal of Official Statistics*, 39(4):571–590, 2023.
- [6] Eric Lesage, Jean-Francois Beaumont, and Cynthia Bocci. Two local diagnostics to evaluate the efficiency of the empirical best predictor under the fay-herriot model. *Survey Methodology*, 47(2):279–298, 2021.
- [7] Shonosuke Sugawara and Tatsuya Kubokawa. *Mixed-Effects Models and Small Area Estimation*. Springer, 2023.
- [8] Jon NK Rao and Mingyu Yu. Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics*, 22(4):511–528, 1994.
- [9] Yegnanew A Shiferaw, Seyifemickael Amare Yilema, and Yikeber Abebaw Moyehodie. A hierarchical bayesian approach to small area estimation of health insurance coverage in ethiopian administrative zones for better policies and programs. *Health Economics Review*, 14(1):29, 2024.
- [10] Gauri S Datta, Partha Lahiri, Tapabrata Maiti, and Kim L Lu. Hierarchical bayes estimation of unemployment rates for the states of the us. *Journal of the American Statistical Association*, 94(448):1074–1082, 1999.
- [11] Seyifemickael Amare Yilema, Yegnanew A. Shiferaw, Temesgen T Zewotir, and Essey K. Muluneh. Spatial small area estimates of undernutrition for under five children in ethiopia via combining survey and census data. *Spatial and Spatio-temporal Epidemiology*, 42:100509, 2022.
- [12] Roberto Benavent and Domingo Morales. Multivariate fay-herriot models for small area estimation. *Computational Statistics & Data Analysis*, 94:372–390, 2016.
- [13] Francesco Branca, Laurence Grummer-Strawn, Elaine Borghi, M de Blössner, and M de Onis. Extension of the who maternal, infant and young child nutrition targets to 2030. 2015.
- [14] GH Choudry and JNK Rao. Small area estimation using models that combine time series and cross sectional data. In *Proceedings of statistics Canada symposium on analysis of data in time*, pages 67–74, 1989.

- [15] Malay Ghosh, Narinder Nangia, and Dal Ho Kim. Estimation of median income of four-person families: a bayesian time series approach. *Journal of the American Statistical Association*, 91(436):1423–1431, 1996.
- [16] Yong You and JNK Rao. Hierarchical bayes estimation of small area means using multi-level models. *Survey Methodology*, 26(2):173–181, 2000.
- [17] MD Esteban, D Morales, A Pérez, and L Santamaría. Two area-level time models for estimating small area poverty indicators. *Journal of the Indian Society of Agricultural Statistics*, 66(1):75–89, 2012.
- [18] G.S. Datta, P. Lahiri, and T. Maiti. Empirical bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, 102(1):83–97, 2002.
- [19] Domingo Morales, Maria Chiara Pagliarella, and Renato Salvatore. Small area estimation of poverty indicators under partitioned area-level time models. *SORT*, 39(1):19–34, 2015.
- [20] Yolanda Marhuenda, Isabel Molina, and Domingo Morales. Small area estimation with spatio-temporal fay–herriot models. *Computational Statistics & Data Analysis*, 58:308–325, 2013.
- [21] L Burck. *Robust small area estimation combining time series and cross-sectional data*. 1990.
- [22] Sharon L Lohr and NG Narasimha Prasad. Small area estimation with auxiliary survey data. *Canadian Journal of Statistics*, 31(4):383–396, 2003.
- [23] Sharon L Lohr and Trivellore E Raghunathan. Combining survey data with other data sources. 2017.
- [24] Junni L Zhang and John Bryant. Combining multiple imperfect data sources for small area estimation: a bayesian model of provincial fertility rates in cambodia. *Statistical Theory and Related Fields*, 3(2):178–185, 2019.
- [25] Jae Kwang Kim, Seunghwan Park, and Seo-young Kim. Small area estimation combining information from several sources. 2015.

- [26] Golshid Chatrchi, Héloïse Gauvin, and Allen Ding. Combining survey data with other data sources using small area estimation: A case study. In *Proceedings of the Survey Research Methods Section*, 2020.
- [27] Michael R Elliott, Trivellore E Raghunathan, and Nathaniel Schenker. Combining estimates from multiple surveys. *Wiley StatsRef: Statistics Reference Online*, 3, 2018.
- [28] Sadikul Islam and Hukum Chandra. Small area estimation combining data from two surveys. *Communications in Statistics-Simulation and computation*, 50(6):1728–1749, 2021.
- [29] Seyifemickael Amare Yilema, Yegnanew A Shiferaw, Temesgen Zewotir, and Essey Kebede Muluneh. Improving survey based estimates of malnutrition using small area estimation. *Statistical Journal of the IAOS*, 38(4):1261–1271, 2022.
- [30] Aldo Gardini, Silvia De Nicolò, and Enrico Fabrizi. A mixture-of-experts model to deal with the rural/urban dichotomy in small area estimation. *Journal of the Royal Statistical Society Series C: Applied Statistics*, page qlaf022, 2025.
- [31] Seyifemickael Amare Yilema. Bayesian small area estimation of anemia prevalence using a generalized linear mixed model. *Statistical Journal of the IAOS*, 41(1):171–179, 2025.
- [32] Jahn timer Daru, Javier Zamora, Borja M Fernández-Félix, Joshua Vogel, Olufemi T Oladapo, Naho Morisaki, Özge Tunçalp, Maria Regina Torloni, Suneeta Mittal, Kapila Jayaratne, et al. Risk of maternal mortality in women with severe anaemia during pregnancy and post partum: a multilevel analysis. *The Lancet Global Health*, 6(5):e548–e554, 2018.
- [33] UNICEF et al. The extension of the 2025 maternal, infant and young child nutrition targets to 2030: Who. 2021.
- [34] Gretchen A Stevens, Christopher J Paciorek, Monica C Flores-Urrutia, Elaine Borghi, Sorrel Namaste, James P Wirth, Parminder S Suchdev, Majid Ezzati, Fabian Rohner, Seth R Flaxman, et al. National, regional, and global estimates of anaemia



- by severity in women and children for 2000–19: a pooled analysis of population-representative data. *The Lancet Global Health*, 10(5):e627–e639, 2022.
- [35] World Health Organization et al. Proposed global targets for maternal, infant and young child nutrition. *Summary of main issues raised and WHO responses*. Geneva: World Health Organization, 2012.
- [36] UNSD UN-Habitat. Sdg indicator metadata, 2021.
- [37] World Health Organization et al. Committing to implementation of the global strategy for women’s, children’s and adolescents’ health (2016–2030): technical report. Technical report, World Health Organization, 2023.
- [38] World Health Organization et al. Maternal, infant and young child nutrition: Comprehensive implementation plan on maternal, infant and young child nutrition: biennial report. 2018.
- [39] Central Statistical Agency (CSA) [Ethiopia] and ICF. Ethiopia demographic and health survey. *Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF*, pages 1–551, 2016.
- [40] Yegnanew Alem Shiferaw. *Improved confidence intervals for a small area mean under the Fay-Herriot model*. PhD thesis, 2016.
- [41] Anoop Kumar, Shashi Bhushan, Rohini Pokhrel, and Walid Emam. Small area estimation using design based direct and synthetic logarithmic estimators. *Ain Shams Engineering Journal*, page 102836, 2024.
- [42] Ruilin Ren. Design-based small area estimation: an application to the dhs surveys. *SAE2021 BIG4small Book of short papers*, page 99, 2021.
- [43] Robert E. Fay III and Roger A. Herriot and. Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366a):269–277, 1979.
- [44] Hukum Chandra. Exploring spatial dependence in area-level random effect model for disaggregate-level crop yield estimation. *Journal of Applied Statistics*, 40(4):823–842, 2013.

- [45] Central Stastical Authority. Population and housing census of ethiopia. *Addis Ababa: Central Statistics Authority*, 2007.
- [46] Domingo Morales, María Dolores Esteban, Agustín Pérez, and Tomáš Hobza. A course on small area estimation and mixed models. *Methods, theory and applications in R*, 2021.
- [47] Debora F. Souza and Fernando A. S. Moura. Multivariate beta regression with application in small area estimation. *Journal of Official Statistics*, 32(3):747–768, 2016.
- [48] Andreea Erciulescu, Jianzhu Li, Tom Krenzke, and Machell Town. Hierarchical bayes small area estimation for county-level health prevalence to having a personal doctor. *Statistical methods & applications*, 33(4):1171–1191, 2024.
- [49] Priyanka Anjoy, Hukum Chandra, and Pradip Basak. Estimation of disaggregate-level poverty incidence in odisha under area-level hierarchical bayes small area model. *Social Indicators Research*, 144:251–273, 2019.
- [50] Seyifemickael Amare Yilema, Yegnanew A Shiferaw, Temesgen Zewotir, and Essey Kebede Muluneh. Multivariate small area estimation of undernutrition for children under five using official statistics. *Statistical Journal of the IAOS*, 38(2):625–636, 2022.
- [51] Seyifemickael Amare Yilema, Yegnanew A Shiferaw, Temesgen Zewotir, and Essey Kebede Muluneh. Small area estimation of undernutrition under age five children based on spatial-temporal models. *The Scientific World Journal*, 2022(1):6882047, 2022.

## Supplementary file (Sup 1)

Sup 1: Summary of CV for children aged 6-59 months  
anemia status in the Ethiopian zones

Zones	Sample Size	DESvy	GDESvy	SAE_Census	GSAE_Census
Addis Ketema	35	81.71	34.57	33.61	15.87
Afder	46	47.42	18.63	65.54	17.56
Agnuak	276	48.18	21.30	25.62	12.56
Akaki Kality	14	54.32	21.83	38.65	16.03
Alaba Sp Woreda	22	70.43	27.14	33.11	13.56
Arada	19	55.23	24.63	34.07	15.62
Arsi	138	46.15	18.70	21.15	11.16
Asosa	318	67.42	27.21	25.09	12.46
Awi	32	100.68	41.67	37.32	16.52
Awsa	180	39.73	17.15	13.00	6.73
Bale	65	56.50	25.47	28.89	13.78
Bench Maji	56	112.69	47.40	54.25	18.49
Bole	23	152.98	68.41	46.20	20.74
Borena	54	44.10	19.10	27.15	12.34
Central Tigray	228	65.36	26.41	23.64	12.46
Dawuro	33	113.08	48.90	29.31	15.80
Degehabur	157	45.10	18.82	18.28	8.44
Derashe	36	103.54	40.02	43.07	18.78
Dire Dawa	350	46.84	19.62	18.29	9.01
East Gojam	87	82.10	34.21	29.04	14.20
East Harerge	150	42.45	17.86	29.94	12.34
East Shewa	44	44.84	22.03	29.85	14.47
East Wellega	49	85.51	35.47	24.08	13.33
Fantana	131	41.42	17.83	20.04	9.45
Gabi	99	45.74	19.24	21.15	8.72
Gamo Gofa	107	60.21	23.31	31.65	13.51

Sup 1: Summary of CV for children aged 6-59 months  
anemia status in the Ethiopian zones

<b>Zones</b>	<b>Sample Size</b>	<b>DESvy</b>	<b>GDESvy</b>	<b>SAE_Census</b>	<b>GSAE_Census</b>
Gedio	67	53.94	21.92	35.34	13.28
Guji	78	45.27	19.25	21.57	11.26
Gulele	28	61.35	25.64	38.33	17.01
Gurage	70	66.28	26.58	27.38	12.20
Hadiya	87	64.79	26.96	29.88	13.72
Hareri	373	51.73	23.19	20.05	10.41
Hari	119	46.23	19.47	12.46	7.33
Huru Guduru wel	9	151.64	63.18	32.56	16.75
Ilubabor	64	58.73	20.21	26.86	12.86
JiJiga	160	37.28	15.45	12.24	5.84
Jimma	137	66.59	24.47	35.17	14.98
Keffa	69	112.37	48.89	45.22	20.93
Kelem Wellega	31	56.90	24.05	26.15	13.02
Kemashi	84	83.74	32.03	39.61	17.73
Kenbata Tembaro	39	68.46	31.04	32.33	13.89
Kilbet	215	48.92	17.96	31.07	11.04
Kirkos	20	99.29	36.57	78.24	24.35
Kolfe Keranyo	45	92.04	36.07	74.87	24.74
Konta sp worda	19	50.43	19.30	24.77	11.67
Korahe	97	38.41	14.75	12.82	6.27
Liben	112	38.74	17.02	17.66	7.42
Lideta	17	98.66	38.26	69.94	25.88
Metekel	257	108.64	45.12	31.32	15.66
Mezhenger	93	97.44	44.98	33.68	16.95
Nefas silk	41	65.37	28.76	35.53	16.64
North Gondar	183	76.18	32.39	29.13	13.95
North Tigray	132	71.10	30.37	31.04	14.29
North Wollo	50	79.55	30.22	28.91	14.31

Sup 1: Summary of CV for children aged 6-59 months  
anemia status in the Ethiopian zones

<b>Zones</b>	<b>Sample Size</b>	<b>DESvy</b>	<b>GDESvy</b>	<b>SAE_Census</b>	<b>GSAE_Census</b>
NShewaR3	74	74.06	29.88	33.30	14.94
NShewaR4	28	57.51	24.60	30.04	13.99
Nuer	135	67.51	32.08	24.67	11.94
NW Tigray	176	60.17	25.16	26.08	13.09
Oromia Sp Zone	16	80.52	35.02	25.40	13.40
Seltie	41	48.26	20.41	21.70	10.48
Shabelle	120	37.29	14.96	11.31	5.68
Shaka	8	136.74	49.11	32.84	17.57
Shinile	11	64.95	29.58	23.83	10.88
Sidama	167	60.03	24.12	27.47	12.50
South Gonder	107	90.65	34.62	27.99	15.09
South Omo	36	78.88	35.21	26.46	13.15
South Tigray	219	67.22	27.61	27.32	12.86
South Wollo	88	85.21	39.22	31.58	15.01
SW Shewa	13	31.83	12.51	19.11	9.29
Waghemira	36	74.34	32.24	34.87	15.99
Welewel	48	35.48	13.82	13.97	5.93
West Arsi sp zon	117	54.77	21.89	30.72	12.61
West Gojam	103	89.51	38.81	48.32	19.12
West Harerge	103	52.21	20.09	23.62	12.38
West Shewa	73	65.55	23.59	20.65	10.84
West Tigray	74	70.09	28.35	32.58	14.25
West Wolega	39	90.55	40.33	25.83	14.04
Wolayita	125	69.48	27.95	29.39	13.54
Yeka	63	59.42	23.95	31.69	14.61
Yem sp woreda	13	112.33	52.33	61.53	22.28