Discover Public Health

Research

Estimating the local-level child full polio vaccination rates in Ethiopia using a hierarchical Bayes small area estimation approach

Seyifemickael Amare Yilema[1,4] · **Yegnanew A. Shiferaw**[2] · **Haile Mekonnen Fenta**[3] · **Alebachew Taye Belay**[1]

Received: 22 February 2025 / Accepted: 15 May 2025
Published online: 24 May 2025

Abstract
Vaccination is one of the most effective, affordable, and life-saving medical interventions ever created. Child vaccination is fundamental to building a healthy and welfare society, which is crucial in 2063 African and 2030 global agendas. This study combines the 2019 Mini Ethiopian Demographic and Health Survey (EDHS) with the 2007 population and housing census datasets to employ the hierarchical Bayes (HB) small area estimation (SAE) approach for estimating local-level child vaccination rates. In the HB SAE framework, the deviance information criterion (DIC) was used to select the best candidate model among the three different models fitted. The logistic normal mixed model with known sampling variance was chosen over the other two models (Fay-Herriot model and log-normal mixed model). The mean coefficient of variation (CV) for direct survey-based estimates is 44.41, which is higher than that for the model-based HB estimates at 36.40. Similarly, the root mean square errors (RMSE) of direct survey estimates are greater than those of the corresponding model-based estimates. Therefore, the results suggest that the HB estimates show improvement over the survey-based estimates. This finding also contributes to the sustainable development goal for health (SDG3), which aims to ensure healthy lives and promote well-being for people around the globe.

Keywords Vaccination · Polio · Childhood · *Hierarchical Bayes* · Small area estimation

1 Introduction

A vaccine-derived poliovirus is a strain that is connected to the attenuated live poliovirus found in the oral polio vaccine (OPV). If it persists in under or unimmunized populations or replicates in an individual with a weakened immune system for an extended period, the weakened virus can mutate into a form that leads to sickness and paralysis [37]. Both the live attenuated OPV and the inactivated polio vaccine (IPV) are efficient means of controlling and curing poliomyelitis, an acute viral disease of the human central nervous system [2, 63].

Vaccination is one of the most effective, affordable, and life-saving medical interventions ever created. Child vaccination is the cornerstone for building healthy and thriving societies. Immunizations and vaccines save the lives

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1186/s12982-025-00695-3.

✉ Seyifemickael Amare Yilema, samarey1981@gmail.com; Yegnanew A. Shiferaw, yegnanews@uj.ac.za; Haile Mekonnen Fenta, hailemekonnen@gmail.com; Alebachew Taye Belay, alex016stat@gmail.com | [1]Department of Statistics, College of Natural and Computational Science, Debre Tabor University, Debre Tabor, Ethiopia. [2]Department of Statistics, University of Johannesburg, Johannesburg, South Africa. [3]Department of Statistics, College of Science, Bahir Dar University, Bahir Dar, Ethiopia. [4]Department of Statistics, University of Pretoria, Pretoria, South Africa.

of almost 4 million people globally each year [16]. An estimated 1.5 million children worldwide die from diseases that can be prevented with vaccinations each year [81]. The common goal of nations and partners worldwide is to accomplish much more with these life-saving tools. As per the African Health Strategy [29], Africans have good health and well-being, a high standard of living, and a high quality of life, which is an African aspiration on agenda 2063 [77]. According to Africa's agenda for children, every child is vaccinated against vaccine-preventable diseases, including polio, by 2040 [3]. However, the COVID-19 pandemic has significantly influenced immunization initiatives, which has negatively impacted the implementation of IA2030 [4]. The Polio Eradication Strategy 2022–2026 of the Global Polio Eradication Initiative (GPEI) is also specifically in line with IA2030, which aims to protect millions of lives from death.

The total coverage of IPV in sub-Saharan Africa is 73%, and the current global coverage of polio immunization is 85%, falling short of the 90% target [7]. As per the criteria set forth by the World Health Organization (WHO), children who have received all prescribed immunization doses by the age of 12–23 months are deemed fully immunized against polio [1, 22].

Like the rest of the world, in Ethiopia, infectious and transmissible diseases are a public problem, which accounts for about 60–80% of the health problems [35]. In addition, 16% of under-five mortality has been attributed to vaccine-preventable diseases [32]. The percentage of children aged 12 to 23 months who are fully vaccinated increased by 15%, from 24% in 2011 to 39% in 2016 [38]. In Ethiopia, only 30% of children living in rural areas were fully vaccinated at the appropriate ages. Furthermore, only 53% of children in rural areas had received the full recommended three doses of oral polio vaccine (OPV 3) [78].

The Ethiopian Demographic and Health Survey (DHS) collected data at both regional and national levels, providing crucial socioeconomic indicators [26]. However, there is a need for socioeconomic indicators at the sub-regional government levels and within the third administration layers, also known as zones, in Ethiopia. These indicators are essential for identifying specific population subgroups that may not be benefiting from particular development targets. Using more detailed and precise statistics can lead to the development of more effective policies and interventions [61, 74].

Conducting surveys in the lower administrative areas of developing countries can be difficult due to logistical issues such as transportation problems, challenging roads, and long distances between homes [36]. Additionally, resource constraints such as limited administrative, financial, time, human, and technical resources can hinder data collection at lower administrative levels, especially in developing countries. These challenges vary across and within countries and regions, with variations in management practices even within local governments in a single state [15].

To improve the accuracy of estimates from surveys for lower administrative areas, various researchers employed Small Area Estimation techniques [12, 19, 42, 45, 49, 53, 59]. These techniques involve borrowing strength from neighboring areas to improve the precision of estimates for target variables in areas with limited sample sizes [23, 59, 64, 86]. In other words, utilizing SAE methods can provide valuable insights at the lower levels of government officials and policymakers by integrating data from various sources [27, 30]. The demand for SAE is high in both developed and developing countries because it provides essential information for policymaking at the local level. SAE approaches are crucial for estimating socioeconomic indicators such as health, education, and environmental indicators, especially in smaller geographic areas like zones and districts [34, 53, 61] For example, the World Health Organization (WHO), the World Bank Group, and the International Monetary Fund (IMF) collaborate to disaggregate DHS data to the third level of government administration [13, 24, 57]. Host countries often ask the DHS Program for sub-regional indicator estimates to aid development planning and policymaking [62].

The importance of small area estimates for policymakers at smaller administrative levels cannot be overstated. There are several reasons why they are crucial:

(i) Small area estimates enable policymakers to target interventions and allocate budgets and resources to disadvantaged areas [79]. For example, our study aimed to identify zones with the highest child vaccination needs.
(ii) They provide detailed estimates for monitoring Sustainable Development Goals (SDGs) [13].
(iii) Small area estimation (SAE) techniques are essential for producing estimates for smaller geographic areas, such as zones in our case, that surveys are not designed to cover. This is achieved by using auxiliary variables from census and other administrative data sources [13, 60]. This information is particularly vital for making decisions at the local level, for example, in our case, zones.

Per our understanding, there is a research gap on applying SAE techniques for fully polio vaccination at the lower administration level in Ethiopia. While some studies have estimated undernutrition rates among children under the age

of five [85, 87, 88], there are also studies on food insecurity and mortality rates [67, 68]. Therefore, obtaining information at the local level is crucial for policy planning and intervention. Our main objective is to estimate fully polio vaccination of children at the zone level using the HB SAE approach.

2 Method and materials

2.1 Data source and study variables

Ethiopia is divided into nine regions and two administration cities, which are further divided into 83 zones (see supplementary file S1). Sample survey design only considers the regional and national levels, with no specific allowance for data collection at the zonal level. This study focuses on the full polio vaccination status of children aged between 12 and 23 months, which was obtained from the 2019 Mini EDHS. The survey used a nationally representative sample to provide estimates of the vaccination status of children in both urban and rural areas, at the national and regional levels [26].

To analyze the data, we followed the WHO protocol. The target variable, which is vaccination status, was categorized as binary. Children were classified as either incompletely vaccinated (missing one or more of the four vaccines or not receiving any vaccines) or fully vaccinated [28, 52].

Suppose Y_{ij} is a binary outcome (zero for no or incomplete vaccine and one for fully vaccinated child) for individual child j ($j = 1, 2, \ldots, N_i$) in zone i ($i = 1, 2, \ldots, m$) with N_i the population size in zone i. Let the sampled values be denoted by y_{ij}, where n_i is the sample size drawn each zone i. Suppose $N = \sum_{i=1}^{m} N_i$ and $n = \sum_{i=1}^{m} n_i$ denote the total population and total sample sizes, respectively. The aim is to estimate the true area proportions P_i in the zone i. Where $P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} Y_{ij}$ that in this case, can be estimated from the EDHS data using the direct survey estimator as follows $p_{iw} = \frac{1}{\sum_{j=1}^{n_i} w_{ij}} \sum_{j=1}^{n_i} w_{ij} y_{ij}$ where w_{ij} represents the survey weight.

The target variable is defined as follows

$$Y_{ij} = \begin{cases} 1, & \textit{If a child is fully vaccinated} \\ 0, & \textit{Otherwise} \end{cases}$$

The 2007 census data and the 2019 MDHS data are combined for the applications of SAE techniques. For the analysis of SAE, auxiliary covariates are taken from the census data to borrow the strength of the estimates [25]. The proportion of full polio vaccination is the variable of interest at the zones level. The auxiliary variables have been taken in two ways from the census data via variables related to 12–23 months and parents. For instance, sex (male and female) and ages (1–2 and 4–5 years) have been taken from children under age five. On the other hand, the auxiliary variables of parents were sex (male and female), place of residence (urban and rural), age (15–24, 25–34, 35–44 and 45–49), source of drinking water (improved and unimproved), educational levels (non-educated, primary and secondary and above), literacy (literate and illiterate), marital status (married, never married and others), type of toilet facility (have toilet facility and doesn't have toilet facility), the number of sons died (no died, one died and two and more died), the number of daughters died (no died, one died and two and more died), the number of families in the household (less than five, and five and more), and disability (disabled and not disabled), and employment status (government-employed, privately employed, self-employed, employer, unemployed and other employment) [85].

The Fay-Herriot model being considered in this study is an aggregate level (or area level) model that connects small area proportions to area-specific auxiliary variables. These models are crucial when unit-level data is not available. As a result, it is common practice to include area-level aggregated auxiliary variables (such as the proportions of individuals in each category within specific zones) derived from population census data and administrative records in the process of fitting SAE models. As highlighted by Rao and Molina [61] and other SAE literatures [34, 74], the success of any model-based method relies on the availability of good auxiliary data.

To ensure the success of our model-based method, we focused on gathering zonal level predictor auxiliary data. We first conducted a correlation analysis to identify any covariates that were significantly correlated with children full polio vaccination, such as proportions of rural place of residence, married marital status, literate, access to improved drinking water, lack of toilet facilities, private and self-employed employment status, age group (15–24 years), and more than five

household sizes at zones level. We then used stepwise regression analysis to select relevant auxiliary variables within the HB-SAE framework for further analysis [41]. To present further evidence of the use of stepwise regression, Chandra [21] conducted a stepwise regression analysis, while Rao and Molina [61] examined the selection of auxiliary predictor variables from a set of 42 census variables using backward stepwise regression analysis.

2.2 Small area models

2.2.1 Fay-Herriot model

The Fay-Herriot (FH) model is a fundamental area-level model that relies on direct aggregate area-level survey estimates. It also includes accessible auxiliary variables that are derived from various secondary sources, primarily census data sets [20, 30, 34]. As a result, the FH model consists of two components [34, 61]:

1. Sampling model for the direct survey estimates, which accounts for the sampling variability of direct survey estimates (p_{iw}) of the population parameter P_i, which is given by

$$p_{iw} = P_i + \varepsilon_i, \ \ i = 1, 2, \cdots, m \tag{1}$$

   where, p_{iw} is the direct survey-based estimator of P_i with $E(p_{iw}|P_i) = P_i$ and $Var(p_{iw}|P_i) = D_i$, P_i denotes the true small area proportion that is to be estimated, $\varepsilon_i \sim N(0, D_i)$ are independent and normally distributed sampling error associated with p_{iw}
2. Linking model to incorporate area-specific auxiliary variables through a linear regression framework.

The linking model, the second component, connects the population parameter P_i to a vector of auxiliary variables ($\mathbf{X}_i$). $\mathbf{X}_i$ is a $k \times 1$ area-specific covariates, often obtained from administrative records or population census. The linking model is expressed as follows

$$P_i = \mathbf{X}_i^T \boldsymbol{\beta} + \boldsymbol{\nu}_\mathbf{i} \tag{2}$$

where, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^T$ is a $p \times 1$ unknown regression coefficients and $\nu_i \sim N(0, \sigma_\nu^2)$ are independent and identically distributed area-specific random effects. Random area-specific effects are included in the linking model to account for dissimilarities between areas. It is common to assume that two random error components follow a normal distribution and are independent within and across areas (zones). However, a possible limitation of the model (1) is that assuming a linear linking model with normal random effects may lead to unreliable and unrealistic estimates when the target variable is proportion. Therefore, the logistic or logit link is preferred to overcome these challenges.

Thus, the linking model is expressed as,

$$logit(P_i) = \boldsymbol{X\beta} + \boldsymbol{\nu} \tag{3}$$

With $P_i = \frac{\exp(\boldsymbol{X\beta}+\boldsymbol{\nu})}{1+\exp(\boldsymbol{X\beta}+\boldsymbol{\nu})}$, when the small area proportion P_i is modeled using the logit function, the estimates always fall within the permissible range of (0, 1].

2.2.2 Hierarchical Bayes Fay-Herriot (FH) model

Many studies have previously focused on frequentist approaches to SAE model [56, 62, 71, 85]. However, we aim to compare the performance of different HB estimators to the direct survey estimator using a Bayesian perspective. In the Bayesian paradigm, model parameters are treated as random variables instead of fixed, unknown values [47, 48]. By applying Bayes' Theorem, we can use prior knowledge about the parameter of interest and model parameters to obtain a posterior distribution for the parameters based on the likelihood function derived from the data [9, 33, 69, 80]. The HB approach can effectively handle complex small area models using the Monte Carlo Markov Chain (MCMC), which overcomes the computational difficulties of high-dimensional integrations of posterior densities. We explored three models for HB estimations: Model 1, which employed the FH model; Model 2, the FH log-normal mixed model; and Model 3, the FH logistic-normal mixed model, under the umbrella of the FH model [10, 69, 72, 76].

The standard SAE under the Fay-Herriot model assumes continuous gaussian response variables, however the outcome variable in this study is whether or not a child has received full polio vaccinations. Therefore, instead of using linear modeling that assumes a gaussian normal, a logistic link function is more suited. To capture variability across the small areas (zones in this case), the HB-SAE method works well for managing logit predictors that can contain random effects in the link function. For binary outcomes, HB-SAE model is used, developed and built upon by the previous studies that used the logit link function [74, 75, 90].The binary outcome variable is suitably modeled using the transformation (logit) function to express the likelihood of success on the log-odds scale. The sampling and unmatching modeling were also adopted and utilized by the researchers [74, 75, 84, 90].

The following are descriptions of the three models:

Model 1: Fay-Herriot model with known sampling variance.

Sampling model: $p_{iw}/P_i \sim^{ind} N(P_i, \sigma^2_{ei}), i = 1, 2, \cdots, m.$

Linking model: $P_i/\boldsymbol{\beta}, \sigma^2_v \sim^{iid} N(x_i\beta, \sigma^2_v), i = 1, 2, \cdots, m.$

Model 2: Log-normal mixed model with known sampling variance.

Sampling model: $p_{iw}/P_i \sim^{ind} N(P_i, \sigma^2_{ei}), i = 1, 2, \cdots, m.$

Linking model: $log(P_i)/\boldsymbol{\beta}, \sigma^2_v \sim^{iid} N(x_i\beta, \sigma^2_v), i = 1, 2, \cdots, m.$

Model 3: Logistic-normal mixed model with known sampling variance.

Sampling model: $p_{iw}/P_i \sim^{ind} N(P_i, \sigma^2_{ei}), i = 1, 2, \cdots, m.$

Linking model: $logit(P_i)/\boldsymbol{\beta}, \sigma^2_v \sim^{iid} N(x_i\beta, \sigma^2_v), i = 1, 2, \cdots, m.$

Both Model 2 and Model 3 are considered unmatched models [54]. The Fay-Herriot model relies on a linear mixed model that assumes a linear relationship between auxiliary variables, which are borrowed from census datasets, and direct survey estimates. However, in the two models examined, which use log and logit links, the actual relationship between these auxiliary variables and the direct estimates is nonlinear. Consequently, these models are referred to as unmatched [55, 61]. In models 1–3, the sampling model accounts for the sampling variability of the survey estimates, while the linking model combines the survey estimates with the auxiliary variables.

In the SAE approach, it is known that direct survey-based estimates at the lower administrative level (such as zones) are often unreliable due to small sample sizes. SAE approaches are typically used to enhance the precision of these estimates by combining survey data with census or administrative records to leverage information from related areas. In the HB SAE approach, the HB posterior samples benefit from the prior parameter specification. This is because priors can help stabilize estimates, particularly in cases where data is limited or noisy, as in SAE. Moreover, priors enable information to be shared across different areas, thereby enhancing the precision of estimates in areas with limited data.

Furthermore, conclusions reached from the posterior densities depend on a wide range of prior distributions, it is important to note that prior distribution selection is a key component of Bayesian analysis. The purpose of non-informative prior distributions is to enable Bayesian inference for parameters that are largely unknown outside of the data used in the current research [70, 89, 92]. In accordance with [39] concept, we have taken into consideration the uniform prior for σ^2_v, that is Uniform (0, 1000) and distribution of $\boldsymbol{\beta}$ has been taken to be N(0, 1,000,000) [11].

The rationale for using a normal prior with a large variance ($\sigma^2 = 10^6$) for the regression coefficients ($\boldsymbol{\beta}$) is that this large variance effectively flattens the prior distribution. This approach approximates a "no prior knowledge" scenario while still ensuring computational stability. Additionally, we opted for a prior for the random effects variance (σ^2_v) that follows a uniform distribution over a wide interval (0, 1000). This choice ensures that all plausible values of σ_v are considered equally likely a priori, avoiding any artificial constraints. This method of selecting diffuse priors is standard in Bayesian SAE when strong prior information is not available [11].

The posterior distributions of $\boldsymbol{\beta}$, $\boldsymbol{\sigma^2_v}$ and $\boldsymbol{\nu_1, \nu_2, \cdots, \nu_m}$ generated from the MCMC. The HB small area proportion estimates are computed for all three models using the MCMC techniques, drawing random samples from the full conditional distributions of posterior quantities [89]. To acquire the results of each HB estimate of the true proportions P_i, we employ 50,000 sampling iterations (the length of each Markov Chain) [89, 91]. The deviance information criterion (DIC) is used when selecting the best candidate model. The smaller the DIC value, the better the model. The analysis is performed using the BayesSAE R software package [66].

2.2.3 Generalized variance function

The generalized variance function (GVF) is used to reduce the uncertainty of the design-based variance estimate in complex situations [31, 34, 82, 83]. Generally, the usual designed-based estimators have as big variabilities because of the small sample sizes. For smaller domains (zones), stable estimators play a crucial role. Compared to direct design-based variance estimators, the GVF produces domain-level variance estimators that are more trustworthy [68]. The bias-corrected GVF was taken into account for this study by fitting the simple linear regression model which is given as follows:

$y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, 2, \cdots, m$, where the dependent variable y_i is $log(Var(P_i))$, and the independent variable x_i is, the $log(P_i)$ success probability at the zone level. β_0 and β_1 are the constant and slope parameters estimates, and $e_i \sim^{iid} N(0, \sigma^2)$ and therefore, the GVF function is $GVF = \exp\left(\frac{\hat{\sigma}^2}{2}\right)\exp(\hat{\beta}_1 + \hat{\beta}_1 x_i)$ [67, 68, 82, 83, 85].

In Fig. 1, we can see the comparison between the estimated sampling variance and predicted variances (based on GVF) versus the direct survey estimates for full polio vaccination rates across the Ethiopian zones. The plot illustrates that the GVF effectively smooths out the unreliable and noisy estimated variance by reducing the direct survey estimates. The GVF variance estimates (blue color) and the direct survey variance estimates (red color) shows that the GVF variance estimates are smaller than the direct survey variance estimates.

2.3 Ethical review

The research used data from the DHS Program. The DHS surveys are conducted following rigorous ethical standards and protocols. Specifically, the survey instruments and methodologies employed by the DHS have been thoroughly reviewed and approved by the ICF Institutional Review Board (IRB). Additionally, the IRB of the host country ensures that these methods comply with local laws and the U.S. Department of Health and Human Services' requirements for the protection of human subjects (45 CFR 46). Given that this data has already undergone ethical review, no additional IRB approval was required for our analysis.

3 Results

The auxiliary variables for the HB-FH model were derived from the 2007 Ethiopian population and housing census datasets. The response variable used was the proportion of children that had received full polio vaccination obtained from the 2019 mini-EDHS. The mixed logistic model with known sampling variance was found to be the best model with the smallest DIC value according to the goodness-of-fit statistics displayed in Table 1.

Table 2 shows the posterior mean, standard deviations and 95% credible intervals for the model parameters of the mixed logistic model with HB inference. The auxiliary variables, such as the proportions having more than five families in the household and families living in rural areas were significant predictors for children full polio vaccination.

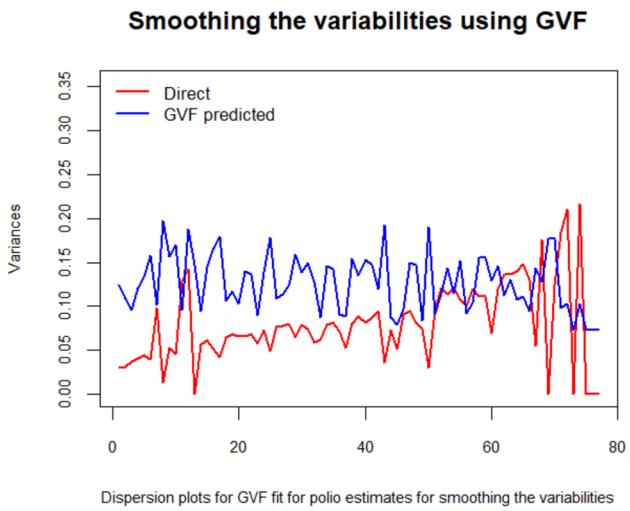

Fig. 1 GVF for stabilizing the variabilities of direct estimates

Table 1 DIC values for the three models

Model	DIC
FH model	40.66
log FH model	38.65
Logit FH model	36.560

Table 2 The posterior means with standard deviations and 95% credible intervals for the model parameters of the mixed logistic model with HB inference

Auxiliary variables	Posterior mean	Std. dev	Credible interval	
			Lower (2.5%)	Upper (97.5%)
Intercept	−6.69	4.27	−14.65	2.68
Rural	1.99	1.06	0.14	4.31
Improved	1.59	1.50	−1.19	4.08
Literate	4.29	3.81	−4.63	10.97
Married	−1.14	1.53	−4.14	1.81
No toilet	−0.17	1.54	−3.24	3.05
Private employed	2.34	2.05	−1.54	6.40
Self employed	0.91	1.67	−2.44	3.95
15–24 years age	3.11	2.51	−1.83	7.62
More than five family size	−2.64	1.47	−5.88	−1.16
Variance of residual in the linking model σ_v^2				
	0.043	0.033	0.006	0.13

This study used the goodness of fit diagnostic test to determine the statistical significance of HB and survey-based direct estimates. The null hypothesis was that HB estimates are not significantly different from the direct estimates. The alternative hypothesis stated that HB and direct survey estimates are statistically different. The goodness of fit diagnostic test was computed using the Wald statistic (W) test for estimating children full polio vaccination [14, 17]. The Wald statistic can be calculated as follows:

$$W = \sum_i \frac{(directestimate_i - HBestimate_i)^2}{Var(directestimate_i) + mse(HBestimate_i)} \sim \chi_m^2$$

where mse(HBestimate) is the mean square error (posterior variance estimate). The Wald statistic has a chi-square distribution (χ_m^2) with m degrees of freedom (m is the number of zones). The p-value obtained from the Wald statistic (W) is 0.199, which suggests that we don't have enough evidence to reject the null hypothesis. Hence, the HB and the direct survey estimates are statistically equivalent, which indicates that the fitted HB SAE model is considered to have a good fit. When the two estimates, particularly the mean HB estimates and direct survey estimates, are considered statistically equivalent, it typically means that their statistical mean attributes are comparable. This often indicates that there should be a reasonable amount of agreement between the direct estimates and the HB estimates.

Figure 2 shows the direct survey and model-based estimates of the children full polio vaccination rates in Ethiopian zones. The bias diagnostic measure validates the accuracy of the model-based estimates, as outlined by different literatures [14, 17, 21]. We performed a regression of direct estimates on model-based small area estimates to conduct the bias diagnostic. This helps us assess the similarity of the model-based estimates (i.e., HB SAE-based estimate) to the true population values. By placing the model-based estimates on the y-axis and the direct estimates on the x-axis, we can observe how the small area estimates deviate from the values that the regression line predicts.

As shown in Fig. 2, the small area estimates from the HB model are not as extreme as the direct estimates, which is the expected result of HB-SAE that tends to reduce extreme values towards the average. Compared to direct estimates, HB estimates are intended to be consistently unbiased estimate.

Figure 3 displays the credible interval and HB estimates across the Ethiopian administrative zones with increasing the HB estimates. There is a high degree of assurance in the HB estimations since the HB credible interval for children's full polio vaccination is not very wide throughout the zones. This result also shows that there are significant variances between the administrative zones given that there are no intervals that coincide.

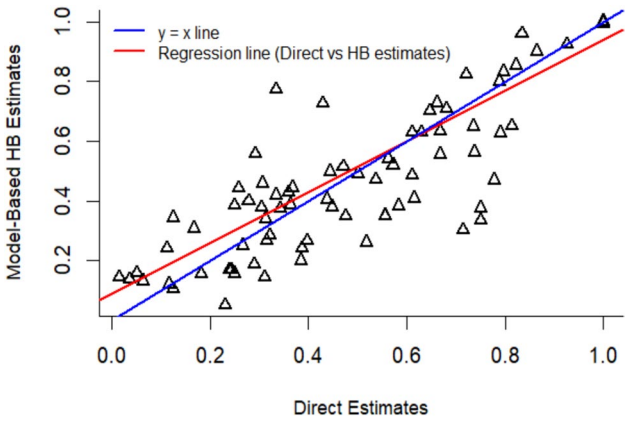

Fig. 2 The direct (*red line*) and model-based HB (*blue lines*) estimates of children full polio vaccination coverage to indicate the consistency of the two estimates

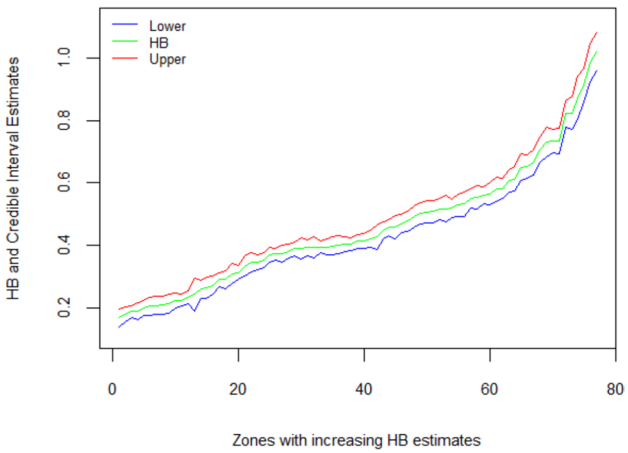

Fig. 3 Plots of HB estimates with 95% credible intervals across the Ethiopian administrative zones, where the upper boundary of the credible interval is marked in red, the HB estimates in *green*, and the lower boundary in *blue*

Table 3 Summary statistics of the CV and root mean square error for full polio vaccination coverage among children, comparing direct and HB estimates

	Sample size	Root mean square error		CV	
		Direct	HB	Direct	HB
Min	5	0.00201	0.0009	4.47	0.99
1st Q1	16	0.00270	0.0018	31.52	13.54
Median	31	0.00551	0.0021	44.90	14.81
Mean	41	0.00930	0.0023	43.87	14.71
3rd Q3	53	0.01167	0.0028	55.60	16.89
Max	251	0.11103	0.0083	79.61	28.78

The CVs are used to compare the accuracy of different estimates. The CVs display the sampling variability as a percentage of the estimate. Smaller CVs are preferable and estimates with larger CVs are considered untrustworthy. However, there are no universally recognized guidelines for determining what is "too large" or "too small." Different organizations use varying cut-off points for CVs when releasing estimates for public use. For example, the UK's Office for National Statistics considers estimates with a CV value of 20% or less to be appropriate [18]. The CV values for the HB and the direct children full polio vaccination percentage estimations were calculated.

We examined the accuracy of small area estimates generated by the HB-SAE method for children full polio vaccination. Table 3 summarizes the percentage coefficient of variations (CVs) for both the direct and HB estimates. In addition, we presented the zonal-wise CVs for both the direct and HB estimates of children full polio vaccination in Fig. 4 (left). These findings indicate that the CVs of the direct estimates are higher than the CVs of the HB estimates. Therefore, the estimates acquired through HB are deemed to be more reliable than the direct estimates.

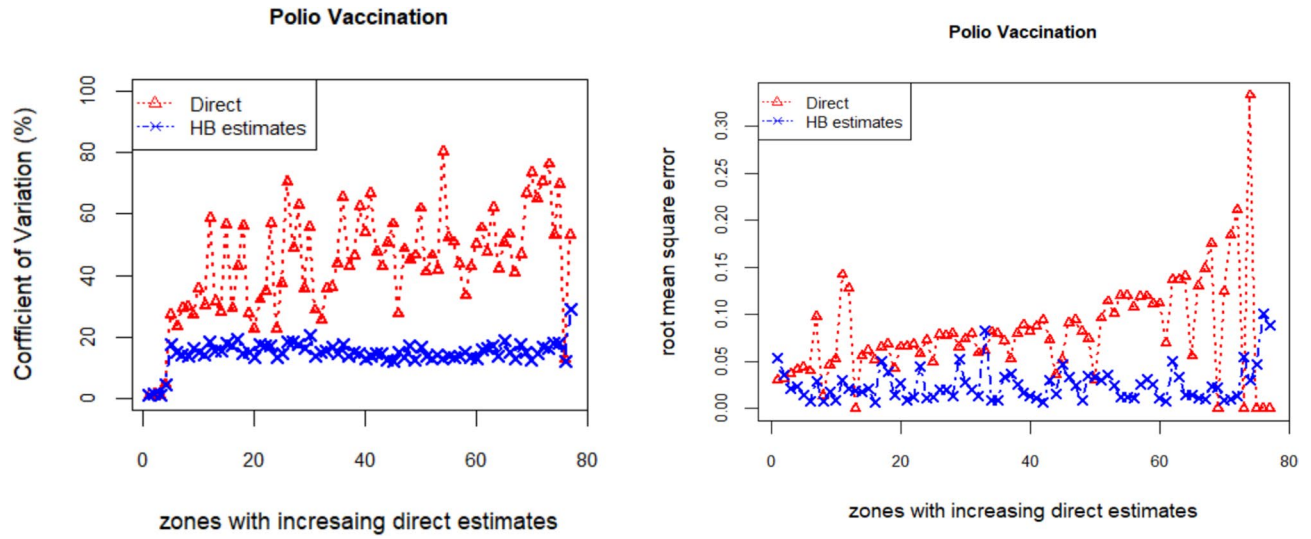

Fig. 4 The model-based HB estimates (*blue*) and direct estimates (*red*) for CV and root means square errors across the zones for measuring the precisions of the estimates

Table 3 and Fig. 4 show that direct survey estimates of the target variables are unstable, with CVs ranging from 4.47 to 79.61%% and an average of 44.90% for direct estimates. However, the CVs for HB estimates have a slight narrow range, varying from 0.99 to 28.78%. These results suggest that HB estimates are more precise than direct survey estimates.

The root mean square error in the HB-SAE method is used to quantify the uncertainty of model-based estimates. The smaller root means square error presented in Fig. 4 (right) shows the HB estimates are smaller than the corresponding direct estimates across the zones. Therefore, the HB estimates are more precise and reliable over the direct estimates, The model-based HB-SAE methods successfully reduced the uncertainty by using auxiliary variables from the census data. Therefore, the HB-SAE model estimates has improved in precision compared to the direct survey estimates. The detailed zonal wise SAE estimations are presented in the supplementary file (S2).

4 Discussion

Over the past 40 years, the DHS has been collecting data at the local level rather than just at the provincial and national policy levels [34]. This has led to an increased demand from countries looking to access local-level data for policy interventions [50, 62]. To meet this need, researchers have used Bayesian approaches, which are more computationally feasible than the maximum likelihood approach due to the complexity of the models used in SAE techniques [8, 66]. This study utilized the HB approach of SAE to produce reliable and representative zonal-level children full polio vaccination rates for Ethiopia. The analysis used data from the 2019 DHS linked with the 2007 Population Census.

The primary objective of this study was to estimate the proportion of children full polio vaccinations at the zone level in Ethiopia using the HB SAE approach. The study combined the 2007 population and household census with the 2019 Ethiopian DHS datasets using the HB-SAE modeling framework [25, 26]. This method offers a cost-effective and state-of-the-art monitoring system for child vaccinations, which can greatly benefit the Ethiopian government and stakeholders.

Our research has aimed to ensure that the estimates of child vaccination obtained through the HB method are consistent with those obtained through survey-based methods. To achieve this, we conducted a Wald statistic test, known as the goodness fit test. The results of the test showed that the estimates obtained through the HB method and direct survey estimates were statistically equivalent [68, 88].

In small area statistics measuring uncertainty of the estimates are crucial. Uncertainty is the degree of variation within or between the random area effects (zones in this case). To quantify the uncertainty in SAE the root mean square and confidence intervals are used based on the literatures [17, 61, 73]. Model-based HB-SAE approaches are encouraged to be used because direct survey based estimations methods are often unstable and unreliable due to small sample sizes [17, 61, 73, 74]. To assess the uncertainties of the estimates, we employed the root mean square error and model diagnostics tests, which are in line with earlier study findings for small area estimates [17, 58, 88]. The HB-SAE is employed

to reduce the uncertainties and produce more precise estimates by borrowing strength from the auxiliary variables and neighboring areas [17, 46, 60, 61, 73, 74]. In the current study shown in Fig. 4 (right), the root mean square error for model-based HB estimates are much lower than the corresponding direct estimates indicating more precise estimates of the mode-based estimates. This is also indicating that the HB estimates of child vaccination are more precise and stable than survey-based estimates. Similar findings were found in literatures; Anjoy and Chandra [8] conducted a study to estimate the incidence of poverty in Odisha at a disaggregated level using an area-level HB-SAE model. The results indicated that the estimates provided by the HB-SAE approach were more accurate than those obtained from direct survey estimates. For Ethiopian undernutrition survey data for children under five, similar studies employing multivariate, spatial, and spatiotemporal SAE techniques showed that these model estimates are more accurate and reliable than the equivalent traditional SAE estimates [87, 88]. According to other research that closely matched the current study on the HB-SAE model to estimate Community-Based Health Insurance (CBHI) coverage at the zone levels in Ethiopia, the HB-SAE model is more reliable and precise than direct survey estimates and traditional SAE estimates [70] since the CV for the HB-SAE is smaller than the corresponding SAE estimates. Our findings are supported by additional research showing that direct survey estimates typically had higher CVs and root mean square errors than HB estimations [40, 41, 44, 85]. In general, using auxiliary variables from the 2007 population census significantly improves the precision and accuracy of HB zonal level estimates of children anemia [61].

Model-based estimates typically enhance precision; however, they can introduce bias due to factors such as model misspecification. For example, a model like the FH Model 1, which assumes a linear relationship, may result in bias if the true relationship is nonlinear. We assessed three linking models: normal, log-normal, and logistic-normal, and observed consistent results, as detailed in Table 1. Furthermore, our analysis revealed that model-based estimates yielded lower CVs and RMEs compared to direct estimates. Our results suggest that any potential bias is likely minimal and is overshadowed by the gains in precision. However, we advise caution in relying solely on point estimates without considering uncertainty intervals. Therefore, we strongly recommend considering uncertainty intervals, such as 95% credible intervals (as illustrated in Fig. 3), alongside point estimates in decision-making processes.

According to Table 2, the 95% credible interval for the coefficient of the proportional auxiliary variable, which is the proportions of households in zones having higher proportion of big family, indicating a significant effect on the vaccination rates of children. This implies that the size of a household plays a vital role in determining the proportion of households that have received full polio vaccination. Other related studies have shown that as a family's size increases, the duration of breastfeeding also increases, but there is a decrease in vaccination coverage [6]. In families with three or more children, the percentage of unvaccinated children, as per the National Preventive Vaccination Schedule, was higher than that of families with fewer children [6]. Furthermore, the spread of poliovirus within unvaccinated households varied greatly across communities with differing vaccination coverage. The community with 70% OPV vaccination coverage had the highest transmissions, while the 10% coverage had the lowest [5]. These findings imply that the number of children in a household can affect breastfeeding duration and vaccination coverage. Larger households are associated with lower vaccination rates. Additionally, the coefficient of the auxiliary variables, which is the proportion rural place of residence has a 95% credible range that does not include zero. This indicates that living in a rural area has a significant impact on the vaccination rates of children. Therefore, the rural place of residence is a crucial factor in determining the percentage of households that have completed full polio vaccination. Studies conducted in different regions have shown significant differences in vaccination coverage based on the place of residence. For instance, in Ethiopia, vaccination coverage was significantly higher in the Mettu district compared to the Sinana district [51]. Similarly, the northern states of Uttar Pradesh and Bihar in India had different associations between polio program exposure and children's full immunization status [43, 47]. In Malaysia, the actual vaccination coverage in Kuala Lumpur was much higher than the estimated coverage, reflecting the success of the immunization campaign in that specific place [65]. These findings suggest that the place of residence plays a crucial role in determining the percentage of households that have received full polio vaccination, with variations observed across different regions.

5 Conclusion

In this study, we used the HB-SAE area level model to analyze the variation in children full polio vaccination across different zones of Ethiopia. For this purpose, we were obtained the target variable from the 2019 DHS dataset, while the auxiliary variables from the 2007 census dataset. The findings of our study reveal that the zonal level HB estimates outperform direct estimates in terms of the root mean square error and the CVs of children full polio vaccination.

Based on the findings of the study, the HB-SAE approach is significantly more efficient in generating estimates of children full polio vaccination at the zonal level. Our research has highlighted the usefulness of this approach in tackling the problem of small sample sizes to produce precise and less expensive estimates at a disaggregated level from the current survey.

The Ethiopian government places great emphasis on local administrative planning to tackle health-related concerns. As zones are a crucial planning domain, having statistics at the zonal level is imperative for policy monitoring and intervention throughout the country. This outcome contributes to the sustainable development goal for health (SDG3), which aims to ensure healthy lives and promote well-being for people of all ages.

5.1 Policy recommendations

We recommend targeted strategies for policymakers based on our findings, emphasizing areas of low coverage identified in our model-based estimates as follows:

Geographically targeted interventions: Our model-based estimates (refer to Supplementary File Two (S2)) have pinpointed specific zones in the Somali and Afar regions, including Afder, Nogob, Korahe, Shabelle, Shinile, Degehabur among others, where full polio vaccination rates consistently fall below 20%. To address this issue, policymakers should consider establishing temporary clinics in these areas, particularly in remote and hard-to-reach locations, especially within pastoralist communities that lack adequate health infrastructure. Furthermore, collaborating with local leaders, such as religious institutions and kebele administrators, will be crucial in overcoming vaccine hesitancy and fostering trust in vaccination initiatives.

Strengthening health systems: Health extension workers in Ethiopia are vital to the primary healthcare system, particularly in rural and underserved regions. The government should prioritize these workers to conduct door-to-door vaccination campaigns in high-risk areas, using our estimates to allocate resources effectively.

Conflict-sensitive approaches: Since 2020, Ethiopia has experienced internal conflicts. In the conflict-affected areas, such as all of Amhara and parts of Oromia, Tigray, and other regions, the government should collaborate with NGOs like the Ethiopian Red Cross to ensure the safe delivery of vaccines.

5.2 Limitations of the study

While our study advances the understanding of polio vaccination rates, there are few limitations that warrant caution. First, the 2007 Ethiopian Census and the 2019 Mini-EDHS may not fully capture recent demographic changes, such as urbanization and displacement due to conflict. We recommend that future research explore alternative data sources, including satellite-derived nighttime lights and mobility data, to more accurately assess these population shifts. Second, vaccination reporting biases within the DHS may influence direct estimates. Future studies should consider integrating spatial random effects to reduce bias from unobserved spatial heterogeneity.

Acknowledgements We are grateful to the DHS for making their data available for our research.

Author contributions SAY has been responsible for data management, writing the R code, analysis, and writing the draft manuscript. YAS, HMF, and AT were involved in critically revising and reviewing the manuscript. All authors have read and approved the manuscript.

Funding The authors did not receive support from any organization for the submitted work.

Data availability The 2019 mini-EDHS children full polio vaccination data was accessed from the DHS repository https://www.dhsprogram.com. In addition, 10% of the 2007 census data are freely accessible for research purpose by requested. In addition, the census data is accessed freely from the Integrated Public Use Microdata Series (IPUMS) by registering on their webpage.

Clinical trial number Not applicable.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Abebe AM, Mengistu T, Mekuria AD. Measles case, immunization coverage and its determinant factors among 12–23 month children, in Bassona Worena Woreda, Amhara Region, Ethiopia, 2018. BMC Res Notes. 2019;12(1):71. https://doi.org/10.1186/s13104-019-4104-8.
2. Adedeji AO, Okonko IO, Adu FD. Sabin and wild type polioviruses from children who presented with acute flaccid paralysis in Nigeria. Afr Health Sci. 2012;12(3):345–54. https://doi.org/10.4314/ahs.v12i3.16.
3. African Agenda. Africa's Agenda for Children 2040 Fostering an Africa Fit for Children Background to the Agenda. 2015.
4. Agenda I. Immunization Agenda 2030 Global Report. 2021.
5. Alekseeva AV, Berezkina EN, Moiseeva KE, Kharbediya SD. The impact of the number of children in the family on the breastfeeding duration and vaccination coverage. Kazan Medical J. 2020;101(4):538–43. https://doi.org/10.17816/KMJ2020-538.
6. Altamirano J, Purington N, Behl R, Sarnquist C, Holubar M, García-García L, Ferreyra-Reyes L, Montero-Campos R, Cruz-Hervert LP, Boyle S, Modlin J, Van Hoorebeke C, Leary S, Huang C, Sommer M, Ferreira-Guerrero E, Delgado-Sanchez G, Canizales-Quintero S, Díaz Ortega JL, Desai M, Maldonado YA. Characterization of household and community shedding and transmission of oral polio vaccine in Mexican communities with varying vaccination coverage. Clin Infect Dis. 2018;67:121–6. https://doi.org/10.1093/cid/ciy650.
7. Anand A, Pallansch MA, Estivariz CF, Gary H, Wassilak SG. Estimating the likely coverage of inactivated poliovirus vaccine in routine immunization: Evidence from demographic and health surveys. J Infect Dis. 2014;210 Suppl 1:S465–74. https://doi.org/10.1093/infdis/jiu343.
8. Anjoy P, Chandra H. Hierarchical Bayes aggregated level spatial model for crop yield estimation. J Indian Soc Agric Statistics. 2019;73(2):143–52.
9. Anjoy P, Chandra H. Hierarchical Bayes estimation of small area means under a spatial nonstationary Fay-Herriot model. Commun Stat Simul Comput. 2021;52(7):1–19. https://doi.org/10.1080/03610918.2021.1926501.
10. Anjoy P, Chandra H, Basak P. Estimation of disaggregate-level poverty incidence in Odisha under area-level Hierarchical Bayes small area model. Soc Indic Res. 2018;144(1):251–73. https://doi.org/10.1007/s11205-018-2050-9.
11. Anjoy P, Chandra H, Basak P. Estimation of disaggregate-level poverty incidence in Odisha under area-level Hierarchical Bayes small area model. Soc Indic Res. 2019;144(1):251–73. https://doi.org/10.1007/s11205-018-2050-9.
12. Arima S, Bell WR, Datta GS, Franco C, Liseo B. Multivariate Fay-Herriot Bayesian estimation of small area means under functional measurement error. J Royal Stat Soc Ser A Stat Soc. 2017;180(4):1191–209. https://doi.org/10.1111/rssa.12321.
13. Asian Development Bank. 2020. Introduction to small area estimation techniques: a practical guide for national statistics offices (Issue May). https://www.adb.org/sites/default/files/publication/609476/small-area-estimation-guide-nsos.pdf.
14. Baffour B, Silva D, Veiga A, Sexton C, Brown JJ. Small area estimation strategy for the 2011 Census in England and Wales. Stat J IAOS. 2018;34(3):1–13. https://doi.org/10.3233/SJI-180427.
15. Baniamin HM, Ramasamy R. Key challenges of studying public administration in developing countries: An exploration from Bangladesh and Sri Lanka. J Public Affairs Educ. 2022;28(4):1–8. https://doi.org/10.1080/15236803.2022.2116262.
16. Brien KLO, Lemango E, Nandy R, Lindstrand A. 2022. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information. Vaccine, January.
17. Brown, G., Chambers, R., Heady, P., & Heasman, D. (2001). Evaluation of small area estimation methods - an application to unemployement estimats from the UK LFS. In: Proceedings of Statistics Canada Symposium, 1–10. https://doi.org/10.1201/9780203166314.ch1
18. Chandra H. Exploring spatial dependence in area-level random effect model for disaggregate-level crop yield estimation. J Appl Stat. 2013;40(4):823–42. https://doi.org/10.1080/02664763.2012.756858.
19. Chandra H, Salvati N, Chambers R. Small area prediction of counts under a non-stationary spatial model. Spatial Stat. 2017;20:30–56. https://doi.org/10.1016/j.spasta.2017.01.004.
20. Chandra H, Salvati N, Chambers R. Small area estimation under a spatially non-linear model. Comput Stat Data Anal. 2018;126:19–38. https://doi.org/10.1016/j.csda.2018.04.002.
21. Chandra H, Salvati N, Sud UC. Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India: an application of small-area estimation technique. J Appl Stat. 2011;38(11):2413–32. https://doi.org/10.1080/02664763.2011.559202.
22. Chanie MG, Ewunetie GE, Molla A, Muche A. Determinants of vaccination dropout among children 12–23 months age in north Gondar zone, northwest Ethiopia, 2019. PLoS ONE. 2021;16(2):e0246018. https://doi.org/10.1371/journal.pone.0246018.
23. Chen S, Lahiri P. Inferences on small area proportions. Pediatr Neuro. 2013;49(4):243–54.
24. Corral P, Molina I, Cojocaru A, Segovia S. Guidelines to Small Area Estimation for Poverty Mapping. 2022. https://doi.org/10.1596/37728.
25. CSA. Population and Housing Census of Ethiopia. 2007.
26. CSA, ICF. 2019. Ethiopia mini demographic and health survey: key indicators. http://www.ephi.gov.et
27. Das S, van den Brakel J, Boonstra HJ, Haslett S. Multilevel time series modelling of antenatal care coverage in Bangladesh at disaggregated administrative levels. Surv Methodol. 2022;48(2):401–37.

28. Dhama K, Sharun K, Tiwari R, Dadar M, Malik YS, Singh KP, Chaicumpa W. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. Human Vaccin Immunother. 2020;16(6):1232–8. https://doi.org/10.1080/21645515.2020.1735227.
29. Doctor HV, Nkhana-Salimu S, Abdulsalam-Anibilowo M. Health facility delivery in sub-Saharan Africa: successes, challenges, and implications for the 2030 development agenda. BMC Public Health. 2018;18(1):1–12. https://doi.org/10.1186/s12889-018-5695-z.
30. Esteban MD, Lombardía MJ, López-Vizcaíno E, Morales D, Pérez A. Small area estimation of proportions under area-level compositional mixed models. TEST. 2020;29(3):793–818. https://doi.org/10.1007/s11749-019-00688-w.
31. Esteban MD, Morales D, Pérez A, Santamaría L. Small area estimation of poverty proportions under area-level time models. Comput Stat Data Anal. 2012;56(10):2840–55. https://doi.org/10.1016/j.csda.2011.10.015.
32. Etana B, Deressa W. Factors associated with complete immunization coverage in children aged 12–23 months in Ambo Woreda, Central Ethiopia. BMC Public Health. 2012;12(1):566. https://doi.org/10.1186/1471-2458-12-566.
33. Fabrizi E, Salvati N, Trivisano C. Robust Bayesian small area estimation based on quantile regression. Comput Stat Data Anal. 2020;145:106900. https://doi.org/10.1016/j.csda.2019.106900.
34. Fay RE, Herriot RA. Estimates of income for small places: an application of james-stein procedures to census data. J Am Stat Assoc. 1979;74(366):269–77. https://doi.org/10.1080/01621459.1979.10482505.
35. FDRE. Federal Democratic Republic of Ethiopia Ministry of Health: Health Sector Development Programme IV 2010/11 – 2014/15. In Federal Democratic Republic of Ethiopia Ministry of Health (Issue October 2010). 2014.
36. Finsterbusch K. 1990. Methods for social analysis in developing countries, 1st edition. https://doi.org/10.4324/9780429043239
37. Garcia M, Lipskiy N, Tyson J, Watkins R, Stein Esser E, Kinley T. Centers for disease control and prevention 2019 novel coronavirus disease (COVID-19) information management: Addressing national health-care and public health needs for standardized data definitions and codified vocabulary for data exchange. J Am Med Inform Assoc. 2020;27(9):1476–87. https://doi.org/10.1093/jamia/ocaa141.
38. Gebrekidan K, Worku A. Factors associated with late ANC initiation among pregnant women in select public health centers of Addis Ababa, Ethiopia: unmatched case–control study design. Pragmat Obs Res. 2017;8:223–30. https://doi.org/10.2147/por.s140733.
39. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. Boca Raton: CRC Press Company; 2006.
40. Guha S, Chandra H. Measuring disaggregate level food insecurity via multivariate small area modelling: evidence from rural districts of Uttar Pradesh. India Food Security. 2021;13(3):597–615. https://doi.org/10.1007/s12571-021-01143-1.
41. Guha S, Chandra H. Multivariate small area modelling for measuring micro level earning inequality: evidence from periodic labour force survey of India. Soc Indic Res. 2022. https://doi.org/10.1007/s11205-021-02857-7.
42. Guha S, Chandra H. Small area estimation under a spatially correlated multivariate area-level model. J Royal Stat Soc Ser A Stat Soc. 2023;187(1):62–84.
43. Haenssgen M. Impact of high-intensity polio eradication activities on children's routine immunization status in Northern India. Health Policy Plan. 2017. https://doi.org/10.1093/HEAPOL/CZX022.
44. Hossain MJ, Das S, Chandra H, Islam MA. Disaggregate level estimates and spatial mapping of food insecurity in Bangladesh by linking survey and census data. PLoS ONE. 2020;15(4):1–16. https://doi.org/10.1371/journal.pone.0230906.
45. Islam S, Chandra H. Small area estimation combining data from two surveys. Commun Stat Simul Comput. 2019;36(4):1–22. https://doi.org/10.1080/03610918.2019.1588308.
46. Liu B, Lahiri P. Hierarchical Bayes estimation and empirical best prediction of small- area proportions. 2009.
47. Liu B, Lahiri P. Adaptive Hierarchical Bayes estimation of small area proportions. Calcutta Statist Assoc Bull. 2017. https://doi.org/10.1177/0008068317722293.
48. Liu B, Lahiri P, Kalton G. Hierarchical Bayes modeling of survey-weighted small area proportions. Surv Methodol. 2014;40(1):1–13.
49. Maples JJ. Improving small area estimates of disability : combining the American Community Survey with the Survey of Income and Program Participation. J Royal Stat Soc Series A Stat Soc. 2017;180(4):1211–27. https://doi.org/10.1111/rssa.12310.
50. Masaki T, Newhouse D, Silwal AR, Bedada A, Engstrom R. Small area estimation of non-monetary poverty with geospatial data. 2020. https://doi.org/10.1596/1813-9450-9383
51. Mebrate M, Workicho A, Alemu S, Gelan E. Vaccination status and its determinants among children aged 12 to 23 months in Mettu and Sinana Districts, Oromia Region, Ethiopia: a comparative cross sectional study. Pediatric Health Med Ther. 2022;13:335–48. https://doi.org/10.2147/phmt.s380303.
52. Mekonnen ZA, Gelaye KA, Were MC, Tilahun B. Timely completion of vaccination and its determinants among children in northwest, Ethiopia: a multilevel analysis. BMC Public Health. 2020;20(1):908. https://doi.org/10.1186/s12889-020-08935-8.
53. Morales D, Esteban MD, Pérez A, Hobza T. A Course on Small Area Estimation and Mixed Models Methods, Theory and Applications in R. In: A course on small area estimation and mixed models. 2021.
54. Mukhopadhyay PK, McDowell A. 2011. Small area estimation for survey data analysis using SAS software. In: SAS Global Forum (Vol. 2011, p. 96).
55. Mukhopadhyay PK, McDowell A. Small area estimation for survey data analysis using SAS software. SAS Global Forum. 2011;2011:1–19.
56. Mweemba C, Hangoma P, Fwemba I, Mutale W, Masiye F. Estimating district HIV prevalence in Zambia using small-area estimation methods (SAE). Popul Health Metrics. 2022;20(1):8. https://doi.org/10.1186/s12963-022-00286-3.
57. Newhouse DL, Merfeld JD, Ramakrishnan A, Swartz T, Lahiri P. Small area estimation of monetary poverty in Mexico using satellite imagery and machine learning. SSRN Electr J. 2022. https://doi.org/10.2139/ssrn.4235976.
58. Nguyen T, Jiang J. Estimating the mean squared prediction error of the observed best predictor associated with small area counts: a computationally oriented approach. Can J Stat. 2024;1:1–9. https://doi.org/10.1002/cjs.11810.
59. Pratesi M. Analysis of poverty data by small area estimation. 2016. https://doi.org/10.1002/9781118814963.
60. Rahman A. A review of small area estimation problems and methodological developments. In: NATSEM-University of Canberra Discussion paper: Vol. October (Issue October). 2008. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Review+of+Small+Area+Estimation+Problems+and+Methodological+Developments#0
61. Rao JNK, Molina I. Small area estimation. 2nd ed. New York: Wiley; 2015.

62. Ren R. 2021. Design-based small area estimation: an application to the DHS surveys. DHS Working Papers No. 180. Rockville, Maryland, USA: ICF.
63. Sabin AB. Pathogenesis of poliomyelitis reappraisal in the light of new data. Science. 1956;123(3209):1151–7. https://doi.org/10.1126/science.123.3209.1151.
64. Saei A, Chambers R. Small area estimation under linear and generalized linear mixed models with time and area effects. S3RI Methodology Working Paper M03/15. S3RI Methodology Working Paper M03/15, January. 2003.
65. Senan P, Loe Y, Grupreet K, Hayati A, Haliza A, Novia K, Odhayakumar N, Fadzilah K, Chong C, Nirmal S, Balachandran S, Devan K. Poliomyelitis vaccination status among children in the Federal Territory of Kuala Lumpur 2007. Western Pac Surveill Response J. 2011;2(3):19–24. https://doi.org/10.5365/wpsar.2011.1.1.014.
66. Shi C. Package ' BayesSAE'. 2018.
67. Shiferaw YA. Analysis of the spatial distribution of under-5 mortality rate in local areas of South Africa. Stat J IAOS. 2020;36(4):1161–73. https://doi.org/10.3233/SJI-200650.
68. Shiferaw YA. Model-based estimation of small area food insecurity measures in Ethiopia using the Fay-Herriot EBLUP estimator. Stat J IAOS. 2020;36(S1):S177–87. https://doi.org/10.3233/SJI-200673.
69. Shiferaw YA. Mapping disaggregate-level agricultural households in South Africa using a Hierarchical Bayes small area estimation approach. Agriculture. 2023;13:1–17. https://doi.org/10.3390/agriculture13030631.
70. Shiferaw YA, Yilema SA, Moyehodie YA. A Hierarchical Bayesian approach to small area estimation of health insurance coverage in Ethiopian administrative zones for better policies and programs. Heal Econ Rev. 2024;14(1):1–14. https://doi.org/10.1186/s13561-024-00498-3.
71. Shiferaw Y, Galpin J. A corrected confidence interval for a small area parameter through the weighted estimator under the basic area level model. J Iran Statist Soc. 2019;18(1):17–51. https://doi.org/10.29252/JIRSS.18.1.17.
72. Srivastava S, Chandra H, Kant S, Kumar A. SSM - Population Health Mapping changes in district level prevalence of childhood stunting in India 1998–2016: an application of small area estimation techniques. SSM - Population Health. 2021;14(January): 100748. https://doi.org/10.1016/j.ssmph.2021.100748.
73. Sugasawa S, Kubokawa T. Small area estimation with mixed models: a review. Jpn J Stat Data Sci. 2020;3(2):693–720. https://doi.org/10.1007/s42081-020-00076-x.
74. Sugasawa S, Kubokawa T. Mixed-effects models and small area estimation. 2023.
75. Sugasawa S, Kubokawa T, Rao J. Small area estimation via unmatched sampling and linking models. TEST. 2017. https://doi.org/10.1007/s11749-017-0551-5.
76. Sunandi E. Logit-normal spatial model for small area estimation : case study of poverty in Bengkulu. IOP Conf Ser. 2018. https://doi.org/10.1088/1755-1315/187/1/012051.
77. Tella O. Agenda 2063 and its implications for Africa's soft power. J Black Stud. 2018;49(7):714–30. https://doi.org/10.1177/0021934718780489.
78. Tessema F, Bisrat F, Kidane L, Assres M, Tadesse T, Asegedew B. Improvements in polio vaccination status and knowledge about polio vaccination in the CORE group polio project implementation areas in pastoralist and semi-pastoralist regions in Ethiopia. Am J Trop Med Hyg. 2019;101(4):52–8. https://doi.org/10.4269/ajtmh.19-0022.
79. UNFPA. Better Data for More Effective Policies and Programmes: Small area estimation. www.398unfpa.org/publications/small-area-estimation. 2022.
80. White GW, McConville KS, Moisen GG, Frescino TS. Hierarchical Bayesian small area estimation using weakly informative priors in ecologically homogeneous areas of the interior western forests. Front Forest Global Change. 2021;4:1–15. https://doi.org/10.3389/ffgc.2021.752911.
81. Wiysonge CS. How ending polio in Africa has had positive spinoffs for public health. Conversation; 2020.
82. Wolter KM. Introduction to variance estimation. 2nd ed. Cham: Springer; 2007.
83. Wolter L. Introduction. In: Nature-based tourism in Mallorca's natural areas. https://doi.org/10.1007/978-3-658-04536-4_1
84. Yilema SA. Bayesian small area estimation of anemia prevalence using a generalized linear mixed model. Stat J IAOS. 2025;41(1):171–9. https://doi.org/10.1177/18747655241308394.
85. Yilema SA, Shiferaw YA, Zewotir T, Muluneh EK. Improving survey based estimates of malnutrition using small area estimation. Stat J IAOS. 2022;38:1–11. https://doi.org/10.3233/SJI-210892.
86. Yilema SA, Shiferaw YA, Zewotir T, Muluneh EK. Multivariate small area estimation of undernutrition for children under five using official statistics. Stat J IAOS. 2022;38(2):1–12. https://doi.org/10.3233/SJI-220935.
87. Yilema SA, Shiferaw YA, Zewotir T, Muluneh EK. Small area estimation of undernutrition under age five children based on spatial-temporal models. Sci World J. 2022. https://doi.org/10.1155/2022/6882047.
88. Yilema SA, Shiferaw YA, Zewotir T, Muluneh EK. Spatial small area estimates of undernutrition for under five children in Ethiopia via combining survey and census data. Spat Spatio-Temporal Epidemiol. 2022;42:100509. https://doi.org/10.1016/j.sste.2022.100509.
89. You Y, Chapman B. Small area estimation using area level models and estimated sampling variances. Surv Methodol. 2006;32(1):97–103.
90. You Y, Rao JNK. Small area estimation using unmatched sampling and linking models1. Zb Canadian J Srntinics. 2002;30(1):3–15.
91. You Y, Rao JNK, Hidiroglou M. On the performance of self benchmarked small area estimators under the Fay-Herriot area level model. Survey Methodol. 2012;39(1):217–29.
92. You Y, Zhou QM. Hierarchical Bayes small area estimation under a spatial model with application to health survey data. Surv Methodol. 2011;37(1):25–37.