BAREKENG: Journal of Mathematics and Its Applications
December 2026 Volume 20 Issue 4 Page 3341-3354
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss4pp3341-3354

THREEFOLD HIERARCHICAL SMALL AREA ESTIMATION MODEL FOR ESTIMATING PREVALENCE OF STUNTING IN WEST NUSA TENGGARA

Logananta Puja Kusuma[1*], **Kusman Sadik**[2], **Anang Kurnia**[3]

[1,2,3]*Program in Statistics and Data Science, School of Data Science, Mathematics, and Informatics, IPB University Jln. Raya Dramaga, Bogor, 16680, Indonesia*

*Corresponding author's e-mail: * **loganpuja@apps.ipb.ac.id***

Article Info

Article History:
Received: 17th October 2025
Revised: 28th April 2026
Accepted: 07th June 2026
Available online: 24th August 2026

Keywords:
Hierarchical Bayes;
Poisson-Gamma;
Small area estimation;
Stunting;
Threefold model.

ABSTRACT

Stunting is a condition of growth failure in toddlers due to chronic malnutrition, becoming an issue in various regions of Indonesia. West Nusa Tenggara Province has one of Indonesia's highest stunting prevalence rates, calculated at 29.8% in 2024, which is more than twice the national target of 14%. Appropriate and efficient policy-making for stunting reduction requires reliable estimates for small areas like districts. This study aims to produce more reliable district-level estimates of stunting prevalence for districts level area, while direct estimation from the Survei Kesehatan Indonesia (SKI) is unreliable due to limited sample size. This research develops a threefold hierarchical Bayesian (HB) Small Area Estimation (SAE) model based on the Poisson-Gamma distribution to model the number of stunted toddlers as discrete count data with overdispersion. The proposed model incorporates auxiliary variables from official sources and includes three hierarchical random effects representing district, regency/municipality, and grouped regency/municipality levels based on geographical structure. The results show that the threefold HB SAE model achieves convergent parameter estimates and provides more stable and precise district-level stunting prevalence estimates compared to direct estimators. The multilevel model performs better than one random effect model as reflected by the lower LOOIC. The findings also suggest that districts in Sumbawa Island, particularly in the eastern part and areas located farther from regency/municipality capitals, tend to have higher stunting prevalence. However, this study is limited by the assumption of independence among area-level random effects and restricted availability of auxiliary variables. This study contributes methodologically by extending Poisson-based SAE literature through the application of a threefold HB framework in stunting estimation and provides district-level stunting statistics that support evidence-based policymaking and targeted interventions aligned with SDG Target 2.2.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License (https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

L. P. Kusuma, K. Sadik, and A. Kurnia, "THREEFOLD HIERARCHICAL SMALL AREA ESTIMATION MODEL FOR ESTIMATING PREVALENCE OF STUNTING IN WEST NUSA TENGGARA", *BAREKENG: J. Math. & App.*, vol. 20, no. 4, pp. 3341-3354, Dec, 2026.

Copyright © 2026 Author(s)
Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Stunting is a condition of growth failure in toddler due to chronic malnutrition, which remains a prevalent issue in various regions of Indonesia. Differing with common malnutrition, stunting has chronic, long-term effects for the patients. Mortality and severity on patient toddlers and children age is the short-term effect of stunting, and they also suffered long-term effect, such as vulnerability to degenerative diseases as adults [1]. Stunting patients also tend to have a lack of cognitive function. This condition leads to multiple implications. Patients' low cognitive function results in potential economic losses when they are in productive age (15-64), concurrently impacting a decrease in GDRP and GDP [2].

West Nusa Tenggara/Nusa Tenggara Barat (NTB) is one of the provinces with highest prevalence of stunting in Indonesia, calculated at 29.8% in 2024 [3]. This is more than twice the national target of 14%. A prevalence above threshold 20% is considered very high [4]. The high number of stunting patients in the province could lead to multiple implications for the region. Therefore, it is crucial for the government to design appropriate policies for addressing stunting. Stunting reduction is also directly related to Target 2.2 of the Sustainable Development Goals (SDGs), which aims to end all forms of malnutrition [5]. Policies related to stunting reduction require integrated intervention of nutrition specific (e.g., additional food or supplement for toddler, pregnant and breastfeeding mothers) and nutrition related (e.g., sanitation and access to clear water) programs [6]. These policies would be more efficient if implemented in a specific small area. Even within same regency, each district may possess distinct characteristics. Those policies should be addressed precisely to subarea which needs it the most.

However, stunting prevalence data are obtained from survey results, which often do not provide reliable estimates for small areas such as districts. Even for districts with large geographic coverage, a limited sample size is the main reason they are still considered small areas. *Survei Kesehatan Indonesia* (Indonesia's Health Survey) uses stratified sampling method with regencies/municipalities as strata. Statistics for the regencies/municipalities level area is available on the survey report, while statistics for districts level area is not provided. Direct estimation using statistics calculated from sample data may not give precise results for sub-areas because of limited number of samples [7]. That limitation could lead to high *relative standard error* (RSE) of the estimator.

To address the limitation of direct estimation in small area, the Small Area Estimation (SAE) method is often used as an indirect estimation approach that utilizes additional information. The concept of SAE is to include the information of larger area and auxiliary variables. The auxiliary variables are additional variables collected from other surveys or administrations. Basic area level SAE model is Fay-Herriot model which consists of fixed effect and random effect [8], [9]. Variance in the small area is captured by additional information of auxiliary variables in the fixed effect and small area specific random effect [10].

While the Fay-Herriot model and its subsequent extensions have provided important methodological advances, challenges remain when handling response variables distribution and hierarchical population structures. Model based on Gaussian distribution are suitable for continuous response variable. In contrast, in modelling discrete response variable, a model based on a discrete distribution is required, particularly a Poisson-based model for count data. [11], [12], [13]. In this study, model based on Poisson distribution is used to handle the discrete number of stunted toddlers. Specifically, the model developed is Poisson-Gamma model and the parameter calculated using hierarchical Bayesian method. Poisson-Gamma is commonly used in Poisson based SAE models [14]. Inverse-Gamma as conjugate prior distribution is applied to handle overdispersion of the response variable.

One of the challenges of SAE modelling is that the response variable may vary between population geographical and administration structures. NTB consists of 117 districts, distributed on 8 regencies and 2 municipalities. Each regency/municipality has their own local government whose different policies in reducing prevalence of stunting. For example, the East Lombok Regency Government implemented the *Gerakan 1000 Hari Pertama Kehidupan* (Life's First 1000 Days Movement), a collaborative program involving the Health Department, Agriculture Department, and village offices, specifically aimed at supporting newborn infants [15]. In addition, the North Lombok Regency Government implemented the *Program Aksi Konvergensi* (Convergence Action Program), which engaged several other departments to assist the Health Department in monitoring the implementation of stunting-related policies. [16]. Therefore, the SAE model should include a random effect of regency/municipality. NTB consists of several parts; Lombok Island, Western part of Sumbawa Island, and Eastern part of Sumbawa Island. Those parts are not only separated geographically, but also inhabited by different ethnic groups. Regencies located on the same island tend to share similar characteristics across multiple aspects, including health.

Therefore, variance of stunting prevalence may be related to regency/municipality and grouped regencies/municipalities based on island. Three-fold SAE can be used to overcome this challenge, with each level is represented by a random effect in the model [17], [18]. In this case, the model consists of three random effects, those are districts level, regencies/municipalities level, and grouped regencies/municipalities level.

Building multilevel model encounters a problem, namely complexity of the models. Empirical Best Linear Unbiased Prediction (EBLUP), the common SAE method, is not suitable for complex model, because the result of parameter estimates is not a closed form [19]. To handle the complexity of the model, Bayesian approach should be implemented. There are two Bayesian method in SAE modelling, namely Empirical Bayes (EB) and Hierarchical Bayes (HB). HB method utilizes prior distribution to estimate parameter of the posterior distribution, making it more flexible in handling complex structures such as overdispersion. Several studies have applied the HB method to construct multilevel models [20], [21].

Despite these methodological developments, a research gap remains in the application of Poisson-based multilevel SAE for stunting prevalence at the district level in Indonesia, particularly in NTB. Existing stunting-related SAE studies focus on single-level random effects or rely on Fay-Herriot models [22], [23], which may be less appropriate for count outcomes and may not sufficiently represent hierarchical administrative variation.

This study aims to address this gap by developing a hierarchical Bayesian Poisson-Gamma threefold SAE model for estimating district-level stunting prevalence in NTB. The main objectives of this research are development of threefold hierarchical Bayesian SAE model based on the Poisson-Gamma distribution for estimating district-level stunting prevalence in NTB, utilization of auxiliary variables from official sources to improve the precision of district-level estimates, and evaluation of the model performance in providing more reliable district-level stunting estimates compared to direct survey-based estimation.

The benefits of this research are threefold. First, this study provides more reliable and precise district-level stunting prevalence estimates, which are not available from direct survey publication. Second, this study contributes methodologically by extending Poisson-based SAE literature through the application of a threefold hierarchical Bayesian framework in the context of stunting prevalence estimation. Third, the results support local governments in identifying priority districts and designing more targeted and efficient stunting interventions, as well as identifying significant auxiliary variables. The improved small area level stunting estimates produced in this study are expected to support evidence-based policymaking and strengthen local interventions aligned with Sustainable Development Goal 2.2 on ending all forms of malnutrition.

2. RESEARCH METHODS

The data used as response variable in this research is secondary data collected by Indonesia's Ministry of Health in *Survei Kesehatan Indonesia* 2023 (https://www.badankebijakan.kemkes.go.id/hasil-ski-2023) [3]. The number of stunting is calculated from survey raw data with 6,809 total respondents in West Nusa Tenggara Province. These sample are from 113 districts, 10 regencies/municipalities. List of regencies/municipalities and their code:

1. Lombok Island: West Lombok Regency (5201), Middle Lombok Regency (5202), East Lombok Regency (5203), North Lombok Regency (5208), and Mataram Municipality (5271).
2. Western Sumbawa Island: Sumbawa Regency (5204) and West Sumbawa Regency (5207).
3. Eastern Sumbawa Island: Dompu Regency (5205), Bima Regency (5206), and Bima Municipality (5272).

As auxiliary variables, some variables from Village Potential Statistics (*potensi desa*) are used. Village Potential Statistics is administrative data gathered by Bureau of Statistics at the village or ward level, commonly used as auxiliary variable in SAE researches [24], [25]. The distance of each district to the regency/municipality capital is also used as auxiliary variable to capture difference between the district near the regency/municipality capital and those which located in remote area. The distance is calculated using Euclidean distance between district office and the regent's/mayor's office. Coordinates of each office is collected from Google Maps and the distance is converted to kilometers using Rstudio. The distance is later transformed using a logarithmic transformation to ensure that its scale is comparable to the other variables, which are mostly expressed as percentages. List of auxiliary variables:

Table 1. List of Auxiliary Variables

Variable	Description	Source
x1	Percentage of village/sub-district which gives incentive for Community Health Workers (*kader posyandu*) in each district	*Village Potential Statistics*
x2	Percentage of village/sub-district which provides health insurance for infant in each district	*Village Potential Statistics*
x3	Percentage of village/sub-district with sanitation and clear water facilities and infrastructure in each district	*Village Potential Statistics*
x4	Percentage of village/sub-district with monthly **i**ntegrated health post (*posyandu*) service in each district	*Village Potential Statistics*
x5	Log of distance between district and capital of regency/municipality	Google maps

2.1 Threefold Hierarchical Bayes Model

Threefold hierarchical model is area level SAE model with three-area random effects. We want to estimate parameters at the area level θ_i, subarea θ_{ij}, and sub-subarea θ_{ijk}. These parameters respectively correspond to groups of regencies/municipalities, regencies/municipalities, and districts, and are used to capture heterogeneity across the geographical and administrative structure. The model is built in multiple level.

1. Sampling model.

$$y_{ijk}|\theta_{ijk} \sim Poisson(\theta_{ijk}), \tag{1}$$

where y_{ijk} is the response variable in *i*-th area, *j*-th subarea, and *k*-th sub-subarea. In this study, the response variable is the stunting prevalence based on direct estimates from the survey result.

2. Link function

$$\eta_{ijk} = \log(\theta_{ijk}) = \boldsymbol{x_{ijk}\beta} + u_{1,i} + u_{2,ij} + u_{3,ijk}, \tag{2}$$

where $\boldsymbol{x_{ijk}\beta}$ is the fixed effect constructed from auxiliary variables matrix $\boldsymbol{x_{ijk}}$ and regression parameters vector $\boldsymbol{\beta}$. Dimension of matrix $\boldsymbol{x_{ijk}}$ is $(n \times p + 1)$, where n is the total number of subsubareas in the population and p is the number of auxiliary variables used. The random effect $u_{1,i}$, represents the random effect for area i where $i = 1, 2, \ldots, l$ and l is the total number of areas in the population. The random effect $u_{2,ij}$ is the represents the random effect for subarea j within area i, where $j = 1, 2, \ldots, m$ and m is the total number of subareas in the population. The random effect $u_{1,ijk}$ represents the random effect for subarea k within area i within subarea j, where $k = 1, 2, \ldots, n$. Each of the random effects u is normally distributed with zero mean and variance distributed inverse-gamma.

3. Prior distribution for parameter

$$\boldsymbol{\beta} \sim N(\boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta}), \boldsymbol{u_{1,i}} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2_{u_1}), \boldsymbol{u_{2,ij}} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2_{u_2}), \boldsymbol{u_{3,ijk}} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2_{u_3}),$$
$$\sigma^2_{u_1} \sim IG(a_{v_1}, b_{v_1}), \sigma^2_{u_2} \sim IG(a_{v_2}, b_{v_2}), \sigma^2_{u_3} \sim IG(a_{v_3}, b_{v_3}), \tag{3}$$

where:

$\boldsymbol{\mu_\beta}$: mean prior distribution of regression parameter.
$\boldsymbol{\Sigma_\beta}$: variance prior distribution of regression parameter.
$\sigma^2_{u_1}$: random effects variance of area $u_{1,i}$.
$\sigma^2_{u_2}$: random effects variance of subarea $u_{2,ij}$.
$\sigma^2_{u_3}$: random effects variance of sub-subarea $u_{3,ijk}$.
a_{v_1}, b_{v_1} : prior distribution parameters for $\sigma^2_{u_1}$.
a_{v_2}, b_{v_2} : prior distribution parameters for $\sigma^2_{u_2}$.
a_{v_3}, b_{v_3} : prior distribution parameters for $\sigma^2_{u_3}$.

4. Posterior distribution

$$P(\boldsymbol{\beta}, \boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3} | y, \boldsymbol{X}) \propto P(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{X}). P(\boldsymbol{\beta}). P(\boldsymbol{u_1}). P(\boldsymbol{u_2}). P(\boldsymbol{u_3}). \tag{4}$$

2.2 Accuracy

Root Mean Square Error (RMSE) is calculated from the error of the prediction of each iteration and the average. Relative RMSE (RRMSE) is the percentage of RMSE scaled to the prediction. The calculation of RRMSE is

$$RRMSE(\hat{\theta}) = \frac{\sqrt{\widehat{MSE(\theta)}}}{\hat{\theta}} \times 100\%, \tag{5}$$

$$\widehat{MSE}(\hat{\theta}) = \widehat{Var(\theta}|y), \tag{6}$$

where $\hat{\theta}$ is the prediction of the model in a particular domain and $MSE(\hat{\theta})$ is the mean squared error.

2.3 Empirical Analysis Procedure

In the empirical analysis of this research, we construct a threefold HB SAE model to estimate the number of stunted toddlers in district level. The procedure of analysis including data pre-process and model estimate.

2.3.1 Data Pre-Processing Stage

The data pre-processing stage is carried out for response and auxiliary variable as follows:

1. Response variable is the number of stunted toddlers in each district, calculated from the stunting prevalence from *Survei Kesehatan Indonesia* 2023 multiplied by projection number of toddlers population from Bureau of Statistics.
2. Auxiliary variables from Village Potential Statistics are aggregated from village/ward level raw data into district level. Then, the percentage is calculated by dividing the aggregated data with the count of villages/wards within the district.
3. Distance between district and capital of regency/municipality auxiliary variable is measured by calculating Euclidean distance between district office and regent's/mayor's office of the regency/municipality using coordinates (latitude, longitude). The distance is converted into kilometer and transformed by logarithmic scale. Formula of Euclidean distance is as follow [26]:

$$d(a, b) = \sqrt{(\boldsymbol{a} - \boldsymbol{b})'(\boldsymbol{a} - \boldsymbol{b})} = \sqrt{\sum_{i=1}^{n} (\boldsymbol{a}_i - \boldsymbol{b}_i)^2}, \tag{7}$$

where $\boldsymbol{a}$ is the coordinate vector of district office and $\boldsymbol{b}$ is the coordinate vector of regent's/mayor's office.

2.3.2 Model Estimation and Evaluation

The stages of model parameter estimation and evaluation are as follows:

1. Estimating variance component.
2. Estimating model parameters.
3. Estimating the number of stunted toddlers for each district.
4. Aggregating the number of stunted toddlers for regency/municipality using the fitted value.
5. Calculating RRMSE.

The software used in this research is R programming language with hbsaems package for estimation (stage 1-3). The package provides hbm function that is flexible in constructing SAE model, including multilevel random effects and adjustment for Poisson distribution model. Estimation is performed using Markov Chain Monte Carlo (MCMC) with 4 chains and 10,000 iterations for each single chain while the number of warm-up iterations remained default; half of the iteration number. The number of chains and iteration is obtained from several trial-and-error processes. The stages of each iteration are as follows:

1. Setting initial parameters.
2. Iteration process:

i. Updating prior parameters.
ii. Updating random effects.
iii. Updating fixed effects.
iv. Storing the results for later computation.

3. Discarding the results of warm-up iterations.
4. Computing posterior summaries, including posterior mean, posterior variance, estimated error, and confidence intervals.

3. RESULTS AND DISCUSSION

3.1 Exploratory Data Analysis

Fig. 1 illustrates the distribution of the number of stunting cases by district in NTB based on the direct estimator calculated from *Survei Kesehatan Indonesia* 2023 survey data. The histogram shows that the number of stunting cases is right-skewed. Outliers are also observed, as in several districts the number of sampled children recorded as stunted is relatively higher than in other districts. Such data characteristics may lead to overdispersion in a Poisson regression model.

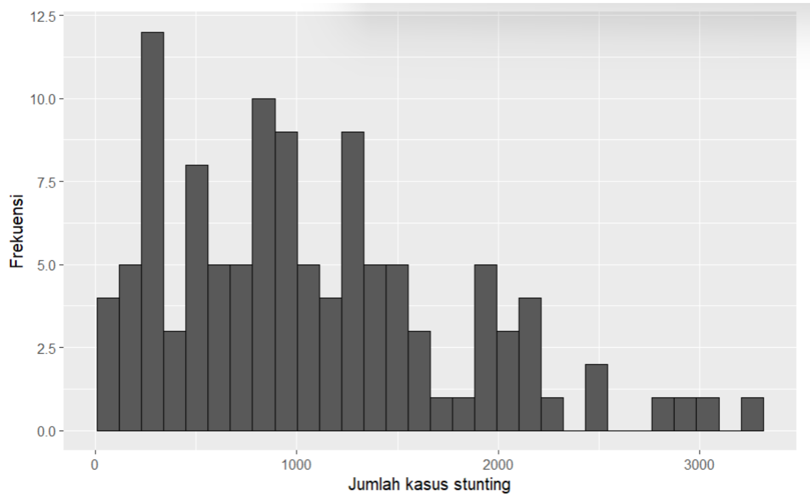


Figure 1. Histogram of Number of Stunting Cases by District

Overdispersion is also indicated by the large difference between the mean and the variance of the response variable. The standard deviation is recorded at 716.08, resulting in a response variance of approximately 512,769.41. This value is much larger than the mean of 1,069.51. Therefore, the response variable is not well suited to be modeled using a standard Poisson model, which assumes that the mean is equal to the variance.

3.2 Parameter Convergence

The statistical significance and convergence of the model's parameters, crucial for the reliability of the Threefold Hierarchical Bayes Small Area Estimation (HB SAE) model, were thoroughly examined. The estimation results for random effect parameters are presented in Table 2.

Table 2. Estimation Result for Random Effect Parameters

Parameter	Mean	Est. Error	Lower-95% CI	Upper-95% CI
σ_{u_1}	0.04	0.04	0.00	0.15
σ_{u_2}	0.08	0.06	0.01	0.24
σ_{u_3}	0.12	0.06	0.04	0.27

Table 2 presents the parameter estimates for the random effect. Posterior mean of each parameter is obtained by averaging the intercept of random effect estimates across the MCMC iterations. The estimated standard errors are calculated from the posterior variances. Using the posterior mean and the estimated standard error, the confidence intervals can be constructed. The Lower-95% CI and Upper-95% CI represent the lower and upper bounds of the 95% confidence interval, respectively.

Based on Table 2, confidence intervals of random effects do not contain 0 in their range indicating that the corresponding random effects are statistically significant. These results suggest the presence of a multilevel structure in the response variable. This finding is consistent with the model comparison results between the three-level HB-SAE model and models with fewer random effects. Table 3 presents a comparison of three models fitted using the same number of iterations and chains. Based on Table 3, the threefold model yields the lowest Leave-One-Out Information Criterion (LOOIC), indicating that it performs best in predicting new, unseen data.

Table 3. Comparison of HB SAE Model with One, Two, and Three Random Effects

HB SAE Model	Random Effect	LOOIC
One Random Effect HB SAE	District	433724
Twofold HB SAE	District, regency/municipality	421750
Threefold HB SAE	District, regency/municipality, grouped regency/municipality	370060

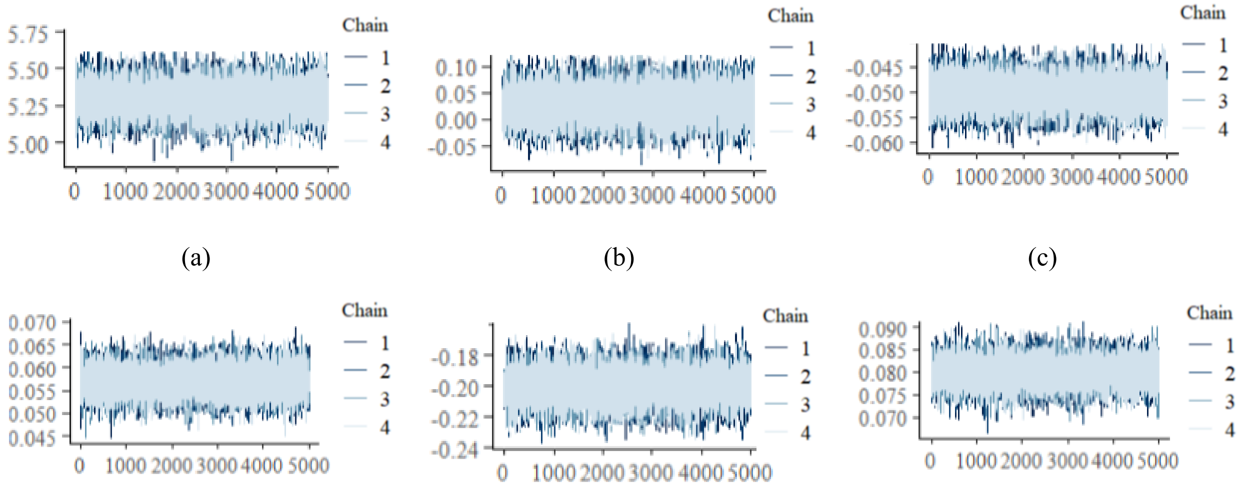
The statistical significance of the auxiliary variable parameters was also examined. The estimation results for the coefficients of the auxiliary variables are presented in Table 4. Unlike the random effect estimates, the posterior means of the auxiliary variable parameters are calculated by averaging the corresponding parameter estimates across the MCMC iterations.

Table 4. Estimation Result for Auxiliary Variable Parameters

Parameter	Mean	Est. Error	Lower-95% CI	Upper-95% CI
β_1	0.03	0.03	-0.03	0.09
β_2	-0.05	0.00	-0.06	-0.04
β_3	0.06	0.00	0.05	0.06
β_4	-0.20	0.01	-0.22	-0.18
β_5	0.08	0.00	0.07	0.09

Based on Table 4, confidence interval of β_2, β_3, β_4, and β_5 do not contain 0 in their range indicating that the corresponding auxiliary variables are statistically significant predictors of the number of stunted toddlers. Conversely, confidence interval of β_1 ranges from -0.03 to 0.09, contains zero in its range. This indicates that percentage of village/sub-district which gives incentive for Community Health Workers is not significant for the number of stunting in a district. Presence of incentives does not strongly correlate with stunting reduction after accounting for other variables.

Tracing with line plot can be used to check the convergence of estimation. Fig. 2 demonstrates that the coefficient estimates for auxiliary variables are stable around a certain value and tend to have a random pattern over iterations. This indicates that the estimation of auxiliary variables posterior distribution is convergent for a long run of the MCMC process.



(a) (b) (c)

(d) (e) (f)

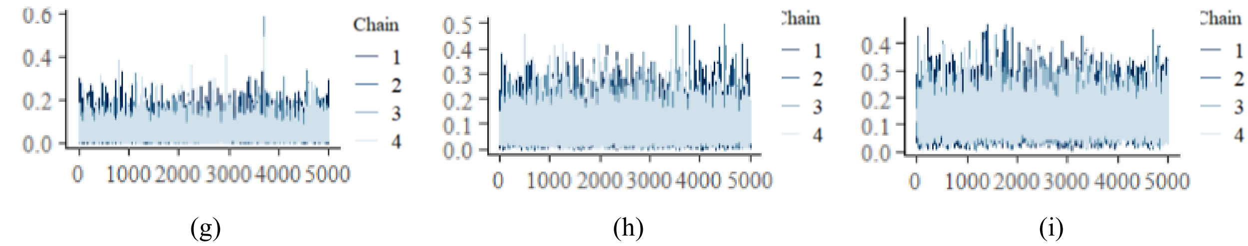


Figure 2. Trace Plot of Parameter Estimates
(a) β_0, (b) β_1, (c) β_2, (d) β_3, (e) β_4, (f) β_5, (g) σ_{u_1}, (h) σ_{u_2}, (i) σ_{u_3}
(Application Source: RStudio)

Fig. 3 shows that density plot of fixed effect parameter estimates, both intercept and auxiliary variables coefficient estimate are symmetrically distributed. The standard deviations of random effect priors are right-tailed, as expected from the Inverse-Gamma distribution used as the prior.

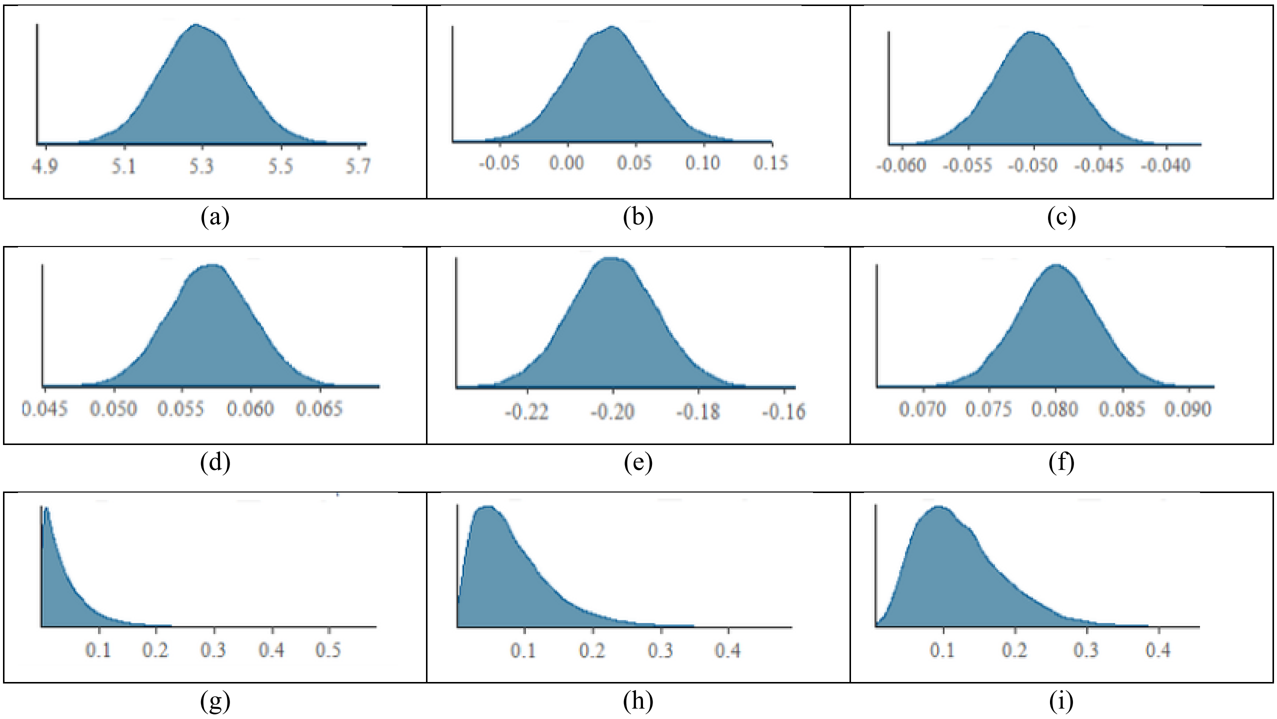


Figure 3. Density of Parameter Estimates
(a) β_0, (b) β_1, (c) β_2, (d) β_3, (e) β_4, (f) β_5, (g) σ_{u_1}, (h) σ_{u_2}, (i) σ_{u_3}
(Application Source: RStudio)

3.3 Estimation for District Level

Direct estimation methods often produce highly unreliable estimates for small areas like districts due to small sample sizes, which results in either excessively high or low prevalence estimates and high sampling variance. The threefold HB SAE model successfully mitigates this issue by borrowing strength from adjacent areas (regency/municipality and island group) and incorporating auxiliary information.

Fig. 4 shows that direct estimators of stunting prevalence may produce percentages that are either too high or too low, making them less plausible in practice. In contrast, the SAE model yields more stable and precise estimates of district-level stunting prevalence. The RRMSE is stable across all districts, regardless of the predicted stunting prevalence. HB SAE model also results on district level stunting prevalence closer to the overall mean of the larger area, resulting in estimates that are neither too high nor too low. The order of the districts is randomized to ensure the confidentiality of the direct estimates.

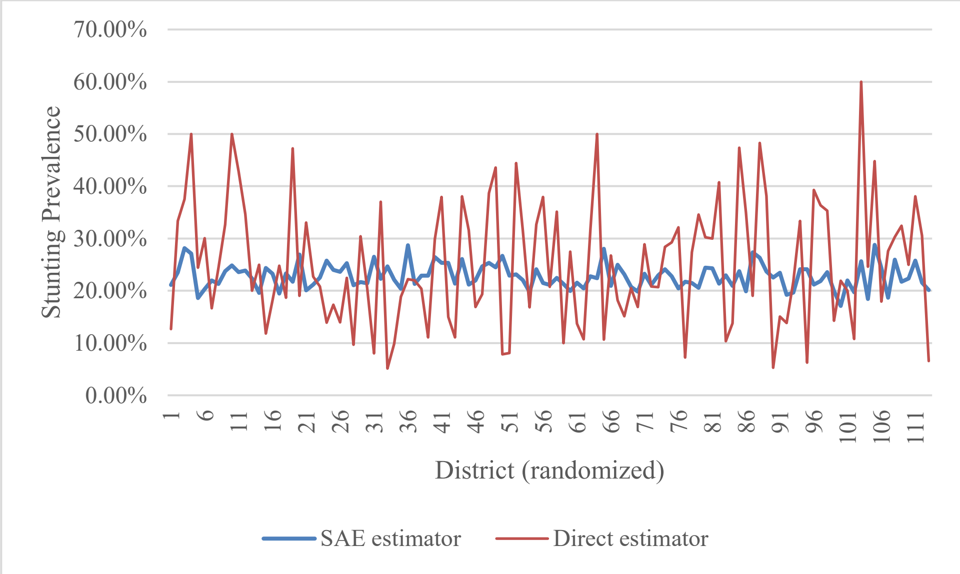


Figure 4. Trace Plot of SAE and Direct Estimators in Each District

The histogram of RRMSE is shown in Fig. 5. Each district level stunting prevalence is estimated with RRMSE lower than 25%. This shows that the use of the t model can provide good precision for estimation [20]. The histogram shows that the RRMSE values of district-level estimators are mostly concentrated between 21.1% and 21.6%, with the highest frequency occurring around 21.3%–21.4%. This indicates that most districts have relatively similar estimation precision, with RRMSE values near the center of the distribution. The shape of the histogram is unimodal and slightly right-skewed, suggesting that while the majority of districts have moderate and consistent accuracy.

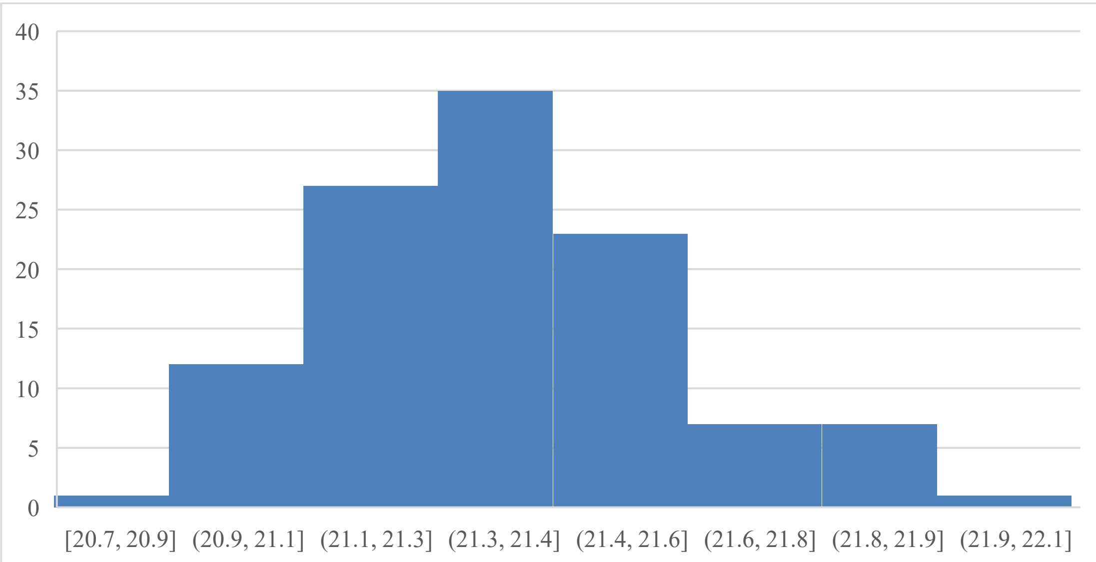


Figure 5. Histogram of RRMSE for District Level Estimators

The detailed RRMSE of district with highest stunting prevalence and lowest stunting prevalence is respectively shown on Tables 5 and 6. Stunting prevalences in both tables are the indirect estimates based on SAE model.

Table 5. RRMSE for Districts with Highest Stunting Prevalence

Regency/Municipality	District	Stunting Prevalence	RRMSE
Sumbawa Regency	Tarano	28.79%	21.19%
Sumbawa Regency	Labangka	28.73%	21.57%
Sumbawa Regency	Alas Barat	28.15%	21.22%
Dompu Regency	Pekat	28.06%	21.46%
Bima Regency	Sanggar	27.35%	21.07%
Bima Regency	Ambalawi	27.11%	21.40%

Regency/Municipality	District	Stunting Prevalence	RRMSE
Sumbawa Regency	Empang	26.92%	21.30%
West Sumbawa Regency	Maluk	26.69%	21.31%
Dompu Regency	Kempo	26.48%	21.42%
Bima Regency	Lambu	26.47%	21.09%

All ten districts with the highest stunting prevalence are located on Sumbawa Island, either the Western part or Eastern part. Most of those districts are far from the capital city of their regency. For example, the three highest ranking districts, Tarano, Labangka, and Alas Barat, are respectively 68 km, 50 km, and 46 km away from the capital of Sumbawa Regency. This pattern underscores the significance of the distance auxiliary variable (x5). However, distance is not the only factors of the high percentage. Maluk for instance, a suburban area near Taliwang (the capital of West Sumbawa), is predicted to have 26.69% stunting prevalence. This high rate is attributed to the lack of sanitation and clean water facilities in Maluk, which corresponds to the significance of auxiliary variable x3 (percentage of village/sub-district with sanitation and clean water facilities).

Table 6. RRMSE for Districts with Lowest Stunting Prevalence

Regency/Municipality	District	Stunting Prevalence	RRMSE
Sumbawa Regency	Sumbawa	17.10%	21.35%
North Lombok Regency	Tanjung	18.45%	21.42%
Mataram Municipality	Ampenan	18.59%	21.38%
Sumbawa Regency	Unter Iwes	18.69%	21.28%
Mataram Municipality	Selaparang	19.21%	21.42%
Mataram Municipality	Cakranegara	19.45%	21.79%
West Sumbawa Regency	Brang Ene	19.61%	21.78%
East Lombok Regency	Selong	19.65%	21.48%
West Sumbawa Regency	Taliwang	19.73%	21.45%
Mataram Municipality	Mataram	19.81%	21.18%

Among 10 districts with lowest stunting prevalence, 6 of them are from Lombok Island. The other 4 are from the Western part of Sumbawa Island. None of those districts are from Eastern part of Sumbawa Island. Districts where the regent's office located tend to have lower stunting prevalence. Sumbawa, Tanjung, Selong, and Taliwang are the Capitals of their respective regency. In line with the previous discussion, districts closer to the capital of their regency tend to have lower stunting prevalence, likely due to better resource allocation and access to health services.

The main purpose of the HB SAE model is to produce precise district-level estimates, but it is equally important that the estimates remain stable when aggregated to larger administrative domains. Table 7 displays average RRMSE for each regency/municipality. This indicates that all subareas have relatively similar estimation precision.

Table 7. Average RRMSE for Regency/Municipality

Regency/Municipality	Average RRMSE
West Lombok Regency	21.23%
Middle Lombok Regency	21.69%
East Lombok Regency	21.28%
North Lombok Regency	21.34%
Mataram Municipality	21.43%
Sumbawa Regency	21.33%
West Sumbawa Regency	21.60%
Dompu Regency	21.35%
Bima Regency	21.18%
Bima Municipality	21.26%

3.4 Discussion

This study aimed to improve the precision of district-level stunting prevalence estimates in NTB Province, where stunting remains far above the national target. Direct estimation from the Survei Kesehatan Indonesia 2023 survey is limited for district-level analysis due to small sample sizes, resulting in unstable estimates with high sampling error. To overcome the limitation of direct estimation, this research applied a threefold HB SAE model using a Poisson-Gamma framework, incorporating auxiliary variables and multilevel random effect.

In Section 3.2, we present the findings of our study, showing that the threefold HB SAE model produces convergent parameter estimates. The model provides more reliable estimates of stunting prevalence than direct estimators, as discussed in Section 3.3. The implementation of the HB approach is important for capturing the complex structure of discrete response variables with overdispersion. Therefore, this approach is more suitable than the conventional Fay–Herriot SAE model, which has been commonly used in previous studies on SAE modeling for stunting prevalence [22].

Other literature using SAE for modeling Poisson-based data literature typically incorporates one or two random effects [20], [21]. In our study, we compare three models: a model with a single random effect at the small-area (district) level, a model with two random effects at the district and regency/municipality levels, and a three-level HB SAE model. Referring to Table 6, the additional random effect in the threefold HB-SAE model further improves model performance compared to the other models and results on better LOOIC. This suggest that multilevel HB SAE model that incorporates random effects at higher administrative levels is important because it is able to capture hierarchical heterogeneity that cannot be explained by a single small-area random effect alone. Districts within the same administrative and geographical area share common latent characteristics, which systematically influence the response variable. By accounting for this clustering effect, the multilevel model provides more realistic borrowing of strength across areas and yields more stable and reliable small area estimates.

However, this study assumes no relationship among small areas, meaning that the random effects of areas are treated as independent. This assumption may not fully reflect real-world conditions. The significance of the distance variable may suggest that geographically connected areas share similar characteristics. Another limitation of this study is the limited availability of auxiliary variables, as only health-related indicators are included. In practice, stunting prevalence may also be influenced by economic and demographic factors [22].

4. CONCLUSION

The threefold HB SAE with Poisson-Gamma distribution creates a model that strengthens direct estimators by including auxiliary variables and random effects across three domains. RRMSE for each district is below 25%. Districts in Sumbawa Island, especially districts in the Eastern part and/or far from the capital of their regency/municipality, tend to have higher stunting prevalence. Limited access to sanitation and clean water facilities contributes to higher stunting prevalence.

The HB SAE estimators of this research can be used to present more precise statistics on district level, where the number of samples is inadequate to present direct estimator. The usage of three random effects is efficient to capture differences between three area domains.

This study is limited by the availability of auxiliary data and the assumption of independence between area-level random effects. Future research could include auxiliary data capturing variables from other sectors related to health (e.g. economic, demography) and/or explore spatial correlation models to improve precision.

Author Contributions

Logananta Puja Kusuma: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Resources, Software, Visualization, Writing - Original Draft, Writing - Review and Editing. Kusman Sadik: Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing - Review and Editing. Anang Kurnia: Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing - Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to express their sincere appreciation to the Ministry of Health of the Republic of Indonesia for providing the health data used in this study. The authors also extend their gratitude to the Bureau of Statistics (Badan Pusat Statistik) for supporting the research with valuable data for auxiliary variables. The assistance of official data request services provided by both institutions contributed to the completion of this research.

Declarations

The authors declare that he has no conflicts of interest to report study.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] A. S. Rambi and Budyanra, “DETERMINANTS OF STUNTING IN THE UNDER-FIVE IN WEST NUSA TENGGARA PROVINCE WITH MULTILEVEL BINARY LOGISTIC REGRESSION,” *Jurnal Matematika, Statistika, dan Komputasi*, vol. 21, no. 1, pp. 103–119, September 2024. doi: https://doi.org/10.20956/j.v21i1.35765.

[2] B. S. Renyoet, D. Martianto and D. Sukandar, “POTENSI KERUGIAN EKONOMI KARENA STUNTING PADA BALITA DI INDONESIA TAHUN 2013,” *J. Gizi Pangan*, vol. 11, no. 3, pp. 247–254, November 2016.

[3] Kementerian Kesehatan Republik Indonesia, *Survei Kesehatan Indonesia*. Jakarta: Kemenkes, 2024.

[4] M. de Onis, E. Borghi, M. Arimond, P. Webb, T. Croft, K. Saha, L. M. De-Regil, F. Thuita, R. Heidkamp, J. Krasevec, C. Hayashi, and R. B. S. Flores-Ayala, “PREVALENCE THRESHOLDS FOR WASTING, OVERWEIGHT AND STUNTING IN CHILDREN UNDER 5 YEARS,” *Public Health Nutrition*, vol. 22, no. 1, pp. 175–179, January 2019. doi: https://doi.org/10.1017/S1368980018002434.

[5] Badan Perencanaan Pembangunan Nasional, *Pilar Pembangunan Ekonomi*. Jakarta: Bapenas, 2022.

[6] D. R. Sari, “ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI PREVALENSI STUNTING PADA BALITA DI JAWA BARAT”. [Thesis]. Bogor, ID: IPB University, 2025. [Online]. Available: IPB Repository.

[7] A. Kurnia, “PREDIKSI TERBAIK EMPIRIK UNTUK MODEL TRANSFORMASI LOGARITMA DI DALAM PENDUGAAN AREA KECIL DENGAN PENERAPAN PADA DATA SUSENAS”. [Dissertation]. Bogor, ID: IPB University, 2009. [Online]. Available: IPB Repository.

[8] R. E. Fay, and R. A. Heriott, “ESTIMATES OF INCOME FOR SMALL PLACES: AN APPLICATION OF JAMES-STEIN PROCEDURES TO CENSUS DATA,” *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 269–277, June 1979. doi: https://doi.org/10.1080/01621459.1979.10482505.

[9] J. N. K. Rao, “SMALL AREA ESTIMATION BY COMBINING TIME SERIES AND CROSS-SECTIONAL DATA,” *Can J Stat*, vol. 22, no. 4, pp. 511–528, December 1994. doi: https://doi.org/10.2307/3315407.

[10] K. Sadik, *Metode Prediksi Tak-Bias Linear Terbaik dan Bayes Berhirarki untuk Pendugaan Area Kecil Berdasarkan Model State Space*. [Dissertation]. Bogor, ID: IPB University, 2009. [Online]. Available: IPB Repository.

[11] E. Berg, “EMPIRICAL BEST PREDICTION OF SMALL AREA MEANS BASED ON A UNIT-LEVEL GAMMA-POISSON MODEL,” *Journal of Survey Statistics and Methodology*, vol. 11, no. 4, pp. 873–894, September 2023. doi: https://doi.org/10.1093/jssam/smac026.

[12] A. C. Cameron, and R. K. Trivedi, *Regression Analysis of Count Data Second Edition*. New York: Cambridge University Press, 2013. doi: https://doi.org/10.1017/CBO9781139013567.

[13] P. R. Sihombing, R. Mastiani, D. A. Sunarjo, and D. Muslianti, “COMPARISON OF GLM, GLMM, AND GEE POISSON MATHEMATICAL MODELING PERFORMANCE (CASE STUDY: NUMBER OF PULMONARY TUBERCULOSIS PATIENTS IN INDONESIA IN 2019-2021),” *Jurnal Tambora*, vol. 6, no 3, pp. 102-106, October 2022. doi: https://doi.org/10.36761/jt.v6i3.2081.

[14] K. Reluga, M. J. Lombardia, and S. Sperlich, “SIMULTANEOUS INFERENCE FOR EMPIRICAL BEST PREDICTORS WITH A POVERTY STUDY IN SMALL AREAS,” *Journal of the American Statistical Association,* vol. 118, no. 541, pp. 583–595, August 2021. doi: https://doi.org/10.1080/01621459.2021.1942014.

[15] P. Wididiati, H. Nurcahyanto, and A. Marom, "IMPLEMENTASI KEBIJAKAN PENANGANAN STUNTING DI KABUPATEN LOMBOK TIMUR (STUDI KASUS DI DESA LENEK DUREN KECAMATAN LENEK)," *Journal of Management and Public Policy*, vol. 11, no. 4, pp. 379 - 394, Oct. 2022.

[16] A. H. Amry, and H. Rowa, “IMPLEMENTASI KEBIJAKAN PENANGULANGAN STUNTING DI KABUPATEN LOMBOK UTARA PROVINSI NUSA TENGGARA BARAT”. [Thesis]. Jatinangor, ID: IPDN, 2025. [Online]. Available: IPDN Repository.

[17] D. B. Rein, “THE PREVALENCE OF BILATERAL HEARING LOSS IN THE UNITED STATES IN 2019: A SMALL AREA ESTIMATION MODELLING APPROACH FOR OBTAINING STATE, AND COUNTY LEVEL ESTIMATES BY DEMOGRAPHIC SUBGROUP,” *The Lancet Regional Health – Americas,* vol 30, no. 100670, February 2024. doi: https://doi.org/10.1016/j.lana.2023.100670.

[18] L. Marcis, D. Morales, M. C. Pagliarella, and R. Salvatore, "THREE-FOLD FAY–HERRIOT MODEL FOR SMALL AREA ESTIMATION AND ITS DIAGNOSTICS," *Statistical Methods & Applications,* vol. 32, no. 5, pp. 1563–1609, May 2023. doi: https://doi.org/10.1007/s10260-023-00700-6.

[19] J. N. K. Rao, "SOME RECENT ADVANCES IN MODEL-BASED SMALL AREA ESTIMATION," *Survey Methodology*, vol. 25, no. 2, pp. 175–186, December 1999.

[20] Y. F. Amin, Indahwati, and A. Kurnia, "TWOFOLD SUBAREA MODEL FOR ESTIMATING COMMUTER PROPORTION IN 10 METROPOLITAN AREAS," *BAREKENG: Journal of Mathematics and Its Application,* vol. 18, no. 2, pp. 1009–1022, June 2024. doi: https://doi.org/10.30598/barekengvol18iss2pp1009-1022.

[21] I. Wulandari, "KAJIAN MODEL POISSON UNTUK PENDUGAAN AREA KECIL DENGAN GALAT PENGUKURAN PADA PEUBAH PENYERTA". [Dissertation]. Bogor, ID: IPB University, 2025. [Online]. Available: IPB Repository.

[22] Tim Percepatan Stunting, *Prediksi Angka Stunting Tahun 2020.* Jakarta: TPPS, 2021.

[23] A. R. Sari, and I. Rahmi. "PENERAPAN METODE SMALL AREA ESTIMATION DENGAN PENDEKATAN KERNEL UNTUK MENDUGA PERSENTASE STUNTING DI INDONESIA," *Jurnal Matematika Unand,* vol. 8 no. 3, pp. 17-24, December 2019. doi: https://doi.org/10.25077/jmu.8.3.17-24.2019.

[24] E. Sunandi, "HIERARCHICAL LIKELIHOOD METHODS FOR SMALL AREA ESTIMATION WITH BETA-BINOMIAL RESPONSE AND MEASUREMENT ERROR IN THE AUXILIARY VARIABLE". [Dissertation]. Bogor, ID: IPB University, 2024. [Online]. Available: IPB Repository.

[25] A. Yanke, Kajian Pendugaan Area Kecil dengan Pendekatan Model Campuran Logistik Kekar pada Data yang Mengandung Pencilan [Thesis]. Bogor, ID: IPB University, 2022. [Online]. Available: IPB Repository.

[26] A. A. Mattjik, and I. M. Sumertajaya, *Sidik Peubah Ganda*. Bogor: IPB Press, 2011.