Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:62
https://doi.org/10.28919/cmbn/8845
ISSN: 2052-2541

SMALL AREA ESTIMATION FOR MORBIDITY RATE PREDICTION

INDRAMAWAN YUSUF ADI PRAYOGA, NOVI HIDAYAT PUSPONEGORO*, SUKIM, WINIH BUDIARTI

Department of Statistics, Politeknik Statistika STIS, Jakarta, 13330, Indonesia

Abstract: Complete and up-to-date data is important for development. However, its provision is costly. Survey is more efficient data collection solution, but the available sample is sometimes unable to produce direct estimate with sufficient precision. Small area estimation is one of the solutions by increasing the effectiveness of the sample. Indicators need to be presented with good precision so that policies are right on target. One of them is the health indicator which is expressed by morbidity rate. Morbidity rate is the percentage of population with health complaints that interfere with activities. Precision morbidity data is needed as criterion for determining non-communicable disease prevention at regional level as mandated by Law No. 17 of 2023. However, direct estimation from the survey shows that there are still districts/cities with Relative Standard Error (RSE) more than 25% like in Papua Island. Even though, especially Papua Province has an increasing morbidity rate. For this reason, in this study, indirect estimation of SAE Hierarchical Bayes (HB) Beta-Logistic was carried out to obtain estimate with good precision (RSE less than 25%). The results emphasize that SAE HB Beta-Logistic provide the estimate of morbidity rate precisely than the direct estimates for all districts/cities on the Papua Island.

Keywords: health complaints; hierarchical bayes; relative standard error; survey.

2010 AMS Subject Classification: 62J05, 62C10, 62P10.

1. INTRODUCTION

Complete and up-to-date data is essential for the implementation of development policies.

*Corresponding author
E-mail address: novie@stis.ac.id
Received August 16, 2024

However, its provision often requires a large amount of resources and costly. Survey can be a solution to data collection at a lower cost based on sample enumeration and certain rules are still able to maintain the representative aspect of population characteristics. However, surveys also have limitations if the number of samples available is not sufficient to represent the direct estimation. This small sample domain is then referred to as "small area" [1]. Small Area Estimation (SAE) is one of the solutions to overcome the insufficient sample in the domain to provide direct estimation.

Relative standard error (RSE) is one of the criteria that express the precision of parameter estimates [2]. Statistics Indonesia classifies the precision of the parameter estimates in three categories by its RSE. The first category is good precision that is appointed by RSE of the parameter estimates less than or equal to 25%. Otherwise, the imprecise parameter estimate has RSE more than 25%. And, the parameter estimate can be accepted to deliver but caution is needed in interpretation if the RSE is more than 25% but less than or equal to 50% [3]. SAE is an indirect estimation method that utilizes information from auxiliary variables and random effect areas so that the effectiveness of the sample size can be increased and reduce the standard error [1]. Some publications have been utilized the SAE method to estimate some indicators or official statistics in Indonesia [4], [5], [6], [7], and [8]. Thus, the indicators can be presented in small domain precisely and some policies are decided to be right on target.

The morbidity rate is the percentage of people with health complaints that interfere with their daily activities [9]. The morbidity rate is important as one of the indicators to monitor the management of non-communicable diseases, which is also a concern of the Sustainable Development Goals in the third goal [10]. In addition, this morbidity rate is also important because it is related to the establishment of the non-communicable disease control program as a regional priority mandated in Law No. 17 of 2023. This is also in line with the RPJMN strategy in the Ministry of Health's Strategic Plan 2020-2024 on disease control [11]. Therefore, the availability of regional level morbidity rate such as districts/cities, is important. However, the RSE of direct estimation of morbidity rate in several districts/cities in Indonesia are more than 25%.

Papua Island is the only region in Indonesia that gain the morbidity rate estimate with high RSE. Several of districts/cities in Papua Island have RSE of the morbidity rate estimate more than 25% and some even more than 50%. Lehtonen & Pahkinen (2004) stated that a relatively small sampling error is always expected in designing sample selection procedures and knowledge of the population structure can help achieve this goal [12]. In 2023, there are 7 districts/cities or 16.7%

of districts/cities on the Papua Island have RSE of the morbidity rate estimate more than 25%. Likewise, in the previous years in 2021, there were 13 districts/cities or around 30.95% of districts/cities had RSE of the morbidity rate estimate more than 25%. Even though the morbidity rate in Papua Island decreases but Papua Province gained higher than the previous year that is 5.73% compared to 5.15%.

The morbidity rate with good precision is important as one of the indicators in supporting the non-communicable disease control program according to the priority areas based on Law No. 17 of 2023. Non-communicable disease management is very important to be a concern because it is the biggest cause of death in Indonesia. Based on the 2020 Long Form Population Census, non-communicable diseases caused 87.14% of the total number of deaths in Indonesia [13]. In this case, the morbidity indicator in percentage form can be analogous to the proportion form by assuming that there is no morbidity rate greater than 100 by dividing it by the number 100 to produce a morbidity rate with a range of [0,1]. In research [8] and [5] using SAE Hierarchical Bayes Beta-Logistic was able to produce indicator estimates in the form of proportions with better precision.

Based on the background that has been described and adjusting data that can be analogous in the form of proportions with a range of [0,1], this study will conduct an indirect estimation of district/cities morbidity rate in Papua Island in 2023 using SAE HB Beta-Logistic with the aim of obtaining an estimation of morbidity rate in good precision for all districts/cities in Papua Island so that the reduction in morbidity rates in Papua Island can be faster and more targeted also as a basis for criteria on determining non-communicable disease prevention programs as regional priorities as mandated by Law No. 17 of 2023.

2. METHODS

Morbidity Rate

The morbidity rate is the percentage of the people with health complaints in the past month, causing disruption to daily activities [9]. In general, the morbidity rate calculation formula can be written as follows:

$$Morbidity\ Rate = \frac{A}{Total\ population} \times 100\% \tag{1}$$

where A is the number of people with health complaints that interfere with daily activities. It also reflects the level of public health [9].

Factor Affecting the Level of Public Health

Hendrik L. Bloom stated that there are several factors that affect the level of public health.

These factors include environmental factors, namely physical aspects (waste, water, air, housing), social (culture, education, economy), and biology (animals, microorganisms, plants); behavioral factors that refer to individual responses to the environment such as behavior towards clean water, dirty water disposal, healthy homes; health service factors influenced by distance to health services and information on the suitability of health service programs to community needs; and heredity [14].

Direct Estimation

Direct estimation is an estimator based on the survey design by considering survey weights that refer to the probability distribution according to the sample design [1]. According to Lehtonen & Pahkinen (2004), a design-based approach is used when estimation based on survey data is done by considering the sampling scheme [12]. Based on the National Socio-Economic Survey (SUSENAS) sampling scheme, the direct estimation of morbidity rate can be written as follows:

$$Morbidity\ Rate = \sum_{ighk} \frac{W_{ighk}^{(adj)}\ A_{ighk}}{W_{ighk}^{(adj)}} \times 100\% \qquad (2)$$

with A_{ighk} is the individual characteristics of the kth household, hth census block, gth stratum and ith district/cities, $W_{ighk}^{(adj)}$ is the final sampling weight with adjustment. As for the variance of direct estimation based on the SUSENAS, it does not have a closed-form by using the Taylor Linearization approach [15].

The quality of an unbiased estimator is usually measured by its variance and the root of the sampling variance is referred to as the standard error which affects the confidence interval of the estimation [2]. The RSE is a more standardized measure to compare the quality between estimators because it is relative to the estimated value. Therefore, a method that can produce a smaller standard error is needed, one of which is indirect estimation.

Indirect Estimation

Based on the model components, there are two types of indirect estimation, namely the implicit model which is a model that only uses auxiliary variables. Then explicit models that utilize auxiliary variables and also random effect areas to explain the variance between domains. One example of the application of the explicit model is in small area estimation [1].

Small Area Estimation

Small Area Estimation (SAE) is a method that can be used to estimate parameters in a small domain. According to Rao & Molina (2015), a domain is said to be small if the number of samples available in that domain is not sufficient for direct estimation with sufficient precision [1]. Therefore, a solution is needed to overcome the inadequacy of the sample, one of which is to use the small area estimation method. One of the more commonly used SAE methods under the assumption of generalized linear mixed models for both continuous and binary variables is Hierarchical Bayes (HB) [1].

Hierarchical Bayes

In general, the goal of Bayesian statistics is to represent prior uncertainty about model parameters with a probability distribution (prior) and update the prior uncertainty using current data (likelihood) to produce a posterior probability distribution that contains less uncertainty [16]. Bayes' theorem can be written as follows:

$$f(\boldsymbol{\theta}|\boldsymbol{data}) = \frac{f(\boldsymbol{data}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{data})} \tag{3}$$

and since the denominator component, $f(\boldsymbol{data})$, only scales the posterior density, Bayes' theorem is also often written as follows:

$$Posterior \propto Likelihood \times Prior \tag{4}$$

In the HB approach, the prior distribution is denoted by $f(\boldsymbol{\lambda})$ and the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ of the small area parameters is obtained by considering the $\boldsymbol{y}$ data. As for the inference, the HB method is based on the posterior mean value and to measure the precision results are based on the posterior variance. However, obtaining the posterior distribution sometimes requires multidimensional integration, so the Markov Chain Monte Carlo (MCMC) method is used by generating samples and simulating to estimate the posterior summary [1].

Markov Chain Monte Carlo

Ntzoufras (2009) mentioned that the MCMC technique is based on the construction of a Markov chain that eventually converges to the target distribution (stationary or equilibrium) which in this case is the posterior distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ [17]. In general, the MCMC algorithm can be written as follows:

a. Select an initial value $\boldsymbol{\theta}^{(0)}$.
b. Generate T value until the equilibrium distribution is reached.
c. Monitor the convergence of the algorithm using convergence diagnostic. If convergence diagnostics fail, then generate more observations.
d. Cut off the first B observations.

e. Consider $\{\boldsymbol{\theta}^{(B+1)}, \boldsymbol{\theta}^{(B+2)}, \dots, \boldsymbol{\theta}^{(T)}\}$ as the sample for the posterior analysis.
f. Plot the posterior distribution (usually focus is on the univariate marginal distribution).
g. Finally, obtain summaries of the posterior distribution (mean, median, standard deviation, quantiles, correlation)

There are several algorithms that can be used in the MCMC technique. Ntzoufras (2009) mentioned that there are two most popular algorithms in the MCMC technique, namely Metropolis-Hastings (M-H) and Gibbs Sampler [17].

Metropolis Hasting (M-H) Algorithm

Suppose the target (posterior) distribution $f(\boldsymbol{\theta}|\boldsymbol{y})$ is assumed, the Metropolis-Hasting Algorithm can be described by the following iterative steps:

1. Set initial values $\boldsymbol{\theta}^{(0)}$
2. For $t = 1, \dots, T$, repeat the following steps:
   a. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t-1)}$
   b. Generate new candidate values $\boldsymbol{\theta}'$ from a proposal distribution $q(\boldsymbol{\theta} \to \boldsymbol{\theta}') = q(\boldsymbol{\theta}'|\boldsymbol{\theta})$
   c. Calculate

$$\alpha = \min\left(1, \frac{f(\boldsymbol{\theta}'|\boldsymbol{y})q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{f(\boldsymbol{\theta}|\boldsymbol{y})q(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right) \tag{5}$$

   d. Update $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}'$ with probability α and $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ with probability $1 - \alpha$.

One special case of the Metropolis-Hastings algorithm is the Gibbs Sampler [17].

Gibbs Sampler

Gilks (1996) mentioned that one of the advantages of the Gibbs sampler is that in each step, the random values have to be generated from a unidimensional distribution which has various computational tools [17]. Often, these conditional distributions have a known form so that the random values can be easily simulated using standard functions in statistical and computational software. In general, the Gibbs sampler algorithm can be written as follows:

1. Set initial values $\boldsymbol{\theta}^{(0)}$.
2. For $t = 1, \dots, T$, repeat the following steps:
   a. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t-1)}$
   b. For $j = 1, \dots, d$, update θ_j from $\theta_j \sim f(\theta_j|\boldsymbol{\theta}_{\setminus j}, \boldsymbol{y})$

c. Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$ and save it as the generated set of values at $(t+1)$ iteration of the algorithm.

Now, there are various computational tools available that can generate random values from univariate distributions, making it possible to implement the Gibbs sampler in many cases, even when the resulting conditional posterior distribution is quite complicated. However, in some cases it is convenient to use the steps of a simple Metropolis-Hasting to generate the conditional univariate posterior distribution. This approach is known as the Metropolis-Hasting within Gibbs algorithm [17].

Metropolis Hasting Within Gibbs

Metropolis Hasting within Gibbs is a special case of Metropolis-Hastings where some components of the parameter vector are generated directly from the corresponding full conditional posterior distribution. The advantage of this algorithm is that it can easily perform blocking which means that the parameter vector is divided into several sub-vectors or blocks, containing correlated components, and each block is updated separately using the Metropolis-Hastings steps. In addition, this algorithm also allows certain fully conditional distributions to be easily generated and is generally capable of building efficient and flexible MCMC algorithms. In this case, the Metropolis-Hasting within Gibbs algorithm follows the component-wise Metropolis-Hasting algorithm which can be written as follows:

1.) Set $\boldsymbol{\theta}^{(0)}$.

2.) For $t = 1, \dots, T$, repeat the following steps:

a. Set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t-1)}$.

b. For $j = 1, \dots, d$,

(1) Generate new candidate parameter values θ_j' for j component of vector $\boldsymbol{\theta}'$ from a proposal distribution $q(\theta_j'|\boldsymbol{\theta})$.

(2) Calculate

$$\begin{aligned}\alpha &= \min\left(1, \frac{f\left(\theta_j'\middle|\boldsymbol{\theta}_{\setminus j}, \boldsymbol{y}\right) q\left(\theta_j\middle|\theta_j', \boldsymbol{\theta}_{\setminus j}\right)}{f\left(\theta_j\middle|\boldsymbol{\theta}_{\setminus j}, \boldsymbol{y}\right) q\left(\theta_j'\middle|\theta_j, \boldsymbol{\theta}_{\setminus j}\right)}\right) \\ &= \min\left(1, \frac{f\left(\boldsymbol{y}\middle|\theta_j', \boldsymbol{\theta}_{\setminus j}\right) f\left(\theta_j', \boldsymbol{\theta}_{\setminus j}\right) q\left(\theta_j\middle|\theta_j', \boldsymbol{\theta}_{\setminus j}\right)}{f\left(\boldsymbol{y}\middle|\theta_j, \boldsymbol{\theta}_{\setminus j}\right) f\left(\theta_j, \boldsymbol{\theta}_{\setminus j}\right) q\left(\theta_j'\middle|\theta_j, \boldsymbol{\theta}_{\setminus j}\right)}\right)\end{aligned} \tag{6}$$

where $\boldsymbol{\theta}_{\backslash j}$ is the vector $\boldsymbol{\theta}$ excluding its jth component θ_j [i.e., $\boldsymbol{\theta}_{\backslash j} = (\theta_1, \theta_2, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_d)$].

(3) Update $\theta_j = \theta_j'$ with probability α.

c. Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$.

Hierarchical Bayes to Estimate Small Area Proportion

Liu, Lahiri, dan Katlon (2014) in Rao & Molina (2015) proposed a beta sampling model to estimate small area proportions which can be written as follows:

$$\hat{\theta}_i|\theta_i \sim Beta(a_i, b_i) \tag{7}$$

with $a_i = \theta_i \left(\frac{n_i}{deff_{iw}} - 1\right)$ and $b_i = (1 - \theta_i)\left(\frac{n_i}{deff_{iw}} - 1\right)$. Then, $E(\hat{\theta}_i) = \frac{a_i}{a_i + b_i}$ is the expected value of $\hat{\theta}_i$ and $Var(\hat{\theta}_i) = \frac{a_i b_i}{(a_i+b_i)^2(a_i+b_i+1)}$ is the variance of $\hat{\theta}_i$. In this case, θ_i is modeled through a logit link function as follows:

$$logit(\theta_i)|\boldsymbol{\beta}, \sigma_v^2 \sim N(\boldsymbol{X}_i^T, \boldsymbol{\beta}, \sigma_v^2) \tag{8}$$

The HB inference on the proportion θ_i is implemented by assuming a flat prior on $\boldsymbol{\beta}$ and a gamma prior on σ_v^{-2}. In this case, $\boldsymbol{\beta}$ and σ_v^2 are independent with $\boldsymbol{\beta} \sim N(\mu_\beta, \sigma_\beta^2)$, with initial value $\mu_\beta = 0$ and $\sigma_\beta^2 = 1$. The MCMC samples of the joint posterior $(\theta_1, \dots, \theta_m, \boldsymbol{\beta}, \sigma_v^2)$ are generated using the Metropolis−Hastings within Gibbs algorithm. The Gibbs conditional of θ_i does not have a closed-form expression with $\boldsymbol{\beta}$ is the vector of regression coefficients, σ_v^2 is the variance of the random effect area, and $\boldsymbol{X}$ is the matrix of auxiliary variables [1].

3. DATA SETS

This paper focuses on the application of indirect estimation with the SAE HB Beta-Logistic method to estimate district/city level morbidity rate in Papua Island with better precision. Indirect estimation requires a dataset that contains at least two components, namely direct estimation and auxiliary variables.

Direct estimation of morbidity rates was obtained from the 2023 Papua and West Papua People's Welfare Statistics Publication produced by the Statistics Indonesia. Meanwhile, the auxiliary variables used in this study come from the 2021 Village Potential Data Collection in the form of raw data and then aggregated to the district/city level. Auxiliary variables that are aggregated include variables that are relevant to morbidity rates based on the literature review, namely the percentage of villages with most families defecating in latrines, percentage of villages with most families using decent water sources for drinking, percentage of villages with most families using decent water sources for bathing/washing, average of public elementary/equivalent school institutions per village, average of public junior high/equivalent school institutions per village, average of public senior high/equivalent school institutions per village, average of hospitals (hospital and maternity hospital) and community health center per village, average of integrated health post with service activities once a month or more per village, average of doctors and midwives per village, and average of other health workers per village.

4. MAIN RESULTS

4.1 Overview of Direct Estimates of Districts/Cities Morbidity Rate in Papua Island in 2023 and The Patterns of Associations with Auxiliary Variables

The results of the direct estimation of the districts/cities morbidity rate in Papua Island in 2023 compiled from the Papua and West Papua People's Welfare Statistics Publication in 2023 can be presented in the descriptive statistics summary table as follows:

Table 1. Descriptive statistics of direct estimation of morbidity rate

Descriptive statistics	Morbidity Rate (%)
Minimum	0.2958
Mean	5.9079
Maximum	13.7658
Variance	8.6201
Range	13.4699

From Table 1, it can be seen that the lowest and highest morbidity rate in Papua Island in 2023 are 0.3% and 13.77% respectively, so the range of morbidity rate in Papua Island is 13.47%. In this case, the district/city with the highest morbidity rate is Kepulauan Yapen District. This is in line with the research of Farsan et al. (2010) that related to the islands, the level of health services is influenced by the level of accessibility (in the spatial dimension), level of education, and income of the people in the region [18]. However, we have to be cautious in the interpretation of these descriptive statistics because there are some morbidity rate estimations of districts/cities that have RSE more than 25%, so SAE modeling needs to be applied to overcome the issue. In conducting SAE modeling, the availability of adequate and relevant auxiliary variables is needed to assist modeling. The following scatterplot is presented to see the relationship pattern between the auxiliary variable from the 2021 Village Potential Data Collection (Podes 2021) and the variable of interest:

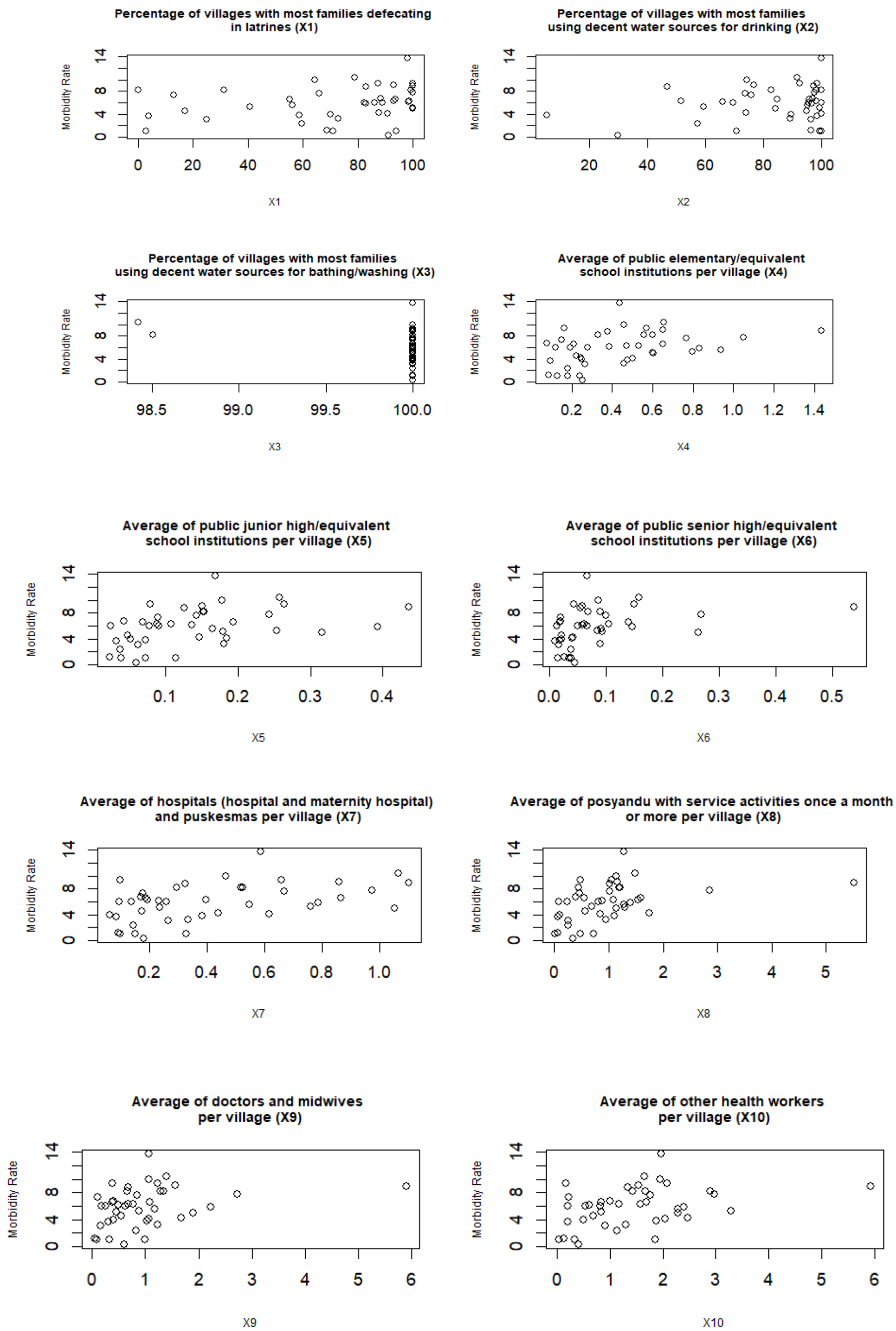

Figure 1. Scatterplot of the relationship between auxiliary variables and morbidity rate

To make it easier to know the relationship between auxiliary variables and variables of interest, the following correlation test is presented using Pearson correlation:

Table 2. Pearson correlation test between auxiliary variables and morbidity rate

Code	Variable	Correlation	P-value	Decision
X_1	Percentage of villages with most families defecating in latrines	0.2328	0.13798	Not significant
X_2	Percentage of villages with most families using decent water sources for drinking	0.1902	0.22770	Not significant
X_3	Percentage of villages with most families using decent water sources for bathing/washing	-0.2649	0.09004	Not significant
X_4	Average of public elementary/equivalent school institutions per village	0.4022	0.00829	**Significant**
X_5	Average of public junior high/equivalent school institutions per village	0.4198	0.00564	**Significant**
X_6	Average of public senior high/equivalent school institutions per village	0.3339	0.03067	**Significant**
X_7	Average of hospitals (hospital and maternity hospital) and community health center per village	0.4612	0.00211	**Significant**
X_8	Average of integrated health post with service activities once a month or more per village	0.4026	0.00821	**Significant**
X_9	Average of doctors and midwives per village	0.3107	0.04522	**Significant**
X_{10}	Average of other health workers per village	0.3614	0.01871	**Significant**

From the results of the correlation analysis, it is obtained that there are 7 variables that have a significant correlation with the interest variable, including the variable $X_4, X_5, X_6, X_7, X_8, X_9,$ and X_{10}. The variables as a whole have a correlation of more than 0.3.

4.2 Indirect Estimation of Districts/Cities Morbidity Rate with SAE HB Beta-Logistic

After testing the correlation and obtaining 7 variables with significant correlation, the next step is to select auxiliary variables using the stepwise regression procedure. By using the R-Studio software, the best three variables were obtained that will be used in building the model, which include variables X_7, X_8, and X_9.

Assuming that there is no percentage of morbidity that is greater than 100, the morbidity rate is analogous to a proportion by dividing the morbidity rate by 100 so that the morbidity rate will be in the range of 0-1. Because the value is in the range of 0-1 and by using the logit link function, the HB Beta-Logistic model will be used in this study. Modeling is done with 500 iterations update, burn in 5000, and 100000 iterations of MCMC. The results can be described through diagnostic plots consisting of trace plot, density plot, and autocorrelation plot as follows:

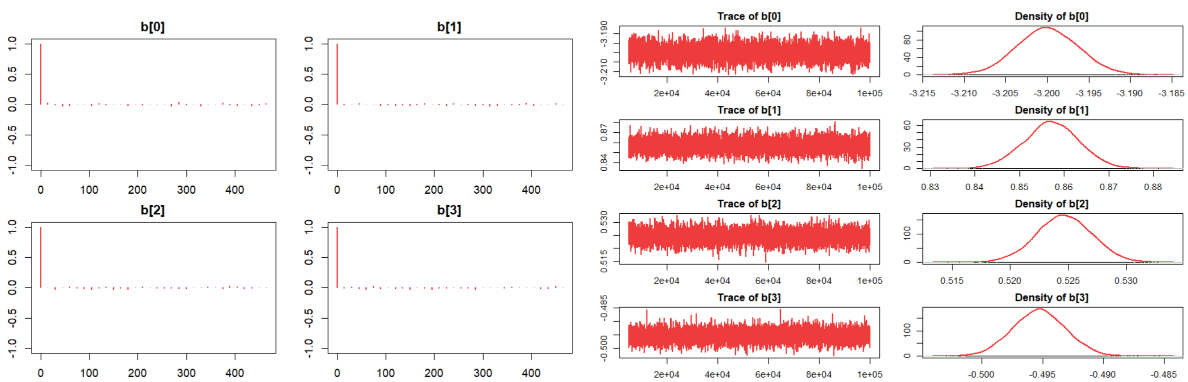

Figure 2. Trace plot, density plot, and autocorrelation plot of SAE HB Beta-Logistic model

From Figure 2, it can be seen that the autocorrelation plot has cut off, the trace plot does not have a periodic pattern, and the density plot is also smooth which shows that the results have converged with the following credible interval:

Table 3. Credible interval of SAE HB Beta-Logistic model coefficient estimates

Variable	Mean	SD	2.5%	97.5%
Intercept	-3.2000	0.0037	-3.2070	-3.1928
Average of hospitals (hospital and maternity hospital) and community health center per village ($\boldsymbol{X_7}$)	0.8572	0.0061	0.8450	0.8693
Average of integrated health post with service activities once a month or more per village ($\boldsymbol{X_8}$)	0.5247	0.0024	0.5199	0.5293
Average of doctors and midwives per village ($\boldsymbol{X_9}$)	-0.4953	0.0022	-0.4995	-0.4911

From Table 3, it can be seen that the results of SAE HB Beta-Logistic modeling show that all variables contribute significantly to the model. This can be seen in the range of 2.5% and 97.5%, none of which passes 0 with the most dominant variable in this model, namely X_7. However, the relationship between this variable and the variable of interest shows a positive relationship with morbidity rate, which means that when these facilities are improved, the morbidity rate will increase. This is thought to be due to the uneven availability of these health facilities on the Papua Island as the results of the Study on Improving Quality Health Services at Health Care Facilities in Papua and West Papua Provinces conducted by the Collaboration of Communities and Services for Prosperity (KOMPAK) in a partnership program of the Government of Indonesia and Australia in 2022 which shows that health facilities in the two provinces are not yet evenly distributed, such as not all districts/cities have Regional General Hospitals (RSUD), community health center which generally tend to be concentrated in urban areas or districts/cities that have long existed so that people must be referred to neighbouring areas that have these health facilities [19].

The following boxplot presents the comparison between the results of the SAE HB Beta-Logistic estimation of the direct estimation results:

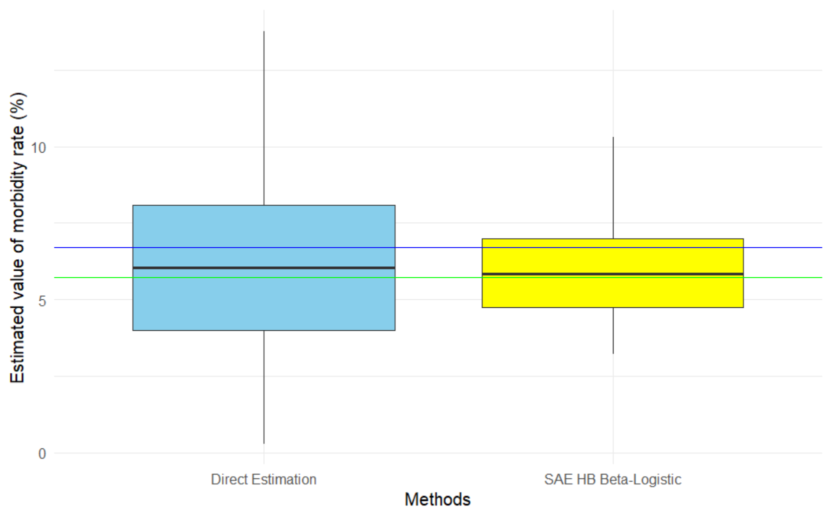

Figure 3. Boxplot of estimation comparison between the two estimators

From Figure 3, it can be seen that the results of estimating the morbidity rate using SAE HB Beta-Logistic indirect estimation show results that are not much different on average. In addition, the Papua (green) and West Papua (blue) provincial figures of 5.73% and 6.70% respectively are still in the boxplot box. This indicates that the estimation results do not deviate far from the morbidity rate at the provincial level. In addition, the estimation results using SAE HB Beta-Logistic tend to be more homogeneous than direct estimation. This indicates that the districts/cities numbers do not spread too far from the average so that the provincial numbers become more representative in representing the districts/cities morbidity rate.

4.3 Comparison of RSE Result of SAE HB Beta-Logistic to Direct Estimation Result

The following table presents a summary table of RSE comparison of indirect estimation results of SAE HB Beta-Logistic against direct estimation by classification:

Table 4. Comparison of RSE results of the two methods according to the classification

RSE	Number of District/Cities	
	Direct Estimation	SAE HB Beta-Logistic
< 25%	35	42
25% - 50%	6	0
> 50%	1	0

From these results, it can be seen that SAE HB Beta-Logistic is able to estimate the districts/cities morbidity rate on Papua Island in 2023 with RSE < 25% for all districts/cities. This provides benefits to the estimation results because it is included in the category of estimation with good precision for all districts/cities on the Papua Island.

4.4 Mapping of Districts/Cities Morbidity Rate based on the Best Estimation Results

From the comparison of the RSE values of the two methods, it can be seen that SAE HB Beta-Logistic is able to produce estimates with the best precision, namely with RSE less than 25% for all districts/cities in Papua Island. The following is a map of the distribution of districts/cities morbidity rate in Papua Island based on the results of indirect estimation of SAE HB Beta-Logistic:

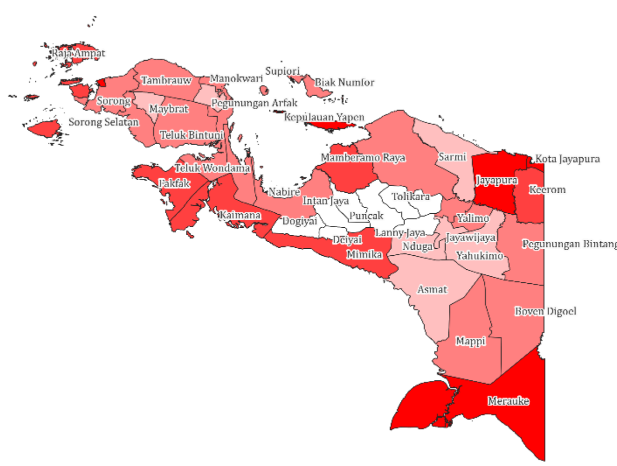

Figure 4. Map of the distribution of morbidity rate from SAE HB Beta-Logistic results

Figure 4 shows that the redder the district, the higher the morbidity rate. From this figure using SAE HB Beta-logistic, it can be obtained that the estimated morbidity rate is similar to the direct estimation results, namely Kepulauan Yapen District is still one of the districts with the highest morbidity rate on Papua Island. In addition, on the map it can also be seen that coastal areas tend to have higher morbidity rates compared to mountainous areas. This is thought to be due to differences in environmental conditions and topography in the two areas. Research conducted by Nurkadri & Hayati (2020), shows that in mountainous areas that have lower temperatures will cause red blood cells that work to bind oxygen to be greater which then causes the desire to exercise to be much greater [20]. In addition, the resilience of communities in highland areas is also better than that of communities in lowland areas. This is in line with research Muis (2016) and Rudi Aldiyanto et al. (2021), each of which sequentially shows that in this case the primary school population and the 16-19 years old population in the highland area have a better fitness level compared to the population in the lowlands [21], [22].

5. CONCLUSION

From the results of the analysis that has been done, the SAE HB Beta-Logistic method is able to produce the results of estimating morbidity rates with good precision categories (RSE ≤ 25%) for all districts/cities in Papua Island in 2023. Thus, it is expected that the resulting data will be better in the context of reducing morbidity rates, especially in Papua Island and as a basis for determining criteria for non-communicable disease prevention program as regional priorities as mandated by Law No. 17 of 2023. In addition, based on the mapping results, coastal areas are expected to be a priority target in the context of reducing morbidity on the island of Papua in the future.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support provided by Politeknik Statistika STIS through the publication funding of this research. This contribution has been instrumental in facilitating the dissemination of research findings and reflects Politeknik Statistika STIS's strong commitment to academic advancement.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] J.N.K. Rao, I. Molina, Small Area Estimation, Wiley, New Jersey, 2015.

[2] R.E. Walpole, R.H. Myers, S.L. Myers, et al. Probability and Statistics for Engineers And Scientists, Pearson, Boston, 2022.

[3] Statistics Indonesia, People's Welfare Statistics 2022, Statistics Indonesia, Jakarta, 2022.

[4] N.H. Pusponegoro, A. Kurnia, K.A. Notodiputro, et al. Small Area Estimation of Sub-District's Per Capita Expenditure through Area Effects Selection Using LASSO Method, Procedia Computer Sci. 179 (2021), 754–761. https://doi.org/10.1016/j.procs.2021.01.064.

[5] I. Fajar, T. Siswantining, S.M. Soemartojo, The Estimated Proportion of Chronic Disease Sufferer in Duren Sawit District, East Jakarta, Using Hierarchical Bayes Method in Small Area Estimation (SAE), J. Phys.: Conf. Ser. 1321 (2019), 022061. https://doi.org/10.1088/1742-6596/1321/2/022061.

[6] Z.O. Briantiko, R. Agustina, Small Area Estimation Application for Complete Basic Immunization Indicator Estimation Calculation on the Island of Java And Bali in 2020, Semin. Nas. Off. Stat. 2022 (2022), 449–458. https://doi.org/10.34123/semnasoffstat.v2022i1.1551.

[7] W.A. Nurizza, Application of the Fay-Herriot Model to Small Area Estimation Simulation Study of Regency/City Level Per Capita Expenditure of East Kalimantan Province 2020, BESTARI: Bull. Stat. Apl. Terkini 1 (2021), 1–7.

[8] N.P. Istiqomah, I.Y. Wulansari, Estimation Gross Enrolment Rate of Higher Education at Regency/City Level in Kalimantan Island using Small Area Estimation Hierarchical Bayes Beta-Logistic, Semin. Nas. Off. Stat. 2022 (2022), 137–146. https://doi.org/10.34123/semnasoffstat.v2022i1.1225.

[9] Statistics Indonesia, People's Welfare Indicators 2023, Statistics Indonesia, Jakarta, 2023.

[10] National Development Planning Agency, Healthy and Prosperous Life, https://sdgs.bappenas.go.id/17-goals/goal-3/, Accessed Aug. 14, 2024.

[11] Ministry of Health Republic Indonesia, Highlights of the Ministry of Health's Strategic Plan 2020-2024, Ministry of Health, Jakarta, 2020.

[12] R. Lehtonen, E. Pahkinen, Practical Methods for Design and Analysis of Complex Surveys, Wiley, 2004. https://doi.org/10.1002/0470091649.

[13] Statistics Indonesia, Mortality in Indonesia from the Long Form of the 2020 Population Census, Statistics Indonesia, Jakarta, 2023.

[14] Irwan, Etika dan Perilaku Kesehatan, CV. Absolute Media, Bantul, 2017.

[15] Statistics Indonesia, Book 1 Guidelines for Heads of Provincial BPS and Heads of Regency/City BPS National Socio-Economic Survey SUSENAS March 2023, Statistics Indonesia, Jakarta, 2023.

[16] S.M. Lynch, Introduction to Applied Bayesian Statistics and Estimation for Social Scientists, Springer, New York, 2007.

[17] I. Ntzoufras, Bayesian Modeling Using WinBUGS, Wiley, 2009.

[18] A. Farsan, Suratman, Spatial Variation in the Level of Utilisation of Health Service Facilities in the Islands Region: The Case of Islands in Towes Sub-district, Muna Regency, Southeast Sulawesi Province, Thesis, Gadjah Mada University, 2010.

[19] Kompak, Study on Improving Quality Health Services at Health Care Facilities in Papua and West Papua Provinces, Kompak, Jakarta, 2022.

[20] N. Nurkadri, R. Hayati, Pengaruh Dataran Tinggi Terhadap Tingkat Kebugaran Jasmani Mahasiswa Universitas Negeri Manado, Kinestetik 4 (2020), 79–84. https://doi.org/10.33369/jk.v4i1.10411.

[21] J. Muis, Perbedaan Tingkat Kesegaran Jasmani Antara Siswa Pesisir Dengan Siswa Pegunungan Pada SD Negeri 46 Kota Parepare, Publ. Pendidikan 6 (2016), 179-186. https://doi.org/10.26858/publikan.v6i3.2273.

[22] R. Aldiyanto, R. Febrianti, A.R. Hakim, Perbandingan Tingkat Norma Kebugaran Jasmani Remaja Usia 16 – 19 Tahun (Usia Sma) Berdasarkan Letak Geografis, J. Ilm. Penjas (Penelit. Pendidik. Pengajaran), 7 (2021), 1–13. https://doi.org/10.36728/jip.v7i1.1383.