Jurnal Biometrika dan Kependudukan


p-ISSN 2302–707X e-ISSN 2540–8828
10(2): 171–180, December 2021
DOI: 10.20473/jbk.v10i2.2021.171–180

# ESTIMATION OF INFANT MORTALITY RATES IN INDONESIA BY USING EMPIRICAL BEST LINEAR UNBIASED PREDICTION

**Easbi Ikhsan[1], Nadra Yudelsa Ratu[2]**
[1]Center for Education and Training, Central Bureau of Statistics, Jakarta, Indonesia
[2]Directorate of Statistical Information Systems, Central Bureau of Statistics, Jakarta, Indonesia
Corresponding Author: Easbi Ikhsan
E-mail: 14.8096@stis.ac.id

## ABSTRACT

Infant Mortality Rate (IMR) is one of the many indicators that can measure the health status of a population in an area. IMR is also part of the third Sustainable Development Goals (SDGs), namely to ensure healthy lives and promote well-being for all of all ages. IMR was produced with direct estimation from the Indonesian Demographics Health Survey (IDHS). However, the result of the 2017 IDHS publication indicated that several direct estimations of IMR in 34 provinces in Indonesia had high relative standard error (RSE) values. Accurate data (from the RSE value) is needed for policy making. Therefore, this paper focused on small area estimation (SAE) by using the empirical best linear unbiased prediction (EBLUP) method and estimated IMR to the provincial level. SAE works by using the strength of several variables from the village potential data (*Potensi Desa*) which correlates strongly with IMR. The results of the analysis with the RSE used as a measure of model accuracy showed that by using the SAE EBLUP method in the IDHS data, an average RSE value of 15.23% was obtained, which is smaller than the direct estimate of the average RSE value of 29.51%. This research paper concludes that SAE using the EBLUP method is good for estimating the Provincial level IMR value in Indonesia in 2017.

**Keywords**: Infant mortality rate, small area estimation, empirical best linear unbiased prediction

## *ABSTRAK*

*Angka kematian bayi (AKB) merupakan satu dari sekian indikator yang dapat mengukur derajat kesehatan populasi penduduk di suatu wilayah. AKB juga merupakan dari program* Sustainable Development Goals *(SDG) yang ketiga, yaitu menjamin kehidupan yang sehat dan meningkatkan kesejahteraan bagi semua usia. AKB dihasilkan melalui estimasi langsung dari Survei Demografi dan Kesehatan Indonesia (SDKI). Akan tetapi, hasil publikasi SDKI 2017, estimasi langsung AKB pada 34 provinsi di Indonesia ditemukan adanya beberapa provinsi dengan nilai standard relatif eror/relative standard error (RSE)yang masih tinggi. Padahal dalam mengambil kebijakan tentunya diperlukan data yang memiliki akurasi (dari nilai RSE) yang baik pula. Penelitian ini membahas penggunaan* Small Area Estimation *(SAE) menggunakan metode* Empirical Best Linear Unbiased Prediction *(EBLUP) untuk mengatasi keterbatasan estimasi AKB di level provinsi tersebut. SAE dilakukan dengan meminjam kekuatan beberapa variabel dari data potensi desa (PODES) yang berkorelasi kuat dengan AKB. Hasil analisis dengan membandingkan ukuran RSE sebagai ukuran akurasi model memperlihatkan bahwa dengan penggunaan SAE metode EBLUP pada data SDKI diperoleh rata-rata nilai RSE sebesar 15,23 persen yang lebih kecil dibandingkan estimasi langsung dengan nilai rata-rata RSE 29,51 persen. Kesimpulan, SAE menggunakan metode EBLUP baik untuk memperkirakan AKB level provinsi di Indonesia pada tahun 2017.*

***Kata kunci:*** *Angka kematian bayi, estimasi berbasis permodelan, relatif standard eror*

***Received: January 25, 2020*** ***Accepted: February 11, 2021***

## INTRODUCTION

The United Nations (UN) stated that various programs could be implemented to improve the welfare of infants in the world and reduce infant mortality rates. The tangible manifestation of this is the implementation of the third goal of the Sustainable Development Goals' (SDGs), namely preventing death of newborns and toddlers. According to this target, every country, including Indonesia, is to reduce their infant mortality rate (IMR) to under 12 in every 1000 live births by 2030 (Santoso et al., 2019).

The IMR can play a role in measuring the health status of a community in an area. This indicator can help measure a community's level of welfare, which includes health and quality of life. In addition, IMR is one of the indicators in the National Medium-Term Development Plan (*Rencana Pembangunan Jangka Menengah Nasional*) in Indonesia (Pinontoan and Tombokan, 2015).

According to the Central Bureau of Statistics (2020), IMR is a number that represents the number of deaths of infants under 1 year old out of every 1000 live births in a given year. IMR can also be defined as the probability of a baby dying before reaching the age of one (expressed by per thousand live births). A high IMR indicates that babies in the area have poor health and are susceptible to disease and even death (Aryuni, 2019).

In Indonesia, to obtain representative values, IMR is calculated from survey results. The results of the survey were obtained by using a direct sampling technique from a specific domain/area (Rao and Molina, 2015). IMR results were obtained through the Indonesian Demographics Health Survey (IDHS). The IDHS is a survey conducted jointly by the Central Bureau of Statistics (BPS), the National Population and Family Planning Commission (BKKBN), and the Ministry of Health. This survey aims to provide up-to-date estimates of basic demographic and health indicators. In addition, the IDHS can provide an overview of a population and public health.

The IMR calculation from the IDHS is generated through a direct estimate known as a design-based estimation approach (Anggreyani et al., 2016). This estimation approach uses the survey design used in the IDHS. In addition, direct estimation involves weighing the value of the survey method and inferential values such as the standard error value based on the probability distribution obtained from the sample design (Rao and Molina, 2015). However, there are several drawbacks to direct estimation, one of which is when the estimate is made on a small sample size, the estimation results obtained can be inaccurate and produce large standard error values (Ikhsan et al., 2019).

Based on the results of the 2017 IDHS report publication, the IDHS was created with a sample design that can produce estimated values at the national and provincial levels. The sampling method used in the IDHS was stratified two-stage sampling. First, a census block was selected by using a systematic probability proportional to size (PPS) method with the size of the number of households obtained from the results of the 2010 population census listing. Second, 25 households were systematically selected from the census block in the first stage. From these 25 households, 8 households with married men (aged 15-24 years) were systematically selected for interviews (National Population and Family Planning Commission, 2018).

The IDHS survey design has limitations as the estimation results are only at the national and provincial levels. This causes difficulty in obtaining data from the 2017 IDHS for lower levels such as districts/cities or sub-districts. If the estimate is forced to a lower level, the estimate obtained is biased and has a high error.

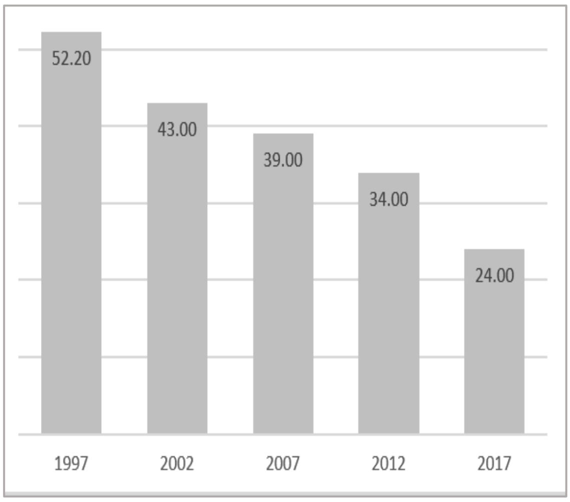


**Figure 1.** The distribution of IMR in Indonesia.

Figure 1 shows that from 1997 to 2017, the IMR value in Indonesia continued to decline. In 1997 the IMR value was 52.20 and in 2017 the IMR value fell to 24.00. This figure indicates that in 1997 there were 52 infant deaths for every 1000 babies born. The lower IMR values in the last three decades suggests an increase in the quality of infant health in Indonesia. Although IMR at the national level has decreased and the sampling design of the presentation of results for the 2017 IDHS data is designed to the provincial level, publication of IMR indicators at the provincial level is still not available. Direct IMR estimation results from the 2017 IDHS are only available at the national level (National Population and Family Planning Commission, 2018). Therefore, there is a lack of data for the provincial government to determine appropriate steps to handle and reduce infant mortality cases.

The IMR from the 2017 IDHS results that are not available at the provincial level is caused by the presence of several provinces with a relative standard error (RSE) value that is still too high at above 25%. RSE itself is an indicator that shows the magnitude of the error caused by the use of sampling techniques for data collection in a survey, which in this case is the RSE from the direct estimation of the IMR indicator from the 2017 IDHS. The RSE value of ≤ 25% is considered accurate for an estimate to be published (Central Bureau of Statistics, 2018). A high RSE certainly does not allow IMR to serve at the provincial level. The value of RSE could be due to the small sample size. In this case, the number of observations or examples in the study are below the standard number of observations.

In the case of the 2017 IDHS, there are several provinces with small sample sizes. If a direct estimate is made for the provincial level, the estimation results will be unrepresentative. Given the limitations of direct estimation in calculating IMR, other statistical methods are needed that can reduce the RSE value (Setyawan, 2016). Therefore, it is necessary to improve the results of the direct estimation of IMR at the provincial level that uses an indirect estimation approach. Improvements could be made by taking an approach that utilizes specific auxiliary/predictor variables at the provincial level. One of the statistical methods that can be used to correct the imprecision of estimation results and high RSE values is the method of small area estimation (SAE) (Muchlisoh, 2017a).

SAE can predict estimation results with relatively small and reliable variances (Rao and Molina, 2015a). The estimated estimation results are indicators at the area level that do not have sufficient samples to produce direct estimates (Muchlisoh, 2017). A simple concept in SAE modeling is estimating the parameter values of an area by using information from accompanying variables in the same area (Notodiputro and Kurnia, 2005). The concept of utilizing power here means that the estimation of the statistical value with SAE takes into account the effect of the surrounding area (Wulansari, 2016). Furthermore, the estimation in SAE is based on an equation involving additional information known as a model-based approach (Rao and Molina, 2015). The model determines that data or other related variables are included in the estimation process so that SAE calculations can be conducted.

Therefore, in this study, an IMR indicator was estimated by using the SAE method. IMR estimation using the SAE method involves additional variables (auxiliary variables) that can explain the model used. The results of this SAE estimation could then determine the RSE performance and direct estimation for the IMR indicator from the 2017 IDHS data. It is expected that the results of this study will show a lower RSE estimation result compared to the results from direct estimates of the IMR indicators from the 2017 IDHS data.

## METHOD

This research applied the statistical analysis method with the cross-sectional data approach. The research areas in this study includes statistical measurements of infant mortality (IMR) through SAE modeling. The unit of analysis in this study are the provinces in Indonesia. The use of SAE was done to determine the accuracy of direct IMR estimation from using the SAE model.

This method was applied to data from the results of the 2017 IDHS and also data on Village Potential in 2018 as secondary data. One of the data generated by the IDHS is the infant mortality rate (IMR) variable per province in Indonesia. This IMR variable was used as the response variable in this study. Several variables from the 2018 Village Potential (PODES) data were used as auxiliary variables.

The stages of analysis used are as follows:

1. The IMR per province was estimated using the sampling design from the IDHS. The IMR estimation process was conducted using the R software with the NHANES (National Health and Nutrition Examination Survey) package.
2. The results of the direct IMR estimation were juxtaposed with the accompanying variables from the 2018 Village Potential data and selected variables that were significantly correlated with the IMR response variable.
3. An SAE model was established using the EBLUP method to estimate the IMR response variables by using information from the accompanying variables.

4. The RSE was calculated by using the SAE model of the EBLUP method from the estimation results of the IMR response variable to determine the accuracy of the direct estimate with the SAE estimation on the IMR response variable.
5. The direct estimated RSE (direct estimation) was compared with the RSE of the SAE estimation of the EBLUP method.
6. A literature review of the linkage of co-variables used in SAE modeling was conducted.

SAE was used because random sampling methods or direct estimation are mostly used in surveys when the sample size is too small or if it is conducted in a small area. This tends to produce low accuracy values or RSE (Notodiputro and Kurnia, 2005). One of the estimates that can be used to improve direct estimation can be through using SAE. SAE works by utilizing the accompanying variables from the observed area. One of the SAE methods commonly used is the empirical best linear unbiased prediction (EBLUP) method. This method was chosen because it is more general and is quite widely used in SAE-related studies in Indonesia.

The EBLUP method is a linear combination method of fixed random effects and fixed effects. The EBLUP method is used in several studies such as Ikhsan et al. (2019) which showed that the measure of the accuracy of the RSE value of the estimate can be minimized. The EBLUP formula is written in the following formula:

$$\tilde{\theta}_i = \boldsymbol{x}_i^T\,\beta + b_i v_i + e_i. \ldots\ldots (1)$$

Where $\tilde{\theta}_i$ is the response variable (estimated) from the *i* area, *i* is the area from 1 to m. This equation ($\boldsymbol{x}_i^T$) is a vector of the accompanying variables, $b_i$ is a constant with a value of 1, $v_i$ is a random influence area that has a normal distribution with an average (mean) parameter of 0 and a variance of $\sigma_v^2$. The variable $e_i$ is the error value of the response variable, which is normally distributed with parameter 0 and variance $\psi_i^2$. The value of $\psi_i^2$ can be obtained from the sampling variance when the direct estimation process of the IMR response variable. Based on equation (1), it can be formulated again as follows (Rao and Molina, 2015):

$$\tilde{\theta}_i^{EBLUP} = \tilde{\gamma}_i\hat{\theta}_i + (1-\tilde{\gamma}_i)\boldsymbol{x}_i^T\,\beta. \ldots (2)$$

Where the component $\tilde{\gamma}_i$ valued $\frac{\sigma_v^2}{\sigma_v^2+\psi_i^2}$. Several parameters such as $\sigma_v^2$ and $\beta$ are estimated using residual maximum likelihood (REML). The results of the SAE estimation using the EBLUP method were then measured for accuracy from the RSE.

RSE itself is a measure of the convergence of the resulting estimates. The RSE value is obtained by dividing the root of the variance of the estimate by the estimated value then multiplied by 100% (Rahman and Harding, 2017). The smaller the RSE value, the larger the estimator. To facilitate the research of the SAE estimation process with the EBLUP method and also to ensure the accuracy of the RSE value, the calculation was conducted by using the SAE package on the R software (Molina and Marhuenda, 2015).

## RESULTS

The IMR has not been published due to the high RSE value in several provinces in Indonesia at the year of 2017. However, when viewed from the processed results of the 2017 IDHS raw data, the distribution of IMR per province can be seen in Figure 2. Based on the publication results, the national IMR obtained is 24 (Central Bureau of Statistics, 2017). Based on the 2017 IDHS data, there are 26 provinces in Indonesia whose IMR is above the national figure. The largest IMR value was in Gorontalo Province at 58.56, this indicates that there are 58 to 59 babies who die every 1000 births in Gorontalo Province. The lowest value is in the Riau Islands Province of 13.46, which means that there are 13 to 14 infant deaths per 1000 births in the Riau Islands Province.

Furthermore, judging from the RSE value, which is a measure of the accuracy of the IMR estimator (Figure 3), the tolerable RSE size from the Central Bureau of Statistics is agreed at 25%. If the RSE value is more than 25%, the survey results tend to not be published. The IMR RSE value in the Riau Islands province, which is 46%, is the largest IMR RSE value in Indonesia in 2017. Furthermore, the lowest IMR RSE value is in Maluku province with 12%. Overall, there are 14 provinces with RSE values greater than 25%. Therefore, the direct estimation results of the 2017 IDHS data for IMR requires further analysis with SAE. The implementation of the SAE method is expected to reduce the RSE value of the direct estimated IMR.

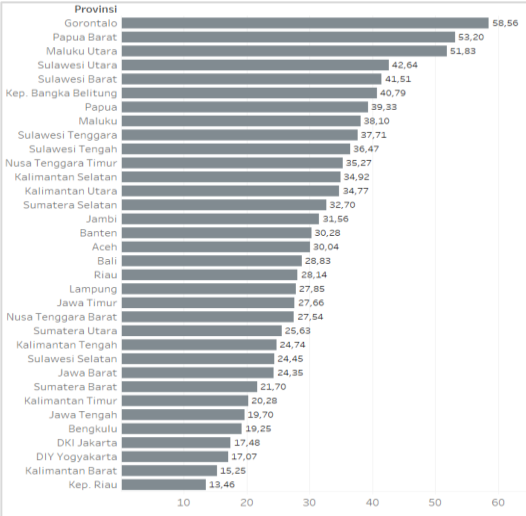


**Figure 2.** IMR per Province in Indonesia in 2017

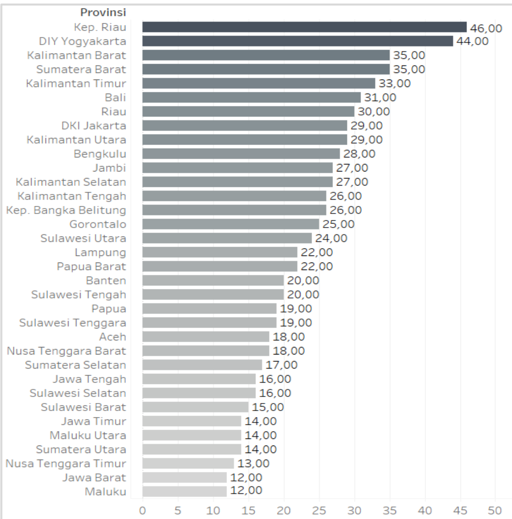


**Figure 3.** Distribution of IMR RSE Values per Province in Indonesia.

The first step in modeling the SAE estimation is to find the accompanying variables that are correlated with the response variable (IMR). The accompanying variables taken from the Village Potential 2018 data are as the number of families in a province ($X_1$), the number of families in a province that uses electricity ($X_2$), the number of villages with health facilities ($X_3$), the percentage of villages in a province that use electricity as the main lighting device ($X_4$), the number of hospitals ($X_5$), the number of integrated healthcare centers/*pos pelayanan terpadu* ($X_6$), the number of doctors ($X_7$), the number of health workers ($X_8$), the percentage of villages that have had outbreaks ($X_9$), the number of villages that have an epidemic ($X_{10}$), the number of cases of malnutrition ($X_{11}$), the number of certificates of incapacity or *Surat Keterangan Tidak Mampu* (SKTM) ($X_{12}$), and the number of midwives ($X_{13}$). The following shows the correlation value between the IMR response variables and the accompanying variables.

**Table 1.** Correlation of Participating Variables from the Village Potential 2018 Data to Direct Estimation of IMR in 2017

| Variable | Correlation (R) |
|---|---|
| $X_1$ | -0.308 |
| $X_2$ | -0.314 |
| $X_3$ | -0.163 |
| $X_4$ | -0.354 |
| $X_5$ | -0.314 |
| $X_6$ | -0.288 |
| $X_7$ | -0.369 |
| $X_8$ | -0.330 |
| $X_9$ | -0.156 |
| $X_{10}$ | -0.139 |
| $X_{11}$ | -0.283 |
| $X_{12}$ | -0.353 |
| $X_{13}$ | -0.301 |

Based on Table 1, the largest correlation of co-variables using absolute values is found in variable $X_7$ (number of doctors) which is 0.369 and the smallest is $X_9$ (percentage of villages that have ever had an outbreak case) at 0.159 (in absolute terms). The accompanying variables from $X_1$ to $X_{13}$ were conducted through stepwise regression to select the appropriate model with IMR. The selection of this model with stepwise regression takes into account the highest R-squared value and also the smallest *Akaike Information Criterion* (AIC) value.

The results from the selection of the model with stepwise regression found that a good accompanying variable was the R-Squared criteria, and the significance of the variable were the model with the accompanying variables $X_7$, $X_{10}$, and $X_{12}$. Details of the magnitude of the R-Squared value are 0.279.

The significant variable at the 5% error rate is the number of doctors ($X_7$).

Furthermore, variables that are significant at an error rate of 10% are the number of villages that have had outbreak cases ($X_{10}$), and the variable number of certificates of incapacity (SKTM) (X12). All of the accompanying variables were obtained from the Village Potential 2018 data.

After obtaining three accompanying variables from the selection of variables through the stepwise regression, modeling was conducted using the SAE EBLUP method. The SAE package in R was used with the *mseFH* function, several parameters from the SAE model were obtained, including *est*, *estcoef*, *refvar*, goodness, and MSE. The parameter *est* produced the estimated value of SAE on IMR data.

The *estcoef* parameter produced the parameter value of the SAE model coefficient which resembles the linear regression coefficient model format, but the value is different because it uses a linear mix model (LMM) model. Furthermore, the *refvar* parameter is the value of the variance of the random effect area or the variance of the randomness of the area. The *refvar* value here is $\sigma$ in the model notation of equation 2. Then, the *goodness* value contains a measure of the goodness of the model which can be seen from the log-likelihood, AIC, *Bayesian Information Criterion* (BIC), and *Kashyap's information criterion* (KIC) values.

The parameter that determines the goodness of the SAE estimation results is called the *mean square error* (MSE). The MSE parameter is a measure of dispersion in SAE. The MSE measure shows the level of accuracy of the model estimated using the SAE method. The smaller the value of the MSE, the better the results of the SAE estimation.The results of SAE modeling through the *est* parameter of the *mseFH* function obtained IMR estimation results from the SAE method. As seen in Figure 4, the value of IMR estimation and SAE with direct estimation of IMR is not much different. Furthermore, several figures are quite influential in the direct estimate while the SAE results are stable, and the outliers are not too far apart. The RSE value as a result of dividing the square root parameter of MSE with the SAE estimated value of the EBLUP method from the *est* parameter, the RSE value is obtained as shown in Figure 5.

The results of the direct estimation and the IMR SAE estimation in Figure 4 are not much different and are still lower than the direct estimation pattern. When viewed from the accuracy of each estimation method based on the RSE value, Figure 5 shows that the IMR RSE value from the SAE EBLUP method is lower than the IMR RSE value from the direct estimate. Figure 5 also shows that when using the direct estimation method, there are still several provinces with RSE values above 25% (shown above the green horizontal line).

The statistical summary shows that the RSE of the SAE estimation with the EBLUP method has an average of 15.23% compared to the direct estimate with an average RSE value of 23.40%. In addition, when viewed from the RSE of the direct estimate of the IMR, the largest RSE value is 45.76%, and the SAE estimate from using the EBLUP method is 18.91%. This also applies to the direct estimated value of IMR in 14 provinces in Indonesia, which was above 25% to now being below 25 percent.

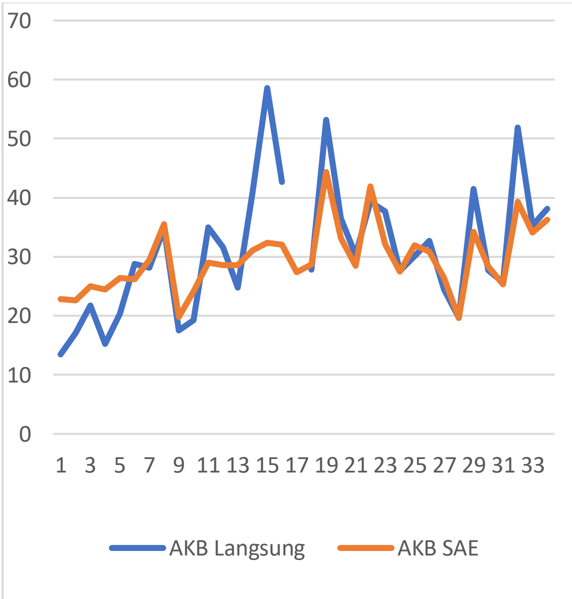


**Figure 4.** Results of the Comparison of IMR estimates and Direct IMR Estimates Using the SAE EBLUP Method

After using the EBLUP method on SAE, there is no longer an RSE value that is above 25%. In addition, the results of other parameters such as the *refvar* value of 24.86, the *goodness* of the model parameter, namely the loglikelihood of -119.614, AIC of 249, 238, BIC of 156, 8706, and KIC of 254, 238.

Moreover, as seen from the *estcoef* parameter in table 2, the value of each beta coefficient is 0.002 for the variable number of doctors, 0.014 for the variable number of epidemic villages, and 0.188 for the variable number of villages that have a certificate of incapacity or SKTM.

In the estimation results, the area randomness variance (*refvar*) is also 24.86. When the gamma value is calculated for each estimation result, it was obtained that the average gamma value for SAE estimation using the EBLUP method is 0.39. Rao and Molina (2015) stated that the gamma value can have a value from 0 to 1. If the gamma value is close to 0 then the SAE estimate tends to lead to synthetic regression estimation, whereas if the gamma value tends to approach 1 then the SAE estimate tends to lead to a direct estimate (estimated based on sampling design). If the estimation results lead to a direct estimate, this indicates that the use of SAE is not suitable to estimate the indicator. This is because the optimal gamma value is expected to be in the middle of 0.5 and in this figure the contribution of direct estimation with synthetic regression estimation will provide an optimal contribution to produce good SAE estimation results (Rao and Molina, 2015).

**Table 2.** Coefficient Estimation for SAE Model Using EBLUP Method

| Variables | Beta Coefficient | Standard Error | P-value |
|---|---|---|---|
| *Intercept* | 47.747 | 8.499 | 0.000 |
| Number of Doctors ($X_7$) | -0.002 | 0.001 | 0.007 |
| Number of Outbreak Villages ($X_{10}$) | 0.014 | 0.007 | 0.041 |
| Number of Villages with SKTM ($X_{12}$) | -0.188 | 0.095 | 0.047 |

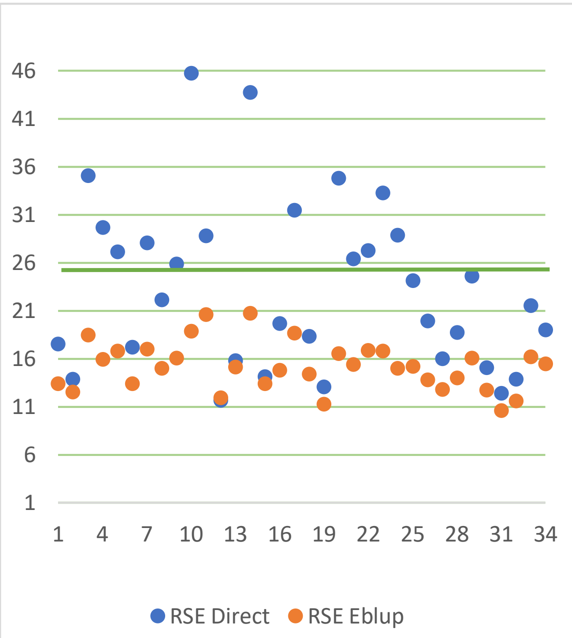


**Figure 5.** Comparison of RSE IMR Direct Estimation with IMR Estimation Using SAE EBLUP Method

## DISCUSSION

The use of SAE estimates for IMR indicators from 2017 IDHS data and 2018 Village Potential data shows better results than the direct estimates obtained from 2017 IDHS data alone. The IMR estimation results at the provincial level from using SAE showed a much lower RSE value than the direct estimation results.

Furthermore, from the results of the research conducted, a comparison of the results of the estimated values of IMR from direct estimates and SAE estimates could be conducted (Figure 4). The comparison of SAE estimation results with direct estimates on IMR shows that there is not too much of a difference between SAE estimation results and direct estimation results. Based on the results of the statistical measure, the average difference between the direct estimate and the SAE estimate is 4.65 points. Furthermore, if we look at the largest difference in absolute value between the two estimation results, it was found that the RSE size value generated by the SAE estimate tends to be smaller than the direct estimate. This shows that the SAE estimation results tend to improve values that look like outliers or values that have a high RSE on direct estimates. The statistical average of the data values in Figure 5 shows that the SAE of the EBLUP method on the IMR indicator from the 2017 IDHS data obtained an average RSE value of 15.23%, which is smaller than the direct estimate with an average RSE value of 29.51%.

The success of the SAE method in improving MSE and RSE values is in line with various previous studies related to the use of SAE estimation in IMR. An example would be Anggreyani, Indahwati, and Kurnia (2016) and their research titled “Small Area Estimation for Estimating the Number of Infant Mortality in West Java, Indonesia”. This study used the SAE method of the generalized linear mix model (GLMM) with quasi likelihood on the IDHS data to estimate the IMR value. The results of research from Anggraeni Indahwati, and Kurnia (2016) showed that the estimated value using SAE can approach the true value of a population on an indicator under study conditions. In addition, the success of the SAE method obtained can be seen from the smaller MSE size compared to the direct estimation model.

The results of the SAE estimation at the provincial level IMR was obtained by the province with the highest IMR at 44.34, namely West Papua. In the case of the highest IMR in Papua based on Papua Provincial Health Office (2017), it was found that the delivery process affected the safety of both mother and baby. The safety of mothers and babies affects the infant mortality rate. This shows that factors related to health workers, one of which is the number of doctors (X7) which is one of the variables in SAE, affects the high IMR, especially in West Papua Province.

The province with the lowest IMR was Central Java at 19.62. According to the Central Java Provincial Health Office (2017), it was found that there was an excess of health workers. Where there should have been 1,195 general practitioners, in 2017 the number of doctors in Central Java was 1,576 general practitioners, resulting in an excess of 399 doctors. For other health workers such as midwives, there was an excess of 9,920 midwives. This shows that the condition of the health workforce has contributed to the low IMR rate in Central Java.

Furthermore, significant concomitant variables used in SAE such as the number of doctors ($X_7$), the number of villages with outbreaks ($X_{10}$), and the number of villages with SKTM ($X_{12}$) in the SAE modeling of the EBLUP method indicate that there is an indication of a relationship between health factors (seen from the availability of medical personnel and outbreak cases) as well as economic factors (as seen from the certificate of incapacity recipients in a village).

The indications of the accompanying variables from the research results are also in line with several other studies that have discussed IMR. Sihite (2017) discussed the factors that affect IMR in North Sumatra. This study used IMR data from North Sumatra Province from 2012 to 2016. The results of this study indicate that factors such as poverty level, health personnel, and also gross domestic product for an area (GRDP) have a significant effect with an effect of 97.9% of IMR in 33 districts/cities in North Sumatra province in 2012-2016.

Then there is a relationship between health factors in the form of an outbreak in an area with IMR as according to a study by Irfan (2018) which discusses IMR in terms of health conditions in the family environment. This study concludes that the level of environmental health is negatively correlated with IMR. This can be related to the accompanying variable on whether or not the presence of an outbreak in an area influences the IMR value.

## CONCLUSION AND SUGGESTIONS

### Conclusion

The results of this study indicate that the provincial level IMR estimation in 2017 which was previously lacking in terms of accuracy can be improved by using the EBLUP method of SAE estimation. The equation model of the SAE EBLUP method with the selected variables in the model is the number of doctors, number of outbreak villages, and number of villages with certificates of incapacity. These variables were selected based on the significance test of the model in multiple linear regression and the *goodness* of the model in the EBLUP method.

The improvement of the provincial IMR estimation by using the SAE EBLUP method can be seen from the RSE value of 23.40% in the direct estimate, with the EBLUP method SAE becoming15.23%. This shows that there is an efficiency level of 35% when using SAE.

### Suggestion

The use of SAE estimation of the EBLUP method on IMR data are expected to be a reference in estimating several indicators of public health status other than IMR in future studies if the condition of the sample size is still small or not representative. In addition, for further research, up-to-date sources or social media big data sources could be used to determine accompanying variables regarding the latest phenomena related to an indicator to produce a better model in estimating IMR by using SAE.

## REFERENCES

Anggreyani, A., Indahwati and Kurnia, A., 2016. Small Area Estimation for Estimating The Number of Infant Mortality in West Java, Indonesia. In: *AIP Conference Proceedings*. American Institute of Physics Inc. pp.1-11.

Aryuni, H.A., 2019. *Evaluasi Pelaksanaan Jejaring Peduli Kesehatan Ibu dan Anak "Rindu KIA" dalam Penurunan AKI dan AKB di Kabupaten Kulon Progo*. Yogyakarta : Universitas 'Aisyiyah.

Central Bureau of Statistics, 2018. *Statistik Kesejahteraan Rakyat Labuhanbatu Utara 2018*. Kabupaten Labuhabatu Utara : BPS Labuhabatu Utara.

Central Java Provincial Health Office, 2018. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2017*. Semarang: Central Java Provincial Health Office.

Central Bureau of Statistics, 2020. *Badan Pusat Statistik*. [online] Available at: <https://www.bps.go.id/> [Accessed Feb 21 . 2020].

Ikhsan, E., Ratu, NY and Nurizza, WA, 2019. Small Area Estimation of Gross Enrollment Rates in Universities in Kalimantan Island in 2018. journal.fmipa.unmul.ac.id, p.pp.82-88.

Ikhsan, E., Ratu, N.Y. and Nurizza, W.A., 2019. Estimasi Small Area Estimation Angka Partisipasi Kasar di Perguruan Tinggi Pulau Kalimantan Tahun 2018. *jurnal.fmipa.unmul.ac.id*, p.pp.82-88.

Irfan, M., 2018. *Analisis Pola Hubungan Kondisi Kesehatan Lingkungan Keluarga Terhadap Kasus Angka Kematian Bayi Di Kota Surakarta Tahun 2016*. Universitas Muhammadiyah Surakarta.

Molina, I. and Marhuenda, Y., 2015. sae: An R Package for Small Area Estimation. *The R Journal*, 7(1), p.pp.81-98.

Muchlisoh, S., 2017a. *Pengembangan Model Pendugaan Area Kecil Dengan Pengaruh Acak Waktu Mengikuti Proses Autoregresif Ordo Pertama*. IPB (Bogor Agricultural University).

Muchlisoh, S., 2017b. *Pengembangan Model Pendugaan Area Kecil Dengan Pengaruh Acak Waktu Mengikuti Proses Autoregresif Ordo Pertama*.

National Population and Family Planning Commission, Central Bureau of Statistics, Ministry of Health and ICF, 2018. *Pedoman Teknis Badan Pusat Statistik dalam Rangka Survei Demografi dan Kesehatan Indonesia 2017*. Jakarta : National Population and Family Planning Commission.

Notodiputro, K.A. and Kurnia, A., 2005. Pendekatan General Linear Mixed Model Pada Small Area Estimation. *Indonesian Journal of Statistics and Its Applications*, 10(2), pp.12-16.

Papua Provincial Health Office, 2018. *Profil Kesehatan Provinsi Papua barat Tahun 2017*. Jayapura : Dinas Kesehatan Provinsi Papua.

Pinontoan, M.V. and Tombokan, S.G., 2015. Hubungan Umur dan Paritas Ibu dengan Kejadian Bayi Berat Lahir Rendah. *Jurnal Ilmiah Bidan*, 3(1), p.pp.20-25.

Rahman, A. and Harding, A., 2017. *Small area estimation and microsimulation modeling*. *Small Area Estimation and Microsimulation Modeling*. Taylor and Francis Group.

Rao, J.N.K. and Molina, I., 2015a. *Small Area Estimation: Second Edition*. *Small Area Estimation: Second Edition*. Hoboken, NJ, USA: John Wiley and Sons.

Rao, J.N.K. and Molina, I., 2015b. *Small Area Estimation*. Second Edi ed. *Wiley Series in Survey Methodology*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Santoso, D., Arifin, K. and Parhusip, B.T.P., 2019. *Panduan Monitoring dan Evaluasi Bagi Pelaksanaan Pencapaian Tujuan Pembangunan Berkelanjutan (TPB/SDGs) di Daerah*. Jakarta: INFID.

Setyawan, A., 2016. *Small Area Estimation Metode Spatial Empirical Best Linear*

*Unbiased Predictor untuk Estimasi Persentase Wanita Usia Subur dengan Fertilitas Tinggi di Kabupaten Mamuju dan Mamuju Tengah*. Surabaya: Institut Teknologi Sepuluh Nopember.

Sihite, D., 2017. *Analisis Faktor- Faktor yang Mempengaruhi Angka Kematian Bayi di Sumatera Utara*. Medan: Universitas Sumatera Utara.

Wulansari, I.Y., 2016. Bayesian Hierarchical Small Area Model For Unmatched Sampling. In: *SEMASTAT 2016*. p.pp.136-144.