



A Thematic Mapping Approach to Stunting Segmentation in Indonesia by Province Using Finite Mixture Partial Least Squares Method

Ayu Setya Permatasari and Rahmawati Erma Standsyah*

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, Indonesia

Abstract

Stunting remains one of Indonesia's most persistent public-health problems, and the national prevalence conceals substantial variation across the country's 38 provinces. This study models seven latent constructs behind stunting, antenatal care, early initiation of breastfeeding, immunization, sanitation, exclusive breastfeeding, low birth weight, and stunting using Partial Least Squares Structural Equation Modeling on 2023–2024 provincial health-profile data published by Indonesia's Ministry of Health, and then applies Finite Mixture Partial Least Squares (FIMIX-PLS) to detect unobserved heterogeneity across provinces and to translate the resulting segments into a thematic map. The measurement model satisfied every recommended validity and reliability threshold, with an Average Variance Extracted between 0.835 and 0.932 and both Cronbach's alpha and composite reliability above 0.80. In the pooled national model, low birth weight was the only construct with a statistically significant direct path to stunting ($\beta = 0.540$, $p < .001$); antenatal care showed no significant direct path to stunting ($\beta = -0.143$, $p = .678$) but exerted a significant total effect ($\beta = -0.365$) entirely through a fully mediated pathway via low birth weight (indirect effect $\beta = -0.263$, $p = .022$), together explaining 63.0% of the variance in stunting. Model-selection criteria favored a three-segment solution comprising 21, 10, and 7 provinces, dominated respectively by sanitation-related, birth-weight-related, and breastfeeding-related structural pathways. Mapping these segments geographically shows that provinces sharing a dominant pathway are not scattered at random but tend to cluster by region, indicating that stunting-reduction programs should be tailored to each segment's dominant pathway rather than applied uniformly across the whole country.

Keywords: FIMIX-PLS; SEM-PLS; stunting; thematic map

Copyright © 2026 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

1. Introduction

Stunting is one of those problems that never has a single cause. Stunted growth means the height of a child being very small compared to their peers. But the root cause behind this issue is generally associated with the most crucial period of one's life – 1,000 days after birth [1]. Analyzing the statistics of different countries, we have noticed the constant tendency, as if this same thing was happening everywhere. It is not just "lack of nutrition", it is something deeper and much more complicated; it is something similar to the chain reaction: poor health care, poor

*Corresponding author. E-mail: rahmawatistandsyah@unesa.ac.id

nutrition, and poor environment – all these factors combined during this crucial period [2], [3]. The most shocking aspect of this case is not merely the children's shorter physical stature. The impact of this nutritional deficiency will affect them throughout their lives: cognitive limitations, brain development issues, and an increased risk infancy resurfaces decades later as a burden on the person and on the health system that has to treat them [4].

Indonesia carries more of that burden than is middle-income status world suggest. The 2023 Indonesia Health Survey put national stunting prevalence at 21.5%, barely down from 21.6% the year before [5], still above the 14% government target and above the 20% mark the World Health Organization treats as a public-health alarm [6]. That national number also hides a lot Papua Tengah, Nusa Tenggara Timur, and Papua Pegunungan sit at the top of the 2023 rankings, part of Java and Bali stay comfortably below the national average, and more than half of Indonesia's 38 provinces are still above it [5], [7]. Some of the inequalities come from differences in money and opportunities for work. Family income and jobs matter a lot too. The schooling the mothers got is another factor [8], [9]. These all play into what food or medical care or sanitation the family can get and it varies by region. If just one thing stood out then the rates should be equal everywhere but they are not. That points to something else being involved. This suggests there might be more going on than what shows up at first [10].

All of the above cannot be measured directly through any standard. Quality of health care services and environmental sanitation are examples of latent constructs which are inferred from the variables that can be directly observed. This may be one of the reasons why structural equation modeling has gained popularity in recent studies on nutrition among children in Indonesia during the last decade. The modeling approach has been used to analyze malnutrition prevalence in locations such as Aceh [11], as well as for determining the changes in inequalities in stunting in the country [8]. Studies have also been carried out in relation to the association between water and sanitation access and health effects on mothers and children throughout Indonesia. The aspect regarding its popularity quite logical considering that direct measurement is impossible in such situations. However, some may disagree [12]. Studies on the impact of sanitation facilities have identified similar negative correlations between stunting in eastern Indonesia and Papua [13], [14]. Also, there are geographic clusters of vaccination rates associated with regions having higher levels of stunting in the area [15]. The PLS regression analysis technique can be used in this situation since it does not assume the multivariate normal distribution [11].

The problem is that applying one SEM-PLS model to each province while assuming all these provinces as one and same population, which has been contradicted by the above mentioned prevalence gap, is indeed a limitation that has been proved by previous studies on stunting based on clustering [16]. The Finite Mixture Partial Least Squares (FIMIX-PLS) model proposed by Hahn et al. [17] and then implemented by Ringle et al. [18], addresses exactly this problem because it estimates both path coefficients and membership probability of segments at the same time.

While there are limited number of studies on applying the technique to analyze Indonesian stunting data, these studies still lack in completeness. Linda and Huda [4] used the technique to classify 34 provinces based on antenatal care, breastfeeding, immunization, and sanitation as the exogenous constructs with the result of 4 segments. However, the authors stopped writing in the middle of their research after presenting the results in table form only. Madeira and Anuraga [10] went finer, down to the district/city level, which is a genuine strength, but they examined only one province, Nusa Tenggara Timur, so there is no sense of how that province compares with the rest of the country. Surayya and Sirait [11] built a solid PLS-SEM model of undernutrition in Aceh, covering health access, food intake, and socioeconomic status, but never added a heterogeneity-detection step, so whatever variation exists within the province is simply averaged into one set of path coefficients. Taken together, these three studies point to a clear gap: no study has combined a national, multi-province structural model of health-service, breastfeeding, immunization, and environmental pathways with a heterogeneity-detection layer

and a map that translates the statistical output into something usable for decision-making. This paper aims to fill that space.

The contribution of this paper is empirical and integrative; no novel estimators or algorithms are developed. To the best of our knowledge, this paper is the first one to apply a methodological pipeline involving outer-model validation, bootstrapped inner-model testing, FIMIX-PLS heterogeneity detection, and thematic mapping to an Indonesian stunting study. Also, this paper uses a set of diagnostic tools that were not applied by other authors who had used FIMIX-PLS to analyze Indonesian stunting data. This diagnostic set involves bootstrapped confidence intervals, out-of-sample prediction performance estimation using *PLSpredict* and segment count selection using multi-criteria optimization approach with sensitivity testing. Segmentation analysis in other papers involved only antenatal care and sanitation, which are considered together in [4] or a single composite construct of health access [10]. In this paper, early initiation of breastfeeding, immunization, and environmental sanitation are taken into account separately along with antenatal care, exclusive breastfeeding, and low birth weight, thus creating a more complex structural model with seven constructs in total, six exogenous determinants together with stunting as the endogenous outcome. Creation of thematic maps based on segment membership is an important part of the results achieved in this paper and cannot be considered just a formal element of analysis.

Analytical information could be obtained from segment probability tables for the methodologist, while for policymakers interested in knowing if the regions share risk profiles and vulnerabilities, the same information could be obtained from the shaded map. The task of identifying whether the FIMIX-PLS segments match regional patterns, that is, if there is a separate region for eastern Indonesia, and if the high prevalence regions like Central and Highland Papua are distinct structural segments is also facilitated through thematic mapping. Solving these issues using segment membership tables is virtually impossible. Which implies that there is a dominant structural path for each segment in the provinces. These are strategic targeting tools and not statistical classifications, as segments associated with antenatal care would require a totally different kind of approach from those associated with sanitation or birth weight. This cannot be accomplished with just one national model. Information could be found in section 2. The Section 2.3 will contain the data and the FIMIX-PLS approach, whereas the Section 2.2 will cover the measurement model, the structural model, the bootstrap results, and the segmentation with provinces and the corresponding thematic map.

2. Methods

This research an applied quantitative approach was employed by using secondary data from the province for the purpose of clustering 38 Indonesian provinces in relation to stunting contributing factors patterns. There was no manipulation or control group used, as this research was not about manipulating and measuring any kind of control group. The value this study adds comes from applying Finite Mixture Partial Least Squares (FIMIX-PLS) at national scale, refreshing both the latent variables and the data period along the way, then pushing things one step further by turning the segmentation output into a thematic map visualization, something neither references study got around to doing.

2.1. Research Design and Data

This is an applied, quantitative study built on secondary data, with each of Indonesia's 38 provinces treated as one unit of analysis. What's being modeled is how seven latent variables line up to produce stunting: antenatal care (ξ_1), early initiation of breastfeeding (ξ_2), immunization (ξ_3), and environmental sanitation (ξ_4) sit on the exogenous side, while exclusive breastfeeding (η_1), low birth weight (η_2), and stunting (η_3) sit on the endogenous side. Each construct is measured through several manifest indicators pulled from routine health-profile statistics, following the same indicator structure used in comparable FIMIX-PLS stunting studies in Indonesia [4], [10].

The analysis itself runs in three stages: outer model evaluation, inner model evaluation with bootstrap-base hypothesis testing, and finally FIMIX-PLS segmentation, each of which is laid out below with its governing equations. The secondary data used in this study came from the Indonesian Health Profiles for 2023 and 2024, which were made accessible by the Republic of Indonesia's Ministry of Health. More details about stunting and its contributing factors can be found in the Thematic Report of the 2023 Indonesian Health Survey: Stunting in Indonesia and its Determinants, which was created by the Ministry of Health's Health Development Policy Agency. Provincial health indicators, such as aspects of immunization, breastfeeding, maternal and child health, sanitation, and other stunting-related concerns, were evaluated in this article. These figures were taken from the official publication repository of the Republic of Indonesia's Ministry of Health on July 25, 2026. ([Ministry Health Development Policy Agency repository](#))

Data Harmonization and Data Cleaning, apart from gathering the published data in a single matrix for analysis, a formatting process was also undertaken. Prior to entering the data in SmartPLS, the data which had been initially provided in Indonesian decimal format was converted into standard decimal format and rounded to the nearest whole number in Excel. The fourteen main indicators have been extracted from the published data in provincial health profiles issued by the Indonesian Ministry of Health without any aggregation in terms of administrative levels, without any changes to the definitions each year, and without carrying out any calculations. This is because all the data points which have been used currently have been published in terms of a provincial percentage.

No missing values exist within the final data matrix. The data matrix of 38 provinces by 14 indicator matrix contained no instances that would have required imputation, deletion, or substitution due to the fact that the fourteen manifest variables used are based on the direct extraction of percentages of all 38 provinces from the health profile statistics produced by the Indonesian Ministry of Health pre-aggregated percentages for all 38 provinces. Thus, other than the aforementioned conversion of decimals into percentages, SmartPLS's requirement of a full data matrix was met. Final dataset for analysis: once cleaning was done, the dataset of this study with 38 provinces and 14 manifest variables, which were loaded on seven latent constructs, was saved as the input data file in SmartPLS ([SmartPLS Raw Dataset](#)).

All statistical analysis was done using SmartPLS 4. PLS-SEM algorithm was run under the path weighting model, standardized indicator, weight of 1.0 for all indicators, up to 3,000 iterations as the maximum number of iterations, and 10^{-7} as the stopping rule. Automatic outer weights was used for all seven constructs. Significance of path coefficients was assessed through bootstrapping with bias-corrected percentile, 5,000 subsamples, two-tailed test at the significance level of 0.05, and fixed seed value. Segmentation FIMIX-PLS for the candidate number of groups 2 to 5 was estimated by up to 5,000 iterations in each run and 10 runs for each K in order to overcome the local optimum problem, with fixed seed value. In case of choosing solution $K = 3$, all 10 repetitions had log-likelihood within -96.2 to -69.8 (from 35 to 187 iterations).

2.2. SEM-PLS Segmentation

The estimation by SEM-PLS and FIMIX-PLS was done in three stages using the seven latent variables and their previously specified manifest indicators. Since the reliability of the structural results rely on the reliability of the underlying variables, the measurement (outer model) was first assessed for indicator reliability, convergent validity, and internal consistency. Second, the 38 provinces were treated as a single population in order to estimate and evaluate the structural (internal model) using bootstrap resampling. Third, the FIMIX-PLS approach was used to identify latent segments and re-estimate the structural links within each segment because grouping can conceal variability between provinces. Below is a detailed description of each stage.

2.2.1. Measurement Model (Outer Model)

Every construct in this model is reflective, meaning the indicators are treated as manifestations of the underlying latent variable rather than the other way around. For an exogenous construct ξ_j with indicator x_j and an endogenous construct η_j with indicator y_j , the outer model is written in Eq. (1) and Eq. (2).

$$x_j = \lambda_j \xi_j + \delta_j \quad (1)$$

$$y_j = \lambda_j \eta_j + \varepsilon_j \quad (2)$$

where λ_j is the loading factor linking the indicator to its construct, and δ_j, ε_j are the corresponding measurement errors. An indicator is kept in the model only if its loading exceeds 0.6; anything below that is dropped and the model is re-estimated [17].

Convergent validity is then checked through the Average Variance Extracted (AVE) in Eq. (3), where a construct is considered valid once AVE exceeds 0.50.

$$AVE = \frac{\sum_{i=1}^n \lambda_i^2}{(\sum_{i=1}^n \lambda_i^2) + (\sum_{i=1}^n \text{Var}(\varepsilon_i))} \quad (3)$$

with λ_i denoting the estimated loading of indicator i and $\text{Var}(\varepsilon_i)$ its measurement-error variance, reliability is checked through composite reliability in Eq. (4) and Cronbach's alpha in Eq. (5). Both should exceed 0.70 for a construct to be considered internally consistent [10].

$$\rho_c = \frac{(\sum_{i=1}^n \lambda_i)^2}{(\sum_{i=1}^n \lambda_i)^2 + \sum_{i=1}^n \text{Var}(\varepsilon_i)} \quad (4)$$

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n S_i^2}{S_t^2} \right) \quad (5)$$

where n is the number of indicators for a construct, S_t^2 is the total variance across those indicators, and S_i^2 is the variance of indicator i . The variance inflation factor in Eq. (6) is used to assess collinearity across indicators of the same construction and between exogenous constructions that predict the same endogenous construction; a VIF more than 10 indicates serious multicollinearity [8].

$$VIF_j = \frac{1}{1 - R_j^2} \quad (6)$$

The PLSpredict procedure [19] is used to determine predictive relevance through Q^2_{predict} , which is calculated for each endogenous construct in Eq. (7).

$$Q^2_{\text{predict}} = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y}_{\text{train}})^2} \quad (7)$$

where $\hat{Y}_i$ is the out-of-sample prediction for observation i from a model trained on the remaining folds and $\bar{Y}_{\text{train}}$ is the training-fold mean used as a naive benchmark; a construct's Q^2_{predict} exceeding zero indicates the model outperforms this benchmark out-of-sample. Moreover, the prediction accuracy at the indicator level was compared to that achieved using the linear regression model (LM) for the same folds, based on the criterion used by Shmueli et al [19], such that when the prediction error (RMSE) of the PLS-SEM is lower than that of the LM for all or most indicators of the construct, it is said to have good predictive power. This study employed 10-fold cross-validation with 10 repetitions and a fixed seed.

2.2.2. Structural Model (Inner Model)

The outer model clears validity and reliability; the structural relationship among the seven latent constructs is estimated in Eq. (8).

$$\eta_j = \sum_{i=1}^I \beta_{ji} \eta_i + \sum_{b=1}^B \gamma_{jb} \xi_b + \zeta_j \quad (8)$$

where β_{ji} is the path coefficient linking endogenous construct η_i to η_j , γ_{jb} is the path coefficient linking exogenous construct ξ_b to η_j , and ζ_j is the structural error term. Model fit is read from the coefficient of determination, R^2 , computed in Eq. (9).

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (9)$$

with Y_i the observed value of the endogenous construct, $\hat{Y}_i$ its model-predicted value, and $\bar{Y}$ its mean, an R^2 of 0.75, 0.50, and 0.25 is conventionally read as substantial, moderate, and weak, respectively [20]. Because PLS has no closed-form standard error, the significance of each path is instead tested through bootstrap resampling, with the t-statistic in Eq. (10).

$$T_{\text{value}} = \frac{b_j}{S(b_j)} \quad (10)$$

where b_j is the bootstrapped estimate of path coefficient γ_j and $S(b_j)$ is its bootstrap standard error. A path is treated as significant at the 5% level once the two-tailed p-value from 5,000 bootstrap resamples falls below 0.05 [11]. Specifically,

$$S(b_j) = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (b_j^{(b)} - \bar{b}_j)^2}$$

where B is the number of bootstrap resamples, $b_j^{(b)}$ is the estimate of path coefficient j from resample b , and $\bar{b}_j$ is the mean of the B bootstrap estimates; confidence intervals for each path coefficient are constructed using the bias-corrected percentile bootstrap method [19].

2.3. FIMIX-PLS Segmentation

The structural model above is fitted once across all 38 provinces, which silently assumes they behave as a single homogeneous population. FIMIX-PLS relaxes that assumption by treating the endogenous construct scores as drawn from a finite mixture of K latent segments, following Eq. (11),

$$\eta_m \sim \sum_{k=1}^K \rho_k f_{(m|k)}(\eta_m | \xi_m, B_k, \Gamma_k, \Psi_k) \quad (11)$$

where ρ_k is the mixing proportion of segment k ($\rho_k > 0$, $\sum_{k=1}^K \rho_k = 1$), and $f_{m|k}(\cdot)$ is the conditional multivariate normal density of case m given segment k , parameterized by the segment-specific path-coefficient matrices B_k and Γ_k and the residual-variance matrix Ψ_k [17, 21]. The mixture is estimated via an EM algorithm that alternates between computing each province's posterior segment-membership probability and re-estimating the segment-specific path coefficients until convergence [18].

Because the number of segments K is not known beforehand, several candidate values ($K = 2, 3, 4, \dots$) are compared using four criteria: the Akaike Information Criterion in Eq. (12), the Bayesian Information Criterion in Eq. (13), the Consistent AIC in Eq. (14), and the Normed Entropy in Eq. (15).

$$AIC_k = -2 \ln L + cN_k \quad (12)$$

$$BIC_k = -2 \ln L + \ln N_k \quad (13)$$

$$CAIC_k = -2 \ln L + (\ln(M) + 1) N_k \quad (14)$$

$$EN = 1 - \frac{\sum_{m=1}^M \sum_{k=1}^K P_{mk} \ln(P_{mk})}{M \ln(K)} \quad (15)$$

where L is the model likelihood, N_k is the number of free parameters for segment solution k , N is the number of observations, M is the number of provinces, K is the number of segments, and P_{mk} is the estimated probability that province m belongs to segment k . Lower AIC, BIC, and CAIC indicate a better-fitting segmentation, while EN closer to 1 indicates cleaner separation between segments; the K that best satisfies all four is retained as the final segmentation [22].

SmartPLS was used for the SEM-PLS estimation, the bootstrapping procedure, and the FIMIX-PLS segmentation, with 5,000 bootstrap subsamples for hypothesis testing, in line with standard practice in PLS-based stunting research [4, 10]. After finding the best value of K , each province is classified to the group having the highest P_{mk} value and the resulting groups are shown on a thematic map, each group corresponding to a color linked to the dominant structural pathway of each group. Therefore, the final segmentation is obtained by two concurrent criteria: the statistical criteria in Eqs. (12)–(15) for the selection of K and the possibility to substantially interpret the dominant path of each segment [22].

3. Results and Discussion

The data used in this study are secondary data obtained from the Ministry of Health of the Republic of Indonesia (Kemenkes RI), specifically the 2023–2024 health-profile publication, which reports several indicators tied to the causes of stunting across Indonesia's provinces. The dataset is publicly accessible through the ministry's official website, kemenkes.go.id. Table 1 lists the latent variables and indicators used in this study.

Table 1: Research variables and indicators

Latent variable	Indicator	Description
Antenatal Care	X_{11}	Percentage of pregnant women receiving antenatal care every trimester, four visits within one year, from a health worker (K4)
	X_{12}	Percentage of pregnant women receiving antenatal care six visits within one year, from a health worker (K6)
Sanitation	X_{21}	Percentage of households with access to proper sanitation
	X_{22}	Percentage of households with access to proper drinking-water facilities
	X_{23}	Percentage of villages practicing open-defecation-free sanitation
Immunization	X_{31}	Percentage of infants who received complete basic immunization
	X_{32}	Percentage of infants who received DPT/HB immunization
	X_{33}	Percentage of infants who received the polio vaccine
	X_{34}	Percentage of infants who received the measles vaccine
Early Initiation of Breastfeeding	X_{41}	Percentage of infants who began breastfeeding immediately after birth (IMD)
Exclusive Breast-feeding	Y_{11}	Percentage of infants who received exclusive breastfeeding for 6 months without any additional food or drink
Low Birth Weight	Y_{22}	Percentage of babies born with low birth weight ($< 2,500$ g)
Stunting	Y_{31}	Proportion of children with stunting
	Y_{32}	Proportion of children with wasting

3.1. Measurement Model Evaluation

The outer model was checked on two counts before any path got interpreted. Convergent validity came first, judged by each indicator's outer loading, kept only above 0.6, and by each construct's Average Variance Extracted (AVE), accepted once above 0.50. Reliability came second, checked through Cronbach's alpha, and composite reliability (ρ_c), all expected to clear 0.70 [17]. Any indicator that failed the loading cutoff was dropped and the model re-estimated before the structural stage began.

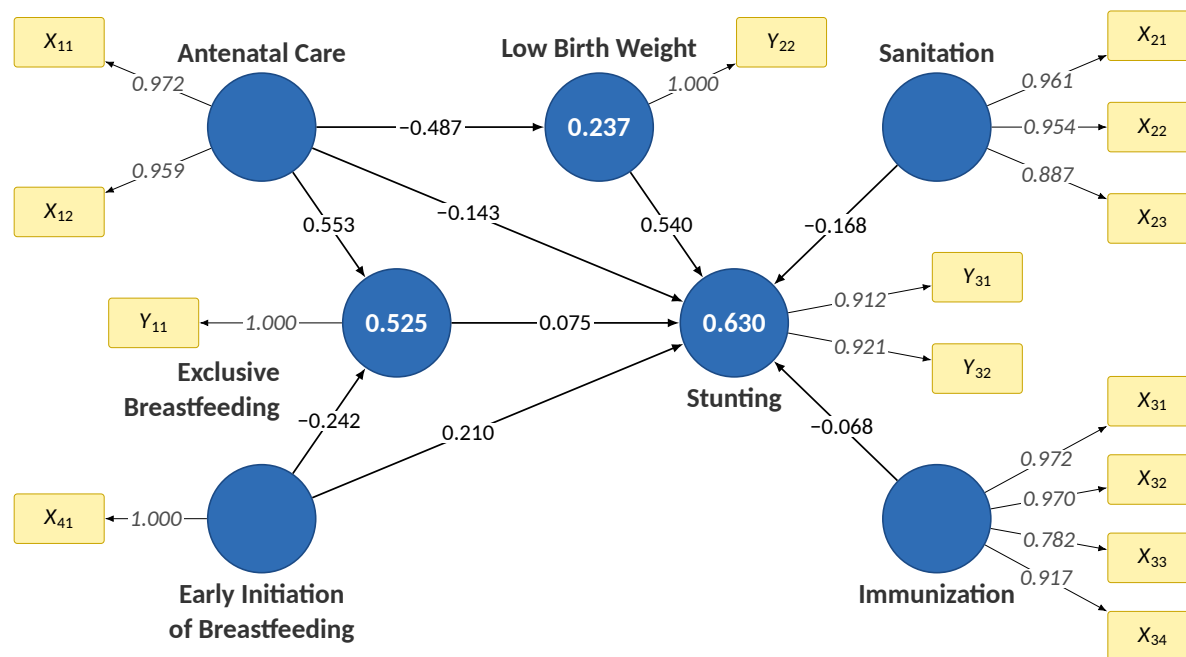


Fig. 1: Path diagram of the stunting SEM-PLS model.

Fig. 1 shows the path diagram that survived that cut. Antenatal care's two indicators, X_{11} and X_{12} , load at 0.972 and 0.959; sanitation's three indicators, X_{21} through X_{23} , sit between 0.887 and 0.961; immunization's four indicators, X_{31} through X_{34} , range from 0.782 to 0.972; stunting is carried by two indicators, Y_{31} and Y_{32} , loading at 0.912 and 0.921; and early initiation of breastfeeding, exclusive breastfeeding, and low birth weight are each measured by a single indicator loading at exactly 1.000, as expected for a single-item construct. Every retained indicator clears the 0.6 convergent-validity bar typically applied in SEM-PLS stunting research [4], [11].

One provincial notion, stunting, is illustrated by the incidence of wasting and stunting. Within the context of ecological analysis, this idea is linked to the overall state of malnutrition among children in each province rather than their specific medical issues. These indicators have high and almost equal loadings on the idea being discussed ($\lambda = 0.912$ for stunting and $\lambda = 0.921$ for wasting). They both show variation at the provincial level and are over the 0.60 cutoff level. This model choice is similar to that of Surayya and Sirait [11], who employ the prevalence of stunting and wasting as district-level indicators of malnutrition. This idea is called Stunting since it is the main focus of this essay, and wasting is a supplementary sign of starvation.

Instead of being two clinical outcomes of the same process in people, stunting and wasting are regarded here as two markers of a single provincial-level construct. In fact, those two physiological outcomes—acute wasting and chronic linear growth halting, respectively—should not be mistaken in a single instance. However, in the ecological setting under consideration, both stunting and wasting are indicators of the same deprivation environment: provinces with inadequate sanitary infrastructure, food insecurity, and poor health care systems are characterized by high rates of both stunting and wasting simultaneously; this is explained by the high and identical factor

loadings for both indicators as well as the acceptable AVE of the construct in question. This choice is made after observing Surayya and Sirait's [11] decision to employ this strategy when addressing undernutrition indicators in Aceh. We acknowledge that it will be fascinating to separate those two outcomes into distinct constructs, but we do not overlook the fact that the selected specification assumes the joint behavior of stunting and wasting at the provincial level.

Table 2 carries the validity and reliability check through to the construct level. Antenatal care, immunization, and sanitation are multi-item constructs, so their AVE, alpha, composite reliability (ρ_c) can all be computed; stunting is a two item construct and follows the same logic; early initiation of breastfeeding, exclusive breastfeeding, and low birth weight are single-item constructs, so their reliability is fixed by construction (loading = 1.000) and SmartPLS does not report a separate coefficient for them.

Table 2: Construct reliability and convergent validity

Construct	AVE	Cronbach's Alpha	Composite Reliability
Antenatal Care	0.932	0.928	0.965
Immunization	0.835	0.932	0.952
Sanitation	0.874	0.927	0.954
Stunting	0.840	0.810	0.913

All multi-item construct in Table 2 clears both bars at once: AVE runs from 0.835 to 0.932, well above then 0.50 floor, and Cronbach's Alpha, Composite Reliability all sit above 0.80, comfortably above the 0.70 floor. Stunting posts the lowest AVE and Cronbach's Alpha of the four, 0.840 and 0.810, which still leaves it well inside the acceptable range; it's simply the construct with the most indicator-to-indicator variation to explain, since Y_{31} and Y_{32} capture two related but not identical severity bands of the same outcome. The measurement model, in short gives no reason to doubt any construct's validity or reliability, which means the structural results that follow are being read off a fairly solid outer model.

Since reflective indicators are naturally highly linked, collinearity between indicators of the same construct was also investigated. For single-indicator structures, external VIF values varied from 1.00 to 14.78 for X_{31} . This outcome is usual for a reflective block and does not always point to an issue with the model specification. The Immunization construct is the source of the two indicators with the greatest external VIF, X_{31} (14.78) and X_{32} (14.30), whose factor loadings are almost the same (0.972 and 0.970).

This demonstrates that they measure the same attribute and are not a source of redundant information that needs to be removed. Collinearity between exogenous constructs that collectively predict the same endogenous construct is more important. Compared to the other five predictors of stunting, which varied from 1.70 to 4.04, the internal VIF value for antenatal care as a predictor of stunting was observed at 6.55, which was significantly higher. This illustrates how, at the province level, immunization, sanitation, and antenatal care are strongly correlated.

The route coefficients and their 95% confidence intervals—obtained by bootstrapping with bias correction and 5,000 resamples—were directly analyzed to determine whether this high VIF value was skewing the structural estimates. In contrast to other stunting predictors whose standard errors fell within a narrower range ($SE = 0.147$ – 0.279), the direct effect of antenatal care on stunting displayed a relatively wide confidence interval ($\beta = -0.143$; $SE = 0.345$; 95% CI $[-0.828; 0.555]$). This was indicative of variance inflation rather than the absence of a real effect. This construct mainly affected stunting indirectly: low birth weight was significantly associated with antenatal care ($\beta = -0.487$; $SE = 0.130$; 95% CI $[-0.690; -0.183]$). which in turn produced a significant specific indirect association ($\beta = -0.263$; $p = 0.022$) and significantly predicted stunting ($\beta = 0.540$; $SE = 0.147$; 95% CI $[0.265; 0.841]$). Therefore, rather than the direct pathway, whose estimate was less accurate, this mediation pathway was primarily responsible for the overall effect of antenatal care on stunting ($\beta = -0.365$).

The measurement model fulfills the recommended criteria of evaluation for reliability of

indicators, convergent validity, reliability of internal consistency, and the appropriate collinearity configuration. The out-of-sample prediction accuracy of the measurement model was further tested through the PLSpredict approach [19], where ten rounds of 10-fold cross-validation were performed. All three endogenous constructs showed positive $Q^2_{predict}$ indices (Exclusive breastfeeding = 0.433; Low birth weight = 0.151; Stunting = 0.203), and, as can be seen from Table 3, the PLS-SEM prediction error (RMSE) is smaller compared to the LM benchmark in all four manifest indicators as criterion for high prediction accuracy [19].

Table 3: Comparison of PLS-SEM and LM prediction error (RMSE) by indicator

Indicator	PLS-SEM RMSE	LM RMSE
Exclusive Breastfeeding	15.46	21.39
Low Birth Weight	1.86	2.30
Stunting	5.02	6.87
Wasting	2.25	2.51

Higher values of VIF in the inner model concerning antenatal care are related to higher confidence intervals of its direct effect on stunting because although the effect of antenatal care on stunting is mainly an indirect one, mediated by low-birth-weight indicator.

3.2. Structural Model Evaluation

With the outer model confirmed as valid and reliable, the structural relationships among the seven latent constructs in Eq. (8) were estimated next, and their explanatory power was read from the coefficient of determination (R^2) in Eq. (9). Table 4 reports R^2 for the three endogenous constructs: exclusive breastfeeding, low birth weight, and stunting. Following the conventional benchmarks of 0.75, 0.50, and 0.25 for substantial, moderate, and weak explanatory power, the model explains stunting and exclusive breastfeeding at a moderate level and low birth weight at a level closer to weak, meaning that a meaningful share of the variation across provinces in low birth weight still traces back to factors outside the four exogenous constructs considered here.

Table 4: Coefficient of determination (R^2) of the endogenous constructs

Endogenous Construct	R^2
Exclusive Breastfeeding	0.525
Low Birth Weight	0.237
Stunting	0.630

The model explains 63.0% of the variance in stunting ($R^2 = 0.630$), which lands in the moderate-to-substantial range; 52.5% in exclusive breastfeeding ($R^2 = 0.525$), also moderate; and 23.7% in low birth weight ($R^2 = 0.237$), which counts as weak on its own but still means almost a quarter of provincial variation in birth weight traces back to antenatal care alone. Table 5 describes the nine structural pathways in terms of the second criterion, now with exact t-statistics and p-values from the bootstrap, no longer integers read from the graph.

As there are no analytical expressions of the standard errors of PLS, the statistical significance of the paths was assessed using bootstrapping with 5,000 samples. In accordance with Eq. (8), the two-sided t-statistics and p-values of the parameters are reported in Table 5, along with the implied fit measures of the model. Model fit is satisfactory, as detailed further below [20].

Only three of the nine hypothesized paths are sustained at the 5 per cent level. Antenatal care pushes exclusive breastfeeding up significantly ($\beta = 0.553$, $t = 3.523$, $p < 0.001$), and it pulls low birth weight down just as significantly ($\beta = -0.487$, $t = 3.758$, $p < 0.001$). This tracks with what other research has already shown regular antenatal contact seems to lower low-birth-weight risk, mostly by keeping closer tabs on maternal nutrition and catching pregnancy complications sooner low birth weight, in turn, is associated with a significant increase in stunting ($\beta = 0.540$,

Table 5: Path coefficients and bootstrap significance testing

Structural path	Coefficient	Standard Error	t-statistic	p-value	Result ($\alpha = 0.05$)
Antenatal Care → Exclusive Breastfeeding	0.553	0.157	3.523	0.0004	Significant
Antenatal Care → Low Birth Weight	-0.487	0.130	3.758	0.0002	Significant
Antenatal Care → Stunting	-0.143	0.345	0.415	0.678	Not significant
Early Initiation of Breastfeeding → Exclusive Breastfeeding	-0.242	0.171	1.414	0.157	Not significant
Early Initiation of Breastfeeding → Stunting	0.210	0.174	1.207	0.228	Not significant
Exclusive Breastfeeding → Stunting	0.075	0.202	0.370	0.712	Not significant
Immunization → Stunting	-0.068	0.202	0.338	0.735	Not significant
Low Birth Weight → Stunting	0.540	0.147	3.662	0.0003	Significant
Sanitation → Stunting	-0.168	0.279	0.602	0.547	Not significant

$t = 3.662$, $p < 0.001$), which fits with earlier Indonesian work naming low birth weight as one of the steadier individual-level predictors of stunting [23].

Interesting although antenatal care, immunization, and sanitation are commonly regarded as key interventions for reducing stunting, none of these variables shows a statistically significant direct effect on stunting once the structural model is estimated simultaneously ($p = 0.678$, 0.735 , and 0.547 , respectively). Likewise, the direct effects of early initiation of breastfeeding and exclusive breastfeeding also fall well short of statistical significance. Therefore, the hypotheses proposing direct effects of early initiation of breastfeeding, immunization, sanitation, and exclusive breastfeeding on stunting were not supported in the pooled national model. Instead, these results in addition to the high VIF of 6.55 for antenatal care on the model internal level may imply that the association of these practices with stunting occurs indirectly via low birth weight and exclusive breastfeeding.

Two potential explanations stand out in connection with the results on breastfeeding. Early initiation of breastfeeding is a discrete variable that stands for a single event that happened during the first hour after delivery as opposed to an ongoing practice. Lack of association between early initiation and exclusive breastfeeding ($p = 0.157$) or stunting ($p = 0.228$) demonstrates that early initiation cannot be used as a predictor of future behaviors. A likely explanation is that the continuation of breastfeeding depends on factors such as maternity leave policies, family support, and exposure to infant formula marketing—factors not accounted for in this model. The absence of a significant relationship between exclusive breastfeeding and stunting ($p = 0.712$) is harder to reconcile with the protective effects observed at the individual level in other studies; however, at the provincial level, these effects may be mitigated by conditions prevailing after the age of six months. Two provinces with identical breastfeeding coverage rates could exhibit vastly different rates of stunting if the quality of complementary feeding differs—an aspect that falls outside the scope of the four exogenous constructs modeled here.

Immunization and sanitation operate according to interdependent yet distinct logics. The non-significant direct link observed for vaccination ($p = 0.735$) can be plausibly explained from a clinical perspective: vaccines primarily protect against morbidity and mortality associated with acute infectious diseases, whereas the link to chronic linear growth retardation involves a longer, indirect chain—where reduced infection frees up energy for growth—that is difficult to detect through simple cross-sectional analysis, especially after accounting for the shared variance among vaccination, prenatal care, and sanitation (as evidenced by the previously noted high VIF value of 6.55). The non-significant combined link regarding sanitation ($p = 0.547$) proves to be the model's most informative null result: demonstrates that the association between sanitation and growth retardation is strongly negative in a segment comprising more than half of the provinces ($\beta = -0.898$) but much weaker in the other regions; thus, the combined coefficient represents an average between a real, pronounced association in certain provinces and a negligible association in others. In other words, a non-significant combined link does not imply that the construct

lacks importance; on the contrary, it may reveal inter-provincial heterogeneity—the very type of variation the FIMIX-PLS method is designed to highlight—which serves as the basis for the subsequent segmentation analysis.

To test whether that explanation actually holds up, the specific indirect effects running through low birth weight and exclusive breastfeeding were bootstrapped separately, with results shown in Table 6.

Table 6: Specific indirect effects

Indirect path	Coefficient	t-statistic	p-value	Result ($\alpha = 0.05$)
Antenatal Care → Low Birth Weight → Stunting	-0.263	2.297	0.022	Significant
Antenatal Care → Exclusive Breastfeeding → Stunting	0.041	0.345	0.730	Not significant
Early Initiation of Breastfeeding → Exclusive Breastfeeding → Stunting	-0.018	0.294	0.769	Not significant

This describes a fully mediated relationship: antenatal care does not move stunting directly so much as it lowers the risk of low birth weight, and it is low birth weight that is most strongly associated with lower stunting. Combining the direct and indirect effects gives antenatal care a total effect on stunting of -0.365 larger in magnitude than its non-significant direct effect alone, and the clearest evidence that antenatal care matters for stunting chiefly by improving birth weight rather than through some other unmodeled channel. Immunization and sanitation, by contrast, showed neither a significant direct effect nor a route through either significant mediator, consistent with their comparatively small effect sizes ($f^2 = 0.004$ and 0.019 , respectively, against 0.464 for the low birth weight to stunting path) and suggesting their independent contribution to explaining cross-province variation in stunting is limited relative to birth weight.

The estimated structural model can be summarized by Eq. (16), Eq. (17), and Eq. (18), one equation for each endogenous construct.

$$\eta_1 = 0.553\xi_1 - 0.242\xi_2 + \zeta_1 \quad (16)$$

$$\eta_2 = -0.487\xi_1 + \zeta_2 \quad (17)$$

$$\eta_3 = -0.143\xi_1 + 0.210\xi_2 - 0.068\xi_3 - 0.168\xi_4 + 0.075\eta_1 + 0.540\eta_2 + \zeta_3 \quad (18)$$

Model fit was acceptable overall: SRMR for the estimated model was 0.080 (against 0.054 for the saturated model) and NFI was 0.719, both within ranges commonly treated as tolerable for a composite-based PLS model of this size, though the SRMR sits right at the conventional 0.08 ceiling and is worth flagging rather than glossing over. Two immunization indicators X_{31} and X_{32} , also carried a variance inflation factor above 10 (14.78 and 14.30, respectively), pointing to appreciable collinearity between BCG and measles coverage rates an unsurprising pairing, since provinces that vaccinate well against one disease tend to vaccinate well against the other, but a limitation worth naming when interpreting immunization's own non-significant path to stunting.

3.3. FIMIX-PLS Segmentation and Thematic Mapping

This structural model, which had already been constructed, combined all 38 provinces as a single population under the assumption that the path coefficients would be the same for all of them. The hypothesis has been tested directly through analyzing segmentation solutions in terms of $K = 2$ up to $K = 5$ using FIMIX-PLS Eqs. (12)–(15). All of the solutions were estimated using SmartPLS 4, running up to 5,000 iterations and performing 10 estimations for each of the values K , with the use of a fixed seed in order to avoid local optima problem. Following the research by Sarstedt et al. [21], CAIC and normalized entropy (EN) were chosen as the main model selection criteria for this analysis because of the ability to provide a more appropriate penalty for the model complexity compared to AIC and BIC.

Table 7 displays the model-selection statistics, log-likelihood values for each K , and the convergence behavior of the ten runs carried out for each candidate value of K , including AIC and BIC.

Table 7: FIMIX-PLS model selection criteria for $K = 2$ to $K = 5$

Criterion	$K = 2$	$K = 3$	$K = 4$	$K = 5$
Log-likelihood (LnL)	-103.646	-69.809	-50.385	-22.388
AIC	257.293	215.617	202.770	172.776
BIC	298.233	277.845	286.287	277.582
CAIC	323.233	315.845	337.287	341.582
EN	0.771	0.981	0.957	0.972

The best iteration for each of the 10 possible values of K , that is, the best iteration selected and kept for segmentation and with the highest log-likelihood value for each K , converged long before reaching 5,000 iterations. Therefore, no artificially truncated runs or convergence problems affected any of the four possible options. Considering each one separately, we may state that both the AIC and the BIC confirm that additional segments are always preferred: The AIC decreases monotonically until $K = 5$, and the BIC minimizes slightly, if at all for $K = 5$. Both these criteria possess a well-known property of finite mixture models: They ignore completely the penalty for model complexity in favor of fit improvement and select, accordingly, the maximum possible value of K , which is known to over-segment FIMIX-PLS models [24]. On the other hand, more reliable results are observed in the case of CAIC and EN – the primary criterion that has been used before.

CAIC reaches its minimum level at $K = 3$ (315.845) and increases at $K = 4$ (337.287) and $K = 5$ (341.582). On the other hand, EN achieves its maximum level at $K = 3$ (0.981) and has the highest level that is close to 1 compared to 0.957–0.972 for $K = 4$ –5 and 0.771 for $K = 2$ (poor separation). There is not a significant difference between two values according to BIC since $K = 3$ (277.845) and $K = 5$ (277.582) differ on two tenths of a point. Fit statistics do not provide sufficient grounds to determine the number of components since AIC and BIC indicate $K = 5$ whereas CAIC and EN suggest $K = 3$. $K = 3$ has been selected as the reference solution based on the previous major criterion with regard to the known tendency of AIC to over-segment finite-mix PLS models [24]. It should be mentioned that the decision about the number of segments has been made by CAIC or EN but not by all four indices.

To assess the practical ramifications of this decision, the major structural path that dominated each segment was analyzed in the solutions for $K = 3$, $K = 4$, and $K = 5$. Two out of three clusters identified in $K = 3$ solution are also represented by separate segments in $K = 4$ and $K = 5$ solutions. First of all, there is the segment defined by the low birth weight with the high positive correlation between the low-birth-weight-to-stunting variables (1.265 at $K = 3$; 1.131 at $K = 4$; 1.221 at $K = 5$). Another segment is characterized by the influence of sanitation on child health (represented by the high negative correlation between sanitation and stunting: -0.898 at $K = 3$; -1.293 at $K = 4$; -0.979 at $K = 5$). However, the third $K = 3$ segment, with high weights for breastfeeding-related variables, does not remain as one single segment beyond higher K . Of the eight segments that constitute the segment at $K = 4$, only three segments were from the segment characterized by breastfeeding coefficients at $K = 3$, while the other five segments were from the sanitation-dominated segment; they merge into a new segment where immunization and early initiation are both dominant ($\beta = 0.944$ and 0.991).

The sanitation and birth-weight-driven structures are therefore stable features of the data regardless of how finely the sample is partitioned, whereas the additional segments favored by AIC and BIC largely arise from splitting the smallest, least homogeneous $K = 3$ segment together with part of the sanitation-dominated group, rather than revealing new structure elsewhere. This three-cluster solution is thus the most parsimonious point at which such structure can be identified, with the breastfeeding cluster being the least stable of the three and the internal

structure of that cluster requiring careful interpretation.

There are 21 provinces (55.3%) in Segment 1, 10 provinces (26.3%) in Segment 2, and 7 provinces (18.4%) in Segment 3, using the allocation of each province to the most probable segment as determined from the highest posterior probability; the probability-weighted (fuzzy) segment size given by SmartPLS is 55.8%, 25.9%, and 18.3%, respectively, which are nearly identical to the discrete counts. The classification quality was excellent, as shown by the fact that in all 38 provinces, the mean highest posterior probability was 0.995 and each province was classified into their retained segment with a posterior probability more than 0.90 (35 out of 38 with posterior probability more than 0.99). There were three provinces — DKI Jakarta, Jawa Timur, and Maluku Utara — which had relatively lower posterior probability (91.5% - 95.0%), but still above the common 0.90 value for classification certainty. Table 8 provides the independent estimates of the path coefficient in each segment compared with the pooled national estimates. The basis of the analysis is based on just 38 province-level data points, and the nine-path structural model is re-estimated using sub-samples of 21, 10, and 7 provinces after segmentation, all far below the minimum number of samples needed for such a complex structural model [20]. This is especially true for Segment 3. Segment-specific standard errors, confidence intervals, and significance tests for these estimates cannot be constructed because they are meant to explain the structural pathway of each segment and not formulate any policy recommendations.

Table 8: Path coefficients relevant to each segment for the kept $K = 3$ solution

Path diagram	Original total effects	Segment 1 (n=21)	Segment 2 (n=10)	Segment 3 (n=7)
Antenatal care → Exclusive breast-feeding	0.553	0.808	0.615	-0.275
Antenatal care → Low birth weight	-0.487	-0.610	-0.746	0.328
Antenatal care → Stunting	-0.143	0.071	-0.027	0.018
Early initiation of breastfeeding → Exclusive breastfeeding	-0.242	0.239	-0.400	-1.068
Early initiation of breastfeeding → Stunting	0.210	-0.077	0.389	1.895
Exclusive breastfeeding → Stunting	0.075	0.786	0.190	0.985
Immunization → Stunting	-0.068	-0.604	0.155	0.315
Low birth weight → Stunting	0.540	0.505	1.265	-0.222
Sanitation → Stunting	-0.168	-0.898	0.365	0.019

The path coefficients that are independently determined for each segment are presented in Table 8 in conjunction with the comparison of the national path coefficients. The nine-path structural model was respecified using a sample of only 21, 10, and 7 provinces after the segmentation process, respectively. These samples are relatively small compared to the recommended size of the sample required for such structural models [20]. The basic data are based on only 38 observations per province.

A very strong negative relationship between sanitation and stunting (-0.898) exists in **Segment 1** (21 provinces), after which a strong positive relationship (0.786) between exclusive breastfeeding and stunting and a negative relationship between vaccination and stunting (-0.604) is seen. It seems that there is a clear indication that improved sanitation can reduce stunting in this segment significantly, in agreement with previous findings in Indonesia [13], [14]. However, it may be hard to understand the positive coefficient associated with exclusive breastfeeding and negative coefficient associated with vaccination as protection; rather, it may suggest reverse targeting of complementary feeding program and vaccinations in segments with higher prevalence of stunting.

The pattern most influenced by low birth weight is the one in **Segment 2** (10 provinces), where the correlation between low birth weight and stunting is 1.265, double the correlation in the overall model, and the influence of prenatal care on low birth weight is also strongest among the

three segments (-0.746). Consistent with the fact that birth weight was the dominant pathway through which the relationship between prenatal care and stunting works in the combined model [24], [23], this shows that antenatal care prevalence is more strongly related to birth weight in this segment than the other two. This is similar to other relationships observed within segments and remains a provincial relationship and is insufficient evidence to conclude beyond reasonable doubt that higher antenatal care is associated with higher birth weight in this particular province.

Rather than the expected positive association, however it is seen that **Segment 3** (7 provinces), which constitutes the smallest number of segments, has strong positive coefficients of stunting associated with both exclusive breastfeeding (0.985) and early initiation of breastfeeding (1.895). These coefficients cannot be used to make policy recommendations specific for the 7 provinces since the size of the segment is very small. The reason could be that provinces constituting this segment have traditionally high levels of stunting, and due to factors other than those captured in the model, breastfeeding and stunting indicators move together. Instead of further analyzing these data, it is essential to conduct an individual-level follow up to prove or disprove this reasoning. This is illustrated in Fig. 2. Table 9 identifies province membership in the cluster map.

Table 9: Province membership by FIMIX-PLS segment ($K = 3$)

Segment	Provinces
Segment 1	Sumatera Utara, Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Riau, Jawa Barat, D I Yogyakarta, Banten, Bali, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Papua Barat Daya, Papua Selatan
Segment 2	Aceh, Sumatera Barat, Bengkulu, DKI Jakarta, Jawa Timur, Kalimantan Timur, Sulawesi Tenggara, Sulawesi Barat, Papua Barat, Papua Tengah
Segment 3	Kep. Bangka Belitung, Jawa Tengah, Nusa Tenggara Barat, Maluku, Maluku Utara, Papua, Papua Pegunungan

Table 9 turns the abstract segment numbers into something a reader can place on a map: **Segment 1's** has 21 provinces run the length of the archipelago, from Sumatera Utara through Kalimantan and into parts of Sulawesi and Papua Barat Daya; **Segment 2's** 10 provinces include Aceh, DKI Jakarta, and Jawa Timur alongside Kalimantan Timur and Papua Tengah; and **Segment 3's** 7 provinces cluster around the eastern periphery, Maluku, Maluku Utara, Papua, and Papua Pegunungan, plus Kep. Bangka Belitung, Jawa Tengah, and Nusa Tenggara Barat. The reason that no island or region can be precisely categorized into a particular group is revealing in its own right: geographical closeness does not guarantee that a particular type of stunting pathway will be the dominant factor in the province.

The reason why Fig. 2 makes it possible to go further than statistical tables is that when the data for 38 provinces are combined and averaged, none of the relationships among sanitation, birth weight, and breastfeeding and stunting in the pooled model of Table 5 meets the 5 percent level of significance, so that any country-wide policy based only on the results from the pooled model would regard the three variables as being equally insignificant in all provinces. As shown by Fig. 2, however, there appears to be a difference in terms of the relationship between the three structural predictors of stunting when 38 provinces are examined separately. For example, in terms of the relationship, **Segment 1** (green) has the strongest association with sanitation, **Segment 2** (blue) with birth weight and antenatal care, and **Segment 3** (red) with breastfeeding predictors.

The segmentation does not prove that any particular intervention will be successful in a particular province, and is open to the ecological fallacy (a finding valid at the provincial level is not necessarily valid at the individual level), reverse causality (the provinces suffering from high levels of stunting could have higher levels of sanitation and immunization programming, rather than the other way around), and confounding due to unobservable differences between

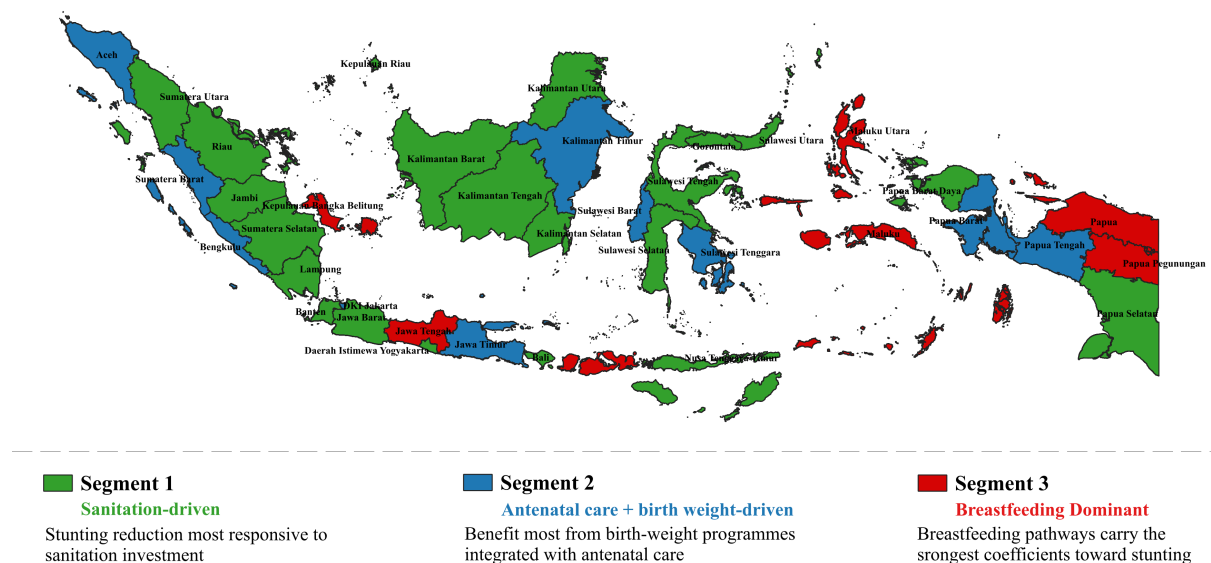


Fig. 2: Provincial thematic map of FIMIX-PLS segmentation in Indonesia. Data source: Analysis based on *Indonesia Health Profile 2023* and *Indonesia Health Profile 2024*, Ministry of Health of the Republic of Indonesia.

the provinces, including poverty, geographic isolation, or health care system capabilities. The segmentation is thus best interpreted as a tool for hypothesis generation regarding where a particular pathway may deserve further investigation at the individual level. The potential usefulness outlined in the Introduction is evident here: condensing nine structural pathways and three provincial segments into one figure usable by health authorities for the purposes of targeting further investigation, rather than assuming that sanitation, birth weight, and breastfeeding are always poor levers for change.

4. Conclusion

The research objective was to segment the Indonesian provinces based on the latent variables responsible for stunting, antenatal care, early breastfeeding initiation, immunization, sanitation, low birth weight, and exclusive breastfeeding through FIMIX-PLS approach, and to create a thematic map from the segmentation obtained. The pooled SEM-PLS model explains 63.0% of the variation in stunting, but only three out of nine theoretical hypotheses can be sustained through bootstrapping: antenatal care is associated with higher exclusive breastfeeding and lower low birth weight, while low birth weight is the only variable that has a direct influence on stunting. The antenatal care influence on stunting can only reach stunting by riding along that birth-weight channel, a fully mediated relationship rather than a direct effect, which is why its own path to stunting looks statistically silent in the pooled model despite being substantively important.

The silence ended up indicating a heterogeneity issue rather than irrelevance. By dividing the 38 provinces into three FIMIX-PLS segments, there was an increase in the explained variance in stunting from 0.63, when pooled, to 0.97 when considering the weighted variance for the three segments. Each segment relied on its own distinct lever that included sanitation for the largest segment, birth weight for the second segment, and breastfeeding for the smallest segment. Mapping those three segments shows the pattern does not simply follow the archipelago's geography, which is precisely why a province-by-province thematic map, rather than a blanket national program or a bare segment table, is the more useful output of an analysis like this one for provincial and district health planners.

Two limitations are worth naming. Segment 3's positive breastfeeding-to-stunting coefficients run counter to the expected protective direction and, given the segment's small size (7 provinces), are read here as a signal for targeted qualitative follow-up rather than a causal finding; and the cross-sectional design cannot rule out reverse targeting, where health programs are intensified precisely in provinces already facing high stunting. Future work could extend the model with household-level socioeconomic and dietary-diversity constructs and test whether these three segments persist against a later year of provincial data.

CRedit Authorship Contribution Statement

Ayu Setya Permatasari: Conceptualization, Methodology, Software, Formal Analysis, Writing of the Original Draft, Data Curation, and Visualization. **Rahmawati Erma Standsyah:** Conceptualization, Methodology, Validation, Writing – Review & Editing, Supervision, and Project Administration.

Declaration of Generative AI and AI-assisted technologies

Generative AI (Claude) was utilized exclusively for language editing, sentence flow enhancement, grammar refinement, and vocabulary suggestions. All authors have reviewed and approved the AI-assisted outputs and accept responsibility for the final manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Ethics Approval and Consent to Participate

This study did not require ethics approval.

Funding and Acknowledgments

This research did not obtain any external funding. The authors would like to thank the Ministry of Health of the Republic of Indonesia for supplying the official health-profile data used in this study, obtained through the 2023 Indonesia Health Survey and its accompanying provincial health-profile publications. The secondary data involved span the period 2023 to 2024 and relate to the province level across all 38 provinces of Indonesia, which are the units used for analysis.

Data and Code Availability

The dataset analyzed during the current study is available from the Indonesia Ministry of Health's official health profile publications and can be provided by the corresponding author upon reasonable request.

References

- [1] Trihono, Atmarita, Dwi Hapsari Tjandrarini, Anies Irawati, Nur Handayani Utami, Teti Tejayanti, and Iin Nurlinawati. *Pendek (Stunting) di Indonesia: Masalah dan Solusinya*. Jakarta: Lembaga Penerbit Balitbangkes, 2015. <https://www.badankebijakan.kemkes.go.id/repositori/id/eprint/3512/>.
- [2] Cristian Ricci, Janet Carboo, Hannah Asare, Cornelius M. Smuts, Robin Dolman, and Martani Lombard. "Nutritional Status as a Central Determinant of Child Mortality in Sub-Saharan Africa: A Quantitative Conceptual Framework". In: *Maternal & Child Nutrition* 15.2, e12722 (2019). DOI: [10.1111/mcn.12722](https://doi.org/10.1111/mcn.12722).

- [3] Aisyah Tri Mulyani, Miski Aghnia Khairinisa, Alfi Khatib, and Anis Yohana Chaerunisaa. “Understanding Stunting: Impact, Causes, and Strategy to Accelerate Stunting Reduction—A Narrative Review”. In: *Nutrients* 17.9, 1493 (2025). DOI: [10.3390/nu17091493](https://doi.org/10.3390/nu17091493).
- [4] Apri Linda Saputri and Nur’ainul Miftahul Huda. “Pengelompokan Provinsi Berdasarkan Faktor Penyebab Stunting Menggunakan Finite Mixture Partial Least Square (FIMIX-PLS)”. In: *BIMASTER: Buletin Ilmiah Matematika, Statistika dan Terapannya* 13.4 (2024). DOI: [10.26418/bbimst.v13i4.78039](https://doi.org/10.26418/bbimst.v13i4.78039).
- [5] Kementerian Kesehatan Republik Indonesia. *Profil Kesehatan Indonesia 2023*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2024. <https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2023>.
- [6] World Health Organization. *Nutrition Landscape Information System (NLiS) Country Profile Indicators: Interpretation Guide*. Geneva: World Health Organization, 2019. <https://www.who.int/publications/i/item/9789241516952>.
- [7] Badan Kebijakan Pembangunan Kesehatan. *Stunting di Indonesia dan Determinannya*. Factsheet Survei Kesehatan Indonesia 2023. Kementerian Kesehatan Republik Indonesia. 2024. <https://www.badankebijakan.kemkes.go.id/repositori/id/eprint/5535/>.
- [8] Muhammad Fikru Rizal and Eddy van Doorslaer. “Explaining the Fall of Socioeconomic Inequality in Childhood Stunting in Indonesia”. In: *SSM - Population Health* 9, 100469 (2019). DOI: [10.1016/j.ssmph.2019.100469](https://doi.org/10.1016/j.ssmph.2019.100469).
- [9] Laili Rahayuwati, Maria Komariah, Citra Windani Mambang Sari, Desy Indra Yani, Yanti Hermayanti, Arlette Setiawan, Hediati Hastuti, Sidik Maulana, and Kelvin Kohar. “The Influence of Mother’s Employment, Family Income, and Expenditure on Stunting Among Children Under Five: A Cross-Sectional Study in Indonesia”. In: *Journal of Multidisciplinary Healthcare* 16 (2023), pp. 2271–2278. DOI: [10.2147/JMDH.S417749](https://doi.org/10.2147/JMDH.S417749).
- [10] Gangga Anuraga and Izequela De Jesus Madeira. “Pengelompokan Stunting di Provinsi Nusa Tenggara Timur Menggunakan Finite Mixture Partial Least Square (FIMIX-PLS)”. In: *Mandalika Mathematics and Educations Journal* 7.2 (June 2025), pp. 598–610. DOI: [10.29303/jm.v7i2.9020](https://doi.org/10.29303/jm.v7i2.9020).
- [11] Ninda Nailis Surayya and Timbang Sirait. “Determinants Modeling of Undernutrition in Toddlers in Aceh Province: A PLS-SEM Approach”. In: *BAREKENG: Jurnal Ilmu Matematika dan Terapan* 18.3 (2024), pp. 1853–1864. DOI: [10.30598/barekengvol18iss3pp1853-1864](https://doi.org/10.30598/barekengvol18iss3pp1853-1864).
- [12] Lisa Cameron, Claire Chase, and Diana Contreras Suarez. “Relationship Between Water and Sanitation and Maternal Health: Evidence from Indonesia”. In: *World Development* 147, 105637 (2021). DOI: [10.1016/j.worlddev.2021.105637](https://doi.org/10.1016/j.worlddev.2021.105637).
- [13] Annita Olo, Henny Suzana Mediani, and Windy Rakhmawati. “Hubungan Faktor Air dan Sanitasi dengan Kejadian Stunting pada Balita di Indonesia”. In: *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini* 5.2 (2021), pp. 1113–1126. DOI: [10.31004/obsesi.v5i2.788](https://doi.org/10.31004/obsesi.v5i2.788).
- [14] Yuni R. Astuti. “Pengaruh Sanitasi dan Air Minum Terhadap Stunting di Papua dan Papua Barat”. In: *Poltekita: Jurnal Ilmu Kesehatan* 16.3 (Nov. 2022), pp. 261–267. DOI: [10.33860/jik.v16i3.1470](https://doi.org/10.33860/jik.v16i3.1470).
- [15] Rizky Yuspita Sari, Untoro Dwi Raharjo, and Siti Nurhayati. “Pemetaan Stunting Berdasarkan Imunisasi Dasar Lengkap dan Rumah Ber-PHBS di Kabupaten Kulon Progo”. In: *Jurnal Kesehatan Tambusai* 5.3 (2024), pp. 8341–8348. DOI: [10.31004/jkt.v5i3.33383](https://doi.org/10.31004/jkt.v5i3.33383).
- [16] Reni Eka Septiani and Mulyanto. “Relevance of Stunting and Economic Studies (Supported with Stunting Mapping Using K-Means Clustering Method in Indonesia)”. In: *International Journal of Economics, Business and Management Research* 6.6 (2022), pp. 21–30. DOI: [10.51505/ijebmr.2022.6703](https://doi.org/10.51505/ijebmr.2022.6703).

- [17] Carsten H. Hahn, Michael D. Johnson, Andreas Herrmann, and Frank Huber. “Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach”. In: *Schmalenbach Business Review* 54.3 (2002), pp. 243–269. DOI: [10.1007/BF03396655](https://doi.org/10.1007/BF03396655).
- [18] Christian M. Ringle, Sven Wende, and Alexander Will. “Finite Mixture Partial Least Squares Analysis: Methodology and Numerical Examples”. In: *Handbook of Partial Least Squares*. Berlin: Springer, 2010, pp. 195–218. DOI: [10.1007/978-3-540-32827-8_9](https://doi.org/10.1007/978-3-540-32827-8_9).
- [19] Galit Shmueli, Marko Sarstedt, Joseph F. Hair, Jun-Hwa Cheah, Hiram Ting, Santha Vaithilingam, and Christian M. Ringle. “Predictive Model Assessment in PLS-SEM: Guidelines for Using PLSpredict”. In: *European Journal of Marketing* 53.11 (2019), pp. 2322–2347. DOI: [10.1108/EJM-02-2019-0189](https://doi.org/10.1108/EJM-02-2019-0189).
- [20] Joseph F. Hair, G. Tomas M. Hult, Christian M. Ringle, and Marko Sarstedt. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd ed. Thousand Oaks, CA: SAGE Publications, 2017. <https://tore.tuhh.de/entities/publication/211e9b72-aa0c-41c2-883b-246f06e659ae>.
- [21] Marko Sarstedt and Christian M. Ringle. “Treating Unobserved Heterogeneity in PLS Path Modelling: A Comparison of FIMIX-PLS with Different Data Analysis Strategies”. In: *Journal of Applied Statistics* 37.8 (2010), pp. 1299–1318. DOI: [10.1080/02664760903030213](https://doi.org/10.1080/02664760903030213).
- [22] Marko Sarstedt, Jan-Michael Becker, Christian M. Ringle, and Manfred Schwaiger. “Uncovering and Treating Unobserved Heterogeneity with FIMIX-PLS: Which Model Selection Criterion Provides an Appropriate Number of Segments?” In: *Schmalenbach Business Review* 63.1 (2011), pp. 34–62. DOI: [10.1007/BF03396886](https://doi.org/10.1007/BF03396886).
- [23] Garedew Tadege Engdaw, Amensisa Hailu Tesfaye, Maru Feleke, Aragaw Negash, Amanuel Yesiwas, Wabiw Addis, Dessie Abebaw Angaw, and Melaku Tadege Engidaw. “Effect of Antenatal Care on Low Birth Weight: A Systematic Review and Meta-analysis in Africa, 2022”. In: *Frontiers in Public Health* 11, 1158809 (2023). DOI: [10.3389/fpubh.2023.1158809](https://doi.org/10.3389/fpubh.2023.1158809).
- [24] Ni Ketut Aryastami, Anuraj Shankar, Nunik Kusumawardani, Besral, Abas Basuni Jahari, and Endang Achadi. “Low Birth Weight Was the Most Dominant Predictor Associated with Stunting Among Children Aged 12–23 Months in Indonesia”. In: *BMC Nutrition* 3, 16 (2017). DOI: [10.1186/s40795-017-0130-x](https://doi.org/10.1186/s40795-017-0130-x).