Jurnal Varian
Vol. 7, No. 2, April 2024, Hal. 145–156
e-ISSN: 2581-2017

Stunting Prevalence Modeling Using Nonparametric Regression of Quadratic Splines

Tutik Handayani[1], **Sifriyani**[1], **Andrea Tri Rian Dani**[1]
[1]Department of Mathematics and Natural Science, Mulawarman University, Indonesia

Article Info

Article history:

Received : 04-16-2024
Revised : 05-15-2024
Accepted : 06-28-2024

Keywords:

Generalized Cross Validation;
Nonparametric Regression;
Quadratic Truncated Spline;
Stunting;
Unbiased Risk.

ABSTRACT

The nonparametric regression approach is used when the shape of the regression curve between the response variable and the predictor variable is assumed to be of unknown shape. The advantages of nonparametric regression have high flexibility. A nonparametric regression approach that is often used is truncated spline which has an excellent ability to handle data whose behavior changes at certain sub-sub intervals. The purpose of this study is to obtain the best model of multivariable nonparametric regression with linear and quadratic truncated spline approaches using the Generalized Cross Validation (GCV) and Unbiased Risk (UBR) methods and to find out the factors that influence the prevalence of stunting in Indonesia in 2021. The data used were the prevalence of stunting as a response variable and the predictor variable used was the percentage of infants receiving exclusive breastfeeding for 6 months, the percentage of households that have proper sanitation, the percentage of toddlers who get Early Initiation of Breastfeeding (IMD), the percentage of poor people, and the percentage of pregnant women at risk of SEZ. The results showed that the best quadratic truncated spline nonparametric regression model in modeling stunting prevalence was quadraic truncated spline using the GCV method with three knot points. This model has a minimum GCV value of 7.29 with an MSE value of 1.82 and a R2 value of 94.07%.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v7i2.2916

Corresponding Author:

Tutik Handayani,
Department of Mathematics and Natural Science, Mulawarman University
Email: tutik.handayani.ttk.110301@gmail.com

A. INTRODUCTION

Indonesia is a developing country, where the problem of poverty is still often found among the community even though economic development in Indonesia itself is classified as having increased rapidly. The problem of poverty that never has a solution will have a negative impact on several fields, one of which is the health sector. Nutritional problems are health problems that arise due to poverty which is characterized by many cases of malnutrition that occur in children under five, both in men and women. The early growth period of toddlers is very vulnerable to the surrounding environment. The growth of toddlers can also be affected by the nutritional intake obtained. Nutritional status is the state of the body caused by food consumption and nutrient use (Wijhati et al., 2021).

Nutritional status is one of the important factors that can determine human resources and quality of life (Desyanti and Nindya, 2017). Therefore, a nutrition improvement program is needed to improve the quality of nutrition for food consumption in order to improve nutritional status in the community. This can be interpreted that nutrition has an important role in the growth period of toddlers (Danaei et al., 2016).

Nutritional status is a condition caused by a balance between nutritional intake from food with nutritional needs needed for the body's metabolism (Liem et al., 2019). Everyone needs different nutritional intake between individuals, this depends on the person, gender, body activity in a matter of days, and body weight. A person's nutritional status indicates how much an individual's physiological needs have been met. A balance between incoming nutrients and nutrients needed for optimal health is important

(Rachmayani et al., 2018). When a person's nutritional needs are met to support the daily needs of the body and increase metabolic needs, then the individual will reach an optimal nutritional state (Indanah et al., 2022). Child growth is the main indicator for assessing nutritional status in children under five years old (da Silva et al., 2018) and s also one of the 6 global nutrition targets set by WHO in 2012 and is a key indicator of the Sustainable Development Goals (SDGs) of 2030 (Dewi et al., 2019). This can be interpreted that nutrition has an important role in the growth period of toddlers.

This is because children will tend to be more susceptible to infectious diseases, If at the age of toddler growth has been hampered it will adversely affect the child's future. The risk of obesity that occurs in children will be very high considering that short children have a low ideal body weight (Nahar et al., 2020). Weight gain alone can cause a child's Body Mass Index (BMI) to rise beyond the normal limit. This situation will continue for a long time until the risk of degenerative diseases occurs.

Sustainable Development Goals (SDGs), one of the goals in the health sector, states that by 2030, the target of community nutrition is to end all forms of malnutrition (Singh et al., 2017), including achieving the 2025 international target to reduce stunting and wasting in children under five, reducing stunting programs globally The government pays appropriate attention to stunting in children under the age of 2-3 years through the National and Internal Nutrition international movement, namely the Scaling Up Nutrition (SUN) movement with a concentration system to border areas (Yulianti et al., 2022).

Stunting is caused by various child development disorders and has affected 21.3% of children under the age of five globally with a total of 144 million cases (Schmidt, 2014). Some studies show that this condition is more common in Asian and African countries. The 2018 Basic Health Research (Riskesdas) and the 2019 Indonesian Toddler Nutrition Status Survey (SSGBI) reported a decrease in stunting rates from 30.8% to 27.7%. However, this is still a health problem because its prevalence is above the WHO standard of 20% (Latuihamallo et al., 2022).

Adequate nutrition, health conditions, protection, and safety factors play an important role in child development, especially at an early age (Remmers et al., 2014). The occurrence of stunting in this period can affect brain structure and function where the reduced number of cells causes growth delay. A survey by the Indonesian Ministry of Health revealed that 16% of children under five years old experienced impaired fine and gross motor development, hearing loss, decreased intelligence and speech delay with a total of 0.4 million cases (Sudfeld et al., 2015).

Some previous studies that conducted research on stunting prevalence cases that became a reference in this study, among others, research are; research on the variable of exclusive breastfeeding (ASI) has a relationship with the incidence of stunting (Hamzah et al., 2021). Research on the percentage of poverty and the percentage of children under five 0-59 months of malnutrition have a significant effect on the incidence of stunting in South Sulawesi in 2020 (Aswi and Sukarna, 2022). In other studies using truncated spline nonparametric regression include nonparametric spline truncated remotion modeling and its application to the crude birth rate in Surabaya (Budiantara, 2014), nonparametric spline truncated regression with knot point selection method generalized cross validation and unbiased risk (Handayani et al., 2023). So that this study has objectives, including obtaining a multivariable nonparametric regression model with a linear and quadratic truncated spline approach in the case of stunting prevalence in Indonesia in 2021 based on optimal knot points using the GCV method, obtaining multivariable nonparametric regression models with linear and quadratic truncated spline approaches in cases of stunting prevalence in Indonesia in 2021 based on optimal knot points using the UBR method and Obtaining the best multivariable nonparametric regression model with linear and quadratic truncated spline approaches in the case of stunting prevalence in Indonesia in 2021.

B. RESEARCH METHOD

The data used in this study is secondary data, which is data that is not collected directly by researchers, but rather through scientific works and government documents. The data used by researchers in conducting this study include the prevalence of stunting, the percentage of households that have access to proper sanitation, the percentage of toddlers who get IMD, the percentage of pregnant women at risk of SEZs obtained from the Ministry of Health of the Republic of Indonesia Indonesian Nutritional Status Study (SSGI). Meanwhile, data on the percentage of infants aged less than 6 months who get exclusive breastfeeding and the number of poor people. The data researchers obtained from the website of the Central Statistics Agency (BPS). The definition of each data that the researchers used is shown in Table 1.

Table 1. Research Variables

Variable Name	Variable Definition
Prevalence of Stunting Toddlers	The condition of growth failure in children under five years old (infants under five years old) due to chronic malnutrition so that the child is too short for his age.

Variable Name	Variable Definition
Percentage of Babies Getting Exclusive Breastfeeding for 6 Months	Babies who only get breast milk from birth to 6 months of age in one working area at a certain period of time.
Percentage of Households That Have Access to Proper Sanitation	Intentional behavior in the cultivation of clean living with the intention of preventing humans from coming into direct contact with dirt or other harmful waste materials in the hope that this effort will maintain and improve human health
Percentage of Toddlers Who Get Early Breastfeeding Initiation (IMD)	is the beginning of a mother giving breast milk to her baby when the baby is born into the world, namely in the first hours or 1 hour after giving birth.
Percentage of Poor People	The percentage of poor people who are below the poverty line.
Percentage of Pregnant Women at Risk of SEZ	SEZ is a condition where the mother experiences malnutrition.

Research steps can be written in each stage to achieve research objectives, so the steps taken are as follows:

1. Collecting secondary data on stunting prevalence cases in Indonesia in 2021 and the factors that influence it.
2. Conducting descriptive statistical analysis on response variables and predictor variables which aims to determine the characteristics of stunting prevalence cases in Indonesia in 2021.
3. Identify data patterns formed from stunting prevalence cases in Indonesia in 2021 with every factor that influences them using scatterplots.
4. Selecting the optimal knot point using the GCV method on quadratic truncated spline nonparametric regression.
5. Selecting the optimal knot point using the UBR method on quadratic truncated spline nonparametric regression.
6. Performs parameter estimation of truncated spline nonparametric regression models.
7. Comparing GCV and UBR methods in optimal knot point selection on quadratic truncated spline nonparametric regression.
8. Selects the best nonparametric spline truncated regression model with optimal knot points obtained from GCV and UBR methods based on MSE and R^2.

C. RESULT AND DISCUSSION

1. Descriptive Statistical Analysis

Before conducting the study, researchers wanted to display a general graph of stunting prevalence in each province in Indonesia. The goal is to be able to display the spread of stunting in every province in Indonesia. The graph of stunting prevalence can be seen in Figure 1.

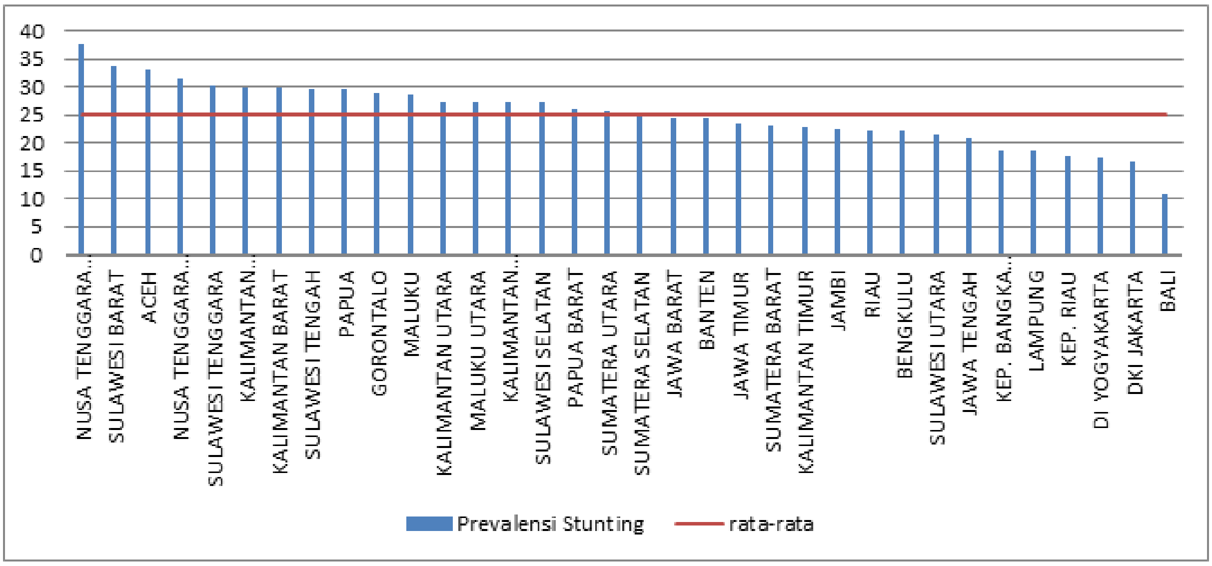

Figure 1. Graph of Stunting Prevalence in Indonesia in 2021

Based on Figure 1, we can see that the prevalence of stunting in Indonesia in 2021 has an average value of 25.21%. The prevalence of stunting can be said to be high if the percentage is above the average value, this shows that eastern Indonesia has a high value of stunting cases when compared to the western part of Indonesia. The provinces with the highest stunting cases are in East Nusa Tenggara Province with stunting cases of more than 35%, followed by West Sulawesi Province and Naggroe Aceh Darussalam Province where stunting cases are more than 30%. The province with the lowest stunting cases is in Bali Province with stunting cases of more than 10%. This data is based on data released by the Indonesian Central Bureau of Statistics (BPS) in 2021.

2. Respon Relationship Patterns Between Predictor Variables and Response Variables

After we look at the data graph of Stunting Prevalence in Indonesia in Figure 1. we can find out about what factors are suspected to be the main causes of stunting cases in Indonesia. In this study, we suspect there are 5 variables that affect the prevalence of stunting in Indonesia. The relationship between stunting prevalence and the five variables that are thought to have an effect can be seen in the scatterplot in Figure 2.

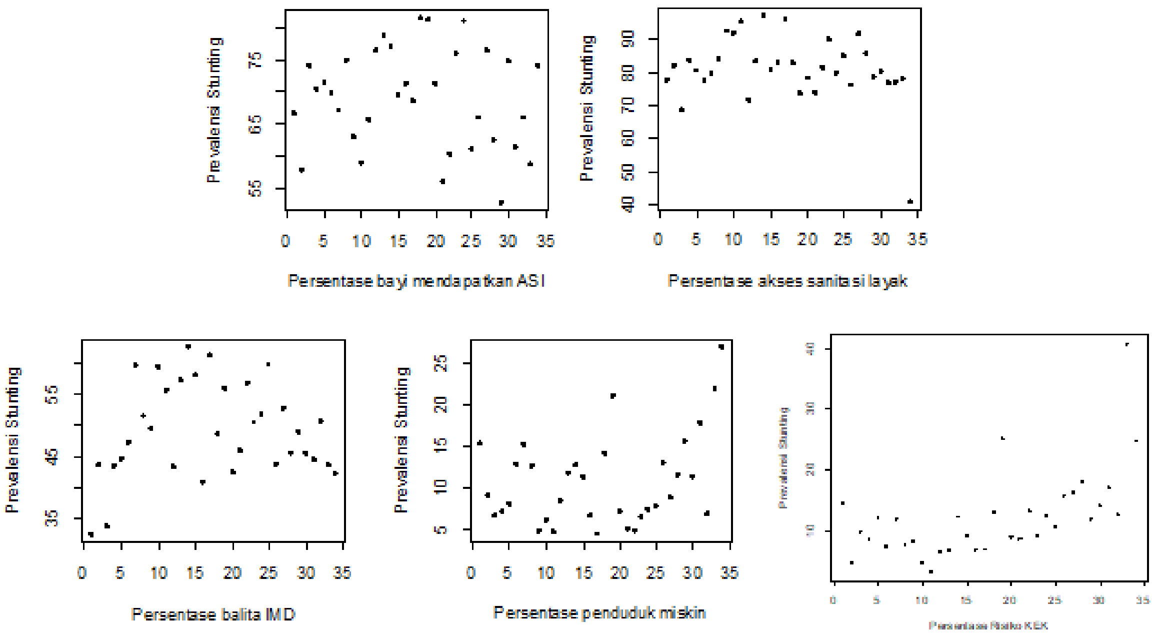

Figure 2. Scatterplot Between Response Variables and 5 Variables Suspected of Influence

Based on Figure 2, it can be seen that the distribution between the Stunting Percentage data and the five variables that are thought to have an effect does not form a certain pattern. The pattern of spreading data makes it difficult to identify how much influence variables are thought to affect the prevalence of stunting in Indonesia. So that the truncated spline nonparametric regression method will be used in order to estimate how much the suspected factor can affect the prevalence in Indonesia. This is because the shape of unknown data patterns is very suitable if using the truncated spline nonparametric regression method.

3. Selection of optimal knot points using GCV method

The first step performed before modeling using truncated spline nonparametric regression is to determine the number of knot points used. In this study, the knot point tried is 3 knots, after which one optimal knot, two optimal knots, and three optimal knots will be searched. Here is the selection of optimal knot points using the GCV method. The spline nonparametric regression model is truncated in stunting prevalence data in Indonesia in 2021 with one knot point shown in Equation (1) below.

$$\begin{aligned} y_i = &\beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - K_{11})_+^2 + \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - K_{21})_+^2 \\ &+ \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - K_{31})_+^2 + \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - K_{41})_+^2 \\ &+ \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - K_{51})_+^2 \end{aligned} \tag{1}$$

In this study in order to obtain optimal results, an experiment of 50 knot points was carried out to get optimal results. After an experiment using 50 knot points to obtain optimal knot points, the smallest 5 GCV values with one knot were obtained. The results of the 5 smallest GCV values are shown in Table 2 as follows.

Table 2. GCV Values with One Point Knots on Quadratic Truncated Spline Nonparametric Regression

Knot Point					GCV
x_1	x_2	x_3	x_4	x_5	
59,19	53,45	39,28	9,54	11,54	12,19
58,60	52,30	38,66	9,09	10,77	12,29
58,02	51,15	38,05	8,63	10,00	12,41
57,44	50,00	37,43	8,17	9,24	12,52
56,85	48,85	36,81	7,72	8,47	12,57

Table 2 shows the minimum GCV values for a multivariable nonparametric regression model on a quadratic truncated spline with a single knot point of 12.19. with the optimal knot point in variable X_1 which is 59.19, in variable X_2 which is 53.45, in variable X_3 which is 39.28, in variable X_4 which is 9.54, and in variable X_5 which is 11.54. The truncated spline nonparametric regression model uses a minimum GCV value with one knot written in Equation (2).

$$\begin{aligned}\widehat{y} = &\widehat{\beta}_0 + \widehat{\beta}_{11}x_1^2 + \widehat{\beta}_{12}x_1^2 + \widehat{\delta}_{13}(x_1 - 59,19)_+^2 + \widehat{\beta}_{21}x_2^2 + \widehat{\beta}_{22}x_2^2 + \widehat{\delta}_{23}(x_2 - 53,45)_+^2 \\ &+ \widehat{\beta}_{31}x_3^2 + \widehat{\beta}_{32}x_3^2 + \widehat{\delta}_{33}(x_3 - 39,28)_+^2 + \widehat{\beta}_{41}x_4^2 + \widehat{\beta}_{42}x_4^2 + \widehat{\delta}_{43}(x_4 - 9,54)_+^2 \\ &+ \widehat{\beta}_{51}x_5^2 + \widehat{\beta}_{52}x_5^2 + \widehat{\delta}_{53}(x_5 - 11,54)_+^2\end{aligned} \tag{2}$$

After obtaining the result of one knot point as shown in Equation (2). So next, the selection of knot points is carried out using two knot points. The spline nonparametric regression model was truncated in stunting prevalence data in Indonesia in 2021. The spline nonparametric regression model is truncated in stunting prevalence data in Indonesia in 2021 with two knot point shown in Equation (3) below.

$$\begin{aligned}y_i = &\beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - K_{11})_+^2 + \delta_{14}(x_1 - K_{12})_+^2 \\ &+ \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - K_{21})_+^2 + \delta_{24}(x_2 - K_{22})_+^2 \\ &+ \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - K_{31})_+^2 + \delta_{34}(x_3 - K_{32})_+^2 \\ &+ \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - K_{41})_+^2 + \delta_{44}(x_4 - K_{42})_+^2 \\ &+ \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - K_{51})_+^2 + \delta_{54}(x_5 - K_{52})_+^2\end{aligned} \tag{3}$$

In this study in order to obtain optimal results, an experiment of 50 knot points was carried out to get optimal results. After an experiment using 50 knot points to obtain optimal knot points, the smallest 5 GCV values with two knot were obtained. The results of the 5 smallest GCV values are shown in table 3 as follows.

Table 3. GCV Values with One Point Knots on Quadratic Truncated Spline Nonparametric Regression

Knot Point					**GCV**
x_1	x_2	x_3	x_4	x_5	
54,51	44,26	34,35	5,88	5,40	11,90
73,84	82,18	54,69	20,93	30,72	
53,92	43,11	33,73	5,44	4,63	11,93
73,84	82,18	54,69	20,93	30,72	
53,33	41,96	33,12	4,98	3,87	11,93
73,84	82,18	54,69	20,93	30,72	
54,51	44,26	34,35	5,90	5,40	11,94
73,26	81,03	54,07	20,48	29,95	
70,91	76,43	51,61	18,66	26,89	11,95
72,08	78,73	52,84	19,57	28,42	

Table 3 shows the minimum GCV values for a multivariable nonparametric regression model on a quadratic truncated spline with a two knots of 11.90. with optimal knot points in variable X_1 which are 54.51 and 73.84, in variable X_2 which is 44.26 and 82.18, in variable X_3 which is 34.35 and 54.69, in variable X_4 which is 5.88 and 20.93, and in variable X_5 which is 5.40 and 30.72. A truncated spline nonparametric regression model using a minimum GCV value with two knots can be written in equation (4).

$$\begin{aligned}\widehat{y}_i = &\widehat{\beta}_0 + \widehat{\beta}_{11}x_1 + \widehat{\beta}_{12}x_1^2 + \widehat{\delta}_{13}(x_1 - 54,51)_+^2 + \widehat{\delta}_{14}(x_1 - 73,84)_+^2 \\ &+ \widehat{\beta}_{21}x_2 + \widehat{\beta}_{22}x_2^2 + \widehat{\delta}_{23}(x_2 - 44,26)_+^2 + \widehat{\delta}_{24}(x_2 - 82,18)_+^2 \\ &+ \widehat{\beta}_{31}x_3 + \widehat{\beta}_{32}x_3^2 + \widehat{\delta}_{33}(x_3 - 34,35)_+^2 + \widehat{\delta}_{34}(x_3 - 54,69)_+^2 \\ &+ \widehat{\beta}_{41}x_4 + \widehat{\beta}_{42}x_4^2 + \widehat{\delta}_{43}(x_4 - 5,88)_+^2 + \widehat{\delta}_{44}(x_4 - 20,93)_+^2 \\ &+ \widehat{\beta}_{51}x_5 + \widehat{\beta}_{52}x_5^2 + \widehat{\delta}_{53}(x_5 - 5,40)_+^2 + \widehat{\delta}_{54}(x_5 - 30,72)_+^2\end{aligned} \tag{4}$$

After obtaining the result of one knot point as shown in equation (4). So next, the selection of knot points is carried out using two knot points. The spline nonparametric regression model was truncated in stunting prevalence data in Indonesia in 2021. The spline nonparametric regression model is truncated in stunting prevalence data in Indonesia in 2021 with two knot point shown in equation (5) below.

$$\begin{aligned}y_i = &\beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - K_{11})_+^2 + \delta_{14}(x_1 - K_{13})_+^2 + \delta_{15}(x_1 - K_{12})_+^2 \\ &+ \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - K_{21})_+^2 + \delta_{24}(x_2 - K_{22})_+^2 + \delta_{25}(x_2 - K_{23})_+^2 \\ &+ \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - K_{31})_+^2 + \delta_{34}(x_3 - K_{32})_+^2 + \delta_{35}(x_3 - K_{33})_+^2 \\ &+ \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - K_{41})_+^2 + \delta_{44}(x_4 - K_{42})_+^2 + \delta_{45}(x_4 - K_{43})_+^2 \\ &+ \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - K_{51})_+^2 + \delta_{54}(x_5 - K_{52})_+^2 + \delta_{55}(x_5 - K_{53})_+^2\end{aligned} \tag{5}$$

In this study in order to obtain optimal results, an experiment of 50 knot points was carried out to get optimal results. After an experiment using 50 knot points to obtain optimal knot points, the smallest 5 GCV values with three knot were obtained. The results of the 5 smallest GCV values are shown in table 4 as follows.

Table 4. GCV Values with three Point Knots on Quadratic Truncated Spline Nonparametric Regression

Knot Point					GCV
x_1	x_2	x_3	x_4	x_5	
70,91	76,43	51,61	18,66	26,88	
72,08	78,73	52,84	19,57	28,42	8,03
80,29	94,82	61,47	25,95	39,16	
71,50	77,58	52,22	19,11	27,65	
72,08	78,73	52,84	19,57	28,42	8,19
80,29	94,82	61,47	25,95	39,16	
70,91	76,43	51,61	18,66	26,89	
72,08	78,73	52,84	19,57	28,42	8,54
80,29	93,67	60,85	25,49	38,40	
70,91	76,43	51,61	18,66	26,89	
71,50	77,58	52,22	19,11	27,65	8,72
80,29	94,82	61,47	25,95	39,16	
71,50	77,58	52,22	19,11	27,65	
72,08	78,73	52,84	19,57	28,42	8,81
79,70	93,67	60,85	25,49	38,40	

Table 4 shows the minimum GCV values for a multivariable nonparametric regression model on a quadratic truncated spline with a three knots of 8.03. with the optimal knot point in variable X_1 of 70.91; 72,08; and 80.29, on variable X_2 which is 76.43; 78,73; and 94.82, on variable X_3 which is 51.61; 52,84; and 61.47, on the variable X_4 which is 18.66; 19,57; and 25.95, and on the variable X_5 which is 26.88; 28,42; and 39.16. The truncated spline nonparametric regression model uses a minimum GCV value with three knots written in equation (6).

$$
\begin{aligned}
y_i = &\widehat{\beta}_0 + \widehat{\beta}_{11}x_1 + \widehat{\beta}_{12}x_1^2 + \widehat{\delta}_{13}(x_1 - 70,91)_+^2 + \widehat{\delta}_{14}(x_1 - 72,08)_+^2 + \widehat{\delta}_{15}(x_1 - K_{12})_+^2 \\
&+ \widehat{\beta}_{21}x_2 + \widehat{\beta}_{22}x_2^2 + \widehat{\delta}_{23}(x_2 - 76,43)_+^2 + \widehat{\delta}_{24}(x_2 - 78,73)_+^2 + \widehat{\delta}_{25}(x_2 - 94,82)_+^2 \\
&+ \widehat{\beta}_{31}x_3 + \widehat{\beta}_{32}x_3^2 + \widehat{\delta}_{33}(x_3 - 51,61)_+^2 + \widehat{\delta}_{34}(x_3 - 52,84)_+^2 + \widehat{\delta}_{35}(x_3 - 61,47)_+^2 \\
&+ \widehat{\beta}_{41}x_4 + \widehat{\beta}_{42}x_4^2 + \widehat{\delta}_{43}(x_4 - 18,66)_+^2 + \widehat{\delta}_{44}(x_4 - 19,57)_+^2 + \widehat{\delta}_{45}(x_4 - 25,95)_+^2 \\
&+ \widehat{\beta}_{51}x_5 + \widehat{\beta}_{52}x_5^2 + \widehat{\delta}_{53}(x_5 - 26,89)_+^2 + \widehat{\delta}_{54}(x_5 - 28,42)_+^2 + \widehat{\delta}_{55}(x_5 - 39,16)_+^2
\end{aligned} \tag{6}
$$

After testing with 50 knot points in the data, the results of selecting optimal knot points were obtained using the GCV method. The GCV values with one knot point, two knot points, and three knot points can be seen in Table 5.

Table 5. GCV Method Knot Point Comparison

Knot Point	GCV Value
1 knot point	12,19
2 knot point	11,90
3 knot point	8,03

Table 5 shows a minimum GCV value of 8.03 in a model with three knot points. This value is the smallest of the three knot points used so it becomes a suitable model for use. The parameter estimation of truncated spline nonparametric regression model using GCV method with three knot points shown in equation (7) below.

$$
\begin{aligned}
\widehat{y} = &0,77 - 1,71x_1 + 0,01x_1^2 + 1,70(x_1 - 70,91)_+^2 - 1,09(x_1 - 72,08)_+^2 - 1,13(x_1 - 80,29)_+^2 \\
&+ 2,22x_2 - 0,01x_2^2 - 0,84(x_2 - 76,43)_+^2 + 2,08(x_2 - 78,73)_+^2 - 2,23(x_2 - 94,82)_+^2 \\
&+ 0,00x_3 - 0,00x_3^2 + 0,00(x_3 - 51,61)_+^2 - 0,26(x_3 - 52,84)_+^2 - 0,07(x_3 - 94,82)_+^2 \\
&+ 0,61x_4 - 0,03x_4^2 + 7,40(x_4 - 18,66)_+^2 + 3,29(x_4 - 19,57)_+^2 + 0,82(x_4 - 25,95)_+^2 \\
&+ 0,03x_5 + 0,03x_5^2 - 6,71(x_5 - 26,89)_+^2 - 2,98(x_5 - 28,42)_+^2 - 0,75(x_5 - 39,16)_+^2
\end{aligned} \tag{7}
$$

From the nonparametric regression model, the spline is truncated with three knot points in equation (7) The coefficient of determination or R^2 value is 0.9557. This shows that the model can explain the factors that influence the stunting prevalence data in Indonesia in 2021 of 95.57%. After testing using the GCV method, it will be continued with the UBR method so that later the results can be compared.

4. Selection of optimal knot points using UBR method

The initial step before modeling using truncated spline nonparametric regression is to determine the number of knot points used. In this study, the knot points tried were 3 knot points, after that one optimal knot, two optimal knots, and three optimal knots would be searched. Here is the selection of optimal knot points using the UBR method. The truncated spline nonparametric regression model on stunting prevalence data in Indonesia in 2021 with one knot point shown in equation (8) below.

$$
\begin{aligned}
y_i = &\beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - K_{11})_+^2 + \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - K_{21})_+^2 \\
&+ \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - K_{31})_+^2 + \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - K_{41})_+^2 \\
&+ \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - K_{51})_+^2
\end{aligned} \tag{8}
$$

In this study in order to obtain optimal results, an experiment of 50 knot points was carried out to get optimal results. After an experiment using 50 knot points to obtain optimal knot points, the smallest 5 UBR values with one knot were obtained. The results of the 5 smallest UBR values are shown in table 6 as follows.

Table 6. UBR Values with one Point Knots on Quadratic Truncated Spline Nonparametric Regression

Knot Point					σ^2	UBR
x_1	x_2	x_3	x_4	x_5		
53,92	43,11	33,73	5,44	4,63	$1,45 \times 10^{-11}$	$1,00 \times 10^{-11}$
58,61	52,30	38,66	9,09	10,77	$1,81 \times 10^{-11}$	$1,15 \times 10^{-11}$
53,33	41,96	33,11	4,98	3,87	$1,73 \times 10^{-11}$	$1,20 \times 10^{-11}$
58,02	51,15	38,05	8,63	10,00	$2,15 \times 10^{-11}$	$1,36 \times 10^{-11}$
59,19	53,45	39,28	9,54	11,54	$2,19 \times 10^{-11}$	$1,39 \times 10^{-11}$

Table 6 shows the minimum UBR values for a multivariable nonparametric regression model on a linear truncated spline with a single knot of . with the optimal knot point in the variable X_1 which is 53.92, in the variable X_2 which is 43.11, in the variable X_3 which is 33.73, in the variable X_4 which is 5.44, and in the variable X_5 which is 4.63. The truncated spline nonparametric regression model uses a minimum UBR value with one knot inscribed in equation (9).

$$\begin{aligned}\widehat{y}_i = \beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - 53,92)_+^2 + \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - 43,11)_+^2 \\ + \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - 33,73)_+^2 + \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - 5,44)_+^2 \\ + \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - 4,63)_+^2\end{aligned} \tag{9}$$

Next, the selection of knot points was carried out using two knot points. The truncated spline nonparametric regression model on stunting prevalence data in Indonesia in 2021 with two knot points shown in equation (10) below.

$$\begin{aligned}y_i = \beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - K_{11})_+^2 + \delta_{14}(x_1 - K_{12})_+^2 \\ + \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - K_{21})_+^2 + \delta_{24}(x_2 - K_{22})_+^2 \\ + \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - K_{31})_+^2 + \delta_{34}(x_3 - K_{32})_+^2 \\ + \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - K_{41})_+^2 + \delta_{44}(x_4 - K_{42})_+^2 \\ + \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - K_{51})_+^2 + \delta_{54}(x_5 - K_{52})_+^2\end{aligned} \tag{10}$$

In this study in order to obtain optimal results, an experiment of 50 knot points was carried out to get optimal results. After an experiment using 50 knot points to obtain optimal knot points, the smallest 5 UBR values with two knot were obtained. The results of the 5 smallest UBR values are shown in table 7 as follows.

Table 7. UBR Values with one Point Knots on Quadratic Truncated Spline Nonparametric Regression

Knot Point					σ^2	UBR
x_1	x_2	x_3	x_4	x_5		
53,92 80,87	43,11 95,97	33,73 62,08	5,44 26,40	4,63 39,93	$1,86 \times 10^{-11}$	$1,24 \times 10^{-11}$
58,02 80,87	51,15 95,97	38,05 62,08	8,63 26,40	10,00 39,93	$2,17 \times 10^{-11}$	$1,31 \times 10^{-11}$
53,33 80,87	41,96 95,97	33,12 62,08	4,98 26,40	3,87 39,93	$2,19 \times 10^{-11}$	$1,46 \times 10^{-11}$
58,61 80,87	52,30 95,97	38,66 62,08	9,09 26,40	10,77 39,93	$2,43 \times 10^{-11}$	$1,47 \times 10^{-11}$
53,33 80,29	41,96 94,82	33,12 61,47	4,98 25,95	3,87 39,16	$2,45 \times 10^{-11}$	$1,48 \times 10^{-11}$

Table 7 shows the minimum GCV values for a multivariable nonparametric regression model on a quadratic truncated spline with a two knots of $1,2410^{-11}$ with optimal knot points on variable X_1 which are 53.92 and 80.87, on variable X_2 which are 43.11 and 95.97, on variable X_3 which are 33.73 and 62.08, on variable X_4 which are 5.44 and 26.40, and on variable X_5 which are 4.63 and 39.93. A truncated spline nonparametric regression model using a minimum GCV value with two knots can be written in equation (11).

$$\begin{aligned}\widehat{y}_i = &\widehat{\beta}_0 + \widehat{\beta}_{11}x_1 + \widehat{\beta}_{12}x_1^2 + \widehat{\delta}_{13}(x_1 - 53,92)_+^2 + \widehat{\delta}_{14}(x_1 - 80,87)_+^2 \\ &+ \widehat{\beta}_{21}x_2 + \widehat{\beta}_{22}x_2^2 + \widehat{\delta}_{23}(x_2 - 43,11)_+^2 + \widehat{\delta}_{24}(x_2 - 95,97)_+^2 \\ &+ \widehat{\beta}_{31}x_3 + \widehat{\beta}_{32}x_3^2 + \widehat{\delta}_{33}(x_3 - 33,73)_+^2 + \widehat{\delta}_{34}(x_3 - 62,08)_+^2 \\ &+ \widehat{\beta}_{41}x_4 + \widehat{\beta}_{42}x_4^2 + \widehat{\delta}_{43}(x_4 - 5,44)_+^2 + \widehat{\delta}_{44}(x_4 - 26,40)_+^2 \\ &+ \widehat{\beta}_{51}x_5 + \widehat{\beta}_{52}x_5^2 + \widehat{\delta}_{53}(x_5 - 4,63)_+^2 + \widehat{\delta}_{54}(x_5 - 39,93)_+^2\end{aligned} \tag{11}$$

After obtaining the result of one knot point as shown in equation (11). So next, the selection of knot points is carried out using two knot points. The spline nonparametric regression model was truncated in stunting prevalence data in Indonesia in 2021. The spline nonparametric regression model is truncated in stunting prevalence data in Indonesia in 2021 with two knot point shown in equation (12) below.

$$\begin{aligned}y_i = &\beta_0 + \beta_{11}x_1 + \beta_{12}x_1^2 + \delta_{13}(x_1 - K_{11})_+^2 + \delta_{14}(x_1 - K_{12})_+^2 + \delta_{15}(x_1 - K_{13})_+^2 \\ &+ \beta_{21}x_2 + \beta_{22}x_2^2 + \delta_{23}(x_2 - K_{21})_+^2 + \delta_{24}(x_2 - K_{22})_+^2 + \delta_{25}(x_2 - K_{23})_+^2 \\ &+ \beta_{31}x_3 + \beta_{32}x_3^2 + \delta_{33}(x_3 - K_{31})_+^2 + \delta_{34}(x_3 - K_{32})_+^2 + \delta_{35}(x_3 - K_{33})_+^2 \\ &+ \beta_{41}x_4 + \beta_{42}x_4^2 + \delta_{43}(x_4 - K_{41})_+^2 + \delta_{44}(x_4 - K_{42})_+^2 + \delta_{45}(x_4 - K_{43})_+^2 \\ &+ \beta_{51}x_5 + \beta_{52}x_5^2 + \delta_{53}(x_5 - K_{51})_+^2 + \delta_{54}(x_5 - K_{52})_+^2 + \delta_{55}(x_5 - K_{53})_+^2\end{aligned} \tag{12}$$

In this study in order to obtain optimal results, an experiment of 50 knot points was carried out to get optimal results. After an experiment using 50 knot points to obtain optimal knot points, the smallest 5 UBR values with three knot were obtained. The results of the 5 smallest UBR values are shown in table 4 as follows.

Table 8. UBR Values with one Point Knots on Quadratic Truncated Spline Nonparametric Regression

Knot Point					σ^2	UBR
x_1	x_2	x_3	x_4	x_5		
61,54	58,05	41,74	11,36	14,61		
68,57	71,84	49,14	16,83	23,82	$2,98 \times 10^{-11}$	$1,28 \times 10^{-11}$
69,74	74,14	50,37	17,74	25,35		
53,92	43,11	33,73	5,44	4,63		
80,29	94,82	61,47	25,95	39,16	$2,63 \times 10^{-11}$	$1,52 \times 10^{-11}$
80,87	95,97	62,08	26,40	39,93		
53,33	41,96	33,12	4,98	3,87		
79,70	93,67	60,85	25,49	38,40	$2,73 \times 10^{-11}$	$1,58 \times 10^{-11}$
80,87	95,97	62,08	26,40	39,93		
59,19	53,45	39,28	9,54	11,54		
79,70	93,67	60,85	25,49	38,40	$3,33 \times 10^{-11}$	$1,73 \times 10^{-11}$
80,29	94,82	62,08	25,95	39,16		
54,51	44,26	34,35	5,90	5,40		
67,40	69,54	47,91	15,92	22,28	$3,79 \times 10^{-11}$	$1,74 \times 10^{-11}$
72,67	79,88	53,45	20,02	29,18		

Table 8 shows the minimum GCV values for a multivariable nonparametric regression model on a quadratic truncated spline with a three knots of $1,5210^{-11}$. with the optimal knot point on the variable X_1 which is 61.54; 68,57; and 69.74, on variable X_2 which is 58.05; 71,84; and 74.14, on variable X_3 which is 41.74; 49,14; and 50.37, on the variable X_4 which is 11.36; 16,83; and 17.74, and on the variable X_5 which is 14.61; 23,82; and 25.35. The truncated spline nonparametric regression model uses a minimum UBR value with three knots written on equation (13).

DOI: https://doi.org/10.30812/varian.v7i2.2916

$$\begin{aligned}\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_{11}x_1 + \widehat{\beta}_{12}x_1^2 + \widehat{\delta}_{13}(x_1 - 61,54)_+^2 + \widehat{\delta}_{14}(x_1 - 68,57)_+^2 + \widehat{\delta}_{15}(x_1 - 69,74)_+^2 \\ + \widehat{\beta}_{21}x_2 + \widehat{\beta}_{22}x_2^2 + \widehat{\delta}_{23}(x_2 - 58,05)_+^2 + \widehat{\delta}_{24}(x_2 - 71,84)_+^2 + \widehat{\delta}_{25}(x_2 - 74,14)_+^2 \\ + \widehat{\beta}_{31}x_3 + \widehat{\beta}_{32}x_3^2 + \widehat{\delta}_{33}(x_3 - 41,74)_+^2 + \widehat{\delta}_{34}(x_3 - 49,14)_+^2 + \widehat{\delta}_{35}(x_3 - 50,37)_+^2 \\ + \widehat{\beta}_{41}x_4 + \widehat{\beta}_{42}x_4^2 + \widehat{\delta}_{43}(x_4 - 11,36)_+^2 + \widehat{\delta}_{44}(x_4 - 16,83)_+^2 + \widehat{\delta}_{45}(x_4 - 17,74)_+^2 \\ + \widehat{\beta}_{51}x_5 + \widehat{\beta}_{52}x_5^2 + \widehat{\delta}_{53}(x_5 - 14,61)_+^2 + \widehat{\delta}_{54}(x_5 - 23,82)_+^2 + \widehat{\delta}_{55}(x_5 - 25,35)_+^2\end{aligned} \quad (13)$$

After testing with 50 knot points in the data, the results of selecting optimal knot points were obtained using the GCV method. The UBR values with one knot point, two knot points, and three knot points can be seen in Table 4.

Table 9. Knot Point Comparison UBR Method

Knot Point	GCV Value
1 knot point	$1,00 \times 10^{-11}$
2 knot point	$1,24 \times 10^{-11}$
3 knot point	$1,28 \times 10^{-11}$

Table 9 shows the minimum UBR value of $1,0010^{-11}$ on models with two knot points. Parameter estimation of a truncated spline nonparametric regression model using the UBR method with two knot points shown in equation (14) below.

$$\begin{aligned}\widehat{y} = -0,01 - 1,87x_1 + 0,01x_1^2 + 0,12(x_1 - 53,92)_+^2 + 1,84x_2 - 0,01x_2^2 - 0,33(x_2 - 43,11)_+^2 \\ + 0,77x_3 - 0,00x_3^2 - 0,09(x_3 - 33,73)_+^2 - 0,84x_4 + 0,04x_4^2 - 0,04(x_4 - 5,44)_+^2 \\ + 1,33x_5 - 0,03x_5^2 - 0,05(x_5 - 4,63)_+^2\end{aligned} \quad (14)$$

From the nonparametric regression model, the spline is truncated with three knot points in equation (14) The coefficient of determination or R^2 value is 0.7909. This shows that the model can explain the factors that influence the stunting prevalence data in Indonesia in 2021 of 79.09%. After selecting the optimal knot point using the GCV and UBR methods, it will be carried out between the two methods.

5. Comparison of GCV Method and UBR Method in Optimal Knot Point selection

After selecting the optimal knot point using the GCV and UBR methods, it will be carried out between the two methods. The following is a comparison table of GCV and UBR methods in selecting optimal knot points in stunting prevalence data in Indonesia in 2021. It is shown in the following table 10.

Table 10. Comparison of GCV Method and UBR Method in Optimal Knot Point selection

Method	Knot Point	MSE R^2	
GCV	three Knot point	1,36	0,9557
UBR	one Knot point	6,43	0,7909

Table 10 shows that the MSE value of 1.36 obtained in nonparametric regression modeling of truncated splines with three knot points using the GCV method is better than the UBR method with an MSE value of 6.43. Clarified by the R^2 value produced by the GCV method of 0.9557 is greater than the R^2 value produced by the CV method of 0.7909. So it can be concluded that the GCV method with three knot points is a more appropriate method in selecting optimal knot points than the UBR method in Indonesia's stunting prevalence data in 2021.

The finding of this study is that the GCV method is better than the UBR method. This is in line with research conducted by (Handayani et al., 2023) in his research entitled Nonparametric spline truncated regression with knot point selection method, generalized cross validation and unbiased risk. The difference is that this study used the quadratic spline nonparametric regression method.

D. CONCLUSION AND SUGGESTION

The best multivariable nonparametric regression model in the case of stunting prevalence in Indonesia was obtained from the squared truncated spline approach using the GCV method with three knots so as to produce an MSE value of 1.36 with an R^2 value of 95.57%, so it can be concluded that the GCV method with three knots is a more appropriate method used for selecting optimal node points compared to using the UBR method in the case of stunting prevalence in Indonesia in 2021. The suggestions that can be given from this study so that future research can develop interval estimation in the parameter estimation process of the truncated spline nonparametric regression model and then use panel data that considers the time aspect for each observation object so that it can use longitudinal truncated spline nonparametric regression and examine other spline bases that can be developed, namely B-Spline or P-Spline.

REFERENCES

Aswi, A. and Sukarna, S. (2022). Pemodelan spasial bayesian dalam menentukan faktor yang mempengaruhi kejadian stunting di provinsi sulawesi selatan. *JMathCoS: Journal of Mathematics, Computations, and Statistics*, 5(1):1–11. https://www.ojs.unm.ac.id/jmathcos.

da Silva, I. C. M., França, G. V., Barros, A. J., Amouzou, A., Krasevec, J., and Victora, C. G. (2018). Socioeconomic inequalities persist despite declining stunting prevalence in low-and middle-income countries. *The Journal of nutrition*, 148(2):254–258. https://doi.org/10.1093/jn/nxx050.

Danaei, G., Andrews, K. G., Sudfeld, C. R., Fink, G., McCoy, D. C., Peet, E., Sania, A., Smith Fawzi, M. C., Ezzati, M., and Fawzi, W. W. (2016). Risk factors for childhood stunting in 137 developing countries: a comparative risk assessment analysis at global, regional, and country levels. *PLoS medicine*, 13(11):e1002164. https://doi.org/10.1371/journal.pmed.1002164.

Desyanti, C. and Nindya, T. S. (2017). Hubungan riwayat penyakit diare dan praktik higiene dengan kejadian stunting pada balita usia 24-59 bulan di wilayah kerja puskesmas simolawang, surabaya. *Amerta Nutrition*, 1(3):243–251. https://doi.org/10.20473/amnt.v1i3.2017.243-251.

Dewi, A. R., Dewi, Y. L. R., and Murti, B. (2019). Life course factors associated with stunting in children aged 2-5 years: a path analysis. *Journal of Maternal and Child Health*, 4(5):348–357. https://doi.org/10.26911/thejmch.2019.04.05.09.

Hamzah, W., Haniarti, H., and Anggraeny, R. (2021). Faktor risiko stunting pada balita. *Jurnal Surya Muda*, 3(1):33–45. https://doi.org/10.38102/jsm.v3i1.77.

Handayani, T., Sifriyani, S., and Dani, A. T. R. (2023). Nonparametric spline truncated regression with knot point selection method generalized cross validation and unbiased risk. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 7(3):848–863. https://doi.org/10.31764/jtam.v7i3.14034.

Indanah, I., Wanda, D., and Nurhaeni, N. (2022). Empowerment of parents with stunting children. *Jurnal Aisyah: Jurnal Ilmu Kesehatan*, 7(4):1013–1022. https://doi.org/10.30604/jika.v7i4.1183.

Latuihamallo, A., Margawati, A., Mexitalia, M., Ediati, A., and Syauqy, A. (2022). Differences in development and diet of stunting and non-stunting children in the rowosari health center work area, semarang, indonesia. *Jurnal Gizi Indonesia (The Indonesian Journal of Nutrition)*, 10(2):161–167. https://doi.org/10.14710/jgi.10.2.161-167.

Liem, S., Marta, R. F., and Panggabean, H. (2019). Sanitation behavior and risk of stunting: Understanding the discourse of a public service announcement. *Jurnal The Messenger*, 11(2):168–181. https://doi.org/10.26623/themessenger.v11i2.1317.

Nahar, B., Hossain, M., Mahfuz, M., Islam, M. M., Hossain, M. I., Murray-Kolb, L. E., Seidman, J. C., and Ahmed, T. (2020). Early childhood development and stunting: Findings from the mal-ed birth cohort study in bangladesh. *Maternal & child nutrition*, 16(1):e12864. https://doi.org/10.1111/mcn.12864.

Rachmayani, S. A., Kuswari, M., and Melani, V. (2018). Hubungan asupan zat gizi dan status gizi remaja putri di smk ciawi bogor. *Indonesian Journal of Human Nutrition*, 5(2):125–130. https://doi.org/10.21776/ub.ijhn.2018.005.02.6.

Remmers, T., van Grieken, A., Renders, C. M., Hirasing, R. A., Broeren, S. M., and Raat, H. (2014). Correlates of parental misperception of their childs weight status: the be active, eat rightstudy. *PLoS One*, 9(2):e88931. https://doi.org/10.1371/journal.pone.0088931.

Schmidt, C. W. (2014). Beyond malnutrition: the role of sanitation in stunted growth. https://doi.org/10.1289/ehp.122-A298.

Singh, A., Upadhyay, A. K., and Kumar, K. (2017). Birth size, stunting and recovery from stunting in andhra pradesh, india: evidence from the young lives study. *Maternal and child health journal*, 21:492–508. https://doi.org/10.1007/s10995-016-2132-8.

Sudfeld, C. R., Charles McCoy, D., Danaei, G., Fink, G., Ezzati, M., Andrews, K. G., and Fawzi, W. W. (2015). Linear growth and child development in low-and middle-income countries: a meta-analysis. *Pediatrics*, 135(5):e1266–e1275. https://doi.org/10.1542/peds.2014-3111.

Wijhati, E. R., Nuzuliana, R., and Pratiwi, M. L. E. (2021). Analisis status gizi pada balita stunting. *Jurnal Kebidanan*, 10(1):1–12. https://doi.org/10.26714/jk.10.1.2021.1-12.

Yulianti, I., Nurasmi, N., and Padlilah, R. (2022). Stunting risk factors in children under 5 years old in indonesian border regions (north kalimantan-malaysia). *International Journal of Health and Pharmaceutical (IJHP)*, 2(2):275–283. https://doi.org/10.51601/ijhp.v2i1.42.