Designing Standard Growth Chart Based on Weight-For-Age Z-Score of Children in East Java Using Least-Square Spline Estimator

Nur Chamidah\textsuperscript{1,*}, Ardi Kurniawan\textsuperscript{1,*} & Toha Saifudin\textsuperscript{1}

\textsuperscript{1}Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Indonesia

*Corresponding author: nur-c@fst.unair.ac.id

Abstract. Children would be categorized as children who have underweight nutritional status, if according to index of anthropometric they have a lack of weight. In Indonesia, this anthropometric index is recorded on a Card Toward Health called as KMS. This card follows the WHO-2005 standard which is designed based on samples from Brazil, Ghana, India, Norway, Oman, and USA. Those samples, of course, physically are very different from Indonesian children. Therefore, in this paper we design weight-for-age Z-score standard growth charts of children by using least-square spline estimator and samples of children from East Java province, Indonesia. Next, the proposed children standard growth charts are used to assess East Java children nutritional status. The results show that the proposed standard growth charts have met the goodness of fit criteria namely the average values of coefficient determination for boy and girl are close to one, and values of mean square errors are close to zero. It means that the proposed growth charts are more suitable to be used to assess the nutritional status of East Java children, because they can better explain the real conditions of children in East Java, Indonesia than the WHO-2005 standard growth charts.

Keywords: least-square spline estimator, nutritional status, standard growth charts, weight-for-Age z-score.

1 Introduction

According to weight-for-age anthropometric index, children would be categorized as children with underweight nutritional status if they have a lack of weight. In terms of malnutrition, Indonesia is ranked seventh in the world, namely in total of 87 million national children there are about 3.8% who suffer malnutrition. The Indonesian government should pay special attention to this problem. In Health Profile of Indonesia 2018 has been presented that there were 16.8% of infants under five years old who still experiencing nutritional problem, namely an average 3.9% and 13.8% caught severely underweight and underweight cases, respectively. Nationally, East Java Province ranks second for malnutrition cases of under five years old children after East Nusa Tenggara Province. According to East Java Province Health Department, the number of malnutrition cases in East Java province have increased by 31.36% from 4,716 cases in 2017 to 6,195 cases in 2018 [1].
KMS, a card toward health for Indonesian children provides weight-for-age anthropometric index presented in a child's normal growth curve. At present, Indonesia uses KMS that follows anthropometric standard of WHO-2005. The KMS used in Indonesia is based on the Z-score curve for weight-for-age. The Z-score identifies a much higher severe undernutrition prevalence compared with conventional [2]. The WHO-2005 standard growth charts used samples of infants aged 0–60 months from Brazil, Ghana, India, Norway, Oman, and USA. These samples used to design WHO-2005 standard growth charts are considered to have represented regions in the world that is recommended as an assessment of global nutritional status [3–4]. However, existence of characteristics differences caused by a race, environment, and physical condition gives a discrepancy in the WHO-2005 standard growth charts patterns, including Indonesia. Therefore, for designing standard growth charts, it is better to use samples from children who have the same physical characteristics as the characteristics of Indonesian children.

Researchers [5] gave suggestion to use the standards growth chart that is carried out locally to be able to describe the actual condition. Although, the standard growth charts for children designed by using a nonparametric regression approach by [6] have carried out locally in Padang, West Sumatera, but they still did not differentiate gender. Nonparametric regression approach is an appropriate approach to show a pattern that is not the same at each stage on the growth chart for children [7]. Nonparametric regression has good flexibility for the observational data curve and results a smooth estimation curve.

Several researches who have designed standard growth charts locally of children in Surabaya by using smoothing techniques are [8–10] who used multi-response local polynomial estimators, [11–13] who used kernel estimators of multi-response nonparametric regression, [14] who used a penalized spline estimator of bi-responses semiparametric regression, [15–16] who used local linear estimators on the median curves for boy and girl infants aged 0–24 months of bi-responses nonparametric and semi-parametric regression. Furthermore, researches in designing standard growth charts for children in East Java were done by [17] that used local linear estimator, and by [18–20] that used the least-square spline and penalized spline estimators as a determination of wasting nutritional status. These standard growth charts produced by those researchers mentioned above were designed by using percentile values and by considering gender differences.

In this paper, we design the standard growth charts of East Java children up to five years old based on weight-for-age Z-score by using least-square spline estimator of a nonparametric regression. The advantage of using this estimator is that it can overcome data that have fluctuated patterns with knots, and results a relative smooth curve [21–23]. These proposed standard growth charts are then able to used to assess the nutritional status of children in East Java of Indonesia. In addition, we will also compare the percentage of
underweight cases between using the proposed standard growth charts and the WHO-2005 standard growth charts.

2 Materials and Methods

Z-Scores
According to [15], to calculate the Z-scores at a value of standard deviation (SD) which follows a Normal distribution, we use the following formula:

\[
SD_i(t) = M(t)[1 + L(t) \times S(t) \times i]^{1/L(t)}
\]

where \(SD_i(t)\) represents Z-scores at the value of standard deviation (SD) for \(t = 0, 1, 2, \ldots, 60\) and \(i = -3, -2, -1, 0, 1, 2, 3\); \(M\) is value of median which estimates mean of population mean; \(L\) is the value of Box-Cox power that is needed for transforming the data and removing skewness; and \(S\) is variation coefficient. Next, according to [14], steps to calculate \(L, M, \) and \(S\) for each age group are as follows:

1. Calculating both mean and SD of natural logarithms of measurements. The antilog of the mean is a geometric mean of measurement notated by \(M_g\). Hence, analogy the SD is called the ‘geometric’ cross validation (CV) notated by \(S_g\).
2. Calculating both mean and SD of the original measurements. This mean is an arithmetic mean of measurement notated by \(M_a\). So, to determine the ‘arithmetic’ CV that is notated by \(S_a\), we divide the SD by the geometric mean (\(M_a\)).
3. Calculating both mean and SD of the reciprocals of measurements. The reciprocal of the mean is to be a harmonic mean of measurement notated by \(M_h\). Next, we multiply the SD by the geometric mean (\(M_h\)) to obtain the ‘harmonic’ CV which is notated by \(S_h\).
4. Substituting values of \(S_a, S_g, S_h\) into the following formulas:
   \[
   A = \log(S_a/S_h)
   \]
   \[
   B = \log((S_aS_h/S_g)^2)
   \]
   So, Box-Cox power \(L\) is given by the following formula:
   \[
   L = -A/(2B)
   \]
5. Calculating the generalized coefficient of variation \((S)\) by using the following formula:
   \[
   S = S_g \exp(AL/4)
   \]
6. By interpolating \(M_a, M_g, M_h\), determining the generalized mean \((M)\) for power \(L\) is as follows:
   \[
   M = M_g + (M_a - M_h)L/2 + (M_a - 2M_g + M_h)L^2/2
   \]

Least-Square Spline Estimator
Least-square spline estimator is a very flexible estimator for estimating a nonparametric regression function, because it lets the data to find their own form of regression function
estimation without intervention of researcher’s subjectivity [17,23–28]. Generally, the nonparametric regression models for \( n \) paired data \( \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\} \) are expressed as follows:

\[ y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n \]  

(7)

where \( f \) is a function of regression assumed to be smooth function namely a continuous and differentiable function, and \( \epsilon_i \) is a zero mean independent random error with variance \( \sigma^2 \).

Next, according to [21], a least-square spline function with the order \( p \) and the knots points \( k_1, k_2, \ldots, k_m \) can be expressed as follows:

\[ y_i = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{j=1}^{m} \beta_{p+j}(x_i - k_j)_+^p, \quad i = 1, 2, \ldots, n \]  

(8)

where \( \beta_j, j = 0, 1, \ldots, p, p+1, \ldots, p+m \) are parameters, and \((x_i - k_j)_+^p\) satisfies the following condition:

\[ (x_i - k)_+^p = \begin{cases} (x_i - k)^p, & x_i \geq k \\ 0, & x_i < k \end{cases} \]  

(9)

Hence, the least-squares spline nonparametric regression model can be written as follows:

\[ \mathbf{y} = \mathbf{X}_k \mathbf{\beta} + \mathbf{\epsilon} \]  

(10)

where \( \mathbf{y} = (y_1, y_2, \ldots, y_n)^T \) is an \((n \times 1)\)-vector of response variable, \( \mathbf{\beta} = (\beta_0, \beta_1, \ldots, \beta_p, \beta_{p+1}, \ldots, \beta_{p+m})^T \) is a \((p + m + 1) \times 1\)-vector of parameters, \( \mathbf{X}_k \) is an \((n \times (p + m + 1))\)-matrix that can be expressed as follows:

\[ \mathbf{X}_k = \begin{bmatrix} 1 & x_1^1 & \cdots & x_1^p & (x_1 - k_1)^p & \cdots & (x_1 - k_m)^p \\ 1 & x_2^1 & \cdots & x_2^p & (x_2 - k_1)^p & \cdots & (x_2 - k_m)^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & \cdots & x_n^p & (x_n - k_1)^p & \cdots & (x_n - k_m)^p \end{bmatrix} \]  

(11)

and \( \mathbf{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_m)^T \) is an \((n \times 1)\)-vector of errors.

Hence, the estimator of the model presented in (10) can be written as follows:

\[ \hat{\mathbf{y}} = \mathbf{X}_k \hat{\mathbf{\beta}} \]  

(12)

where

\[ \hat{\mathbf{\beta}} = (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{y} \]  

(13)

**Selection of Optimal Knot**

In the nonparametric regression, determining the optimal knots is very important. The best estimated spline model will be obtained, if the optimal knots have been obtained. There are some methods used to select the optimal knots, one of them is by minimizing the following generalized cross validation (GCV) function [13]:

\[ \text{GCV}(\lambda) = \frac{\text{MSE}(\lambda)}{\left[ \frac{1}{n} \text{trace}(1 - H(\lambda)) \right]^2} \]  

(14)
Goodness of Fit Criteria
There are two goodness of fit criteria used in this research. They are determination coefficient and mean square errors. The determination coefficient ($R^2$) is to be proportion of variation in the dependent variable or response variable ($y$) that can be explained by some independent variables or predictor variables ($x$). Value of $R^2$ can be determined by using the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$  \hspace{0.5cm} (15)

Furthermore, a mean squared errors (MSE) value is an average value of squared difference between observation values and the estimated values. The MSE value can be determined by using the following formula:

$$\text{MSE} = n^{-1} \sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$  \hspace{0.5cm} (16)

Data Sources
The secondary data used in this research was collected from POSYANDU (Integrated Service Center) and PUSKESMAS (Centers of Public Health) in 21 cities and regencies in East Java province, Indonesia. The data obtained are weight measurements of infants aged 0 up to 60 months based on gender. There are 31,150 data observations of boy infants and 29,371 data observations of girl infants. The dependent variable or response variable in this research is weight (in kg), and the independent variable or predictor variable is age (in month) for each gender.

Steps of Analysis
To analyze the data, the steps are as follow:

a. Calculating the values of $L$, $M$, and $S$ for each age group.

b. Calculating the SD value or Z-score, namely -3SD, -2SD, 0SD, 2SD, and 3SD of weight in each age group.

c. Estimating the weight-for-age standard growth chart in each Z-score by using the least-squares spline estimator of nonparametric regression by creating Open Source Software R (OSS-R) code.

d. Determining the optimal knots based on minimum GCV values at every SD value or Z-score.

e. Calculating the coefficient of determination ($R^2$) value and MSE value.

f. Designing the standard growth charts of weight-for-age children up to five years old by using the least-squares spline estimator of nonparametric regression.

g. Determining nutritional status of children for both boys and girls based on both locally standard growth charts that have been designed and WHO-2005 standard growth charts.
3 Results and Discussion

To get the best model estimate to design the standard growth chart of weight-for-age, by using the least-square spline estimator of nonparametric regression we first determine the SD value or Z-score for each age group. Next, we can obtain the best estimated model by determining the optimal knots that minimize the values of GCV in (14) for each Z-scores, namely -3SD, -2SD, 0SD, 2SD, and 3SD. The optimal order, the optimal knots, and minimum GCV values are given in Table 1.

Table 1 The Optimal Order and Optimal Knots Based on Minimum GCV Values for Each Z-scores.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Z-Score</th>
<th>Order</th>
<th>Optimal Knot</th>
<th>Minimum GCV Value</th>
<th>MSE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boy</td>
<td>-3 SD</td>
<td>1</td>
<td>6</td>
<td>0.119484</td>
<td>0.1080201</td>
<td>97.68</td>
</tr>
<tr>
<td></td>
<td>-2 SD</td>
<td>1</td>
<td>6, 24</td>
<td>0.050572</td>
<td>0.0441571</td>
<td>99.20</td>
</tr>
<tr>
<td></td>
<td>0 SD</td>
<td>2</td>
<td>6, 12</td>
<td>0.019324</td>
<td>0.0162859</td>
<td>99.81</td>
</tr>
<tr>
<td></td>
<td>2 SD</td>
<td>2</td>
<td>6, 24</td>
<td>0.199014</td>
<td>0.1677255</td>
<td>98.86</td>
</tr>
<tr>
<td></td>
<td>3 SD</td>
<td>2</td>
<td>12</td>
<td>0.546425</td>
<td>0.4771119</td>
<td>97.59</td>
</tr>
<tr>
<td>Girl</td>
<td>-3 SD</td>
<td>1</td>
<td>6, 48</td>
<td>0.077203</td>
<td>0.0674096</td>
<td>98.53</td>
</tr>
<tr>
<td></td>
<td>-2 SD</td>
<td>2</td>
<td>6, 12, 36</td>
<td>0.035840</td>
<td>0.0291361</td>
<td>99.48</td>
</tr>
<tr>
<td></td>
<td>0 SD</td>
<td>2</td>
<td>6, 24, 36, 48</td>
<td>0.023022</td>
<td>0.0180413</td>
<td>99.80</td>
</tr>
<tr>
<td></td>
<td>2 SD</td>
<td>2</td>
<td>6</td>
<td>0.175957</td>
<td>0.1536376</td>
<td>99.99</td>
</tr>
<tr>
<td></td>
<td>3 SD</td>
<td>2</td>
<td>6, 48</td>
<td>0.504645</td>
<td>0.4253066</td>
<td>97.90</td>
</tr>
</tbody>
</table>

Table 1 shows that in average, R² values of the estimation of weight-for-age at every Z-score are 98.63% for gender boy and 98.94% for gender girl. Also, in average, MSE values are 0.16266 for gender boy and 0.13871 for gender girl. Hence, the estimated models of the median (0 SD) growth chart of weight-for-age Z-score for boy and girl are given as follows:

For boy: $\hat{y} = 3.736 + 1.016x - 0.060x^2 + 0.051(x - 6)^2 + 0.0088(x - 12)^2$ (17)
For girl: $\hat{y} = 3.453 + 0.995x - 0.063x^2 + 0.0599(x - 6)^2 + 0.00439(x - 24)^2 - 0.00478(x - 36)^2 + 0.00409(x - 48)^2$ (18)

Furthermore, the estimated models of median (0 SD) growth chart of weight-for-age Z-score in (17) and (18) can be expressed as piecewise functions to make them easier to understand. These piecewise functions are:

For boy: $\hat{y} = \begin{cases} 
3.736 + 1.016x - 0.060x^2 & \text{for } 0 \leq x < 6 \\
5.572 + 0.404x - 0.009x^2 & \text{for } 6 \leq x < 12 \\
6.839 + 0.193x - 0.0002x^2 & \text{for } x \geq 12 
\end{cases}$ (19)
For girl: \[ y^* = \begin{cases} 
3.453 + 0.995x - 0.063x^2 & \text{for } 0 \leq x < 6 \\
5.609 + 0.2362x - 0.003x^2 & \text{for } 6 \leq x < 24 \\
8.138 + 0.025x + 0.0013x^2 & \text{for } 24 \leq x < 36 \\
1.943 + 0.369x - 0.0035x^2 & \text{for } 36 \leq x < 48 \\
11.366 - 0.024x + 0.00059x^2 & \text{for } x \geq 48 
\end{cases} \]

Based on equations (19) and (20), the highest average weight gain occurs in infants less than six months, namely 0.653 kg for boy and 0.681 kg for girl. Hence, the interval \( x \geq 12 \) is an interval where the lowest average weight gain for boy occurs, and the interval \( x \geq 48 \) is an interval where the lowest average weight gain for girl occurs. The estimation and observation plots of median growth charts both boys and girls are shown in Fig 1.

Figure 1 Plots of estimated median growth of boy (left) and girl (right) in East Java province.

As we know that the anthropometric index weight-for-age of Indonesian children is recorded on a card toward health called as KMS in which this KMS is guided by anthropometric index of WHO-2005 that used samples of children from Brazil, Ghana, India, Norway, Oman, and USA. Of course, those samples have different physical characteristic from Indonesian children. Therefore, in designing standard growth charts should be used samples of children that physical follow children in Indonesia. The results of designing standard growth charts of children up to five years old for both boy and girl of weight-for-age Z-score in East Java, Indonesia by using least-square estimator are given together in Fig 2.
Figure 2. Least-square spline standard growth charts of weight-for-age Z-score for boy (left) and for girl (right).

Additionally, Figure 3 shows that the median growth charts of weight for both boy and girl in East Java, Indonesia are lower than WHO-2005 standard growth charts with a difference of 1.281 kg for boy and 1.206 kg for girl. It means that the median growth chart for girl is closer to the WHO-2005 standard growth charts.

Figure 3 Comparison between least-square spline median growth charts and WHO-2005 median growth charts of weight-for-age for boy (left) and for girl (right).

To assess the nutritional status of children up to five years old based on these two standard growth charts, we take 31,150 boys and 29,371 girls. Next, the percentages of nutritional status of children based on anthropometric index weight-for-age of standard growth chart designed by using the least-square spline estimator of nonparametric regression and anthropometric index weight-for-age of WHO-2005 standard growth chart are given in Table 2.
Table 2. Percentages of nutritional status between proposed growth charts and WHO-2005

<table>
<thead>
<tr>
<th>Anthropometric index</th>
<th>Nutritional Status</th>
<th>Boy East Java</th>
<th>WHO 2005 East Java</th>
<th>Girl East Java</th>
<th>WHO 2005 East Java</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight-for-age</td>
<td>Underweight</td>
<td>1.86 %</td>
<td>14.16 %</td>
<td>1.98 %</td>
<td>12.29 %</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>50.52 %</td>
<td>60.65 %</td>
<td>50.28 %</td>
<td>60.41 %</td>
</tr>
<tr>
<td></td>
<td>Overweight</td>
<td>47.61 %</td>
<td>25.18 %</td>
<td>47.74 %</td>
<td>27.30 %</td>
</tr>
</tbody>
</table>

Table 2 shows that the percentages of underweight nutritional status based on standard growth charts of weight-for-age Z-score designed by using the least-square spline estimator of nonparametric regression and used samples children from East Java, Indonesia are less than those based on WHO-2005 standard growth charts with a difference of 12.30% for boy and 10.31% for girl. It means that some children who have normal weight based on standard growth charts of weight-for-age Z-score designed by using the least-square spline estimator of nonparametric regression are categorized as underweight or overweight based on WHO-2005 standard growth charts.

4 Conclusion

The standard growth charts of weight-for-age Z-score designed by using the least-square spline estimator of nonparametric regression have satisfied the goodness of fit criteria, namely the average values of coefficient determination for boy and girl are close to one (98.63% for boy and 98.94% for girl), and values of mean square errors are close to zero (0.16266 for boy and 0.13871 for girls). It means that the proposed growth charts are more suitable to be used to assess the nutritional status of East Java children, because they can better explain the real conditions of children in East Java, Indonesia than the WHO-2005 standard growth charts.

5 Acknowledgement

Authors wish to thank Airlangga University Community Service Institution and Faculty of Science and Technology, Airlangga University for funding this work through the Community Service Activities Grant with Rector’s Decision Letter No. 1023/UN3/2022. Authors also thank to editors and peer-reviewers of Contemporary Mathematics and Applications (ConMathA), and to a proof-reader Dr. Budi Lestari, Drs., PG,Dip.Sc., M.Si., from the University of Jember, Indonesia, who have given useful suggestions and criticisms for improving the quality of this manuscript.

6 References


