Journal of Modern Applied Statistical Methods
https://jmasm.com/index.php/jmasm/index

Article

Clean and Healthy Living Behavior Factors That Influence Environmental-Based Diseases in Lampung Province, Indonesia: A SEM-DWLS Analysis

Subian Saidi *, Netti Herawati and Khoirin Nisa

Department of Mathematics, Faculty of Mathematics and Natural Science, University of Lampung, Lampung 35141, Indonesia
* Correspondence: subian.saidi@fmipa.unila.ac.id

How To Cite: Saidi, S.; Herawati, N.; Nisa, K. Clean and Healthy Living Behavior Factors That Influence Environmental-Based Diseases in Lampung Province, Indonesia: A SEM-DWLS Analysis. *Journal of Modern Applied Statistical Methods* **2025**, *24*(2), 1. https://doi.org/10.53941/jmasm.2025.100001.

Abstract: This research aims to determine the factors of clean and healthy living behavior that influence environmental-based diseases in Lampung Province, Indonesia using structural equations with the Weighted Least Square diagonal estimation method. Data were collected from students living in all districts in Lampung Province, Indonesia (n = 8524). Our findings show that environmental-based diseases are influenced by the clean and healthy living behavior of students in Lampung Province, Indonesia. The cleaner and healthier a society is, the smaller the incidence of environmental-based diseases in the community. Factors such as the use of clean water, the habit of washing hands, the use of clean water in latrines and the habit of not smoking, can reduce the number of people suffering from environmental-based diseases by 31%, 4%, 12%, 16%, respectively.

Keywords: SEM-DWLS; clean and healthy living behavior; Lampung Province

1. Introduction

Clean and healthy living behavior (CHLB) can create a better quality of life, this behavior must be pursued and carried out continuously so that it becomes a habitual and cultural pattern in society. Apart from that, clean and healthy living behavior can create conditions for individuals, families, groups and communities, by opening lines of communication, providing information and providing education to improve knowledge, attitudes and behavior, through a leadership approach (advocacy), building an atmosphere (social support) as well as community empowerment as an effort to help people recognize and overcome their own problems, in their respective environments so they can adopt healthy lifestyles. Study by [1] has found that poor clean and healthy living behavior could influence the incidence of diarrhea. In addition, CHLB is also a preventive measure (preventing a disease or health disorder) and can improve a person's health status [2].

The problem of CHLB in the community in Indonesia is an important topic that is of concern to many parties because as we know, individuals who have a healthy lifestyle will achieve higher life goals. The percentage of CHLB in Indonesia during the COVID-19 pandemic in the community, among students and children were 52–77.5%, 52–77.5%, and 50–86.49%, respectively [3] A study by [4] showed that there is a relationship between knowledge and attitudes about clean and healthy living habits to prevent transmission of COVID-19. Apart from that, the higher the public's knowledge, the more positive attitudes and behavior shown to prevent COVID-19, the better it will be [5]. Apart from that, children aged 7–12 years are the right age to acknowledge about CHLB [6]. The importance of CHLB for the family makes the individuals in it oriented toward family health [7] and CHLB problems can actually be handled effectively with a holistic approach with a focus on action for their health [8]. CHLB itself affects health. clean and healthy living behavior and a healthy place to live have an effect on reducing

Copyright: © 2025 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

the incidence of ARI in children [9]. CHLB also affects diarrheal disease. However, diarrhea can be prevented with good hygiene habits. The realm of health behavior consists of knowledge, attitudes, and actions [10].

It is very important to know the importance of healthy living because it can form public awareness to start practicing healthy living. Where consuming healthy food, exercising regularly, and developing healthy social habits, can improve a person's physical and mental health [11]. Clean and Healthy Living Behavior is also important for heart health promotion, disease risk reduction, and disease prevention. CHLB behavior which includes food intake patterns, physical activity and not smoking is used as the main component for children, adolescents and adults from various populations [12]. It is stated that adopting a healthy lifestyle can significantly increase life expectancy regardless of the presence of multimorbidity [13]. A cultural approach can also be used to improve CHLB practices among students [14,15]. Environmental-based diseases such as chronic obstructive pulmonary disease, especially asthma, bronchitis, and emphysema, have been linked to exposure to environmental tobacco smoke [16]. One of the main risk factors for infections in the respiratory tract, digestive tract, reproductive tract and other systems in humans is smoking. Smoking can also increase the prevalence of HIV and tuberculosis. Quitting smoking can reduce the risk of infection. Therefore, smoking cessation education is needed to prevent and reduce disease infections caused by tobacco use [17].

Structural Equation Modeling (SEM) is a combination of regression analysis methods, factor analysis, and path analysis techniques used to build and test statistical models in the form of causal models where changes in one variable have an impact on other variables [18]. There are some estimation methods in SEM such as Partial Least Square (PLS) and Diagonally Weighted Least Square (DWLS) estimation methods with their respective categories and assumptions. The DWLS method is suitable for ordinal size scale data. DWLS itself is a consistent estimator and does not depend on the assumption of normality [19]. DWLS method is known to produce more accurate parameter estimates compared to the maximum likelihood method [20,21]. This method also produces a more robust estimator for data with small categories and abnormalities [22–34] and produces root mean square error (RMSE) and Tucker-Smaller Lewis index [25], also a better RMSE of Approximation, Comparative Fit Index (CFI) and standardized root mean square residual (SRMR) values [26] and as a parameter estimator the DWLS method has complete information [27]. Furthermore, DWLS uses all available information to estimate model parameters, including, for example, not only the assignment of indicators to latent factors, but also whether latent factors are correlated [28].

SEM-DWLS estimation is used to estimate a regression model consisting of one binary dependent variable predicted by ten continuous independent variables that are allowed to correlate with each other. SEM was used to analyze these data and obtain robust DWLS estimates of nonstandard regression coefficients for continuous independent variables, by regressing the dichotomous dependent variable on a set of ten continuous variables [29]. SEM-DWLS can be implemented on health data to construct models of healthy living habits [30]. Most health and survey data use a Likert scale so that the data is not normally distributed. Therefore, DWLS is a solution to that problem [31,32]. In this article the author wants to apply the SEM-DWLS estimation method to determine the factors of clean and healthy living behavior that can influence environmentally based diseases in communities in Lampung Province, Indonesia.

2. Structural Equation Modelling (SEM)

According to [18], SEM modeling is a statistical technique that is capable of analyzing relationship patterns between latent variables and their indicators, one latent variable with other latent variables, as well as direct measurement errors. SEM helps establish relationships between several variables and establish three or more relationships under controlled conditions. Apart from that, SEM is also able to measure how well a phenomenon is connected and furthermore its effectiveness and the path structure resulting from various factors [33,34].

In SEM there are 2 types of latent variables, namely exogenous and endogenous. The exogenous variable (ξ) is the independent variable in all equations in the model, while the endogenous variable ($\boldsymbol{\eta}$) is the dependent variable in at least one equation in the model [35]. Observed variables or measurable variables are variables that can be observed or measured empirically and are often called indicators [36]. Observed variables that are related to or are the influence of the exogenous latent variable (ξ) are labeled **X**, while those related to the endogenous latent variable ($\boldsymbol{\eta}$) are labeled Y.

In general, the Structural Equation Model is defined as follows: Suppose it is a random vector $\eta^T = (\eta_1, \eta_2, \dots, \eta_m)$ dan $\xi^T = (\xi_1, \xi_2, \dots, \xi_n)$ respectively are endogenous and exogenous latent variables which form simultaneous equations with a system of linear equation relationships [37]:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

with $\mathbf{B}$ is $m \times m$ the coefficient matrix of endogenous latent variables $m \times m$, $\mathbf{\Gamma}$: $m \times n$ the coefficient matrix of exogenous latent variables, $\boldsymbol{\eta}$ is $m \times 1$ vector of endogenous latent $\boldsymbol{\xi}$ is $n \times 1$ vector of exogenous latent variables, $\boldsymbol{\zeta}$ is $m \times 1$ random residual vector of the relationship between $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. The assumptions are $E(\boldsymbol{\eta}) = 0$, $E(\boldsymbol{\xi}) = 0$, $E(\boldsymbol{\zeta}) = 0$; $\boldsymbol{\xi}$ is uncorrelated with $\boldsymbol{\zeta}$. In forming the model, SEM is divided into two models, namely measurement models and structural modeling. In the measurement model, each latent variable is modeled as a factor underlying the related observed variable. What is meant by "loading factor (λ)" is connecting latent variables with observed variables. SEM has two different loading factors, namely one matrix on side X (λX) and another matrix on side Y, namely (λY). The random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ cannot be measured directly but through indicator variables, namely the variables $\boldsymbol{Y}' = (y_1, y_2, \dots, y_p)$ and $\boldsymbol{X}' = (x_1, x_2, \dots, x_q)$ which measured by a measurement model, stated as:

$$\mathbf{X} = \mathbf{\Lambda}_X \boldsymbol{\xi} + \boldsymbol{\delta}$$

$$\mathbf{Y} = \mathbf{\Lambda}_Y \boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

A measurement model with three observed variables and one exogenous latent variable can be written as $X_i = \lambda_{Xii}\xi_i$, with 1 = 1, 2, ..., n.

Furthermore, structural modeling is a model that describes the relationship between latent variables [35]. In SEM, exogenous latent variables can "covary" freely and the covariance matrix of these variables is characterized by Φ. According to [37], the form of the structural equation model is obtained by $\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$ or can be written as $\boldsymbol{\eta} = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta})$. In addition, the structural model with two endogenous variables and one exogenous variable can be written as: $\eta_1 = \gamma_{11}\xi_1 + \zeta_1$; $\eta_2 = \gamma_{21}\xi_1 + \beta_{21}\eta_1 + \zeta_2$

In forming a structural model, the relationship between latent variables is similar to the linear regression equation between the latent variables. The parameter that shows the regression of an endogenous latent variable against an exogenous latent variable is given the symbol (γ), while the regression parameter for an endogenous latent variable against another endogenous latent variable is given the symbol (β).

Parameter estimates for the model are used to obtain values for the parameters in the model. In structural equation models, parameter estimation is used to obtain estimates of each parameter specified in the model which forms a matrix $\sum(\boldsymbol{\theta})$ such that the parameter values are as close as possible to the values in the **S** matrix (the sample covariance matrix of the observed variables). The sample covariance matrix (**S**) is used to represent the population covariance matrix **(Σ)** because the population covariance matrix is unknown.

Based on the null hypothesis, efforts are made to ensure that the difference between **S** and $\sum(\boldsymbol{\theta})$ is close to or equal to zero. This can be done by minimizing the function $\mathbf{F}(\mathbf{S}, \sum(\boldsymbol{\theta}))$ through iteration [35]. Estimation of the model can be done using one of the available estimation methods. Estimation methods that can be used in structural equation models are Instrumental Variables (IV), the general least squares method (Generalized Least Square, GLS), Maximum Likelihood (ML), weighted least squares method (WLS) and DWLS.

2.1. Diagonally Weighted Least Square (DWLS) Estimates

The Diagonally Weighted Least Square method or diagonally weighted least squares method is obtained by implementing or using the diagonal weight matrix W from the WLS estimator by minimizing the function:

$$\mathbf{F}_{\mathbf{DWLS}}(\boldsymbol{\theta}) = (\mathbf{s} - \boldsymbol{\sigma})' \, \mathbf{diag}\,(\mathbf{W})^{-1}(\mathbf{s} - \boldsymbol{\sigma}) \tag{1}$$

with $\mathbf{s}'$ is a vector containing the elements of the lower triangle and the diagonal of the sample covariance matrix **S** as a parameter estimate. Meanwhile, $\boldsymbol{\sigma}'$ it is a vector that contains the elements of the lower triangle and the diagonal of the covariance matrix $\sum(\boldsymbol{\theta})$ in the model. Matrix **S** and $\boldsymbol{\sigma}'$ is a symmetric and positive definite matrix. $\mathbf{W}^{-1}$is the inverse of the weighting matrix **W** for the error matrix which is an asymptotic variance matrix whose elements are written $\mathbf{W}_{\mathbf{ii,k,k}}$ [38].

The DWLS method is a method that is not affected by violations of multivariate normality. DWLS can be less stable if used for large models and small samples [38]. The weakness of this method is that the number of variables in the model must be small, namely less than 20 variables [34].

To obtain a diagonally weighted least squares estimator, first from a structural equation model, namely:

$$\boldsymbol{\eta} = \boldsymbol{\beta}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{2}$$

From Equation (2), the error is:

$$\boldsymbol{\zeta} = \boldsymbol{\eta} - \boldsymbol{\beta}\boldsymbol{\xi} \tag{3}$$

Then substitute the error in Equation (3) into Equation (1), so that the sum of the remaining squares is obtained as follows:

$$\begin{aligned} F_{DWLS} &= \zeta' \, diag\,(W^{-1})\zeta \\ &= (\eta - \hat{\beta}\xi)' \, diag\,(W^{-1})(\eta - \hat{\beta}\xi) \\ &= (\eta - \hat{\beta}\xi)'(diag\,(W^{-1})\eta - diag\,(W^{-1})\hat{\beta}\xi) \\ &= \eta' diag\,(W^{-1})\eta - \eta' diag\,(W^{-1})\hat{\beta}\xi - \xi'\hat{\beta}' diag\,(W^{-1})\eta + \xi'\hat{\beta}' \, diag\,(W^{-1})\hat{\beta}\xi \\ &= \eta' diag\,(W^{-1})\eta - 2\xi'\hat{\beta}' diag\,(W^{-1})\eta + \xi'\hat{\beta}' diag\,(W^{-1})\hat{\beta}\xi \end{aligned} \tag{4}$$

Because $\eta' diag\,(W^{-1})\hat{\beta}\xi$ is a scalar, its form is the same as the transpose is

$$\hat{\beta}\xi \, diag\,(W^{-1})\eta.$$

To obtain an estimator so that the residual sum of squares is as small as possible, then differentiate Equation (4) with respect to $\boldsymbol{\beta}$, then we obtain the equation as following:

$$\frac{\partial F_{DWLS}}{\partial \hat{\beta}} = -2\xi' diag\,(W^{-1})\eta + 2\xi' diag\,(W^{-1})\hat{\beta}\xi \tag{5}$$

by minimizing $\frac{\partial F_{DWLS}}{\partial \hat{\beta}} = 0$ then it is obtained:

$$\hat{\beta} = (\xi' diag\,(W^{-1})\xi)^{-1}\xi' diag\,(W^{-1})\eta \tag{6}$$

$\hat{\beta}$ is an unbiased estimator of $\boldsymbol{\beta}$, with $\mathbf{E}(\zeta) = 0$

2.2. Goodness of Fit Test

To find out whether the model obtained is the right model to describe the actual data, a goodness of fit model is carried out. Evaluation of the level of suitability of the data to the model is carried out through several stages, namely overall model suitability, measurement model suitability, and structural model suitability. The degree of Goodness of Fit was assessed using the Adjusted Goodness of Fit Index (AGFI), Goodness of Fit Index (GFI), Root Mean Square Residual (RMSR), Root Mean Square Error of Approximation (RMSEA), Chi-Square statistical test and N Criterion. If the Chi-square value is smaller than 3.0 and the *p* value is greater than 0.05 then the model is fit. In addition, if the GFI and AGFI values are greater than 0.90, it means that the model developed is acceptable. In addition, the model is said to be good if the RMSR or RMSEA is ≤ 0.05 and CN > 200 [18].

3. Materials and Methods

Data was obtained from the results of a survey conducted on 8524 students in Lampung Province, Indonesia using proportional sampling techniques. The frame work of survey instrument was adapted from [39]. The 25-item questionnaire was created in the form of a self-administered questionnaire. Before the survey was conducted, the validity and reliability test of the questioner was first carried out for each indicator (question) using a sample of 200 people. The results of the validity and reliability tests show that the research indicators are valid and have high reliability. The latent variables and indicators used in this research are as shown in Table 1.

To start SEM-DWLS modeling, steps are taken to design the model. Next, model identification is carried out. After that, parameter estimation was carried out and a model suitability test was carried out using the Goodness of Fit model test. The final step is to choose the best model and interpret the model. The design of the research model according to the variables listed in Table 1 is to build an SEM-DWLS model to see what CHLB factors influence environmental-based disease (Y) directly or indirectly. The model design in this study is presented in Figure 1.

In mathematical equations it can be written as:

$$Y = \gamma_{11}X + \zeta_1; \tag{7}$$

$$Y = \gamma_{21}X1 + \gamma_{22}X2 + \gamma_{23}X3 + \gamma_{24}X4 + \zeta_2 \tag{8}$$

where, Y_j is an endogenous latent vector variable, with $j = 2$, X_b are vector of endogenous latent variables, with values $b = 1, 2, 3$ and 4, γ_{jb}: coefficient matrix ξ, with values $j = 1$ and $b = 1, 2, 3$ and 4, ζ_j: structural error, with values $j = 1$.

Table 1. Research variables and indicators.

CHLB (X)		Environmental-Based Diseases (y)	
Latent Variables	Indicators	Latent Variables	Indicators
Using clean water (X1)	Water sources are free from pollutant sources and disturbing factors (X_{11})	Environmental-Based Diseases (Y)	Coughing and colds that take for about 14 days (ARI) (Y_1)
	Use clean water to wash vegetables and fruit as well as to process food ready to eat (X_{12})		The frequency of bowel movements increases and the consistency of stool is soft or runny for less than 14 days (diarrhea) (Y_2)
	Cooking water to drink family until it boils (X_{13})		Suffering from intestinal worms (Y_3)
	Not dipping hands in water that has been processed into drinking water (X_{14})		Cough with phlegm more than 2-3 weeks (tuberculosis)(Y_4)
	Wash eating and drinking utensils with soap and clean water before use (X_{15})		Experience rashes, redness, itching on the skin (skin disease) (Y_5)
	Clean water is stored in a clean place and is always closed (X_{16})		Have a health condition that is sensitive to poor air quality (air sensitivity) (Y_6)
Wash hands with clean water and soap (X2)	Wash your hands with clean water and running water (X_{21})		
	Wash your hands with soap after defecating (X_{22})		
	Wash your hands with soap after sneezing, coughing, throwing snot and after returning from traveling (X_{23})		
	Wash hands with soap after playing and holding animals (X_{24})		
	Wash the soap before and after eating (X_{25})		
Using healthy latrines (X3)	Latrine is available (X_{31})		
	Using a toilet to defecate (X_{32})		
	Using a toilet to urinate (X_{33})		
	Latrines are cleaned every day (X_{34})		
	Available clean water and cleaning tools in the toilet (X_{35})		
	All family members use latrines (X_{36})		
	Inside the latrine there is no visible dirt, no insects and mice roam (X_{37})		
Do not smoke (X4)	Do not smoke (X_{41})		
Sport (X5)	Sporting is regularly (X_{51})		

The measurement model that will be designed consists of 26 indicator variables where the latent variable **Y** consists of 6 indicator variables, the latent variable X1 consists of 6 indicator variables, the latent variable X2 consists of 5 indicator variables, the latent variable consists of 2 indicator variables. The latent variable X consists of 4 indicators. Each indicator variable has a measurement error (ÿ). The design of the measurement model is presented in the following equation:

$X1 = \lambda_{X_{1a}} X_{1a} + \delta_1$, where a = 1, 2, ..., 6
$X2 = \lambda_{X_{2b}} X_{2b} + \delta_2$, where b = 1, 2, ..., 5
$X3 = \lambda_{3c} X_{3c} + \delta_3$, where c = 1, 2, ..., 7
$X4 = \lambda_{X_{4d}} X_{4d} + \delta_2$, where a = 1, 2
$Y = \lambda_{Y_i} Y_i + \varepsilon_i$, where i = 1, 2, …, 6
$X = \lambda_{X_j} X_j + \delta_j$, where j = 1, 2, …, 4

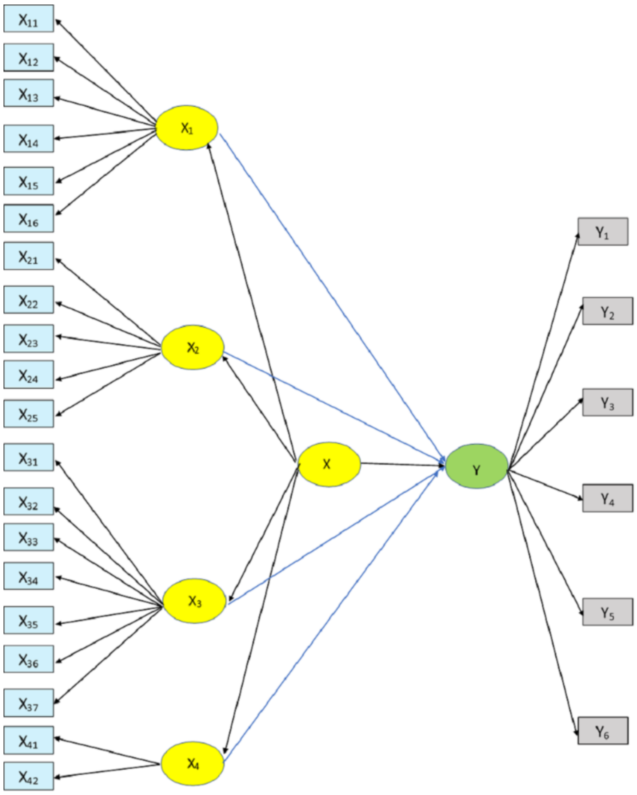

Figure 1. SEM-DWLS Modeling Design.

4. Results and Discussion

Descriptive analysis of data with n = 8524 respondents consisting of the variables age, gender, education, employment, income, place of residence (city/village) and type of water used. The majority of respondents were female with a percentage of 61% while men were 39%. The average age is 24.85 years with the majority aged between 15–20 years. Then there are quite a lot of people aged 20 to 29 years old. This is in line with the majority of jobs surveyed being students, namely 62%. Where the majority of people have no income, namely 66%. The survey results also show that the majority of people live in urban areas. Meanwhile, the results of descriptive analysis of data regarding CHLB and environmental-based diseases are presented in Table 2 below.

Table 2. Descriptive analysis of data regarding CHLB and environmental-based diseases.

	A. CHLB Indicator	**Percentage**
1	Uses clean water	91%
2	Wash hands with clean water and soap	95%
3	Using a healthy toilet	90%
4	Do not consume cigarettes	74%
5	Doing sport regularly	42%
	B. Environmentally based diseases	
1	Acute Respiratory Infections (ARI),	32%
2	Diarrhea	25%
3	Worms	10%
4	Tuberculosis	19%
5	Skin disease	33%
6	Air sensitive	27%

Based on Table 2, the implementation of CHLB among students is quite high, namely above 75%. The highest was washing hands with clean water and soap at 95% and the lowest was doing regular exercise at 42%. Overall CHLB implementation is 78%. The diseases that people often experience are skin diseases at 33% and the rarest are worms at 10%. Overall, 24% suffered from environmental-based illnesses. This condition is inversely proportional to the implementation of CHLB in society.

Based on the model design that has been created, model formation begins with identifying indicators (observed variables) for all late variables. Model identification is carried out through path analysis by looking loading factor coefficient. Loading factor is the magnitude of the correlation coefficient between an indicator and its latent construct. Indicators with high loading factors have a higher contribution to explaining the latent construct. On the other hand, indicators with low factor loadings have a weak contribution to explaining the latent construct. According to [18,40] a factor loading of 0.50 or more has validation that is strong enough to explain the latent construct. Meanwhile, according to [41] the weakest factor loading that can be accepted is 0.40. Using the help of R software version 4.3.2, the following loading factor coefficients are presented in Table 3 as well as the parameter estimation, *p*-value and t-value of indicators for the latent variable.

Table 3. Model Identification Through Path Analysis.

No	Variable	Indicator	Estimate	Std. Err	t-Value	*p*-Value	Load. Factor
1	X1	X11	1.000				0.510
		X12	1.612	0.084	19.157	0.000..	0.823
		X13	0.870	0.054	16.094	0.000..	0.444
		X14	1.176	0.067	17.550	0.000..	0.600
		X15	1.774	0.092	19.257	0.000..	0.906
		X16	1.351	0.073	18.549	0.000..	0.690
2	X2	X21	1.000				0.916
		X22	0.911	0.031	28.991	0.000..	0.835
		X23	0.837	0.027	30.900	0.000..	0.767
		X24	0.852	0.028	30.831	0.000..	0.781
		X25	0.900	0.030	30.371	0.000..	0.825
3	X3	X31	1.000				0.959
		X32	1.023	0.006	160.167	0.000..	0.982
		X33	0.912	0.008	110.020	0.000..	0.875
		X34	0.737	0.013	57.086	0.000..	0.707
		X35	1.000	0.008	132.109	0.000..	0.960
		X36	1.020	0.006	161.173	0.000..	0.979
		X37	0.840	0.011	75.885	0.000..	0.806
9	Y	Y1	1.000				0.750
		Y2	1.045	0.020	51.549	0.000..	0.784
		Y3	1.066	0.024	44.069	0.000..	0.799
		Y4	1.071	0.022	49.140	0.000..	0.804
		Y5	0.888	0.020	43.664	0.000..	0.666
		Y6	0.870	0.021	41.269	0.000..	0.653
11	X	X1	1.000				0.926
		X2	1.507	0.087	17.283	0.000..	0.777
		X3	1.418	0.082	17.317	0.000..	0.699
		X4	1.104	0.075	14.645	0.000..	0.522
		X5	0.179	0.044	4.031	0.000..	0.085

Based on model identification in Table 3, it shows that all indicators can represent latent variables significantly and can be included in the model. Only the X5 indicator has a factor loading coefficient <0.4, meaning that X5 is very weak in representing latent variables so it is not included in forming the best model.

The next step is to carry out a model fit test to see the suitability of the model with the SEM-DWLS estimation method used. The model fit test was carried out on the entire model using the absolute fit test, incremental fit test, and parsimony fit test. Overall model fit tests are presented in Table 4.

Table 4. SEM-DWLS model suitability test.

Goodness of Fit test	Target Match Level	Results	Level Suitable
RMSEA	RMSEA $\leq$ 0.05 close fit 0.05 $<$ RMSEA $\leq$ 0.08 good fit 0.08 $<$ RMSEA $\leq$ 0.1 marginal fit RMSEA $>$ 0.1 poor fit	0.036	Good fit
NFI	NFI $\geq$ 0,90 good fit 0.80 $\leq$ NFI $<$ 0.9 marginal fit	0.97	Good fit
GFI	GFI $\geq$ 0,90 good fit 0.80 $\leq$ GFI $<$ 0.9 marginal fit	0.98	Good fit
AGFI	AGFI $\geq$ 0.90 good fit 0.80 $\leq$ AGFI $<$ 0.90 marginal fit.	0.98	Good fit
IFI	IFI $\geq$ 0.90 good fit 0.80 $\leq$ IFI $<$ 0.90 marginal fit.	0.97	Good fit
CFI	CFI $\geq$ 0.90 good fit 0.80 $\leq$ CFI $<$ 0.90 marginal fit.	0.97	Good fit
TLI	TLI $\geq$ 0.90 good fit 0.80 $\leq$ CFI $<$ 0.90 marginal fit.	0.98	Good fit

Based on Table 4, it can be seen that GFI, RMSEA, AGFI, NFI, IFI, CFI, TLI provide values indicating that the model built is in accordance with the data. In other words, the goodness-of-fit test shows the model fit. The final stage of SEM-DWLSmodeling is forming the best model. The final stage of SEM-DWLS modeling is establishing the best model. Based on Table 4 above, a model was obtained as shown in Figure 2. The best model produced will be used as a basis for analyzing the factors that influence CHLB. As seen in Figure 2, the variables that influence CHLB are variables X1 (Using Clean Water), X2 (Washing hands with clean water and soap), X3 (Using Healthy Latrines).

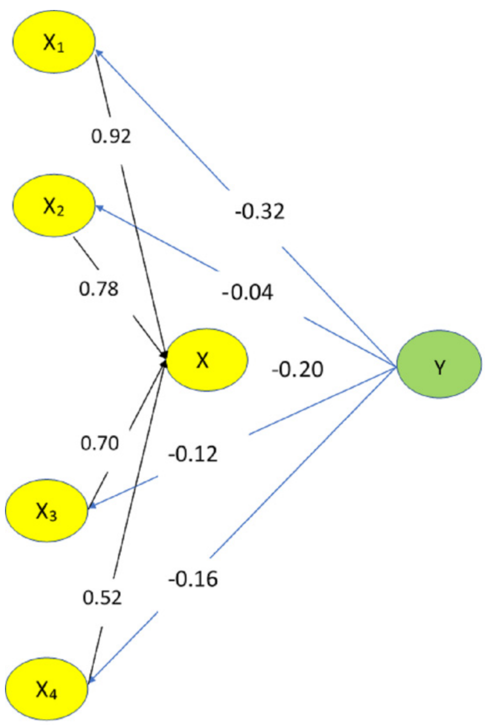

Figure 2. SEM-DWLS Model.

From Figure 2, the final SEM-DWLS model can be written in equation form:

(1) $Y = -0.32X_1 - 0.04X_2 - 0.12X_3 - 0.16X_4$, for direct modeling.

(2) $Y = -20\ X$, for modeling through variable X (CHLB)

(3) $X = 0.92\ X1 + 0.78X2 + 0.70X3 + 0.52X4$, modeling CHLB indicators

The SEM modeling that was formed can be interpreted that by increasing CHLB in the community it can reduce PBL by 20%. If modeled based on CHLB factors, then implementing the use of clean water for daily needs of 1 unit will reduce the number of environmental-based disease cases by 32%. Likewise, increasing the habit of washing hands with clean water and soap will reduce the risk of environmental-based diseases by 4%. The same thing also happens if people use healthy latrines, it will reduce the risk of environmental-based diseases by 12%. Likewise, if people do not consume cigarettes, it will have an impact on reducing the risk of diseases originating from the environment by 16%, in this case respiratory-related diseases. Tobacco is closely related to environmental health, not only having a negative impact on individual health.

Cigarettes and cigarette waste can enter the environment, polluting water, air and soil with toxic chemicals, heavy metals and nicotine residue [42] (Karaman, 2019). As stated by [43] WHO (2023), the products most scattered on the planet are tobacco products which contain more than 7000 toxic chemicals, which will enter our environment if thrown away. About 4.5 trillion cigarette filters pollute oceans, rivers, city sidewalks, parks, land and beaches every year around the world. Apart from that, other products also add to the accumulation of plastic pollution, such as cigarettes, smokeless tobacco and e-cigarettes which contain microplastics and are the second highest plastic pollution in the world. Therefore, to reduce the risk of environmental-based diseases in the community, the most important thing to provide sufficient clean water that the community can use for their daily lives. Then pay attention to the latrines used by the community to support reducing the risk of this environmental-based disease. As well as reducing cigarette consumption and increasing the habit of washing hands.

5. Conclusions

Based on the results of the SEM-DWLS model used in this research data, it can be concluded that clean and healthy living behavior as measured by the indicators of using clean water, washing hands with clean water and soap, using a healthy toilet, and not smoking have a significant influence on behavior. live clean and healthy. influence on clean and healthy living behavior. environmentally based diseases. Where increasing clean and healthy living behavior in society will reduce the number of cases of disease originating from the environment. The use of clean water will be more significant in reducing the risk of environmental-based diseases. Then not smoking is also quite big in reducing the risk of environmental-based diseases. Two other variables may also reduce the risk of environmentally based diseases.

Author Contributions

S.S.: Conceptualization, Methodology, Investigation, Writing—Original draft preparation; N.H.: Conceptualization, Methodology, Validation, Supervision, Writing—Reviewing and Editing; K.N.: Validation, Supervision, Writing—Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data that support the findings of this study are available on request.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Febriza, N.; Tang, U.M.; Maryanti, E. Pengaruh Perilaku Hidup Bersih dan Sehat (Phbs), Pendapatan dan Sanitasi terhadap Kejadian Diare di Kelurahan Meranti Pandak, Rumbai Pesisir Pekanbaru. *J. Ilmu Lingkung.* **2015**, *9*, 12–22.

2. Julianti, R.; Nasirun, M.; Wembrayarli, W. Pelaksanaan Perilaku Hidup Bersih dan Sehat (CHLB) Di Lingkungan Sekolah. *J. Ilm. Potensia* **2018**, *3*, 11–17.
3. Utami, F.A.; Sani, F. Gambaran Perilaku Hidup Bersih dan Sehat (CHLB) di Era Pandemi Corona Virus Disease (COVID-19) di Indonesia Description of Clean and Healthy Living Behavior (CHLB) in The Corona Virus Disease (COVID-19) Pandemi Era in Indonesia COVID-19 perilaku. *Bikfokes* **2021**, *1*, 197–209.
4. Saputri, M.E.; Suharyanto, T.; Khumaera, D. Knowledge and Attitude Relationship on Clean and Healthy Lifestyle towards the Prevention of COVID-19 Transmission in RW.03 Ciganjur Village South Jakarta. *J. Educ. Nurs.* **2021**, *4*, 29–34. https://doi.org/10.37430/jen.v4i1.83.
5. Putri, R.M.; Rosdiana, Y.; Nisa, A.C. Application of Clean and Healthy Living Behavior (CHLB) From the Household Knowledge and Attitude Study. *J. Nurs. Pract.* **2019**, *3*, 39–49. https://doi.org/10.30994/jnp.v3i1.64.
6. Sriwahyuni, S.; Zakiyuddin, Z.; Firdaus, M.R. The Effect of the Promotion of Clean and Healthy Lifestyle in Ujong Tanjung Elementary School through the SAL (Snake and Ladder) Media. *J. Fak. Kesehat. Masy.* **2021**, *8*, 41–44.
7. Gunawan, R.; Napitupulu, H.L.; Abdullah, H. Clean and Healthy Lifestyle Behaviour in the Family in Sena Village. *Glob. Sci. J.* **2021**, *9*, 485–494.
8. Mahadewi, E.P.; Sutawidjaya, A.H.; Asih, D.; et al. Sustainable Marketing of Healthcare in Indonesia with Religiosity and Health Promotion Clean Healthy Lifestyle. *Bp. Int. Res. Crit. Inst. J.* **2020**, *5*, 2751–2762.
9. El Kautsar, M.; Anggraeni, N.; Sari, V.P.; et al. The clean and healthy behaviors and healthy housing influenced the incident of acute respiratory infection in childhood. *EurAsian J. BioSci.* **2020**, *14*, 4837–4842.
10. Mufida, L.; Yunitasar, E.; Ulfiana, E. Health Education of Clean and Healthy Lifestyle using Card Telling Methods towards Diarrhea Prevention Among Children in Elementary School. *Pediomaternal Nurs. J.* **2021**, *7*, 8–15. https://doi.org/10.20473/pmnj.v7i1.21184.
11. Ridzuan, A.R.; Karim, R.A.; Marmaya, N.H.; et al. Public Awareness towards Healthy Lifestyle. *Int. J. Acad. Res. Bus. Soc. Sci.* **2018**, *8*, 927–936. https://doi.org/10.6007/ijarbss/v8-i10/4790
12. Hayman, L.L.; Worel, J.N. Healthy lifestyle behaviors the importance of individual and population approaches. *J. Cardiovasc. Nurs.* **2014**, *29*, 477–478. https://doi.org/10.1097/JCN.0000000000000199.
13. Chudasama, Y.V.; Khunti, K.; Gillies, C.L.; et al. Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: A longitudinal cohort study. *PLoS Med.* **2020**, *17*, e1003332. https://doi.org/10.1371/journal.pmed.1003332.
14. Saidi, S.; Herawati, N.; Nisa, K. Environmental-based diseases of students in urban and rural areas, Lampung province, Indonesia. *Int. J. Innov. Res. Sci. Stud.* **2023**, *6*, 242–249. https://doi.org/10.53894/ijirss.v6i2.1249.
15. Ruslana, F.H.; Mulyono, S. The relationship of cultural values with clean and healthy life behaviour among Islamic boarding school students in Indonesia. *J. Public Health Res.* **2022**, *11*, 83–86. https://doi.org/10.4081/jphr.2021.2739.
16. Vanker, A.; Gie, R.P.; Zar, H.J. The association between environmental tobacco smoke exposure and childhood respiratory disease: A review. *Expert Rev. Respir. Med.* **2017**, *11*, 661–673. https://doi.org/10.1080/17476348.2017.1338949.
17. Jiang, C.; Chen, Q.; Xie, M. Smoking increases the risk of infectious diseases: A narrative review. *Tob. Induc. Dis.* **2020**, *18*, 60. https://doi.org/10.18332/tid/123845.
18. Hair, J.F.; Black, W.C.; Babin, B.J.; et al. M*ultivariate Data Analysis*; Pearson Education Limited: London, UK, 2014.
19. DiStefano, C.; Morgan, G. A Comparison of diagonal weighted least square robust estimation techniques for ordinal data. *Struct. Equ. Model.* **2014**, *21*, 425–438.
20. Li, C.H. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol. Methods* **2016**, *21*, 369–387. https://doi.org/10.1037/met0000093.
21. Baghdarnia, M.; Soreh, F.F.; Gorji, R. The Comparison of Two Methods of Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) in Testing Construct Validity of Achievement Goals. *J. Educ. Manag. Stud.* **2014**, *4*, 22–28.
22. Bandalos, D.L. Relative Performance of Categorical Diagonally Weighted Least Squares and Robust Maximum Likelihood Estimation. *Struct. Equ. Model.* **2014**, *21*, 102–116. https://doi.org/10.1080/10705511.2014.859510.
23. Hox, J.J.; Maas, C.J.M.; Brinkhuis, M.J.S. The effect of estimation method and sample size in multilevel structural equation modeling. *Stat. Neerl.* **2010**, *64*, 157–170. https://doi.org/10.1111/j.1467-9574.2009.00445.x.
24. Mindrila, D. Maximum likelihood (mL) and diagonally weighted least square (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariat non-normal data. *Int. J. Digit. Soc.* **2010**, *1*, 60–66.
25. Xia, Y.; Yang, Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behav. Res. Methods* **2018**, *51*, 409–428. https://doi.org/10.3758/s13428-018-1055-2.
26. Shi, D.; Maydeu-Olivares, A. The Effect of Estimation Methods on SEM Fit Indices. *Educ. Psychol. Meas.* **2020**, *80*, 421–445. https://doi.org/10.1177/0013164419885164.
27. Verhulst, B.; Maes, H.H.; Neale, M.C. GW-SEM: A Statistical Package to Conduct Genome-Wide Structural Equation Modeling. *Behav. Genet.* **2017**, *47*, 345–359. https://doi.org/10.1007/s10519-017-9842-6.

28. Nestler, S. A Monte Carlo study comparing PIV, ULS and DWLS in the estimation of dichotomous confirmatory factor analysis. *Br. J. Math. Stat. Psychol.* **2013**, *66*, 127–143. https://doi.org/10.1111/j.2044-8317.2012.02044.x.
29. Linden, A.; Bryant, F.B.; Yarnold, P.R. Logistic Discriminant Analysis and Structural Equation Modeling Both Identify Effects in Random Data. *Optim. Data Anal.* **2019**, *8*, 97–102. Available online: https://www.researchgate.net/publication/333150079 (accessed on 1 January 2024).
30. Yanuar, F.; Uttaqi, F.N.; Zetra, A.; et al. The Comparison of WLS and DWLS Estimation Methods in SEM to Construct Health Behavior Model. *Sci. Technol. Indones.* **2022**, *7*, 164–169. https://doi.org/10.26554/sti.2022.7.2.164-169.
31. Isnayanti, I.; Abdurakhman, A. Model Persamaan Struktural Dengan Metode Diagonally Weighted Least Square (Dwls) Untuk Data Ordinal. *Media Stat.* **2019**, *12*, 100. https://doi.org/10.14710/medstat.12.1.100-116.
32. Rhemtulla, M.; Brosseau-Liard, P.É.; Savalei, V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* **2012**, *17*, 354–373. https://doi.org/10.1037/a0029315.
33. Creswell, J.W. Research Design: Qualitative, Quantitative, and Mixed Method. In *Research Design Qualitative Quantitative and Mixed Methods Approaches*; Sage Publications: Thousand Oaks, CA, USA, 2014. https://doi.org/10.1007/s13398-014-0173-7.2.
34. Kline, R.B. Principles and Practice of Structural Equation Modeling. *J. Basic Educ.* **2017**, *3*, 445–458.
35. Wijanto, S. *Structural Equation Modeling Dengan Lisrel 8.8*; Graha Ilmu: Yogyakarta, Indonesia, 2008.
36. Setiawan, E.; Pratiwi, A.; Herawati, N.; et al. A Structural Equation Modeling of Factors Affecting Student Motivation in Thesis Preparation. *J. Phys. Conf. Ser.* **2020**, *1751*, 012030. https://doi.org/10.1088/1742-6596/1751/1/012030.
37. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: Hoboken, NJ, USA, 1989. https://doi.org/10.1002/9781118619179.
38. Joreskog, K.G.; Sorbom, D. *LISREL8 User's Reference Guide*; Scientific Software International: Chicago, IL, USA, 1996.
39. Sudarmadji, S.; Haryono, B.; Suhardi. *Analisa Bahan Makanan dan Pertanian*; Liberty Yogyakarta: Yogyakarta, Indonesia, 1998.
40. Ghozali, I. *Structural Equation Modeling, Metode Alternatif Dengan Partial Least Square*; Universitas Diponegoro: Semarang, Indonesia, 2008.
41. Sharma, S. *Applied Multivariate Techniques*; John Wiley and Sons Inc.: New York, NY, USA, 1996.
42. Karaman, M.I. Smoking as an Environmental Health Problem. *Int. J. Hum. Health Sci.* **2019**, *3*, 123–126.
43. World Health Organization. WHO Raises Alarm on Tobacco Industry Environmental Impact. Available online: https://www.who.int/news/item/31-05-2022 (accessed on 1 January 2024).