

Modeling of Binary Logistic Regression in the Event of Childhood Diarrhea in Indonesia

Tofan Pramesi Wibowo¹, Soenarnatalina Melaniani², Lutfi Agus Salim²

¹Bachelor Student in Faculty of Public Health, ²Associate Professor in Public Health Faculty, Universitas Airlangga, Surabaya, Indonesia

Abstract

Logistic regression or logit regression is a statistical technique used to create predictive modeling. To determine the logistic regression equation that will be used to predict the value of the dependent variable, parameter estimates must be sought in the logistic regression model, one of the methods in this study uses the maximum likelihood method. The 2017 Indonesian Demographic and Health Survey (IDHS) collected various information, including about toddler diarrhea. Therefore, the logistic regression method is the right method to use in this condition, considering that the diarrhea data collected in the IDHS are categorical data. In this study, the variables influencing toddler diarrhea were child age, birth size (average), exclusive breastfeeding, maternal age, maternal education (high school), wealth index (top), and latrine facilities. The results of the logistic regression analysis showed that the variable with the highest risk figure for reducing the diarrhea rate was wealth index (top) (OR=1,438).

Keyword : *Logistic regression, maximum likelihood methods, IDHS, diarrhea*

Introduction

Logistic regression or logit regression is a statistical technique used to create predictive modeling. Therefore, researchers predict the dependent variable with a dichotomy scale. The dichotomy scale referred to is a nominal data scale with two categories, for example: Yes and No, Good and Bad or High and Low⁽¹⁾. Logistic regression does not require the assumption that the error variance (residual) is normally distributed because this type of logistic regression follows a logistic distribution. Therefore, Logistic Regression is a type of regression that connects one or more independent variables (independent variables) with the dependent variable in the form of categories; usually 0 and 1⁽²⁾.

In this case like other statistical modeling methods, the goal of logistic regression is to find the most fit

and sensible model to explain the relationship between the dependent variable (outcome or response) and the independent variable (predictor or explanatory)⁽²⁾. What distinguishes the logistic regression model from the linear regression model is that the outcome variable of logistic regression is binary or dichotomous. Because of this difference the y axis on the regression graph is a probability and the line in logistic regression is curved⁽³⁾.

The Indonesian Demographic and Health Survey is a survey to provide the latest estimates of basic demographic and health indicators for Indonesia. One of the specific design objectives of the IDHS is to measure the achievement of targets made by the national health program, particularly maternal and child health (National Population and Family Planning Agency, 2018). The Indonesian Demographic and Health Survey (IDHS) in 2017 collected various information, including about diarrhea in children under five.

Diarrhea is a disease that causes sufferers to have frequent bowel movements, with watery stools. In general, diarrhea occurs due to food and drink being exposed to viruses, bacteria, or parasites⁽⁴⁾. Based on

Corresponding author:

Lutfi Agus Salim,

Faculty of Public Health, Universitas Airlangga, Surabaya, Indonesia, Email: lutfi.as@fkm.unair.ac.id

data on Indonesia’s health profile information in 2017 from the Indonesian Ministry of Health, the number of diarrhea cases throughout Indonesia is around 7 million. Generally occurs in toddlers and children. But it does not rule out diarrhea can affect adults, so that diarrheal disease attacks all age groups⁽⁵⁾.

The prevalence of under-five diarrhea according to the IDHS 2002 - 2003, 2007, 2012, & 2017 is (11%) (IDHS, 2003), (13.7%) (IDHS, 2007), (14.3%) (IDHS, 2013), & (14.1%) (National Population and Family Planning Agency, 2017). Because diarrheal disease is a multifactorial disease, ranging from direct causes such as infectious diseases and food consumption to sanitation to family socio-economic factors⁽⁶⁾. Therefore, the logistic regression method is the right method to use in this condition, considering that the diarrhea data collected in the IDHS are categorical data⁽¹⁾.

Based on the description above, it shows that the death rate due to diarrhea is quite high. There are many factors that can influence the incidence of diarrhea in children under five. Therefore, diarrhea disease in children under five is a multifactorial disease. So that the researcher intends to conduct a research on Binary Logistic Regression Modeling in the Incidence of Toddler Diarrhea based on the 2017 Indonesian Demographic and Health Survey.

Methods

This study uses quantitative research data analysis with descriptive methods using secondary data from the 2017 IDHS results. Population data recorded as diarrhea sufferers aged under five in Indonesia in 2017 are research subjects that are divided into each province

throughout Indonesia in 2017.

The population of the 2017 IDHS is households and women of reproductive age (WUS) throughout Indonesia. The sampling technique from the 2017 IDHS is a systematic sampling technique. From the results of the sample selection, a sample of 49,261 households with WUS who met the requirements to be interviewed was 50,730 women. The sample in this study was selected from the 2017 IDHS sample based on the criteria for giving birth within the five years preceding the IDHS interview. From these criteria, a sample size of 17,848 respondents was obtained. Then data cleaning was carried out and produced a total of 13,921 respondents.

The variables in this study are divided into two, namely the dependent variable and the independent variable. The dependent variable in this study is the data on the incidence of toddler diarrhea (Y), while the independent variable in this study is the child’s age (X1), birth size (X2), exclusive breastfeeding (X3), maternal age (X4), maternal education (X5), wealth index (X6), and latrine facilities (X7). The analysis step carried out in the study was to describe the characteristics of the toddler’s diarrhea rate and the variables that influence it, then to see whether there was an effect of the independent variable on the dependent variable, the Binary Logistic Regression test was used.

Results

Descriptive statistical calculations (Table 1) were used to see the characteristics of the incidence of diarrhea that occurred in Indonesia during 2017 and each variable that affected diarrhea.

Table 1. Overview of the incidence of diarrhea

Variable	N	Mean	Variance	Minimum	Maximum
Y	13921	0,14	0,122	0	1
X1	13921	2,95	1,998	1	5
X2	13921	2,78	0,663	1	5
X3	13921	0,47	0,249	0	1
X4	13921	3,77	1,745	1	7
X5	13921	1,9	0,48	0	3
X6	13921	2,76	2,1	1	5
X7	13921	0,19	0,154	0	1

Based on table 1, the distribution of diarrhea incidence rates in children under five in 2017 with an N value of 13921. All respondents on average have experienced diarrhea. The average age of the child who had diarrhea most frequently was 3 years of age. The average measure of body weight at birth that most often had diarrhea was a very small measure of body weight. The average child who experiences diarrhea is not exclusively breastfed. The average age of having children with diarrhea is 20-24 years old, with relatively

low education. The average wealth index of families with children suffering from diarrhea are middle to lower class families, and on average they still do not have latrine facilities.

Logistic Regression Test

Factors affecting toddler diarrhea in Indonesia in 2017 can be analyzed using logistic regression. Following are the results of the Logistic Regression Test (Table 2).

Table 2. Results of Logistic Regression Test Analysis

No.	Variable	B	Wald	P-value	Significance
1.	Children’s Age	-0,006	16.672	0,001	Significant
2	Birth size		9,952	0,041	Significant
3	Exclusive breastfeeding	-0,158	10.909	0,001	Significant
4	Mom’s Age	-0,023	42.741	0,0001	Significant
5	Mother’s education		13,538	0,004	Significant
6	Wealth Index		18,701	0,001	Significant
7	Latrine Facilities	-0,268	19,256	0,0001	Significant

Logistic Regression Test is used to see whether there is an effect of each independent variable on the dependent variable. Based on Table 2, it is known that the value of each of the independent variables has been declared significant because the resulting p-value is less than α (0.05). So that the variables of children’s age, birth size, exclusive breastfeeding, maternal age, maternal education, wealth index, and latrine facilities have an influence on the incidence of diarrhea.

Logistic regression analysis in this study has resulted in a known odds ratio from the Exp (B) value. The odds ratio formed from the results of the regression analysis in Table 3 is the odds ratio for the incidence of diarrhea. The magnitude of the risk of toddler diarrhea can be seen in each independent variable which has a significant effect on the dependent variable, which can be seen from the value in the Exp (B) column.

Table 3. The results of the Odds Ratio value that affect toddler diarrhea

No	Variable	Exp (B)	$\frac{1}{Exp(B)}$
1.	Children’s Age	0,994	1,006
2.	Birth size	0,791	1,264
3.	Exclusive breastfeeding	0,854	1,171
4.	Mom’s Age	0,978	1,023
5.	Mother’s education	0,721	1,388
6.	Wealth Index	0,695	1,438
7.	Latrine Facilities	0,765	1,308

Based on Table 3, that diarrhea can occur due to several factors. The variable age of children can increase the risk of diarrhea by 0.994 ., For the variable size of the baby’s weight at birth can increase the risk of diarrhea by 0.791 ., For variables not given exclusive breastfeeding can increase the risk of diarrhea by 0.854 ., For the variable maternal age can increase the risk of diarrhea by 0.978 ., For maternal education variables can increase the risk of diarrhea by 0.721 ., The wealth index variable can increase the risk of diarrhea by 0.695., For the variable not having latrine facilities can increase the

risk of diarrhea by 0.765.

Logistic Regression Model

The independent variable which is significant with the dependent variable will be entered into the binary logistic regression model. From the β coefficient value the influencing variable can be used to determine a person’s chances of being included in the diarrhea category. Based on the β coefficient value in table 3, the equation model formed is as follows:

$$f(z) = \frac{1}{1 + e^{-[-0,006(x_1) - 0,234(x_2(2)) - 0,158(x_3) - 0,023(x_4) - 0,328(x_5(2)) - 0,428(x_5(3)) - 0,364(x_7) - 0,268(x_8)]}}$$

Explanation : $f(z)$ = Probability of children under five getting diarrhea

- X₁ = Child’s age
- X₂ = Birth size (average)
- X₃ = Exclusive breastfeeding
- X₄ = Mother’s age
- X₅₍₂₎ = Maternal education (SMP & SMA)
- X₅₍₃₎ =Mother’s education (College)
- X₇₍₄₎ =Wealth Index (Top)
- X₈ = Latrine

The logistic regression model that has been formed can be interpreted as follows:

If the mother's age is 15 years, the mother's last education is not in school / elementary school, the child does not receive exclusive breastfeeding, the child is 1 year old, the size of the child at birth is large / greater

than average / smaller than average / very small, index the lowest / middle upper / middle wealthy people, and do not have proper toilet facilities are;

$$\begin{aligned}
 f(z) &= \frac{1}{1 + e^{-[-0,023(15) - 0,328(0) - 0,428(0) - 0,158(0) - 0,006(1) - 0,234(0) - 0,0364(0) - 0,268(0)]}} \\
 &= \frac{1}{1 + e^{-[-0,345 - 0 - 0 - 0 - 0,006 - 0 - 0 - 0]}} \\
 &= \frac{1}{1 + e^{-[-0,351]}} \\
 &= 0,41314
 \end{aligned}$$

Discussion

The resulting variance value in the incidence of diarrhea in children under five is 0.122, where this value is considered quite large considering that in 2017 Indonesia experienced Diarrhea Extraordinary Events (KLB). The prevalence of toddler diarrhea in Indonesia in 2017 is (14.1%).

None of the independent variables included in modeling the incidence of daire in children under five also resulted in a P-value > 0.05. So that H0 is rejected, which means that there is an influence from the independent variable on the dependent variable.

It is also confirmed that the accuracy of the resulting logistic regression model can be seen through the Hosmer and Lemeshow test. The Hosmer and Lemeshow significance number is greater than 0.05, indicating that the logistic regression model formed is in accordance with the data or in other words there is no difference between the results of the observations and the results of the prediction of the model⁽¹⁾.

The regression model formed from this study is a logistic regression model by including the variables of maternal age, maternal education, exclusive breastfeeding, child age, birth size, wealth index, and latrine facilities and produces a Hosmer and Lemeshow test of 0.18. So that the logistic regression model formed

is a regression model that fits the data.

The Influence of Child Age on the Incidence of Toddler Diarrhea

The results of the logistic regression analysis show that the child's age has a significant value of 0.001, which means that the child's age has an influence on the incidence of toddler diarrhea in Indonesia in 2017. The age of children has an effect on diarrhea because at a low age it is very susceptible to disease. According to Wibisono, Marchianti and Dharmawan, (2020) in their study stated that younger people are prone to disease because the immune system has not developed properly⁽⁷⁾.

Most diarrhea occurs in the first 2 years of life. In the group aged 6 to 11 months, the highest incidence occurs, when complementary feeding is given to breastfeeding, this is because the natural immunity of children under one year of age has not yet been formed. Children aged 0-24 months are at higher risk for diarrhea than children aged more than 24 months⁽⁸⁾.

The odds ratio for under-fives is 1.006 (CI95%; 1.009-1.003), which means that every 1 year increase in the child's age will increase the likelihood of the child being included in the category of non-toddler diarrhea by 1.006 times. In accordance with research conducted by Wibisono, Marchianti and Dharmawan, (2020) states

that the older you get, the less you experience diarrhea⁽⁷⁾.

The Effect of Weight Measurement at Birth on the Incidence of Toddler Diarrhea

The results of logistic regression analysis on the weight variable at birth showed that a significant value in the category of child size, namely, the average with a significance value of 0.017, which means that the size of the child (average) has an influence on the incidence of toddler diarrhea in Indonesia in 2017. The size of the child's birth, both small and large have an influence on the occurrence of diarrhea because children are still susceptible to disease. The resulting odd ratio was 1.264 (CI95%; 1.531-1.043), meaning that underfives whose birth weight had an average chance of not having diarrhea was greater than 1.264 than children with very small, small, large and very large birth weight.

Infectious disease that is at risk in the first 2 years of stunting is diarrheal disease. Moreover, if the baby is born prematurely, where babies born prematurely have low birth weight (LBW), it is possible to get diarrhea⁽⁹⁾.

A cohort study conducted by Monalisa et al., (2020) with a sample of 13912 children concluded that a history of low birth weight babies had a higher risk of diarrhea, namely 1.45 (0.74-2.28) times greater than those born with normal weight⁽¹⁰⁾.

The Effect of Exclusive Breastfeeding on the Incidence of Toddler Diarrhea

The results of logistic regression analysis show that exclusive breastfeeding has a significant value of 0.001, which means that exclusive breastfeeding has an influence on the incidence of toddler diarrhea in Indonesia in 2017. The resulting odds ratio is 1.171 (CI 95%; 1.285-1.066), this means that children who are given Exclusive breastfeeding is likely not to have diarrhea greater than 1,171 times than under five who are not exclusively breastfed.

Exclusive breastfeeding contains colostrum which produces antibodies to prevent the growth of viruses and bacteria in the baby's body so that exclusive breastfeeding has an effect on preventing diarrhea. Exclusive breastfeeding for babies aged 4 - 6 months

will provide immunity to babies against various diseases because breast milk is a solution that contains immune substances that can protect babies from various bacterial, viral, fungal, and parasite infections. With breastfeeding, babies who are exclusively breastfed can be protected from diarrhea⁽¹¹⁾.

The Effect of Maternal Age on the Incidence of Toddler Diarrhea

The results of logistic regression analysis show that maternal age has a significant value of 0.0001, which means that maternal age has an influence on the incidence of toddler diarrhea in Indonesia in 2017 with the odds ratio of maternal age to toddler diarrhea is 1.023 (CI 95%; 1.035-1.021), this means that every 1 year increase in the mother's age will increase the possibility of the mother to be categorized as non-toddler diarrhea by 1,023 times. Maternal age has an influence on the incidence of toddler diarrhea due to the lack of knowledge of mothers in fulfilling child nutrition and also because the mother's age is still relatively young. Susanti, Novrikasari and Sunarsih, (2016) stated that there is a relationship between maternal age and the incidence of under-five diarrhea in Indonesia⁽¹²⁾.

The younger age group of mothers getting is associated with the likelihood of conceiving a child who is underweight, and a lack of experience in caring for children, as well as difficulties in providing adequate food, including those who tend to wean early⁽⁸⁾. Meanwhile, mothers who enter young adulthood (> 25 years) are likely to have completed their education at least at the middle or university level so that it is possible that the mother has prepared herself both knowledgeably and economically⁽¹²⁾.

The Effect of Mother's Education on the Incidence of Toddler Diarrhea

The results of logistic regression analysis on the variable maternal education on toddler diarrhea resulted in significant results. The significance value for the last education of category mothers was 0.015 with an odds ratio value of 1.388 (95% CI 1.805-1.066), which means that mothers with junior high school & high school education are likely to have their children not get diarrhea. greater 1,388 times compared to mothers who

do not go to school.

Mother’s education that does not go to school has an influence on toddler diarrhea because the lack of knowledge about caring for children is still lacking, especially in fulfilling nutrition. Parents’ education is a very important factor in the successful management of diarrhea in children. Parents with low levels of education, especially illiteracy will not be able to provide proper care to children with diarrhea due to lack of knowledge and lack of ability to receive information ⁽¹³⁾.

The Effect of Family Wealth Index on the Incidence of Toddler Diarrhea

The results of logistic regression analysis showed significant results in the top wealth index category with a significance value of 0.0001, which means that the wealth index has an influence on the incidence of under-five diarrhea in Indonesia in 2017. The resulting odds ratio is 1.438 (CI 95%; 1.730-1.196). Children under five with the highest family wealth index were likely not to have diarrhea 1.438 times greater than those with the lowest, middle, lower, middle and upper middle wealth indexes. Economic factors have an influence on various aspects of life, for example the community in meeting nutritional needs, sanitation needs, and other needs ⁽¹⁴⁾.

The socioeconomic status of the family is closely related to the purchasing power of the family which in turn affects the nutritional status of children under five. Low nutritional status of children under five or even malnutrition can increase risk factors for disease,

including diarrhea ⁽¹⁵⁾.

The Effect of Latrine Facility Ownership on the Incidence of Toddler Diarrhea

The results of the logistic regression analysis show that latrine facilities have a significant value of 0.001, which means that the latrine facilities have an influence on the incidence of under-five diarrhea in Indonesia in 2017. 1.308 children who do not have diarrhea are more than families who do not have inadequate latrine facilities. There are still many people who do not have latrines, this causes a greater risk of diarrhea. Prakoso, (2020) which states that the variable of the type of latrine used by the family has a significant relationship with the frequency of diarrhea ⁽¹⁶⁾.

Diarrheal disease is an environmentally based disease, the two dominant factors that can cause diarrhea are clean water facilities and feces disposal. These two factors will interact together with human behavior. If environmental factors are not healthy because they are contaminated with diarrhea germs and accumulate with unhealthy behaviors, such as food and drinks, it can cause diarrhea ⁽¹⁷⁾.

Conclusion

The prevalence of diarrhea under five in Indonesia in 2017 is 14.1 percent. The results of the logistic regression analysis showed that the variable with the highest risk figure for experiencing diarrhea under five was the Family Wealth Index (OR = 1.438). The logistic regression model that is formed is

$$f(z) = \frac{1}{1 + e^{-[-0,006(x_1) - 0,234(x_2(2)) - 0,158(x_3) - 0,023(x_4) - 0,328(x_5(2)) - 0,428(x_5(3)) - 0,364(x_7) - 0,268(x_8)]}}$$

Explanation : $f(z)$ = Probability of children under five getting diarrhea

- X₁ = Child’s age
- X₂ = Birth size (average)
- X₃ = Exclusive breastfeeding
- X₄ = Mother’s age
- X₅₍₂₎ = Maternal education (SMP & SMA)
- X₅₍₃₎ =Mother’s education (College)
- X₇₍₄₎ =Wealth Index (Top)
- X₈ = Latrine

The probability value of a 1-year-old toddler, very small weight at birth, mother aged 15 years and uneducated not in school, do not have hand washing facilities, the lowest wealth index, do not have proper latrines, as well as unsuitable drinking water by 41.3%.

Suggestion

Based on the above conclusions, it is necessary to increase public awareness about the importance of maintaining health because of not getting diarrhea. In addition, the Government of Indonesia can increase efforts to reduce the incidence of diarrhea by conducting counseling and other programs as a measure to reduce the incidence of diarrhea.

Conflict of Interest: The authors have no conflict of interest with the material presented in this paper.

Source of Funding: None.

Ethical Clearance: None. This research was conducted without treating informants

References

1. Azizah SM, Chandra NE. Model Regresi Logistik Pada Faktor-Faktor Yang Mempengaruhi Imunisasi Lengkap Balita. *J Ilm Teknosains*. 2017;3(2):3–6.
2. Wibowo A, Ridha MR. Comparison of Logistic Regression Model and MARS Using Multicollinearity Data Simulation. *JTAM | J Teor dan Apl Mat*. 2020;4(1):39.
3. Alzen JL, Langdon LS, Otero VK. A Logistic Regression Investigation of The Relationship Between The Learning Assistant Model and Failure Rates in Introductory STEM courses. *Int J STEM Educ*. 2018;5(1):1–12.
4. Kemenkes RI. Profil Kesehatan Indonesia Tahun 2016. In: *Profil Kesehatan Indonesia Tahun 2016*. Jakarta; 2017.
5. Kemenkes RI. Profil Kesehatan Indonesia Tahun 2017. In: *Profil Kesehatan Indonesia Tahun 2017* [Internet]. Jakarta; 2018. p. 107–8. Available from: website: <http://www.kemkes.go.id>
6. Kemenkes RI. Profil Kesehatan Indonesia Tahun 2015. In: *Profil Kesehatan Indonesia Tahun 2015*. Jakarta; 2016.
7. Wibisono A margaret, Marchianti ACN, Dharmawan DK. Analisis Faktor Risiko Kejadian Diare Berulang Pada Balita Di Puskesmas Sumberjambe Kabupaten Jember. *J Agromedicine Med Sci*. 2020;6(1):43–52.
8. Vasconcelos MJ de OB, Rissin A, Figueiroa JN, de Lira PIC, Filho MB. Factors associated with diarrhea in children under five years old in the state of Pernambuco, according to surveys conducted in 1997 and 2006. *Rev Saude Publica*. 2018;52:1–11.
9. Dewi NT, Widari D. Hubungan Berat Badan Lahir Rendah dan Penyakit Infeksi dengan Kejadian Stunting pada Baduta di Desa Maron Kidul Kecamatan Maron Kabupaten Probolinggo. *Amerta Nutr*. 2018;2(4):373.
10. Monalisa SR, Achadi E, Sartika RAD, Ningsih WM. Risiko Diare Pada Balita Usia 6-59 Bulan di Pulau Sumatera Indonesia (Analisis Data IFLS 2014). *J Ilmu Kesehat Masy STIKIM* [Internet]. 2020;9(2):129–36. Available from: <http://journals.stikim.ac.id/index.php/jikm%0ARisiko>
11. Ramadani RI. Pengaruh Pemberian ASI Eksklusif, PHBS dan Kepadatan Penduduk terhadap Kejadian Diare pada Balita di Kota Surabaya Tahun 2018. *J Media Gizi dan Kesehat Masy*. 2019;8(2):40–8.
12. Susanti WE, Novrikasari, Sunarsih E. Determinant of Diarrhea on Children Under Five Years in Indonesia (Advanced Analysis Idhs 2012). *J Ilmu Kesehat Masy*. 2016;7(1):64–72.
13. Febriyanti AP, Nugrahini M. Hubungan Tingkat Pendidikan Ibu Terhadap Penggunaan Zinc Dalam Terapi Diare Pada Anak Balita Di Apotek Platak Jaya Surabaya. *J UIN Alauddin* [Internet]. 2017;5(2):140–51. Available from: http://journal.uin-alauddin.ac.id/index.php/jurnal_farmasi/article/viewFile/3421/3220
14. Indrayani T, Rifiana AJ, Novitasari T. Faktor-Faktor Yang Berhubungan Dengan Kejadian Diare Pada Balita di Rumah Sakit Islam Bogor Jawa Barat Tahun 2017. *J Ilm Kesehat dan Kebidanan*. 2017;VII(2):1–12.
15. Ginting S. Hubungan Sosial Ekonomi dan Perilaku Hidup Bersih dan sehat dengan Kejadian Diare pada Balita di Desa Baru Kecamatan Pancur Batu.

- J Penelit Kesehat Masy. 2019;1(2).
16. Prakoso ID. Correlation Between Access of Drinking Water and Sanitation With Diarrhea Incidence in East Java. *J Berk Epidemiol.* 2020;8(1):42.
 17. Wahyudi F, Indah MF, Agustina N. Hubungan Sarana Ketersediaan Air Bersih, Perilaku Ibu, Kepemilikan Jamban dengan Diare pada Balita di Wilayah Puskesmas Tamiyang Layang Tahun 2020. *J FKM Uniska.* 2020;42(13201):1–10.