Jurnal Pijar MIPA
https://jurnalfkip.unram.ac.id/index.php/JPM

ISSN 1907-1744 (Print)
ISSN 2460-1500 (Online)

Modelling the Number of Pneumonia Cases among Toddlers in East Java Province with the Generalized Additive Model Approach

Verina Tita Nabila[1*], Naufal Ramadhan Al Akhwal Siregar[2]

[1]Faculty of Information Technology, Universitas Catur Insan Cendekia, Cirebon, Indonesia
[2] Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia
[*]e-mail: verina.titanabila@cic.ac.id

Received: February 11, 2026. Accepted: May 19, 2026. Published: May 25, 2026

Abstract: Pneumonia is a major respiratory infection that significantly affects toddlers. Based on Riskesdas 2021, East Java recorded the highest number of pneumonia cases in Indonesia. This study aims to model the number of toddler pneumonia cases in East Java using the Generalized Additive Model (GAM) to address the issue of overdispersion, comparing the performance of Poisson and Negative Binomial distributions. This study utilized secondary data sourced from the East Java Health Profile 2024, with a total of 38 districts/cities as the unit of observation. The analysis technique employed GAM with smoothing functions to capture non-linear relationships between the response and predictor variables, which include nutritional status, breastfeeding, immunization, vitamin A, and household air quality, followed by an overdispersion test. The results indicate that the GAM with a Negative Binomial distribution is superior to the Poisson model in handling overdispersion, as evidenced by a significantly lower deviance of 507.7 versus 8932.6. In conclusion, the Negative Binomial GAM provides a more accurate model for this case. Furthermore, improving public health awareness and upgrading health facilities are crucial steps to manage toddler pneumonia cases more effectively.

Keywords: Generalized Additive Model; Negative Binomial; Overdispersion; Pneumonia; Toddler.

Introduction

Pneumonia is an acute infection of the respiratory tract caused by bacteria, viruses, and fungi [1]. Pneumonia is the largest cause of death in toddlers worldwide, killing 740,180 toddlers or 14% of all deaths in toddlers [2]. Based on the 2021 Riskesdas data, 278,261 toddlers in Indonesia suffered from pneumonia, with the highest number of cases in East Java province, which was 74,071 cases [3]. Focusing on East Java is not only crucial due to it having the highest national caseload, but its vast demographic and geographical diversity across 38 districts/cities creates highly heterogeneous public health profiles. This complexity provides an ideal empirical setting for testing advanced statistical modelling and formulating region-specific health policy interventions. The high mortality rate in toddlers due to pneumonia makes it a strategic issue in the Sustainable Development Goals 2030, especially in goal number three, which is good health and well-being.

Pneumonia in toddlers is influenced by various risk factors such as nutritional status, exclusive breastfeeding, basic immunization, vitamin A administration, exposure to cigarette smoke, and air pollution [4]. Most healthy toddlers can fight infections with their immune system. However, toddlers with weakened immune systems have a higher risk of developing pneumonia. One of the causes of a weakened immune system is malnutrition, especially in toddlers who are not exclusively breastfed [5]. Basic immunizations, such as measles and DPT-HB-Hib 4 immunizations, can protect toddlers from comorbidities that cause pneumonia, namely measles and diphtheria [6]. Giving the right dose of vitamin A can also maintain the immune system and reduce the risk of pneumonia [7]. In addition, environmental factors such as air pollution and exposure to cigarette smoke also affect the risk of pneumonia in toddlers [8].

Recent developments in epidemiological modelling highlight that disease incidence data frequently exhibit irregular patterns and heterogeneous variability across regions. These characteristics limit the applicability of strictly parametric regression models, particularly when the relationship between response and predictor variables does not follow a predefined functional form. Consequently, flexible statistical approaches have gained attention because they allow complex non-linear relationships to be analysed while maintaining interpretability in public health research. Empirical health modelling studies show that flexible regression frameworks are increasingly used to capture irregular patterns in epidemiological data and improve model adaptability to real-world variability [9].

In addition, epidemiological count data commonly exhibit overdispersion, in which the variance exceeds the mean. This condition violates the equidispersion assumption of classical Poisson regression models, leading to biased parameter estimates and underestimated standard errors. Consequently, methodological studies in health statistics emphasize the need for modelling strategies that accommodate more flexible variance structures when analysing heterogeneous disease incidence data [10].

In the statistical approach, relationships between response and predictor variables that do not follow a specific data pattern can be analysed using nonparametric or semiparametric frameworks [11]. One widely used approach for modelling count-based health outcomes is the Generalized Additive Model (GAM) [12], which extends

How to Cite:
V. T. Nabila and N. R. A. A. Siregar, "Modelling the Number of Pneumonia Cases among Toddlers in East Java Province with the Generalized Additive Model Approach," *Jurnal Pijar MIPA*, vol. 21, no. 3, pp. 472–478, 2026. https://doi.org/10.29303/jpm.v21i3.11595

generalized linear models by incorporating smoothing functions to capture complex non-linear patterns. The GAM framework allows response variables to be modelled as discrete or continuous, making it suitable for epidemiological count data [13]. When applied to count responses, GAM is often combined with a Poisson distribution under the equidispersion assumption, where the mean equals the variance. However, disease incidence data often exhibit overdispersion, suggesting that the Poisson assumption is not fully met. To address this limitation, the negative binomial distribution provides a more flexible alternative by accommodating additional variance in the data [14].

Recent studies have shown that GAM has strong capabilities for modelling health data with non-linear patterns and temporal variability. Sitanggang et al. [15] applied a GAM to analyse acute respiratory infection incidence among children under five in Jambi City and demonstrated that GAM can capture complex relationships between environmental factors and disease occurrence through flexible smoothing functions. However, a significant research gap remains: the application of GAM, specifically integrated with a negative binomial distribution to address overdispersion in regional pneumonia surveillance data in Indonesia, remains highly limited. Recent epidemiological studies over the last five years have predominantly relied on classical Poisson regression or standard spatial models, which often fail to simultaneously accommodate both complex non-linear covariate relationships and variance inflation [16], [17]. This study addresses that gap. The novelty of this research lies in combining the GAM framework with rigorous overdispersion handling (Negative Binomial distribution), specifically tailored for macro-level epidemiological surveillance data.

The study highlights that GAM is particularly useful in epidemiological analysis because it can reveal non-linear exposure–response relationships that are difficult to detect using conventional parametric regression. These methodological advantages support the use of the GAM framework in analysing pneumonia incidence data that exhibit irregular patterns and overdispersion. Therefore, this study aims to model the number of pneumonia cases among toddlers in East Java Province using the GAM approach.

Research Methods

Nonparametric regression is a flexible statistical framework employed to investigate the association between a response variable and one or more predictor variables without imposing a predetermined functional form. This approach is particularly useful when the relationship's structure is unclear or when prior information about the underlying model is limited. The general form of nonparametric regression is as follows

$$y_i = m(x_i) + \varepsilon_i, i = 1,2,\dots,n \tag{1}$$

where ε_i denotes a random error term assumed to be independent with a mean of zero and variance σ^2, m is a regression function of unknown shape [18].

The additive model has a response variable Y that depends on the sum of the functions of the predictor variables X, so the additive model is in the form of

$$y_i = \sum_{j=1}^{p} m_j(x_{ij}) + \varepsilon_i, i = 1,2,\dots,n. \tag{2}$$

Generalized Additive Model (GAM) is an extension of the additive model equation form in equation (2), by connecting the expected value of a response variable to additive predictor variables through a link function so as to allow the distribution of the response variable to be exponential. The GAM model is given in the following equation

$$g(\mu_i) = \eta_i = \sum_{j=1}^{p} f_j(x_{ij}), i = 1,2,\dots,n \tag{3}$$

The g function is a link function whose functional form is different for each distribution of exponential form members [19].

This study utilizes secondary data obtained from the East Java Health Profile 2024, which was officially released by the East Java Provincial Health Office. There are 38 districts or cities and seven predictor variables to model the number of cases of penumonia among toddlers in East Java Province. This research employs a quantitative modelling design to analyze the non-linear relationship between the incidence of toddler pneumonia and each indicator. The population of this study encompasses all administrative regions in East Java Province. A total sampling technique was applied, meaning that all 38 districts/cities in East Java were used as the unit of observation, with no exclusions. The variables examined in this study are summarized in Table 1.

Table 1. Description of research variables

Var.	Description	Unit of Measurement	Source
y	Total number of pneumonia cases among toddlers	Cases (Count)	East Java Provincial Health Office 2024
x_1	Percentage of exclusive breastfeeding among children aged 0-23 months	Percentage (%)	East Java Provincial Health Office 2024
x_2	Percentage of toddlers with malnutrition	Percentage (%)	East Java Provincial Health Office 2024
x_3	Percentage of households that implement household air quality management	Percentage (%)	East Java Provincial Health Office 2024
x_4	Percentage of toddlers who received DPT-HB-Hib 4 immunization	Percentage (%)	East Java Provincial Health Office 2024
x_5	Percentage of toddlers immunized against measles	Percentage (%)	East Java Provincial Health Office 2024

x_6	Percentage of toddlers who receive vitamin A	Percentage (%)	East Java Provincial Health Office 2024

The data analysis was computationally performed using the R software, specifically utilizing the "mgcv" package for GAM estimation. The procedures for data analysis were carried out as follows:

1. Calculating the summary statistics for all variables to understand the data distribution.
2. Plotting the number of pneumonia cases in toddlers with each predictor variable
3. Fitting the Generalized Additive Model (GAM) to the data using smoothing splines to accommodate non-linear patterns. The smoothing parameters were estimated automatically using the Generalized Cross-Validation (GCV) method.
4. Evaluating the equidispersion assumption of the Poisson model.
5. Applying the GAM framework with a Negative Binomial link function to handle the detected overdispersion.
6. Comparing the performance of the Poisson GAM and Negative Binomial GAM based on their deviance values. A model with a significantly lower deviance value indicates a better goodness-of-fit.

Results and Discussion

The characteristics of the number of pneumonia cases among toddlers in districts in East Java Province, along with the seven predictor variables, are shown in Table 2 below.

Table 2. Characteristics of the number of pneumonia cases among under-fives in East Java Province in 2024

Var.	N	Mean	Variance	Minimum	Maximum
y	38	2607	5687298	275	11246
x_1	38	32.82	115.49	13.95	70.09
x_2	38	5.045	3.59	1.6	10.8
x_3	38	4328	770.99	0	86.29
x_4	38	85.19	238.17	28	100
x_5	38	85.66	222.45	34	100
x_6	38	87.78	89.89	56	100

The first step before modelling the data is to plot the number of pneumonia cases in toddlers for each predictor variable, holding the other predictor variables constant. The plotting results are shown in Figures 1 and 2 below.

Based on Figures 1 and 2, the plot between each predictor variable, namely x_1, x_2, x_3, x_4, x_5, and x_6 against the response variable (y) is irregular and does not form a specific data pattern, so the data on the number of pneumonia cases among toddlers in East Java Province can be modelled using GAM.

Based on Table 2, the variance of the number of pneumonia cases among toddlers exceeds the mean. This indicates overdispersion in the data. Overdispersion can also be detected from the ratio of the deviance to the degrees of freedom. The hypothesis used to test the occurrence of overdispersion in the response variable of the number of pneumonia cases in toddlers is as follows

$H_0 : \alpha = 1$ (indicating that the distribution of pneumonia case counts among toddlers exhibits equidispersion)
$H_1 : \alpha > 1$ (indicating the presence of overdispersion in the pneumonia case count data among toddlers)

The results of the overdispersion test for the number of pneumonia cases in toddlers are presented in Table 3, which includes the deviance value, the degrees of freedom, and the overdispersion value.

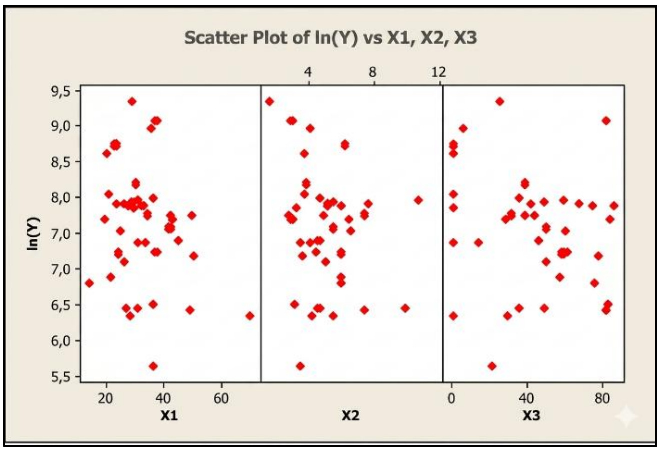

Figure 1. Scatter plot of x_1, x_2, and x_3 variables with the response variable.

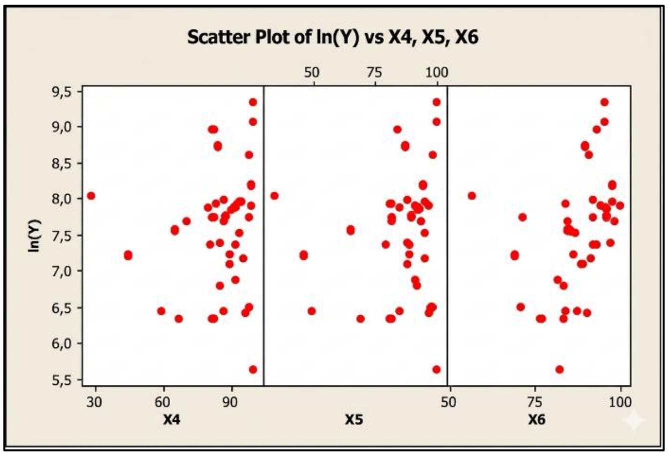

Figure 2. Scatter plot of x_4, x_5, and x_6 variables with the response variable

Table 3. Overdispersion test results on data

Deviance value	Degree of freedom value	Overdispersion value
35290.65	29	1216.919

Based on Table 3, the overdispersion value is more than 1, indicating overdispersion in the data on the number of pneumonia cases among children under five. The test results indicate that the data on the number of pneumonia cases among children under five years of age per city in East Java Province can be analyzed using GAM with a negative binomial distribution.

Furthermore, Table 4 presents the results of testing the significance of the modulus function, including the p-value and Effective Degrees of Freedom (EDF) for each predictor variable.

Table 4. The results of testing the significance of the stimulus function

Predictor variable	p-value	EDF value
x_1	0.86081	0.0000549
x_2	0.92196	0.0000689
x_3	0.00864	1.808
x_4	0.71698	0.0000228
x_5	0.87649	0.0000226
x_6	0.01703	1.565

Based on Table 4, the smoothing functions for variables $x_1, x_2, x_4,$ and x_5 produce p-values greater than $\alpha = 0.05$. This indicates that the four predictor variables do not have a significant effect on the number of pneumonia cases among toddlers. In addition, the EDF values for the five predictor variables are less than 1. These findings suggest that the association between the four predictor variables and the response variable follows a linear pattern. Meanwhile, variables x_3 and x_6 produced p-values less than $\alpha = 0.05$. This indicates that both predictor variables have a significant effect on the number of pneumonia cases among toddlers. In addition, the EDFs of both predictor variables exceed 1, indicating that the relationship between the two and the number of pneumonia cases among toddlers is non-linear. Therefore, the predictor variables used to model the number of pneumonia cases among under-fives using the GAM approach are the variables x_3 and x_6.

In this study, the optimal smoothing parameter for the GAM model is determined using the default estimator in the gam package in R. The GAM model can be written as follows:

$$y = \log(\mu) = 7.8127 + S(x_3) + S(x_6) \qquad (4)$$

with

$$\begin{aligned} S(x_3) = &-0.0695s(x_3)_1 - 0.2337s(x_3)_2 - 0.0627s(x_3)_3 \\ &-0.1518s(x_3)_4 + 0.11(x_3)_5 + 0.1527s(x_3)_6 \\ &-0.0637s(x_3)_7 - 0.8032s(x_3)_8 - 0.000008s(x_3)_9. \end{aligned}$$

$$\begin{aligned} S(x_6) = &\ 0.1189s(x_6)_1 - 0.0188s(x_6)_2 - 0.0733s(x_6)_3 \\ &+0.0912s(x_6)_4 + 0.0866s(x_6)_5 + 0.0954s(x_3)_6 \\ &-0.0892s(x_6)_7 - 0.6349s(x_6)_8 + 0.000029s(x_6)_9. \end{aligned}$$

The GAM model in equation (4) cannot be interpreted explicitly, so the contribution of variables x_3 and x_6 to the response variable is presented through the curve in Figure 3.

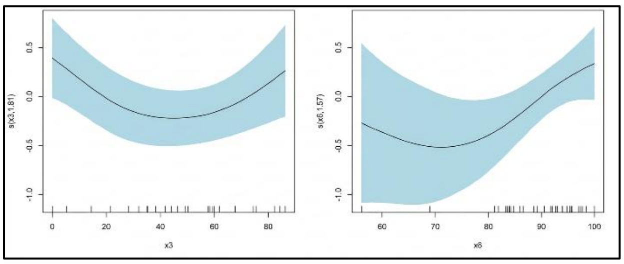

Figure 3. Curves of x_3 and x_6 variables on the number of pneumonia cases in toddlers

Based on Figure 3, if the functions of the other predictor variables are held constant, the curve of the variable x_3 shows that when the percentage of households with air quality management is below 46.2%, the number of pneumonia cases among toddlers in East Java Province will decrease. Conversely, if the percentage of households with air quality management exceeds 46.2%, it will increase the number of pneumonia cases among toddlers in East Java Province. The decreasing trend in pneumonia cases at lower levels of x_3 coverage indicates that improvements in household air quality management reduce children's exposure to respiratory risk factors within indoor environments. Practices such as improving ventilation, minimizing exposure to cooking smoke, reducing indoor particulate matter, and maintaining cleaner living spaces contribute to lowering the concentration of airborne pathogens and irritants [20]. From a biological perspective, reduced exposure to pollutants supports better respiratory defence mechanisms in toddlers, including improved mucosal function and reduced airway inflammation [21]. As a result, the transmission and progression of respiratory infections decline, which is reflected in the reduced number of reported pneumonia cases. This finding aligns with environmental health principles, stating that healthier indoor air conditions directly influence respiratory outcomes in early childhood populations [22].

Figure 3 additionally shows that if the functions of the other predictor variables are held constant, the x_6 variable curve shows that when the percentage of toddlers given vitamin A is in the range of 56% to 72.4%, it will decrease the number of pneumonia cases among toddlers while if the percentage of toddlers given vitamin A is more than 72.4%, it will increase the number of pneumonia cases. The response curve for variable x_6 demonstrates a non-linear relationship between vitamin A supplementation coverage and pneumonia incidence among toddlers. When supplementation coverage ranges between 56% and 72.4%, the declining trend in pneumonia cases reflects the biological role of vitamin A in strengthening immune defence and maintaining respiratory epithelial integrity. Adequate vitamin A status supports mucosal immunity and reduces inflammatory responses in the respiratory tract, which contributes to lower susceptibility to respiratory infections among young children [23]. In addition, clinical studies show that micronutrient supplementation, including vitamin A, is associated with improved pneumonia outcomes and reduced disease severity in pediatric populations [24]. However, when vitamin A coverage exceeds 72.4%, the curve indicates an increase in reported pneumonia cases. From an epidemiological perspective, this pattern reflects the influence of strengthened child health programmes and surveillance activities rather than a harmful biological effect of supplementation. Regions with higher vitamin A coverage generally show better access to health services and more systematic monitoring of child health, leading to increased case detection and reporting. Evidence shows that vitamin A programmes are closely linked to maternal education, health service utilization, and information access, all of which contribute to higher recorded morbidity through improved surveillance sensitivity [25]. Furthermore, studies have identified a significant association between vitamin A supplementation and pneumonia reporting patterns among toddlers, indicating that programme coverage interacts with social and healthcare factors to shape epidemiological data [26]. These findings indicate that the association between vitamin A supplementation and pneumonia incidence is driven by both biological mechanisms and health system dynamics. The non-linear pattern captured by the GAM approach reflects the complex interaction between nutritional interventions, immune function, and surveillance performance in determining observed disease patterns among toddlers.

Furthermore, the plot between the actual data and the estimation results on the response variable is presented in Figure 4.

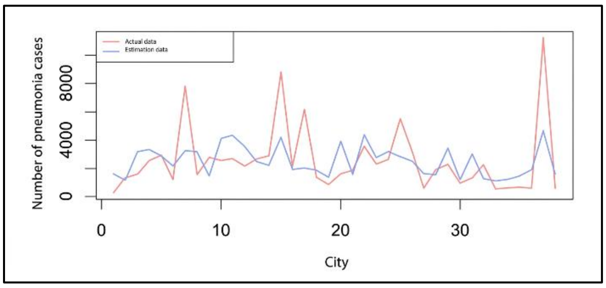

Figure 4. Plot of actual data and GAM model estimation results

Based on Figure 4, the comparison between the actual data and the GAM estimation results shows that the predicted values generally follow the overall pattern of pneumonia cases across cities, although several extreme peaks in the observed data are not fully captured by the model. The proximity between the estimated and actual curves indicates that the GAM approach can capture the population-level trend in pneumonia incidence, suggesting that the selected predictors explain a substantial portion of regional variation in pneumonia cases among toddlers. From a scientific perspective, the differences observed in certain cities reflect epidemiological heterogeneity and surveillance dynamics rather than solely modelling limitations. Variability in reporting systems, healthcare access, and diagnostic practices can influence the number of recorded pneumonia cases, since surveillance sensitivity and case-finding strategies play an important role in determining observed incidence patterns [27]. Scientifically, the ability of GAM to approximate the overall trend while allowing deviations across cities reflects the non-linear nature of epidemiological processes. GAM has been widely applied in disease modelling because its smoothing functions can capture complex exposure–response relationships that arise from environmental and health system factors [28].

The GAM model fit criteria are presented in Table 5, which includes the deviance values for the Poisson and negative binomial distributions.

Table 5. Comparison of model fit criteria between GAM with poisson distribution and negative binomial distribution

Model fit criteria	Poisson distribution	Negative binomial distribution
Deviance value	8932.6	507.7

Based on Table 5, it is known that the deviance value of the negative binomial distribution GAM model is 507.7, this value is less than the deviance value of the Poisson distribution GAM model of 8932.6, so this shows that the GAM model with a negative binomial distribution provides better results than the Poisson distribution GAM model. Therefore, it can be concluded that the GAM model with a negative binomial distribution is appropriate for the number of pneumonia cases among toddlers in East Java Province.

The findings of this study corroborate previous epidemiological research demonstrating the superiority of the Negative Binomial distribution over the Poisson model in addressing overdispersion in health count data. Furthermore, the ability of the GAM to uncover non-linear relationships provides significant theoretical implications for epidemiological modelling. Unlike standard parametric regressions that assume constant, linear effects, the GAM framework reveals the dynamic nature of predictor variables. This indicates the presence of potential threshold effects, suggesting that the effectiveness of public health interventions, such as immunization coverage or improvements in nutritional status, may not be strictly linear and could exhibit varying impacts once certain saturation points are reached within a population.

This study has certain limitations, primarily regarding the scope of the independent variables analyzed. The formulated model incorporated only six predictor variables related to health and sanitation indicators. However, toddler pneumonia is a multifactorial condition that is likely to be influenced by other dimensions not accounted for in this model, such as household socioeconomic status, residential crowding, and climatic factors.

Conclusion

This study concludes that the Generalized Additive Model (GAM) with a negative binomial distribution provides a more appropriate framework for modelling pneumonia incidence among toddlers in East Java Province, as indicated by a deviance value of 507.7, which is substantially less than the deviance value of 8932.6 obtained from the Poisson-based GAM. This difference confirms the presence of overdispersion and highlights the complex epidemiological variability inherent in pneumonia case data. From a scientific perspective, the model reveals that pneumonia dynamics are influenced by non-linear interactions between environmental conditions, nutritional interventions, and regional health system characteristics. Increased public awareness and improved access to healthcare services contribute to earlier detection and more effective management of pneumonia, shaping the observed distribution of reported cases. Therefore, the findings emphasize that flexible non-linear modelling approaches are essential for capturing the biological and epidemiological complexity of respiratory disease patterns in early childhood populations and for supporting evidence-based public health strategies. For future research, it is highly recommended to incorporate a more comprehensive set of predictor variables. The addition of variables from demographic, socioeconomic, and macro-environmental aspects is expected to improve the model's goodness of fit and yield statistical modelling that better represents the true dynamics of pneumonia cases.

Author's Contribution

V.T. Nabila: contributed to the study design, methodological development, data analysis, initial manuscript preparation, and the revision and refinement of the paper. N.R.A.A Siregar: was responsible for software implementation, visual presentation of the results, manuscript revision and improvement.

Acknowledgements
The authors sincerely appreciate the East Java Provincial Health Office for their valuable support and for providing the data resources that contributed to the completion of this study.

References

[1] G. Waterer, "What is pneumonia?," *Breathe*, vol. 17, no. 3, 2021, doi: https://doi.org/10.1183/20734735.0087-2021.
[2] WHO, "Pneumonia in Children," 2022. https://www.who.int/news-room/fact-sheets/detail/pneumonia
[3] Ministry of Health of the Republic of Indonesia, "Riset Kesehatan Dasar (Riskesdas)," 2021. https://www.badankebijakan.kemkes.go.id/laporan-hasil-survei/
[4] Y. F. Falah, Suhartono, M. S. Adi, and A. Mukhibin, "Faktor Risiko Terjadinya Pneumonia Pada Balita Di Puskesmas Sruweng Kabupaten Kebumen," *J. Epidemiol. Kesehat. Komunitas*, vol. 10, no. 3, pp. 19–25, 2025, doi: https://doi.org/10.14710/jekk.v10i3.26735.
[5] B. B. Abate, B. S. Tusa, A. K. Sendekie, and F. G. Araya, "Non-exclusive Breastfeeding is Associated with Pneumonia and Asthma in under-five Children : An Umbrella Review of Systematic Review and Meta-analysis," *Int. Breastfeed. J.*, vol. 20, no. 18, 2025, doi: https://doi.org/10.1186/s13006-025-00712-w.
[6] M. Turaiche *et al.*, "Disease Progression, Clinical Features, and Risk Factors for Pneumonia in Unvaccinated Children and Adolescents with Measles: A Re-Emerging Disease in Romania," *Int. J. Environ. Res. Public Health*, vol. 19, no. 20, 2022, doi: 10.3390/ijerph192013165.
[7] T. Indarwati, A. Y. Salam, and Roisah, "Hubungan Riwayat Imunisasi dan Pemberian Vitamin A dengan Kejadian Pneumonia pada Balita di Rawat Inap RSUD Pasirian Lumajang," *J. Nurs. Res. Publ. Media*, vol. 2, no. 2, pp. 92–102, 2023, doi: 10.55887/nrpm.v2i2.40.
[8] E. Yanuaringsih, D. G. Pradian, R. Pangaribuan, O. Nina, P. F. Gurning, and M. Y. Pratama, "Smoke Exposure at Home to the Incidence of Pneumonia in Children Under 5 Years Old," *Indian J. Forensic Med. Toxicol.*, vol. 15, no. 1, 2021, doi: 10.37506/ijfmt.v15i1.13683.
[9] Y. Teng, Y. Jian, X. Chen, Y. Li, B. Han, and L. Wang, "Comparison of Three Prediction Models for Predicting Chronic Obstructive Pulmonary Disease in China," *Int. J. Chron. Obstruct. Pulmon. Dis.*, vol. 18, pp. 2961–2969, 2023, doi: 10.2147/COPD.S431115.
[10] O. I. Guneri and B. Durmus, "Models for Overdispersion Count Data with Generalized Distribution: An Application to Parasites Intensity," *J. New Theory*, vol. 35, pp. 48–61, 2021, doi: 10.53570/jnt.902066.
[11] T. A. Reddy and G. P. Henze, "Parametric and Non-Parametric Regression Methods," in *Applied Data Analysis and Modeling for Energy Engineers and Scientists*, Springer, Cham, 2023, pp. 355–407. doi: https://doi.org/10.1007/978-3-031-34869-3_9.
[12] S. N. Wood, "Generalized Additive Models," *Annu. Rev. Stat. Its Appl.*, vol. 12, no. 1, pp. 497–526, 2025, doi: https://doi.org/10.1146/annurev-statistics-112723-034249.
[13] A. M. B. de Oliveira, J. M. Binner, A. Mandal, L. Kelly, and G. J. Power, "Using GAM Functions and Markov-Switching Models in an Evaluation Framework to Assess Countries' Performance in Controlling the COVID-19 Pandemic," *BMC Public Health*, vol. 21, no. 2173, 2021, doi: https://doi.org/10.1186/s12889-021-11891-6.
[14] J. Stoklosa, R. V. Blakey, and F. K. C. Hui, "An Overview of Modern Applications of Negative Binomial Modelling in Ecology and Biodiversity," *Diversity*, vol. 14, no. 5, 2022, doi: https://doi.org/10.3390/d14050320.
[15] H. D. Sitanggang, U. Kalsum, and A. N. Putra, "Air Quality and Its Effects on Acute Respiratory Infections Among Children Under Five in Jambi City: A Generalized Additive Model Approach," *J. Heal. Nutr. Res.*, vol. 4, no. 3, 2025, doi: https://doi.org/10.56303/jhnresearch.v4i3.573.
[16] S. Annas, S. Side, M. A. S. Tabbu, Ahmadin, and A. M. R. Y. S. Andi, "Modelling the Impact of Pneumonia Influenced by Air Pollution and Climate Change: A Numerical Approach in South Sulawesi," *Int. J. Simul. Process Model.*, vol. 22, no. 3–4, pp. 220–231, 2025, doi: https://doi.org/10.1504/IJSPM.2025.150532.
[17] F. A. Farisa, S. N. H. Salby, A. A. Rahman, and Purhadi, "Modeling the Number of Pneumonia in Toddlers in East Java Province in 2021 with Generalized Poisson Regression," *Inferensi*, vol. 6, no. 2, pp. 91–96, 2023, doi: http://dx.doi.org/10.12962%2Fj27213862.v6i2.15339.
[18] T. Saifudin *et al.*, "Analysis Indonesian Toddlers Using Nonparametric Regression with Least Square Spline and Fourier Series Methods," *BAREKENG J. Math. Its Appl.*, vol. 20, no. 1, pp. 87–104, 2026, doi: https://doi.org/10.30598/barekengvol20iss1pp0087-0104.
[19] N. Chamidah, M. Rifada, and D. Amelia, "A Theoretical Discussion on Modeling The Number of Covid-19 Death Cases Using Penalized Spline Negative Binomial Regression Approach," *Commun. Math. Biol. Neurosci*, vol. 2, 2022, doi: https://doi.org/10.28919/cmbn/7518.
[20] B. Desye, A. K. Geto, C. Daba, G. Berihun, and L. Berhanu, "Indoor Air Pollution Exposure and Acute Respiratory Infection among Under-five Children in Low- and Middle-income Countries: A Systematic Review and Meta-analysis of Epidemiological Studies," *BMC Infect. Dis.*, vol. 25, no. 1, 2025, doi: 10.1186/s12879-025-12148-y.
[21] K. A. Holden, A. R. Lee, D. B. Hawcutt, and I. P. Sinha, "The Impact of Poor Housing and Indoor Air Quality on Respiratory Health in Children," *Breathe*, vol. 19, no. 2, 2023, doi: 10.1183/20734735.0058-2023.
[22] S. M and S. Vaithilingan, "Childhood Pneumonia in Low- and Middle-Income Countries: A Systematic Review of Prevalence, Risk Factors, and Healthcare-Seeking Behaviors," *Cureus*, vol. 16, no. 4, 2024, doi: 10.7759/cureus.57636.
[23] S. Rao, T. Li, X. Li, L. Hou, and W. Sun, "Vitamin A and Its Related Diseases," *Food Sci. Nutr.*, vol. 13, no.

7, 2025, doi: https://doi.org/10.1002/fsn3.70630.
[24] A. Atta *et al.*, "Investigating the Efficacy of Zinc and Vitamin A in Treating Pediatric Community-Acquired Pneumonia," *Cureus*, vol. 16, no. 1, 2024, doi: 10.7759/cureus.52197.
[25] B. W. Yirdam *et al.*, "Vitamin A Supplementation Coverage and Associated Factors among Ethiopian Children under Five Years: A Systematic Review and Meta-analysis," *BMC Pediatr.*, vol. 25, no. 1, 2025, doi: 10.1186/s12887-025-06030-8.
[26] N. H. Purwati, D. Natashia, and B. F. Marsyah, "The Relationship between Maternal Socioeconomic Status and Vitamin A Provision with Pneumonia Occurrence in Toddlers at Pulogadung Sub-District Health Center, East Jakarta," *J. Ilm. Ilmu Keperawatan Indones.*, vol. 15, no. 1, 2025, doi: https://doi.org/10.33221/jiiki.v15i01.3660.
[27] C. Adamson, Alexander J. Kallis, I. Douglas, and J. K. Quint, "Accuracy of The Recording of Pneumonia Events in English Electronic Healthcare Record Data in Patients with Chronic Obstructive Pulmonary Disease," *Pneumonia*, vol. 16, no. 8, 2024, doi: https://doi.org/10.1186/s41479-024-00130-2.
[28] D. Nerurkar, A. M. Bleichrodt, R. Luo, and G. Chowell, "Predicting Global Pneumonia Mortality Trends by Age Group: A Modeling Study," *Public Health*, vol. 247, 2025, doi: https://doi.org/10.1016/j.puhe.2025.105820.