Poisson regression application for sosiodemography Acute Respiratory Infections (ARI)

I O Hamarandji, A Atti* and M A Kleden
Department of Mathematics, University of Nusa Cendana, Jl. Adisucipto no 10, Kupang, Nusa Tenggara Timur, Indonesia.

*astri_atti@yahoo.com

Abstract. Acute respiratory infection (ARI) is an infectious disease caused by bacteria or viruses that attack the respiratory organs. The case of ARI in toddlers is called Pneumonia. This research aims to determine regression Poisson and negative binomial regression models and significant factors affecting ARI in toddlers in East Nusa Tenggara province based on sociodemography aspects. The method used is the Poisson regression model and the negative binomial regression model. Poisson regression is used to form ARI case models in toddlers in East Nusa Tenggara province based on sociodemography aspects and negative binomial regression to overcome the overdispersion case in Poisson regression. The results showed that the Poisson regression model over dispersed, so the Poisson regression model was not suitable to be used suitable model for ARI cases in toddlers in East Nusa Tenggara based on sociodemography aspects is negative binomial regression models. The result of the data processing obtained a negative binomial regression model: $\mu_i = \exp(3.5260 + 0.021X_2)$. The result of this study indicates that a significant sociodemographic aspect affecting ARI cases in toddlers at NTT is poverty ($X_2$).

1. Introduction
One of the main problems in the health sector facing the world is infectious diseases problem, both those caused by factors that cause them or those caused by lifestyle. Acute respiratory infection (ARI) is a contagious disease. The problems of ARI has tended to increase in the last few decades, both globally and nationally. Nationally, the prevalence of ARI in Indonesia increased from 2010 with a percentage of 23% until 2016 with a percentage of 65.27% [1]. ARI has become the leading killer of toddlers in the world. This disease has become a public health problem in both developed and developing countries [2].

ARI cases in toddlers are called pneumonia caused by an influenza-A virus, adenovirus, and parainfluenza virus. ARI is a disease that often occurs in children. The incidence by age group toddlers is estimated to be 96.7% in developing countries. Most cases occurred in India (43 million), China (21 million), and Pakistan (10 million) as well as Bangladesh, Indonesia, and Nigeria with 6 million episodes each [2].

According to basic health research, the prevalence (period prevalence) of ARI in Indonesia in 2013 was 25.0% [2]. Based on the province, the prevalence of pneumonia in toddlers is 18.5% per mile with five provinces that have the highest of pneumonia in toddlers are East Nusa Tenggara (35.5%), Aceh (35.6%), Bangka Belitung (34.8%), West Sulawesi (34.8%) and Central Kalimantan (32.7%) [2].

Based on basic health research in 2018, East Nusa Tenggara (NTT) ranks 4th out of 34 provinces in this problem with 7.5% of ARI patients [1]. There are 22 districts in NTT with prevalence rates of ARI
from the lowest to the highest is Ngada, Sikka, Malacca, Sabu Raijua, Rote Ndao, Nagekeo, East Manggarai, Alor, Kupang City, North Central Timor, Central Sumba, Lembata, West Sumba, Ende, West Manggarai, South Tengan Timor, Belu, Manggarai, East Flores, Southwest Sumba, East Sumba, and Kupang District. Kupang District is an area with the highest ARI prevalence rate with a percentage of 61.07%. Until now, ARI is also the main cause of infant and toddler mortality in NTT with a percentage of 80% to 90% [3].

Research by M. Habibi Syahidi shows that ARI in toddlers is caused by various factors including environmental factors, individual child factors, and behavioral factors. These factors are related to population characteristics. Sociodemography is a scientific study about the structure and processes of the population in an area where structural changes of the population are also influenced by social processes and social changes in the society. Demography does not study the population as individuals, but the population as a group [4]. The three sociodemography aspects studied relate to factors causing ARI, where population density is a part of environmental factors, the poverty is a part of behavioral factors, and malnutrition state is a part of individual child factors.

This research aims to determine regression Poisson and negative binomial regression models and significant factors affecting ARI in toddlers in East Nusa Tenggara province based on sociodemography aspect. The method used is Poisson regression model and negative binomial regression model. Poisson regression is used to form ARI case models in toddlers in East Nusa Tenggara province based on sociodemographic aspects and negative binomial regression to overcome the overdispersion case in Poisson regression.

Poisson regression is a nonlinear regression analysis modeled by the Generalized Linear Model (GLM). Poisson regression analysis is used when dependent variable data to be analyzed is form count data, and independent variables data are categorical, interval, or count data. The Poisson regression analysis has to fulfill the equidispersion assumption, which is the assumption of the similarity between mean and variance of the dependent variable. In practice, there are often differences in variance and mean values of data. Differences in variance value and mean indicates an overdispersion condition. Overdispersion is a condition of variance value greater than mean. If there is a violation of equidispersion assumption, where the variance greater than the mean in this study, then the overdispersion case will be resolved with a negative binomial regression model [5].

2. Research methods
The variables in this research are:
   a. Dependent Variable (Y) is the number of ARI cases in toddler in East Nusa Tenggara Province in 2018.
   b. Independent Variable (X) is the cause of ARI in toddlers in East Nusa Tenggara Province based on sociodemography aspects, includes:
      • Population Density ($X_1$) in people/$km^2$
      • Poverty ($X_2$) in thousand inhabitants
      • Malnutrition States ($X_3$) in number of people

The steps taken in this research method are:
   • Performed Kolmogorov-Smirnov test on dependent variable Y
   • Determine the Poisson Regression model
   • Identify overdispersion by calculating the dispersion coefficient
   • Determine the binomial negative regression model
   • Significance parameters Test
   • Interpret the model
3. Results and discussion

3.1. Kolmogorov-Smirnov test

Kolmogorov-Smirnov test was conducted to test the suitability of Poisson distribution on dependent variable $Y$, that is number of cases of acute respiratory infections (ARI) data, with the hypothesis:

$H_0$: Poisson distribution data
$H_1$: not Poisson distribution data

The criterion for this test is to reject $H_0$ if $D_{count}$ greater than $D_{table}$ value with statistic test $D = \max \left| F_0(x) - S_n(x) \right|$. The Kolmogorov-Smirnov test results are presented in Table 1.

**Table 1. The result of Kolmogorov-Smirnov test.**

| One-Sample Kolmogorov-Smirnov Test | Y   |
|------------------------------------|-----|
| N                                 | 22  |
| Most Extreme Differences           | Absolute 0.188 |

Based on Table 1, the result of can Kolmogorov-Smirnov test obtained the value of $D_{count} = 0.188$ which is contained in Most Extreme Differences in the Absolute value. Meanwhile, $D_{table}$ value of Kolmogorov-Smirnov test is 0.281 with a significant level of $\alpha = 5\%$. The value of $D_{count} < D_{table}$ then $H_0$ is accepted and it can be concluded that the number of cases of acute respiratory infections (ISPA) data in toddlers in East Nusa Tenggara Province follows the Poisson distribution.

3.2. Regression poisson models

Generally, Poisson regression models formed for acute respiratory infections (ISPA) cases in toddlers in East Nusa Tenggara Province based on sociodemography aspects is (see equation 1 and table 2):

$$\mu_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$$ (1)

With:
- $\beta_0$ : intercept or parameter when the value of $X_1, X_2,$ and $X_3$ are 0 (constant)
- $\beta_1$ : parameter variable $x_1$
- $\beta_2$ : parameter variable $x_2$
- $\beta_3$ : parameter variable $x_3$

**Table 2. The results of poisson regression analysis in modeling ARI case.**

| Parameter | $\beta$ | P-Value |
|-----------|---------|---------|
| (intercept) | 3.689 | 0.000 |
| $X_1$ | -0.001 | 0.000 |
| $X_2$ | 0.018 | 0.000 |
| $X_3$ | 0.000 | 0.000 |

Poisson regression models formed from ARI cases in toddlers in NTT based on sociodemography aspects is:

$$\mu_i = \exp(3.689 - 0.001x_1 + 0.018x_2 + 0.000X_3)$$ (2)

The results of Poisson regression analysis indicate that the Poisson regression model can explain acute respiratory infections (ARI) in toddlers in East Nusa Tenggara Province based on sociodemography
aspects significantly with significance value on omnibus test output lesser than 0.05. This result shows that the sociodemography aspect in NTT significantly affects ARI cases in toddlers.

Regression Poisson models indicate that ARI cases in toddlers in NTT Province are influenced by the sociodemography aspects, that is population density factors \(X_1\), poverty \(X_2\) and malnutrition states \(X_3\), with each of their \(P\)-Value lesser than 0.05.

3.3. Poisson overdispersion test

Overdispersion Poisson in table 3 shows that deviance and Pearson Chi-Square value greater than 1. This result indicates that the Poisson regression model does not meet the equidispersion assumption or is overdispersed then Poisson regression model is not suitable for modeling ARI cases in toddlers in NTT based on sociodemographic aspects. Therefore, a further analysis was carried out to overcome the overdispersion case is using the negative binomial regression model.

| ISPA              | Value  | df  | Value/df  |
|-------------------|--------|-----|-----------|
| Deviance          | 2436.025 | 18  | 135.335   |
| Pearson Chi-Square| 2445.778 | 18  | 135.877   |

3.4. Binomial negative regression models

Binomial Negative regression is typically used when there is overdispersion in Poisson regression models. Generally, binomial negative regression model formed for ARI cases in toddlers in East Nusa Tenggara Province based on sociodemography aspects is:

\[ \mu_i = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3) \]  

With:
- \(\beta_0\): intercept or parameter when the value of \(X_1, X_2,\) and \(X_3\) are 0 (constant)
- \(\beta_1\): parameter variable \(x_1\)
- \(\beta_2\): parameter variable \(x_2\)
- \(\beta_3\): parameter variable \(x_3\)

| Parameter Estimate | Parameter | Wald Chi-Square | P-Value |
|--------------------|-----------|----------------|---------|
| (intercept)        | 3.526     | 42.570         | 0.000   |
| \(X_1\)            | 0.000     | 0.641          | 0.425   |
| \(X_2\)            | 0.021     | 5.887          | 0.015   |
| \(X_3\)            | 0.000     | 2.839          | 0.092   |

Table 4 presents the results of binomial negative regression analysis which shows that binomial negative regression model formed from ARI cases in toddlers in NTT based on sociodemography aspects is only influenced by the poverty factor \(x_2\) (p-value < 0.05) with the models \(\mu_i = \exp(3.5260 + 0.021x_2)\).

Negative binomial regression model showed that the sociodemographic aspect that significantly affected on ARI cases in toddlers in NTT was poverty factor \(x_2\). The model states that every increase in poverty rate by one thousand people will increase the ARI cases in toddlers in NTT by an average of \(\exp(0.021)\) or 1.0212220516 cases in each district. In other cases, the total number of ARI cases toddlers in NTT increased by around 22 cases if poverty in NTT increased by a thousand people. Twenty-two
cases of ARI in toddlers occurred in NTT, which means that there were districts that had no ARI cases and there were districts that had more than one ARI case.

3.5. Poverty and ARI cases in toddler in NTT
Poverty in East Nusa Tenggara Province is measured by calculating the poverty line which consists of two components, that is the Food Poverty Line and non-Food Poverty Line. Based on the poverty profile in 2018, the number of poor people in East Nusa Tenggara in March 2018 was 1,142.17 thousand people (21.35 percent), an increase of around 7,430 people compared to the poor population in September 2017 which amounts to 1,134.74 thousand people (21.38 percent) [6]. Poverty in NTT has increased.

East Nusa Tenggara is a province that is very thick with the culture that contributes to the level of poverty which is called Cultural poverty [7]. Based on the residence, from September 2017 to March 2018, the number of poor people in rural areas increased by 4,510 people, from 1,015.70 thousand people to 1,020.21 thousand people and urban areas also experienced an increase of 2,910 people, from 119.04 thousand people to 121.95 thousand people [6].

Poverty culture affects to low education level and public knowledge about healthy and hygerines. Many toddlers in direct contact with kitchen smoke without proper air ventilation, toddlers in direct contact with outside dust which can infect the toddler's respiratory tract. Home cleanliness is a risk factor for ARI in toddlers. The unhealthy house can interfere with the health of its occupants and increase the risk of various diseases [8].

Based on WHO data (2000), more than two million poor people in the world still depend on to use biomass (wood, charcoal, animal dung, coconut dregs) and they use of coal as household energy needs. These materials have an impact on increasing indoor air pollution that exceeds the applicable international air quality standards, which caused to exposure of children who live in poverty in their daily lives. This exposure increases the risk of diseases such as acute respiratory infections and lung cancer.

4. Conclusion
The best model for ARI cases in toddlers in East Nusa Tenggara province based on sociodemography aspects is a negative binomial regression model and a significant factor affecting ARI is poverty. The binomial regression model that is formed is:

$$\mu_i = exp(3.5260 + 0.021x_2)$$  \hspace{1cm} (4)

5. Recommendation
The next researchers can study Poisson regression with overdispersion settlement using the following method: Generalized Poisson Regression, Zero Inflated Poisson Regression, Regresi Poisson Inverse Gaussian, Zero-Inflated Negative Binomial.

References
[1] Ministry of Health RI 2018 Main Results of Basic Health Research (Jakarta)
[2] Waworuntu W 2016 Guidelines for the Prevention and Control of Acute Respiratory Infections (Jakarta: Disease Control and Prevention Team)
[3] East Nusa Tenggara Provincial Health Office 2018 Health Profile (Kupang)
[4] Hauser P M and Duncan O D, eds. 1959 The Study of Population: An Investory and Appraisal (Chicago: The University of Chicago Press)
[5] Zahro J, Caraka R, Herliansah R 2018 Aplikasi generalized linear model pada R (Yogyakarta: Innosain)
[6] East Nusa Tenggara Statistic Center Agency 2018 poverty and inequality [Online] available on the page https://ntt.bps.go.id/dynamictable/2015/03/03/12/jumlah-penduduk-miskin-menurut-kabupaten-kota-di-provinsi-nusa-tenggara-timur-2002-2019-ribu-jiwa-.html (accessed on 27 February 2020, 15:45:27)
[7] Ministry of Public Works and Public Housing 2007 *Cultural poverty and Focus Discussion Group* (Jakarta)

[8] Atti A, Dodo D O, Purimahua S L 2018 Relationship between Unhealthy Homes with the Incidence of Acute Respiratory Infections (ARI) *IOSR J Nurs Health Sci IOSR-JNHS* 7(3 Ver. VII): 4. pp 21-24