Statistical Modeling on the Risk Factors of Pneumonia

ISSN 2672-3107 (Print) • ISSN 2704-288X (Online)
PHILIPPINE SOCIAL SCIENCE JOURNAL
Advancing Social Science Research in the Philippines and Beyond

Virgelio M. Alao[1*] and Anthony D. Bartolore[2]
[1,2]Department of Statistics, Visayas State University, Baybay City, Leyte, Philippines

Article history:
Submitted: 25 February 2022
Revised: 22 March 2022
Accepted: 25 March 2022

Keywords:
Binary logistic regression
Pneumonia
Risk factors

****Corresponding Author:***
Virgelio M. Alao
virgilmonter@gmail.com

ABSTRACT. In our world today, pneumonia is among the common diseases that causes difficulty in breathing for human. In the Philippines, it is one of the deadliest diseases that cut off the life of the citizens. Its direct or indirect negative effects pose threat to the health and life expectancy of human beings. This study determines the risk factors of pneumonia by applying Binary Logistic Regression, then develop a model to estimate the probability that an individual may acquire pneumonia based on his or her medical and non-medical records. Results showed that the best fitting model contains eight risk factors, namely, Pulmonary Tuberculosis (PTB), Bronchitis, Chronic Obstructive Pulmonary Disease (COPD), duration of hospital stay, age, acute gastritis, cerebral infarction and dehydration. The first four risk factors and age (of at least 65 years) increase the odds of having pneumonia while the last three risk factors decrease the odds of having pneumonia.

Introduction

In 2013, pneumonia remained to be the top leading cause of death in children worldwide. Males were more prone to pneumonia having 18.6% incidence compared to females having only 13.9% incidence worldwide (World Health Organization (WHO), 2013). Moreover, it is the seventh leading cause of death in people at least 65 years old (American Lung Association, 2013). In 2016, Pneumonia is one of the leading cause of mortality in children under 5 years old, and specifically a child dies in every 20 seconds worldwide (WHO, 2016).

In the Philippines, pneumonia ranked third as the leading cause of mortality in 2016. Moreover, it is ranked fourth as the leading cause of death in both males and females and ranked first as the leading cause of death of infants. In Region VIII (Eastern Visayas), Leyte province has the highest number of deaths of pneumonia with the rate of 73.2 per 100,000 individuals (Philippine Statistics Authority (PSA), 2013). In 2016, pneumonia ranked first in Region VIII (PSA, 2016).

Tong (2013) stated several factors that affect pneumonia and these are from the environment which are overcrowded living conditions, and exposure to indoor and outdoor air pollutions. For infants, these are Auto Immune Deficiency Syndrome (AIDS), measles, and malaria. On the other hand, Comfort Keepers (2020) mentioned that for elderly people, these are weakened immune system, diabetes, Parkinson's disease, chemotherapy, HIV, cystic fibrosis, asthma, COPD (Chronic Obstructive Pulmonary Disease), bronchiectasis. Meanwhile, Coupland et al. (2009) emphasized that immune system and the age of an individual greatly affect pneumonia because the individuals that are at highest risk of this disease are adults aged at least 65 years old and also infants and young children who do have weak immune system. Furthermore, Srivastava et al. (2015) emphasized that overcrowding and malnutrition significantly affects pneumonia in infants and children aged younger than five years old. Given the limited literature of the risk factors of pneumonia and its grave threat to the people, thus this study was conducted.

The study focused on determining risk factors of pneumonia incidence in Baybay City, Leyte, Philippines. This study specifically aimed to: (1) describe pneumonia patients with respect to sex, age, address, and other medical diagnosis such as hypertension, heart disease, asthma, upper respiratory tract infection (URTI), bronchitis, chronic obstructive pulmonary disease (COPD), dehydration,

urinary tract infection (UTI), acute gastritis, seizure disorder, diabetes, cerebral infarction, pulmonary tuberculosis, osteoarthritis, intestinal parasitism, and anemia; (2) develop a binary logistic regression model which can be used to estimate the probability that an individual may acquire pneumonia based on his or her medical and non-medical records; and (3) validate the binary logistic regression model in terms of sensitivity, specificity, precision, accuracy, and overall misclassification percentage in classifying pneumonia incidence.

The result of this study would help increase the awareness of people about pneumonia and its risk factors and improve prevention measures against pneumonia. Medical practitioners could gain more knowledge about the risk factors of pneumonia incidence and plan out interventions, policies, and guidelines in reducing its incidence. The local government unit could also improve the implementation of primary health care services (Gella and Caelian, 2021) to minimize the incidence of pneumonia affected by the different risk factors.

2.0. Methodology

A quantitative research design was used in this study. Descriptive statistics was utilized to describe and summarize the pneumonia patients with respect to sex, age, address, and other medical diagnosis such as hypertension, heart disease, asthma, upper respiratory tract infection (URTI), bronchitis, chronic obstructive pulmonary disease (COPD), dehydration, urinary tract infection (UTI), acute gastritis, seizure disorder, diabetes, cerebral infarction, pulmonary tuberculosis, osteoarthritis, intestinal parasitism, and anemia. On the other hand, inferential statistics was employed to estimate the probability that an individual may acquire pneumonia based on his or her medical and non-medical records, validate the develop model, and draw conclusions from the validated model along with the significant variables.

A secondary data was used in this study from the medical and non-medical records of 1,790 patients admitted at Western Leyte Provincial Hospital (WLPH), Baybay City, Leyte, Philippines and residing at Baybay City, Leyte, Philippines from January 2015 to July 2018. These records consist of patients' age, sex, address, duration of the hospital stay, and final diagnoses (autoimmune diseases) that has something to do with pneumonia such as heart diseases, hypertension, chronic obstructive pulmonary disease (COPD), bronchitis, asthma, upper respiratory tract infection (URTI), dehydration, urinary tract infection (UTI), acute gastritis, seizure disorder, diabetes, cerebral infarction, pulmonary tuberculosis, osteoarthritis, intestinal parasitism, and anemia.

In this study, age was categorized based on the preferred age group by Department of Health (DOH) (0 to 4, 5 to 14, 15 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and above) to determine what age group has the highest pneumonia incidence. In addition, the address of patients was grouped by district, namely; north, east, poblacion, and south, to compare the magnitude in terms of environment that affects pneumonia. Another risk factor was the duration of the hospital stay which is calculated as the number of days from admission to hospital discharge. Other potential risk factors are autoimmune diseases identified from the final diagnosis.

Data were analyzed using descriptive statistics including pie chart, frequency polygon, and horizontal bar chart. A binary logistic regression model (Tranmer and Elliot, 2008) was also fitted to estimate the probability that an individual may acquire pneumonia based on his or her medical and non-medical records. Hosmer-Lemeshow statistics, Likelihood Ratio Chi-Square Test, Cox and Snell R^2, Akaike information criterion (AIC), and Bayesian information criterion (BIC) were used to test the goodness-of-fit of the model derived consisting of the risk factors of pneumonia (Hosmer and Lemeshow, 1989; Cox and Snell, 1989). Wald's test was used to assess the statistical significance of the potential predictors or risk factors of pneumonia. Sensitivity, specificity, precision, and accuracy were also calculated to validate the model (Giancristofaro and Salmaso, 2003) with its significant predictors or risk factors of pneumonia.

This study has observed an ethical procedure. Prior to the conduct of the study, a consent/permission letter was sent to the head of the hospital. Upon approval of the consent letter, the researchers coordinated and asked the person in charge of the data in the hospital for data collection. The hospital together with its data officer is well informed that the study respects the patients' privacy of their medical records and conformed the Data Privacy Act of 2012 (RA 10173).

3.0. Results and Discussion

Profile of patients confined at WLPH

Out of 1,790 patients confined at Western Leyte Provincial Hospital (WLPH) from January 2015 to July 2018, more than half (56%) were diagnosed with pneumonia.

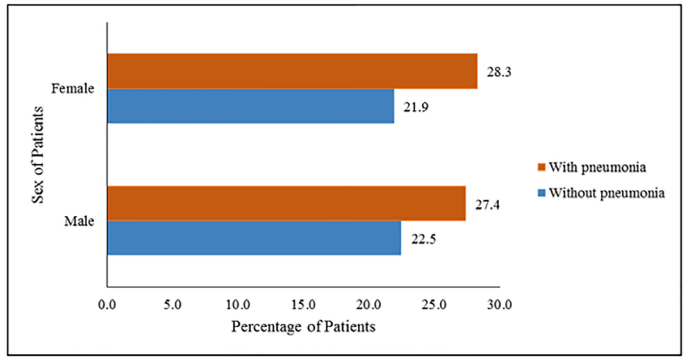

Figure 1. Percentage distribution of patients by sex and pneumonia incidence (n=1,790)

Figure 1. Percentage distribution of patients by sex and pneumonia incidence (n=1,790)

Figure 1 shows that the pneumonia incidence in both sexes have almost equal percentages (28.3% and 27.4%). These percentages was contradicted by a study of Evaristo-Mendez and Rocha-Calderon (2015) that majority of pneumonia patients are female (62%) and 38% are male, but still they found that patient's sex was not a significant risk factor of having pneumonia.

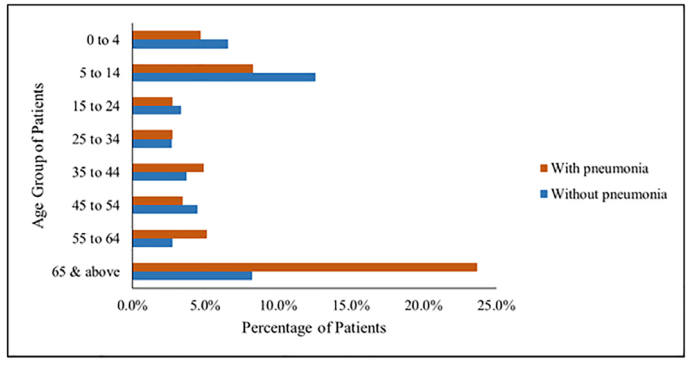

Figure 2. Percentage distribution of patients by age groups and pneumonia incidence (n=1,790)

Figure 2. Percentage distribution of patients by age groups and pneumonia incidence (n=1,790)

In Figure 2, patient's age were grouped into eight and ages 65 and above has the highest pneumonia incidence (23.7 %), followed by ages 5 to 14 years (8.3%). This result was supported by Racaza (2015), she found that older individuals (60 years old and above) have the highest pneumonia incidence among age groups having 38.5%. Furthermore, she concluded that younger individuals were less likely to have pneumonia than older individuals. Maybe because older individuals suffer from weakening and life-threatening conditions, such as heart disease which makes them more vulnerable to pneumonia (Mylotte et al., 2007).

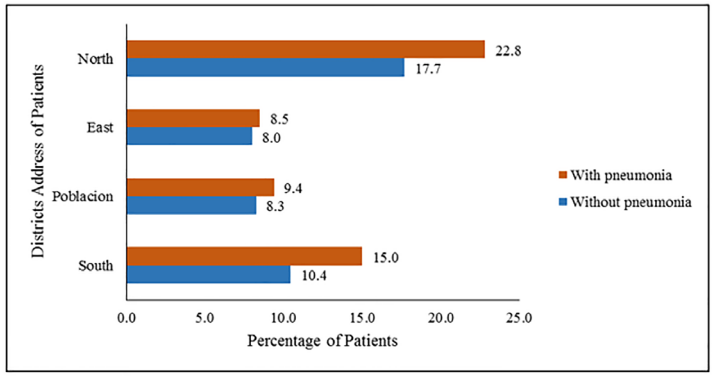

Figure 3. Percentage distribution of patients' address by district and pneumonia incidence (n=1,790)

Figure 3. Percentage distribution of patients' address by district and pneumonia incidence (n=1,790)

Patients' address were divided into four districts as shown in Figure 3. Many of the pneumonia cases occurred in North District (22.8%), followed by South District (15.0%). This is due to the fact that the North District has the highest average population from the year 2013 to 2017. Also, there were environmental risk factors that cause pneumonia (WHO, 2016) which might be present in north district. These are indoor air pollution that came from cooking using biomass fuels like woods and (or) dungs, crowded area, first and second hand smokes just like smoke emissions from cigarette smokers.

In Figure 4, pneumonia incidence is high in patients with UTI (10.7%), heart disease (8.4%), hypertension (7.7%), pulmonary tuberculosis (5.1%), COPD (5.0), bronchitis (4.1%), asthma (3.8%), and acute gastritis (2.8). While pneumonia incidence is low in patients with URTI (1.4%), dehydration (1.2%), intestinal parasitism (0.8%), cerebral infarction (0.7%), anemia (0.6%), diabetes (0.6%), seizure disorder (0.6%), and osteoarthritis (0.6%). This result was supported by the study of Racaza (2015) that having UTI was not a significant risk factor of having pneumonia. Moreover, in her study, lung diseases such as pulmonary tuberculosis, bronchitis, COPD, and asthma were significant risk factors of having pneumonia.

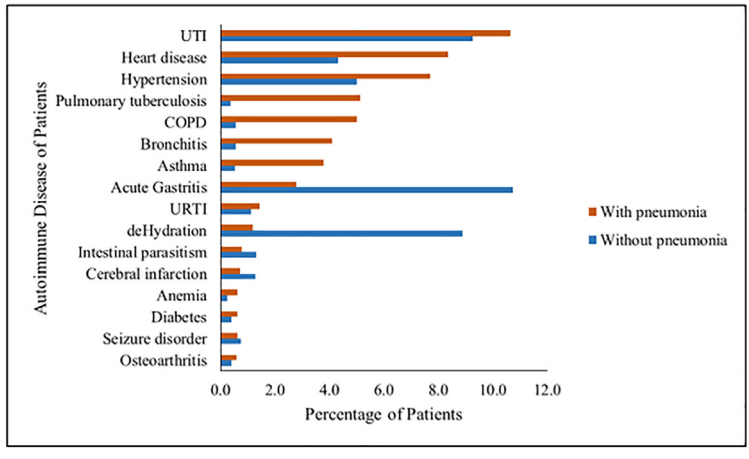

Figure 4. Percentage distribution of patients by final diagnosis (autoimmune diseases) and pneumonia incidence (n=1,790)

Figure 4. Percentage distribution of patients by final diagnosis (autoimmune diseases) and pneumonia incidence (n=1,790)

Figure 5 implies that more pneumonia patients stayed longer in the hospital compared to non-pneumonia patients. This scenario was supported by Huang et al. (2006) that the length of hospital stay of pneumonia patients was associated with functional status (walking independently, walking with assistance, and walking with a prosthesis), time to receive a first dose of antibiotic therapy, use of certain antibiotics, presence of urinary catheter and the importance of time to physiological stability.

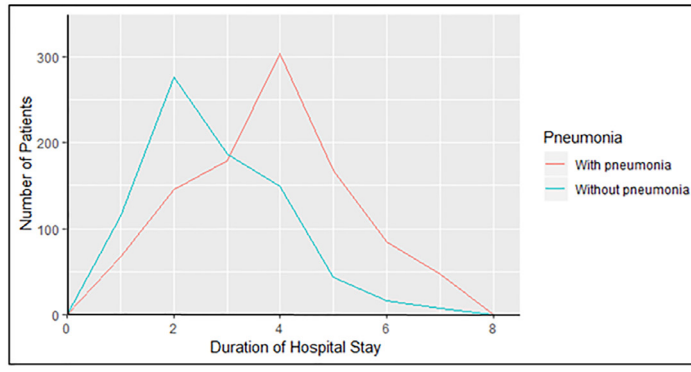

Figure 5. Frequency distribution of the patients by duration of hospital stay and pneumonia incidence (n=1,790)

Figure 5. Frequency distribution of the patients by duration of hospital stay and pneumonia incidence (n=1,790)

Binary logistic regression model

Prior to data splitting, assumptions of binary logistic regression were checked and there were no assumptions violated since the response is binary (having pneumonia or not), there were no influential values (extreme values) present, there was also low multicollinearity among predictors, and there was a linear relationship between duration of hospital stay and the logit of having pneumonia or not.

Data set was split into training set (70%) and validation set (30%). Forward selection was used in generating binary logistic regression models from the training set. Table 1 shows the AIC and BIC of the generated model. Model 8 is considered as the best model because it has the smallest AIC and BIC value among the eight binary logistic regression models generated. This implies that this model is the closest to the true model of predicting pneumonia incidence.

Table 1. Information Criterion of Models 1 to 8

Model	AIC	BIC
1	1464.373	1466.569
2	1342.233	1345.527
3	1286.64	1291.032
4	1232.483	1237.973
5	1191.038	1197.626
6	1163.206	1170.892
7	1151.756	1160.54
8	1144.612	1154.494

The best model was then tested for goodness-of-fit as shown in Table 2 and it shows goodness-of-fit based on Hosmer-Lemeshow statistics wherein the p-value is greater than 0.05 and the likelihood ratio chi square wherein the p-value < 0.001. Moreover, 37.5% of the variation in the probability of having pneumonia can be explained by or attributed to the risk factors in the best model.

Table 2. Goodness-of-fit statistics for the best-fitted binary logistic regression model

No. of Risk Factors	Hosmer and Lemeshow 's statistics	H-L stat. p-value	Likelihood Ratio Chi-Square Test	χ^2 test p-value	Cox & Snell R^2	AIC	BIC
8	8.31	0.404	588.62	0.000	0.375	1144.61	1154.50

The best-fitted model consists of eight significant risk factors of having pneumonia (Table 3). Four of them and age at least 65 years old will increase the odds of having pneumonia while the other three risk factors will decrease the odds of having it.

Increase in odds of having pneumonia were observed in the patients diagnosed with lung diseases such as pulmonary tuberculosis, bronchitis, chronic obstructive pulmonary disease (COPD), and staying longer in the hospital. The mentioned lung diseases were also significant risk factors of having pneumonia in a study of Racaza (2015). The odds of having pneumonia is 15.43 times higher for patients with pulmonary tuberculosis (PTB) than those without PTB assuming all else are equal. Patients with bronchitis has 9.63 times higher odds of having pneumonia than those without bronchitis assuming all else are constant. The odds of having pneumonia is 5.0 times higher for patients with COPD than those without COPD assuming all else are equal. In terms of patient's hospital stay, there is 0.611 times higher odds of having pneumonia for every successive days of hospital stay assuming all else are equal.

Table 3. Coefficients of best-fitted binary logistic regression model

Risk Factors	β	S.E.	Wald	Sig.	Exp(β)
Pulmonary Tuberculosis(1)	2.737	0.483	32.079	0.000	15.433**
Bronchitis(1)	2.265	0.372	37.133	0.000	9.628**
COPD(1)	1.610	0.381	17.886	0.000	5.003**
Duration of Hospital Stay	0.477	0.053	79.839	0.000	1.611**
Acute Gastritis(1)	-0.774	0.256	9.115	0.003	0.461**
Cerebral Infarction(1)	-1.347	0.353	14.523	0.000	0.260**
Dehydration(1)	-1.733	0.335	26.698	0.000	0.177**
Age Group			32.977	0.000	**
0 to 4 years old	-0.455	0.274	2.761	0.097	0.635ns
5 to 14 years old	-0.611	0.225	7.410	0.006	0.543**
15 to 24 years old	-1.100	0.326	11.352	0.001	0.333**
25 to 34 years old	-0.499	0.330	2.283	0.131	0.607ns
35 to 44 years old	-0.605	0.274	4.869	0.027	0.546*
45 to 54 years old	-1.495	0.288	27.021	0.000	0.224**
55 to 64 years old	-0.466	0.290	2.573	0.109	0.628ns
Constant	-0.721	0.223	10.437	0.001	0.486**

Reference notes: Pulmonary Tuberculosis(0) – patients without pulmonary tuberculosis, Bronchitis(0) – patients without bronchitis, COPD(0) – patients without COPD, Dehydration(0) – patients without dehydration, Acute Gastritis(0) – patients without acute gastritis, Cerebral Infarction(0) – patients without cerebral infarction, reference age group – 65 years old and above

* - significant at 5% level of significance

** - significant at 1% level of significance

ns - not significant

Patient's age (65 years old and above – the reference group) was also a significant risk factor of acquiring pneumonia. The odds of having pneumonia for patients aged five to 14 years old is 0.543 times lower than those patients aged 65 years old or greater assuming other things held constant. Patients aged 15 to 24 years old has 0.333 times lower odds of having pneumonia than patients aged 65 years old and above assuming all other things held constant. The odds of having pneumonia for patients aged 35 to 44 years old is 0.546 times lower than patients aged 65 years old and above assuming all other things held constant. And lastly, patients aged 45 to 54 years old has 0.224 times lower odds of having pneumonia compared to patients aged 65 years old or greater assuming all other things remain unchanged. On the other hand, patients with acute gastritis has 0.461 lower odds to acquire pneumonia assuming all other things held constant, other risk factors that decrease the odds of having pneumonia were Cerebral Infarction (0.260), and Dehydration (0.177) ceteris paribus.

Validation of the Model

The calculated minimum cut-off point was 57% based on the mean of the predicted probabilities of the best-fitted model. On the average, the derived model has 80.5 percent (Specificity) ability of correctly classifying patients without pneumonia and 74.6 percent (Sensitivity) in classifying pneumonia patients correctly as shown in Tables 4 and 5. Moreover, this model is 70.9 percent accurate and 83.3 percent precise as shown in Table 4.

Table 4. Average Performance Evaluation of the Best-Fitted Model

Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)
74.6	80.5	70.9	83.3

Table 5. Classification Rates of the Best-Fitted Model

Diagnosis of Pneumonia		Predicted Membership	
		Without Pneumonia	With Pneumonia
ObservedMembership	WithoutPneumonia	80.5%	19.5%
	WithPneumonia	25.4	74.6%
	Overall Correct Classification		77.2%
	Overall Misclassification		22.8%

4.0. Conclusion

The best-fitted model which explains the probability of having pneumonia consists of eight risk factors, namely; pulmonary tuberculosis, bronchitis, COPD, staying longer in the hospital, age, acute gastritis, cerebral infarction, and dehydration. Moreover, patient having pulmonary tuberculosis, bronchitis, COPD, staying longer in the hospital, and aged 65 and above were at the highest risk of acquiring pneumonia. On the other hand, patients having acute gastritis, cerebral infarction, and dehydration would less likely to have pneumonia.

It is highly recommended that patients who are 65 years old or greater and have been diagnosed with pulmonary tuberculosis, bronchitis, COPD must be given more medical attention and care because they are more vulnerable to pneumonia disease. It is also suggested that future research in pneumonia may include information about monthly or quarterly number of pneumonia patients where presence of seasonal variation of pneumonia incidence can be evaluated as one of the potential risk factors. Another is to perform modelling about the three types of pneumonia (community acquired, hospital acquired, and ventilator acquired pneumonia) by using multinomial logistic regression.

5.0. Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

6.0. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

American Lung Association. (2013). Five Facts You Should Know About Pneumonia. Retrieved February 3, 2019 from https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/pneumonia/five-facts-you-should-know.html

Comfort Keepers. (2020). *Pneumonia in Seniors: Prevention and Treatments*. Published on September 21, 2020. Retrieved from: https://www.comfortkeepers.com/articles/info-center/senior-health-and-wellbeing/pneumonia-in-seniors-prevention-and-treatments

Coupland, C., Hippisley-Cox, J., & Vinogradova, Y. (2009). Identification of new risk factors for pneumonia: population-based case-control study. *British Journal of General Practice*. https://doi.org/10.3399/bjgp09X472629.
Cox, D. R., & Snell, E. J. (1989). The Analysis of Binary Data, 2nd ed. London: Chapman and Hall.
Evaristo-Mendez, G., & Rocha-Calderon, C.H. (2015). *Risk factors for nosocomial pneumonia in patients with abdominal surgery*. https://doi.org/10.1016/j.circir.2015.05.051
Gella, H. B., & Caelian, M. V. (2021). Implementation of primary healthcare services in community health stations in the highly urbanized city. Philippine Social Science Journal, 4(3), 123-132. https://doi.org/10.52006/main.v4i3.385
Giancristofaro, R.A., & Salmaso, L. (2003). *Model Performance Analysis and Model Validation in Logistic Regression.* Statistica, anno LXIII, n. 2, 2003. https://doi.org/10.6092/issn.1973-2201/358
Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.
Huang, J. Q. Hooper, P. M. & Marrie, T. J. (2006). Factors Associated with Length of Stay in Hospital for Suspected Community-Acquired Pneumonia. Can Respir J 2006;13(6):317-324.
Mylotte, J. M., Raghavendran, K., & Scannapieco, F. A. (2007). *Nursing Home-Associated Pneumonia, Hospital-Acquired Pneumonia, and Ventilator-Associated Pneumonia: The Contribution of Dental Biofilms and Periodontal Inflammation*. Periodontology 2000, 44(1), 164-177.
Philippine Statistics Authority (2013). *2017 Regional Social and Economic Trends Eastern Visayas Region*. Philippine Statistics Authority Regional Statistical Services Office VIII Leyte SR Building, Artemio Mate Road, Brgy. Abucay 6500 Tacloban City Philippines.
Philippine Statistics Authority (2016). *Death Statistics in the Philippines*. Civil Registration and Vital Statistics Fact Sheet.
Racaza, J. (2015). *Determinants of the Incidence of Pneumonia at Gatchalian Hospital in Ormoc City*, Leyte. Special Problem. Visayas State University
Srivastava, P., Mishra, A. K. & Roy, A. K. (2015). *Predisposing Factors of Community Acquired Pneumonia in Under-Five Children*. Lung Dis Treat 1: 101. DOI: 10.4172/2472-1018.1000101
Tong, N. (2013). *A Public Health Approach to Innovation.* Institute of Health Metrics and Evaluation 2010. https://www.who.int/medicines/areas/priority_medicines/BP6_22Pneumo.pdf
Tranmer, M. & Elliot, M. (2008). *Binary Logistic Regression*. Cathie Marsh Centre for Census and Survey Research
World Health Organization (2013). *Pneumonia still responsible for one fifth of child deaths*.https://www.who.int/mediacentre/news/releases/2013/world-pneumonia-day-20131112/en/
World Health Organization (2016), *Pneumonia,* Retrieved from https://www.who.int/news-room/fact-sheets/detail/pneumonia

Additional Authors' Information:

VIRGELIO M. ALAO*
virgilmonter@gmail.com
https://orcid.org/0000-0002-7345-8958

ANTHONY D. BARTOLORE
bartoloreanthony@gmail.com
https://orcid.org/0000-0003-2695-2168