Research Article

Risk factors of pneumonia among elderly with robust Poisson regression - A study on mimic III data

Kalesh M. Karun[1], Amitha Puranik[2], Lintu M. K.[3] Deepthy M. S.[4]

[1]ICMR- National Institute of Traditional Medicine, Dept. of Health Research (Govt. of India), Nehru Nagar, Belagavi, 590010, Karnataka, India

[2]Centre for Health Informatics, Computing and Statistics, Lancaster Medical School, Lancaster University, LA1 4YW, United Kingdom

[3]Department of Data Science, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India

[4]Department of Biostatistics, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, 605006, India

(Received: November 2022 Revised: March 2023 Accepted: April 2023)

Corresponding author: **Kalesh M. Karun.** Email: karunkmk@gmail.com

ABSTRACT

Introduction and Aim: Pneumonia is a common and serious illness among the elderly. Early identification of the risk factors for pneumonia is essential for improving the survival outcomes among elderly. The present study aimed to identify an optimal regression approach to determine the risk factors for pneumonia among elderly patients.

Materials and Methods: The present study utilized data from the Medical Information Mart for Intensive Care (MIMIC III) to evaluate the use of alternative generalized linear models to identify the risk factors for pneumonia. The regression model with the smallest AIC, BIC and SE was considered as the appropriate regression model for the data. Robust Poisson model was considered the best fit for the current data as it had the lowest AIC, BIC and standard error compared to other regression models.

Results: Variables such as BMI, renal failure, hypertension, diabetes and asthma were identified as the significant risk factors for pneumonia. The risk of pneumonia was found to be significantly higher in the underweight category of BMI [RR_{adj}=1.70; 95% CI=1.38, 2.08]; diabetic patients [RR_{adj} =1.29; 95% CI=1.03, 1.61); asthmatic patients [RR_{adj} =1.35; 95% CI=1.15, 1.58] and patients with renal failure [RR_{adj} =1.16; 95% CI= 1.05, 1.29].

Conclusion: Among various binary regression models, Poisson regression with robust variance (sandwich Poisson regression) provided unbiased estimates of the relationship. In the present study, variables such as BMI, renal failure, diabetics, hypertension and asthma were identified as the significant risk factors for pneumonia in the elderly using robust Poisson regression.

Keywords: Regression models; robust Poisson regression; negative binomial; elderly; pneumonia; risk factors; MIMIC III.

INTRODUCTION

Pneumonia is an inflammation and infection of the lung tissue caused by infectious agents such as bacteria, viruses, or fungi (1). Community-acquired pneumonia (CAP) has a global incidence of 1.5 to 14 cases per 1000 person-years (2). This condition is highly common among persons at the extremes of age due to their weakness and vulnerability. A US-based retrospective study reported that CAP had higher hospitalization rates and total expenditures than stroke, myocardial infarction, and osteoporotic fractures in the elderly population (3). Elderly patients with pneumonia who require hospital care are more apparent and are more likely to develop complications that require longer hospital stays (4). Studies have reported that different factors such as gender, age, diabetes, hypertension, body mass index, dementia, history of heart failure, history of stroke, chronic respiratory disease, and chronic liver disease are associated with hospitalization for pneumonia among the elderly (5-8). The selection of an appropriate regression technique to model the data is very important in any field of research for the proper identification of the risk factors for events of interest. Several types of generalized linear regression models are available for modelling the binary event in cohort studies such as Poisson, log binomial, robust (sandwich) Poisson and negative binomial regression models (9). The present study aimed to identify an optimal regression approach to determine the risk factors for pneumonia among elderly patients in the intensive care unit.

MATERIALS AND METHODS

Data source

The current study utilized data from MIMIC III to evaluate the use of alternative generalized linear models for the analysis of binary outcome i.e. pneumonia status and compared the results of these selected models to identify the most suitable model for this data (10). The MIMIC III database consists of de-identified clinical data of 49785 patients, among which a total of 5214 patients aged greater than 65 years and BMI between 15 and 50 years were included

in the current study. Other variables included in the study were gender, heart disease, renal failure, hypertension, diabetes, asthma, dementia and ethnicity. Data extraction was carried out using Python version 3.7.6.

Statistical analyses

All categorical variables like Pneumonia, BMI, Gender, heart failure etc., were summarized as frequency and percentage. Different regression models including log binomial, Poisson, robust (sandwich) Poisson and negative binomial regressions were fitted to the data. To identify the best suited regression model for the current data, AIC and BIC (Akaike's and Bayesian Information Criteria) and the Standard error (SE) of the regression coefficients were estimated separately for all regression models. The regression model with the smallest AIC, BIC and SE was considered as the appropriate regression model for the data. The simple (univariate) and multiple regressions were fitted to ascertain the risk factors associated with pneumonia based on the identified best suitable regression model. $p<0.05$ was considered as statistically significant. The regression analyses were carried out using the *glm* package in R version 4.0.2./SPSS.

RESULTS

A total of 5214 ICU admitted patients from the MIMIC registry were considered for the present study. The prevalence of pneumonia was 24.1% (1255 out of 5214). The most common chronic condition in the study population was hypertension (55.8%) followed by renal failure (26.5%), diabetes (3.5%), asthma (7.2%), dementia (2.4%) and heart disease (0.3%). More details about these variables are provided in Table 1.

Regression model identification for the data

Table 2 provides the AIC, BIC and SE estimates of the regression coefficients obtained from different regression models. Due to failure in convergence and a large number of categorical independent variables, the log binomial model failed to estimate the coefficients for the current data and hence it was not included in the comparison.

Table 1: Baseline characteristics of the elderly subjects included in the study (n=5214)

Variables		**Frequency (n)**	**Percentage (%)**
Pneumonia	Absent	3959	75.9
	Present	1255	24.1
Gender	Female	2226	42.7
	Male	2988	57.3
BMI	Normal	1542	29.6
	Underweight	140	2.7
	Overweight	1912	36.7
	Obesity	1620	31.1
Age	Youngest-old	2502	48.0
	Middle-old	2147	41.2
	Oldest-old	565	10.8
Heart disease	Absent	5197	99.7
	Present	17	0.3
Renal failure	Absent	3831	73.5
	Present	1383	26.5
Hypertension	Absent	2305	44.2
	Present	2909	55.8
Diabetes	Absent	5030	96.5
	Present	184	3.5
Asthma	Absent	4837	92.8
	Present	377	7.2
Dementia	Absent	5089	97.6
	Present	125	2.4
Ethnicity	Asian	101	1.9
	African American	357	6.8
	Latino	120	2.3
	White	4172	80.0
	Other/unknown	464	8.9

Table 2: Standard error (SE) of the regression coefficients for the regression models

Variables		Standard error of regression coefficients		
		Standard Poisson model	**Robust Poisson model**	**Negative binomial model**
Gender	Female†	.	.	.
	Male	0.0582	0.0503	0.0651
BMI	Normal †	.	.	.
	Underweight	0.1377	0.1047	0.1643
	Overweight	0.0711	0.0615	0.0793
	Obesity	0.0740	0.0638	0.0830
Age	youngest-old	.	.	.
	middle-old	0.0608	0.0525	0.0681
	oldest-old	0.0977	0.0853	0.1091
Heart disease	Absent †	.	.	.
	Present	0.4529	0.3439	0.5212
Renal failure	Absent †	.	.	.
	Present	0.0623	0.0532	0.0702
Hypertension	Absent †	.	.	.
	Present	0.0572	0.0496	0.0639
Diabetes	Absent †	.	.	.
	Present	0.1359	0.1133	0.1565
Asthma	Absent †	.	.	.
	Present	0.0972	0.0812	0.1117
Dementia	Absent †	.	.	.
	Present	0.1729	0.1526	0.1966
Ethnicity	Asian †	.	.	.
	African American	0.2071	0.1711	0.2375
	Latino	0.2555	0.2137	0.2909
	White	0.1868	0.1545	0.2137
	Other/unknown	0.2081	0.1740	0.2368
AIC		6054	6054	6347
BIC		6165	6165	6459

Note. † - Reference category, AIC: Akaike information criterion; BIC: Bayesian Information Criteria

AIC and BIC were found to be lesser for the Poisson regression and robust Poisson regression models compared to the negative binomial regression model. The robust Poisson regression model had the least SE compared to other regression models for all the regression coefficients. Based on the aforementioned criterion, robust Poisson model was considered the best fit for the current data.

Risk factors of pneumonia based on robust Poisson regression

To identify the risk factors for pneumonia in the MIMIC III data, both univariate and multiple robust Poisson regressions were used. Adjusted relative risk (RR_{adj}) with 95% confidence intervals for the covariates was estimated based on robust Poisson regression using predictors such as BMI, age, gender, diabetes, heart disease, renal failure, hypertension, asthma, dementia and ethnicity (Table 3).

The multiple robust Poisson regression revealed that variables such as BMI, renal failure, hypertension, diabetes and asthma were significant risk factors of pneumonia. It is observed that compared to the normal category of BMI, the underweight category had 1.70 times more risk [RR_{adj}=1.70; 95% CI=1.38, 2.08) of having pneumonia and this was found to be statistically significant (p=0.006). The risk of pneumonia was found to be less for the overweight category [RR_{adj} =0.88; 95% CI=0.78, 0.99]. It was observed that patients with hypertension had lesser risk for pneumonia than those without hypertension [RR_{adj} =0.84; 95% CI= 0.76, 0.92]. Diabetic patients were shown to have a considerably greater risk of pneumonia [RR_{adj} =1.29; 95% CI=1.03, 1.61); asthmatic patients [RR_{adj} =1.35; 95% CI=1.15, 1.58] and patients with renal failure [RR_{adj} =1.16; 95% CI= 1.05, 1.29]. Around 83 % [4331 out of 5213] of the study population and 82% [1030 out of 1255] of the pneumonia patients had any one or more than one of these risk factors. It was also revealed that pneumonia was not significantly associated with variables such as gender, age categories such as youngest old (65 to 74 years), middle old (75 to 84 years) and oldest old (≥85 years), heart disease, dementia and ethnicity (Table 3).

Table 3: Risk factors of pneumonia based on robust (sandwich) Poisson regression

Variables		Univariate robust Poisson regression		Multiple robust Poisson regression	
		Unadjusted RR (95% CI)	p value	Adjusted RR (95% CI)	p value
Gender	Female†	1		1	
	Male	0.94 (0.85, 1.03)	0.18	0.98(0.89,1.08)	0.63
BMI	Normal †	1		1	
	Underweight	1.77 (1.44, 2.17)	<0.001*	1.70(1.38,2.08)	<0.001*
	Overweight	0.88 (0.78, 1.00)	0.04*	0.88(0.78,0.99)	0.03*
	Obesity	0.95 (0.84, 1.07)	0.41	0.93(0.82,1.05)	0.23
Age	youngest-old	1		1	
	middle-old	1.06 (0.96, 1.18)	0.23	1.08(0.97,1.19)	0.17
	oldest-old	1.01 (0.86, 1.19)	0.88	1.01(0.85,1.19)	0.92
Heart disease	Absent †	1		1	
	Present	1.22 (0.59, 2.56)	0.59	1.19(0.61,2.34)	0.61
Renal failure	Absent †	1		.	
	Present	1.19 (1.07, 1.32)	<0.001*	1.16(1.05,1.29)	0.01*
Hypertension	Absent †	1		1	
	Present	0.81 (0.73, 0.89)	<0.001*	0.84(0.76,0.92)	<0.001*
Diabetes	Absent †	1		1	
	Present	1.35 (1.09, 1.67)	<0.001*	1.29(1.03,1.61)	0.03*
Asthma	Absent †	1		1	
	Present	1.36 (1.16, 1.59)	<0.001*	1.35(1.15,1.58)	<0.001*
Dementia	Absent †	1		1	
	Present	1.17 (0.88, 1.55)	0.29	1.04(0.77,1.41)	0.79
Ethnicity	Asian †	1		1	
	African American	1.08 (0.77, 1.51)	0.67	1.10(0.79,1.54)	0.58
	Latino	0.90 (0.59, 1.37)	0.62	0.90(0.60,1.37)	0.64
	White	0.97 (0.58, 1.07)	0.12	0.84(0.62,1.14)	0.25
	Other/unknown	0.77 (0.55, 1.08)	0.13	0.84(0.60,1.19)	0.33

Note. †: Reference category; *: Significant (p<0.05), RR: Relative Risk, CI: Confidence Interval

DISCUSSION

This study compared the performance of various binary regression models such as Poisson regression, log binomial regression, robust Poisson regression and negative binomial regression in estimating the relative risk in cohort studies using MIMIC III dataset. Robust Poisson and negative binomial regression models can be considered as a generalization of Poisson regression with an additional parameter to model the over-dispersion. The log binomial model failed to estimate the coefficients for the present data due to failure in convergence. The log-binomial regression usually fails to converge due to restricted range of the parameter space and when there are too many categories and continuous independent variables, or the incidence rate is high (11). Based on AIC and BIC, the robust Poisson model was considered as the best suited model for the MIMIC III data to identify risk factors of pneumonia. Occasionally, the standard Poisson regression model overestimates the standard errors of relative risk which results in a wider confidence interval. In such situations, the overestimation of standard error can be corrected using Poisson regression with robust covariance estimation (12).

The confidence intervals for the Poisson regression model are likely to be wider than those for the robust Poisson/negative binomial regression if the conditional distribution of the outcome variable is over-dispersed. Compared to the log-binomial method, the robust Poisson method provides less biased estimates of the prevalence ratios for a moderate sample size and very high prevalence. However, if there is no convergence problem, for a moderate sample size and moderate prevalence, the log-binomial method provides little less biased estimates than the robust Poisson method (13). Comparison of various binary regression models using simulated data revealed that Poisson regression with a robust variance estimator provides unbiased estimates of the relationship (13).

Based on the results from robust Poisson regression, we found that the occurrence of pneumonia among the elderly was more likely in the middle-old (75 to 84 years) age group and this occurrence was higher among the underweight. However, a study by *Tongo et al.,* reported that BMI did not show any association with pneumonia but patients who had a risk for malnutrition had 6 times more odds of developing intermediate risk pneumonia (14). Similar finding was reported in a meta-analysis in which a U shaped

association was observed between risk of influenza-related pneumonia and BMI (15). A 5-year prospective follow-up study reported that cardiovascular disease increased the risk of pneumonia by 2.63-fold and a 98% higher chance of mortality compared to those who didn't have these conditions (16). We found similar effects of heart disease and renal failure based on the MIMIC data in our study. Several studies have identified chronic diseases such as dementia, diabetes and asthma to have a greater risk of contracting or dying from pneumonia (17-19). These factors were also significantly associated with pneumonia in the present study. A few previous studies also reported that African Americans had a greater incidence of pneumonia than other ethnic groups (20-22).

CONCLUSION

In the present study, variables such as BMI, renal failure, hypertension, diabetes and asthma were found to be significant risk factors of pneumonia in the elderly using Sandwich Poisson regression. Among the various binary regression models, Poisson regression with robust variance (Sandwich Poisson regression) gives unbiased estimates of the relationship.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Atkinson, T.P., Balish, M.F., Waites, K.B. Epidemiology, clinical manifestations, pathogenesis, and laboratory detection of *Mycoplasma pneumoniae* infections. FEMS Microbiol Rev. 2020; 32(6): 956-973.
2. Regunath, H., Oba, Y. Community-Acquired Pneumonia. StatPearls. 2020.
3. Brown, J.D., Harnett, J., Chambers, R., Sato, R. The relative burden of community-acquired pneumonia hospitalizations in older adults: a retrospective observational study in the United States. BMC Geriatr.2018; 18(1):1-1.
4. Fein, A., Niederman, M. Severe pneumonia in the elderly. Clin Geriatr Med.1994; 10:121-143.
5. Matsuo, A., Takamori, A., Kawaura, F., Iwanaga, Y., Ono, H., Kobayashi-Watanabe, N., *et al.,* Risk for prolonged hospitalization and mortality in aged community acquired pneumonia patients: a retrospective study in Japan. J Clin Biochem Nutr.2020;67(3):302-306.
6. Ishisaka, T., Igarashi, Y., Kodera, K., Okuno, T., Morita, T., Himeno, T., *et al.,* Relationship between blood pressure levels on admission and the onset of acute pneumonia in elderly patients with cerebral hemorrhage. J. Clin. Med. Res .2020;12(11):1-6.
7. Lin, C.J., Chang, Y.C., Tsou, M.T., Chan, H.L., Chen, Y.J., Hwang, L.C. Factors associated with hospitalization for community-acquired pneumonia in home health care patients in Taiwan. Aging Clin Exp Res. 2020; 32(1):149-155.
8. Kikutani, T., Tamura, F., Tashiro, H., Yoshida, M., Konishi, K., Hamada, R. Relationship between oral bacteria count and pneumonia onset in elderly nursing home residents. Geriatr Gerontol Int 2015;15(4):417-421.
9. McNutt, L.A., Wu, C., Xue, X., Hafner, J.P. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol.2003;157(10):940-943.
10. Johnson, A.E., Pollard, T.J., Shen L., Lehman, L.W., Feng, M., Ghassemi, M., *et al.,* MIMIC-III, a freely accessible critical care database. Scientific data. 2016;3(1):1-9.
11. Knol, M. J., Le Cessie, S., Algra, A., Vandenbroucke, J. P. and Groenwold, R. H. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. CMAJ 2012;184(8), 895-899.
12. Huang, F. L. Alternatives to logistic regression models in experimental studies. J. Exp. Educ. 2019;1-16.
13. Chen, W., Qian, L., Shi, J. and Franklin, M. Comparing performance between log-binomial and robust poisson regression models for estimating risk ratios under model misspecification. BMCMed. Res. Methodol 2018;18(1), 1-12.
14. Tongo, M.A., Sy, R.A. The relation of nutritional assessment and pneumonia severity index among elderly patients with community acquired pneumonia admitted at Cardinal Santos Medical Center. J Nutr Diet Suppl 2017;1(1).
15. Phung, D.T., Wang, Z., Rutherford, S., Huang, C., Chu, C. Body mass index and risk of pneumonia: a systematic review and meta-analysis. Obes Rev 2013;14(10):839-857.
16. Holter, J.C., Ueland, T., Jenum, P.A., Müller, F., Brunborg, C., Frøland, S.S., *et al.,* Risk factors for long-term mortality after hospitalization for community-acquired pneumonia: a 5-year prospective follow-up study. PLoS One 2016;11(2):1-16.
17. Kim, M.H., Rhee, C.K., Shim, J.S., Park, S.Y., Yoo, K.H., Kim, B.Y., *et al.,* Inhaled corticosteroids in asthma and the risk of pneumonia. Allergy Asthma Immunol Res. 2019;11(6):795-805.
18. Hespanhol, V., Bárbara, C. Pneumonia mortality, comorbidities matter? Pulmonology. 2020; 26(3):123-129.
19. Konomura, K., Nagai, H., Akazawa, M. Economic burden of community-acquired pneumonia among elderly patients: a Japanese perspective. Pneumonia. 2017;9(1):1-10.
20. Burton, D.C., Flannery, B., Bennett, N.M., Farley, M.M., Gershman, K., Harrison, L.H., *et al.,* Socioeconomic and racial/ethnic disparities in the incidence of bacteremic pneumonia among US adults. Am J Public Health. 2010;100(10):1904-1911.
21. Hausmann, L.R., Ibrahim, S.A., Mehrotra, A., Nsa, W., Bratzler, D.W., Mor, M.K., *et al.,* Racial and ethnic disparities in pneumonia treatment and mortality. Med Care. 2009;47(9):1-17.
22. Frei, C.R., Mortensen, E.M., Copeland, L.A., Attridge, R.T., Pugh, M.J., Restrepo, M.I., *et al.,* Disparities of care for African-Americans and Caucasians with community-acquired pneumonia: a retrospective cohort study. BMC Health Serv Res. 2010;10(1):1-11.