Archives of Academic Emergency Medicine. 2024; 12(1): e13

OPEN ACCESS

ORIGINAL RESEARCH

Zero-Inflated Count Regression Models in Solving Challenges Posed by Outlier-Prone Data; an Application to Length of Hospital Stay

Saeed Shahsavari[1], Abbas Moghimbeigi[2], Rohollah Kalhor[3], Ali Moghadas Jafari[4], Mehrdad Bagherpour-kalo[1], Mehdi Yaseri[1 *], Mostafa Hosseini[1 †]

1. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

2. Department of Biostatistics and Epidemiology, School of Health, Research Center for Health, Safety and Environment, Alborz University of Medical Sciences, Karaj, Iran.

3. Social Determinants of Health Research Center, Research Institute for Prevention of Non-Communicable Diseases, Qazvin University of Medical Sciences, Qazvin, Iran.

4. Tehran Medical Sciences Azad University, Farhikhtegan Hospital, Tehran, Iran.

Received: September 2023; Accepted: October 2023; Published online: 21 November 2023

Abstract: **Introduction:** Ignoring outliers in data may lead to misleading results. Length of stay (LOS) is often considered a count variable with a high frequency of outliers. This study exemplifies the potential of robust methodologies in enhancing the accuracy and reliability of analyses conducted on skewed and outlier-prone count data of LOS. **Methods:** The application of Zero-Inflated Poisson (ZIP) and robust Zero-Inflated Poisson (RZIP) models in solving challenges posed by outlier LOS data were evaluated. The ZIP model incorporates two components, tackling excess zeros with a zero-inflation component and modeling positive counts with a Poisson component. The RZIP model introduces the Robust Expectation-Solution (RES) algorithm to enhance parameter estimation and address the impact of outliers on the model's performance. **Results:** Data from 254 intensive care unit patients were analyzed (62.2% male). Patients aged 65 or older accounted for 58.3% of the sample. Notably, 38.6% of patients exhibited zero LOS. The overall mean LOS was 5.89 (± 9.81) days, and 9.45% of cases displayed outliers. Our analysis using the RZIP model revealed significant predictors of LOS, including age, underlying comorbidities (p<0.001), and insurance status (p=0.013). Model comparison demonstrated the RZIP model's superiority over ZIP, as evidenced by lower Akaike information criteria (AIC) and Bayesians information criteria (BIC) values. **Conclusion:** The application of the RZIP model allowed us to uncover meaningful insights into the factors influencing LOS, paving the way for more informed decision-making in hospital management.

Keywords: Length of stay; intensive care units; outliers; robust; excess zeros

Cite this article as: Shahsavari S, Moghimbeigi A, Kalhor R, et al. Zero-Inflated Regression Models in Solving Challenges Posed by Outlier-Prone Data; an Application to Length of Hospital Stay. Arch Acad Emerg Med. 2024; 12(1): e13. https://doi.org/10.22037/aaem.v12i1.2074.

1. Introduction

The analysis of count variables is a common task in fields such as medical and social sciences. Traditional statistical techniques, like the Poisson regression model, are often employed for count data analysis. However, these techniques frequently encounter limitations when applied to datasets from these domains. This is primarily attributed to the distinct characteristics of empirical count datasets, such as over-dispersion and an excessive prevalence of zero values. Consider the example of modeling the Length of Stay (LOS) in a hospital setting, where a substantial portion of the data comprises zero values. These zeros correspond to instances where patients have not stayed overnight, representing scenarios where individuals either opt out of hospital stays or refrain from utilizing inpatient services. In such datasets, the excessive frequency of zeros cannot be appropriately accommodated by conventional count distributions, including Poisson, binomial, or negative binomial.

To address this challenge, Zero-Inflated (ZI) models have

* **Corresponding Author:** Mehdi Yaseri; Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. E-mail: m.yaseri@gmail.com, Tel: +989122244492, ORCID: https://orcid.org/0000-0002-4066-873X.

† **Corresponding Author:** Mostafa Hosseini; Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. E-mail: mhossein110@yahoo.com, Tel: +989121904102, ORCID: https://orcid.org/0000-0002-1334-246X.

been developed to handle datasets with an overabundance of zeros (1-3).

While ZI models have gained popularity in addressing count data complexities, their utilization in contexts like hospital LOS and other right-skewed data has raised questions. The application of linear regression, suitable for continuous or normally distributed outcomes, to non-transformed or logarithm-transformed count variables has been explored (4, 5). Another approach to address the non-normality of LOS data involves dichotomization and logistic regression; however, this may result in information loss (6). In light of simulated and empirical data analyses, count regression models employing a log-odds link function have demonstrated superior odds ratio estimation precision, as underscored by Sroka (7).

Despite the increasing adoption of ZI models for LOS data (8-10), a gap remains in their treatment of outliers within the dataset. Numerous studies utilizing ZI models to analyze LOS data have disregarded potential outlier presence. Additionally, the impact of outliers on model performance has been explored in the literature (11, 12). While Maximum Likelihood (ML) estimation is often favored for model fitting, its sensitivity to outliers and data contamination is well-documented. Among the plethora of available models, ZI count models like Zero-Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB) models (8, 13, 14), have garnered significant attention.

However, these models can be significantly influenced by outlier data, potentially affecting inferences about observed variables. Given the susceptibility of ZI models to outlier influence and the presence of outliers in LOS data, the recommendation is to employ the Robust Zero-Inflated Poisson (RZIP) model.

The RZIP model, introduced by Hall and colleagues (15), incorporates a novel parameter estimation technique known as the Robust Expectation-Solution (RES) algorithm. A distinct advantage over traditional ZIP models lies in the incorporation of outlier data point weights. By introducing these weights, the model effectively manages the impact of outlier data on outcomes. Simulation results highlight that robust estimators consistently outperform their ML counterparts in the presence of outlying values, irrespective of sample size and dispersion value (11, 12).

Furthermore, in practical applications, addressing both zero-inflation and outliers simultaneously is a crucial consideration. Despite this, a strategy for robust modeling of count data in the context of LOS remains absent. Hence, this study aims to illustrate such a method through a practical example, bridging the gap between theoretical advancements and real-world clinical applications.

2. Methods

2.1. Study design and setting

This cross-sectional study is based on a retrospective analysis of data from hospital records. The study population consisted of patients aged 18 years and older admitted to the surgical intensive care unit (ICU) of Bouali Sina hospital in Qazvin city, within a 12-month period from September 23, 2019 to September 22, 2020. The study included both male and female patients, covering a wide age range. Ethical clearance was obtained from the Ethics Committee of Tehran University of Medical Sciences (ethics code: IR.TUMS.SPH.REC.1401.037). All methods were performed in accordance with the relevant guidelines and regulations. Informed consent was deemed unnecessary by the ethics committee of Tehran University of Medical Sciences.

2.2. Participants

The focus of this study is solely on patients admitted to the surgical ICU of Bouali Sina hospital due to variations in illness severity across different ICU units. This unit, comprising 12 beds, primarily caters to internal medical cases, with an average monthly admission rate of 35 patients.

2.3. Data gathering

The study samples were obtained from the Hospital Information System (HIS) and underwent data cleaning and preprocessing using Microsoft SQL Server. The data encompassed patient identity, clinical and managerial details, and attending physician records.

2.4. Outcomes measurement

The dependent variable of interest was LOS, calculated as the time between admittance and discharge. Given a higher-than-expected occurrence of zeros in the LOS data, indicating zero inflation, the study examined two sources for the excess zeros: "sampling zeros" arising from patients opting not to stay overnight and "structured zeros" from individuals not utilizing inpatient services.

2.5. Statistical analysis

Descriptive statistics included mean and standard deviation for continuous variables and numbers (percentages) for categorical variables. ZIP and RZIP models were employed to assess the impact of variables on the LOS, with statistical analyses conducted using R and Stata 17. Model selection was based on Akaike information criteria (AIC) or Bayesians information criteria (BIC), with the Vuong test (17) used to compare the zero-inflated model against count regression.

The presented methodology provides a comprehensive framework for analyzing the impact of variables on Length of Stay using the RZIP model, addressing the challenges posed

by zero inflation and outliers.
In the realm of discrete independent random variables, the ZIP regression model is often employed (14, 16). The ZIP model is defined as follows:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i} & \text{for } y_i = 0 \\ (1 - p_i)e^{-\mu_i}\mu_i^{y_i}/y_i! & \text{for } y_i = 1,2,.. \end{cases}$$

Here, μ_i represents the mean, the first state is called the zero state (with probability p_i), and the second state is called counts state (with the Poisson distribution with probability (1 - p_i)). The ZIP regression model is comprised of two components: the zero-inflation component for p_i and the Poisson component for μ_i.
The zero-inflation component accounts for the excess zeros in the data and is defined as:
$\text{logit}(p_i) = z^T{}_i\, \alpha$
And the Poisson component models the positive counts and is defined as:
$\ln(_i) = x^T{}_i\, \beta$
Here, α and β are the regression parameters, while $z_i{}^T$ and $x_i{}^T$ represent the explanatory variables. The Maximum likelihood estimation (MLE) method is commonly employed for parameter estimation of ZIP models (16). It is crucial to note that MLE estimates can be sensitive to outliers, potentially leading to biased parameter estimates. Hence, a robust approach is essential.

2.6. Robust Zero-Inflated Poisson (RZIP) model

To overcome the potential bias caused by outliers, the Robust Zero-Inflated Poisson (RZIP) model employs the RES approach. This involves replacing the estimators from the M step of the Expectation-Maximization (EM) algorithm with robust estimators, as outlined by Hall (15). The RES algorithm assigns lower weights to observations in the extreme tails of the Poisson distribution during estimation. For RZIP, the parameters of the ZIP model (α and β) are determined by solving equations during the M-step of the EM algorithm:

$$1/n\ \Sigma w(G_i) z_i{}^{(}r) - \text{logit}^-1\ (G^T{}_i \alpha) G_i = 0$$

$$1/n\ \Sigma(1 - -z_i{}^r) w(B_i) \psi_c(Y_i) - o_i(\beta, c) B_i = 0$$

In RZIP, the ψ-Huber function, which was introduced by Hall (15), is chosen for weighting the response variable:

$$\psi_c(y) = \begin{cases} j_1, & y < j_1 \\ y, & y \varepsilon [j_1, j_2] \\ j_2, & y > j_2 \end{cases}$$

Where j_1 and j_2 are the c and (1-c) quintiles of the Poisson component, respectively. Here, $o_i(\beta, c) = E\psi_c(Y_i)|Y_i$ represents an expected value for ψ-Huber function and $w(G_i)$ is

$$\sqrt{1 - h_i}$$

with h_i denoting the ith diagonal element of $H = G(G^T G)^{-1} G^T$ a similar definition applies to $w(B_i)$.

3. Results

3.1. Baseline characteristics of studied cases

The analysis encompassed 254 patients who were admitted to the ICU (62.2% male). Among male patients, the mean LOS was 8.81 days (± 8.25), and 40.5% experienced LOS=0. The majority of patients (58.3%) were aged 65 years or older, with patients aged >65 years having a mean LOS of 8.79 days (± 8.33) and 64.9% experiencing LOS>0. In the context of local place of residence, 58.8% of patients had a mean LOS of 8.90 days (± 9.04), while 41.2% experienced LOS=0. Similarly, patients without insurance exhibited a mean LOS of 7.23 days (± 7.78), with 41.7% having LOS=0. Descriptive statistics for LOS and different variables are summarized in Table 1. The total mean (±SD) LOS for all patients was 5.89 (± 9.81) days. Notably, 98 patients (38.6%) experienced zero LOS, highlighting the need for a zero-inflated model. Figure 1 displays the distribution of LOS in patients, revealing the presence of numerous outliers, notably 28 patients (9.45%) with more than 17 days of LOS.

3.2. Comparative analysis of studied outcomes

Table 2 presents the outcomes of the ZIP and RZIP models concerning ICU LOS. The RZIP model findings indicate that factors including age, underlying comorbidities (P<0.001), and having insurance (p=0.013) significantly impact LOS. Model comparison, as guided by AIC and BIC values, highlights the superior fit of the RZIP model. The notable difference between the AIC and BIC values of the RZIP and ZIP models (exceeding 170) further confirms RZIP as the preferred model due to its lower AIC and BIC values. Importantly, the lower AIC value in the RZIP model signifies better goodness of fit. Thus, the RZIP regression model effectively addresses zero inflation and outlier challenges in LOS data, shedding light on factors influencing LOS.

3.3. Odds ratios and regression coefficients

Table 2 provides Odds Ratios (ORs) from the logistic component of the RZIP model. In the logistic part, underlying comorbidities emerge as significant (OR = 5.2), signifying 5.2 times higher odds of longer LOS for patients with underlying comorbidities compared to those without. In the RZIP model's count component, age, underlying comorbidities, and insurance significantly influence the LOS. Importantly, sex and place of residence are not significantly associated with LOS. Risk Ratios (RRs) greater than 1 in the second, third, and fourth age groups imply an increased risk of longer LOS compared to the reference group (age < 20). Particularly, the RR for the fourth age group (>65) is 5.5, with a 95% confidence interval (CI) of 1.2 to 25, indicating a statistically significant association (p=0.028) between this age group and extended LOS. Similarly, individuals with underlying comorbidities exhibit 2.1 times higher risk of longer LOS, with a narrower confidence interval (1.7 to 2.5), and a significant p-value (P<0.001). Additionally, patients with insurance have 1.4 times higher risk of longer LOS, with a relatively precise estimate (95% CI: 1.1 to 1.6) and a significant p-value (P<0.001).

3.4. Model comparison

Based on the Vuong non-nested hypothesis test, the ZIP model significantly outperforms the count regression model. Vuong z-statistic values of 4.50, 4.37, and 4.17 for the raw, AIC-corrected, and BIC-corrected models, respectively, all indicate p-values less than 0.05, providing strong evidence for the superiority of the ZIP model. Figure 2 visually depicts the 95% confidence intervals for the count and logistic components of all variables in ZIP and RZIP models, enabling the determination of the variables' statistical significance.
Through these comprehensive results, the RZIP model demonstrates its efficacy in capturing the nuanced associations between patient characteristics and Length of Stay in the ICU setting.

4. Discussion

Our study underscores the importance of robust modeling approaches in addressing the challenges posed by outlier data in LOS analysis. As evident from the initial inspection of our dataset, the distribution of LOS was significantly skewed, indicating the presence of outliers. Conventional statistical methods are susceptible to the influence of outliers, potentially leading to biased parameter estimates and erroneous conclusions. By adopting the RZIP model, we effectively accounted for the impact of outliers and improved the accuracy of parameter estimation. This method is in accordance with the findings of Zandkarimi (2020) (12) and Abonazel (2021) (11, 12), both of whom highlighted the superior performance of robust estimators when dealing with outlier presence.
Our study identified several key factors that significantly influence the length of stay in the surgical ICU. Notably, age, underlying comorbidities, and insurance status emerged as significant predictors of LOS. These findings resonate with existing literature, highlighting the importance of these variables in shaping patients' hospitalization durations (8, 18-20). Our robust modeling approach enabled a more accurate assessment of the impact of these factors, as it effectively addressed the challenges posed by outliers and excess zeros.
The outcomes of our study bear practical implications for hospital management and policy-makers. The robust modeling approach, exemplified by the RZIP model, demonstrates its effectiveness in capturing the complex interactions between predictor variables and LOS, even in the presence of outlier data. Hospitals and healthcare institutions can benefit from such modeling techniques to optimize resource allocation, manage patient flow, and enhance quality of care. Furthermore, our study highlights the need for future research to delve deeper into the interplay of various clinical and non-clinical factors and hospital LOS. Longitudinal studies and larger datasets could provide more comprehensive insights into the dynamics of LOS and inform targeted interventions to reduce hospitalization durations.

5. Limitations

While our study provides valuable insights into the realm of hospital LOS analysis, certain limitations should be acknowledged. Our analysis is based on retrospective data from a single surgical ICU, which might limit the generalizability of our findings to different clinical settings. Additionally, other factors not included in our dataset, such as patient preferences and socioeconomic factors, could further influence LOS. Future studies could incorporate these dimensions to enhance the accuracy and applicability of LOS models.

6. Conclusion

Our study underscores the significance of robust modeling approaches in addressing the challenges posed by outliers in length of stay analysis. The application of the RZIP model allowed us to uncover meaningful insights into the factors influencing LOS, paving the way for more informed decision-making in hospital management. This study exemplifies the potential of robust methodologies in enhancing the accuracy and reliability of analyses conducted on skewed and outlier-prone count data.

7. Declarations

7.1. Acknowledgments

This article is a part of the PhD thesis in Biostatistics at the Faculty of Health, Tehran University of Medical Sciences. The authors of this paper would like to thank the officials and managers of the hospitals who facilitated the conduct of the study.

7.2. Conflict of interest

The authors declare that they have no conflicting interests.

7.3. Funding and supports

This work was supported by Tehran University of Medical Sciences, Tehran, Iran.

7.4. Authors' contribution

The authors confirm contribution to the paper as follows: study conception and design: SS, MY, AM, MH; data collection: SS, MB and RK; analysis and interpretation of results: SS, MY, AM, MH; draft manuscript preparation: SS, MY, AM and AMJ. All authors reviewed the results and approved the final version of the manuscript.

7.5. Data availability

The datasets used during the current study are available from the corresponding author on reasonable request.

References

1. Sarul LS, Sahin S. An application of claim frequency data using zero inflated and hurdle models in general insurance. Journal of Business Economics and Finance. 2015;4(4).
2. Hilbe JM. Negative binomial regression: Cambridge University Press; 2011.
3. Workie MS, Azene AG. Bayesian zero-inflated regression model with application to under-five child mortality. Journal of big data. 2021;8(1):1-23.
4. O'Hara R, Kotze J. Do not log-transform count data. Nature Precedings. 2010:1-.
5. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological bulletin. 1995;118(3):392.
6. Huang JQ, Hooper PM, Marrie TJ. Factors associated with length of stay in hospital for suspected community-acquired pneumonia. Canadian respiratory journal. 2006;13:317-24.
7. Sroka CJ, Nagaraja HN. Odds ratios from logistic, geometric, Poisson, and negative binomial regression models. BMC medical research methodology. 2018;18(1):1-11.
8. Fernandez GA, Vatcheva KP. A comparison of statistical methods for modeling count data with an application to hospital length of stay. BMC Medical Research Methodology. 2022;22(1):1-21.
9. Farhadi Hassankiadeh R, Kazemnejad A, Gholami Fesharaki M, Kargar Jahromi S. Efficiency of zero-inflated generalized poisson regression model on hospital length of stay using real data and simulation study. Caspian Journal of Health Research. 2018;3(1):5-9.
10. SONG JX. Zero-inflated Poisson regression to analyze lengths of hospital stays adjusting for intra-center correlation. Communications in Statistics—Simulation and Computation®. 2005;34(1):235-41.
11. Abonazel MR, El-sayed SM, Saber OM. Performance of robust count regression estimators in the case of overdispersion, zero inflated, and outliers: simulation study and application to German health data. Commun Math Biol Neurosci. 2021;2021:Article ID 55.
12. Zandkarimi E, Moghimbeigi A, Mahjub H, Majdzadeh R. Robust inference in the multilevel zero-inflated negative binomial model. Journal of Applied Statistics. 2020 2020/01/25;47(2):287-305.
13. Mullahy J. Specification and testing of some modified count data models. Journal of econometrics. 1986;33(3):341-65.
14. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992;34(1):1-14.
15. Hall DB, Shen J. Robust estimation for zero-inflated Poisson regression. Scandinavian Journal of Statistics. 2010;37(2):237-52.
16. Jansakul N, Hinde J. Score tests for zero-inflated Poisson models. Computational statistics data analysis. 2002;40(1):75-96.
17. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: journal of the Econometric Society. 1989:307-33.
18. Feng CX, Li L. Modeling zero inflation and overdispersion in the length of hospital stay for patients with ischaemic heart disease. Advanced Statistical Methods in Data Science. 2016:35-53.
19. Nanni L. Modeling Zero-inflated and overdispersed count data: Application to IN-Hospital mortality data. 2019.
20. Zeleke AJ, Moscato S, Miglio R, Chiari L. Length of stay analysis of COVID-19 hospitalizations using a count regression model and Quantile regression: a study in Bologna, Italy. International journal of environmental research and public health. 2022;19(4):2224.

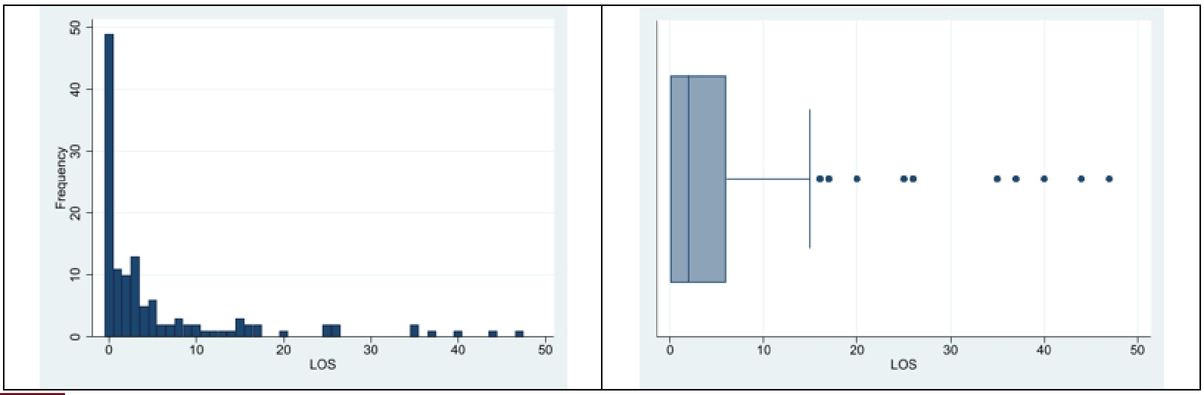

Figure 1: Frequency distribution and boxplot of the length of stay (LOS) in the studied surgical intensive care unit.

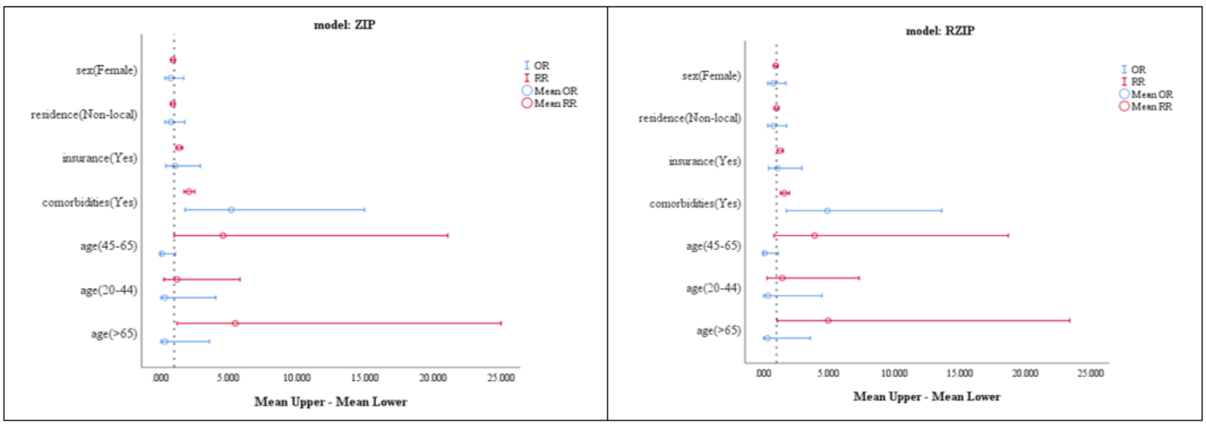

Figure 2: 95% confidence interval (CI) of odds ratio (OR) and risk ratio (RR) for variables in Zero-Inflated Poisson (ZIP) and robust Zero-Inflated Poisson (RZIP) models.

Table 1: Descriptive statistics, frequency, and characteristics of the study sample

Variables	category	n (%)	Length of stay (day)		
			= 0 n (%)	> 0 n (%)	Mean ± SD
Sex	Male	158 (62.2)	64 (40.5)	94 (59.5)	8.81 ± 8.25
	Female	96 (37.8)	34 (35.4)	62 (64.6)	7.90 ± 8.02
Age (years)	<20	10 (3.9)	8 (80)	2 (20)	2.5 ± 0.25
	20-44	38 (15.0)	24 (63.2)	12 (37.3)	4.14 ± 2.03
	45-65	58 (22.8)	14 (24.1)	44 (75.9)	8.45 ± 8.36
	>65	148(58.3)	52(35.1)	96(64.9)	8.79 ± 8.33
Place of residence	Local	136 (53.5)	56 (41.2)	80 (58.8)	8.90 ± 9.04
	Non-local	118 (46.5)	42 (35.6)	76 (64.4)	7.97 ± 7.16
Comorbidity	Yes	160 (63.0)	42 (26.3)	118 (73.7)	9.42 ± 9.29
	No	94 (37.0)	56 (59.6)	38 (40.4)	8.14 ± 7.76
Having insurance	Yes	192 (75.6)	80 (41.7)	112 (58.3)	7.23 ± 7.78
	No	62 (24.4)	18 (29.0)	44 (71.0)	8.93 ± 8.26

SD: standard deviation.

Table 2: Estimates of Zero-Inflated Poisson (ZIP) and robust Zero-Inflated Poisson (RZIP)models for length of stay (LOS)

Predictor	ZIP				RZIP			
	Logistic part		Count part		Logistic part		Count part	
	OR	P	RR	P	OR	P	RR	P
Sex								
Male*	1		1		1		1	
Female	0.74(0.32,1.70)	0.477	0.91(0.78,1.06)	0.235	0.75(0.33,1.72)	0.500	0.93(0.79,1.10)	0.406
Age								
<20*	1		1		1		1	
20-44	0.32 (0.03, 4.04)	0.379	1.20 (0.25, 5.83)	0.824	0.35 (0.03, 4.47)	0.422	1.44 (0.28, 7.31)	0.659
45-65	0.09 (0.01, 1.10)	0.059	4.58 (1.01, 21.06)	0.049	0.10 (0.01, 1.14)	0.063	3.91 (0.82, 18.68)	0.088
>65	0.32 (0.03, 3.59)	0.356	5.48 (1.20, 24.97)	0.028	0.32 (0.03, 3.58)	0.356	4.95 (1.05, 23.38)	0.044
Place of residence								
Local*	1		1		1		1	
Non-local	0.77 (0.33, 1.77)	0.533	0.90 (0.76, 1.05)	0.178	0.77 (0.33, 1.78)	0.534	0.99 (0.84, 1.18)	0.942
Underlying comorbidities								
No*	1		1		1		1	
Yes	5.19 (1.80, 14.96)	0.002	2.07 (1.72, 2.50)	P<0.001	4.88 (1.75, 13.61)	0.002	1.61 (1.31, 1.99)	P<0.001
Insurance								
No*	1		1		1		1	
Yes	1.06 (0.39, 2.92)	0.907	1.35 (1.14, 1.61)	0.001	1.07 (0.39, 2.94)	0.891	1.27 (1.05, 1.52)	0.013
Log-likelihood	-538.5				-452.6			
AIC	1091				919.2			
BIC	1093.8				922			
Vuong (uncorrected)	4.50**							
AIC-corrected	4.37**							
BIC-corrected	4.17**							

Data are presented with 95% confidence interval. OR: odds ratio; RR: risk ratio; AIC: Akaike information criteria; BIC: Bayesians information criteria. *reference group ** p < 0.05. Positive Vuong test statistic values indicate a using the ZIP model instead of the standard count regression model.