Available online at https://journal.rescollacomm.com/index.php/ijqrm/index

International Journal of Quantitative Research and Modeling

e-ISSN 2721-477X
p-ISSN 2722-5046

Vol. 5, No. 4, pp. 370-377, 2024

Risk Analysis Using Poisson-Pareto Models to Estimate Reserve Funds for Catastrophic Diseases in National Health Insurance

Setyo Luthfi Okta Yohandoko[1*], Almira Ajeng Pangestika[2], Yasir Salih[3]

[1]*Master Program in Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jalan Raya Bandung Sumedang Km 21, Hegarmanah, Jatinangor, Sumedang Regency, West Java 45361, Indonesia*
[2]*Badan Pusat Statistik Provinsi Sulawesi Selatan, Jalan Haji Bau No.6 Makassar 90125 Sulawesi Selatan, Indonesia*
[3]*Department of Mathematics, Faculty of Education, Red Sea University, SUDAN*

**Corresponding author email:* Setyo23001@mail.unpad.ac.id

Abstract

Catastrophic diseases such as heart disease, cancer, stroke, and kidney failure pose significant financial burdens on national health insurance systems due to their high treatment costs and frequency. This study utilizes the Poisson-Pareto model to analyze aggregate claims and determine premium loading for these diseases, ensuring the financial sustainability of the National Health Insurance program. Using secondary data from 2018 to 2023, we estimate the parameters for frequency and severity distributions, calculate the expected aggregate claims, and derive the required premium loading at various confidence levels. The results show that heart disease accounts for the highest reserve fund allocation, while kidney failure requires the lowest. These findings emphasize the importance of preparing sufficient reserve funds to manage financial risks associated with catastrophic diseases. The proposed approach provides a robust framework for national health insurance providers to maintain financial stability and optimize resource allocation for high-cost diseases.

Keywords: Catastrophic diseases, Poisson-Pareto model, aggregate claims, premium loading, national health insurance, reserve fund estimation

1. Introduction

Risk management in national health insurance programs is essential to ensure the sustainability of universal healthcare coverage and the financial stability of the system**.** National health insurance programs face the challenge of balancing premium affordability for participants while maintaining sufficient reserves to cover claims (Tawai, et al., 2023), particularly those arising from catastrophic diseases. Catastrophic diseases, such as cancer, cardiovascular diseases, and chronic kidney disease, often require prolonged and costly treatments, significantly impacting healthcare budgets (Heniwati & Thabrany, 2017). These diseases are characterized by their rarity yet high severity, leading to substantial claims that can strain the financial resources of health insurance programs if not managed properly. As a result, effective risk management strategies, including the accurate estimation of claim reserves and premium adjustments, are critical to addressing the financial risks posed by these high-cost events.

The determination of insurance premiums is a crucial aspect of risk management to ensure the sustainability of national health insurance (Muttaqin, et al., 2015). Inadequate premiums to cover unexpected claims can result in significant financial losses for national health insurance provider (Maf'ula, & Mi'raj, 2022). Therefore, premium determination must be based on comprehensive risk analysis that accounts for the uncertainties in both the frequency and severity of claims. Probabilistic approaches are often employed to estimate aggregate claims by simultaneously considering the distributions of claim frequency and claim severity (Embrechts, 1997). The Pareto distribution is frequently used to model claim severity due to its ability to capture heavy-tailed events rare occurrences with large values that have significant impacts. This distribution is particularly flexible in capturing the pattern of large claims in insurance data, especially in sectors exposed to high-loss risks, such as fire insurance and natural disaster coverage (Kleiber & Kotz, 2003). Parameters of the Pareto distribution, such as α (shape parameter) and θ (scale parameter), provide effective control in modeling the probabilities of large claims. To address the risks posed by variability in large claims, a more comprehensive approach is needed in premium calculation. One such approach involves the inclusion of security loading, an additional component added to the pure premium to account for variability in aggregate claims. This variability encompasses risks related to large claims or claim frequencies exceeding the expected average. In practice, the value of the security loading is determined based on the standard deviation of aggregate claims and a certain confidence level, often assuming a standard normal distribution. This adjustment allows national health insurance provider to maintain sufficient financial reserves to manage claim uncertainties (Wang, 2000).

This study is significant due to the importance of understanding risk management for national healthcare providers to ensure the financial stability of the system. Effective management of financial reserves is essential for sustaining the program, particularly in the face of uncertainties arising from claim variability and the high costs associated with catastrophic diseases. Without proper reserve estimation, the national health insurance program risks financial strain, which could compromise its ability to provide comprehensive coverage to participants. The objective of this research is to determine the reserve financing for the national health insurance program, which may be allocated to participants, by analyzing aggregate claims distributed as Poisson-Pareto. This approach provides a probabilistic framework for estimating claims and reserves, helping policymakers and administrators design fair and sustainable funding strategies while safeguarding the program's financial resilience.

2. Literature Review

Previous studies have investigated various approaches to premium loading calculations, each offering distinct methodologies to address the challenges posed by claim variability and extreme risks in insurance. Wang (2000) introduced a distortion transformation framework, which focuses on calculating fair premiums by incorporating extreme risks into the pricing structure. This approach applies a mathematical distortion to the probability distribution of aggregate claims, effectively emphasizing the tail-end risks. By adjusting the weight assigned to extreme claim values, Wang's framework ensures that premiums reflect the heightened financial risks associated with rare but significant loss events, offering a robust tool for risk-averse insurers.

Bühlmann (1985) explored the equilibrium principle as a foundation for premium determination, grounded in actuarial science. This principle balances the insurer’s expected payouts with a suitable profit margin, while accounting for the variability and uncertainty inherent in aggregate claims. Bühlmann emphasized the need for a systematic and fair allocation of premiums to policyholders, ensuring both the financial sustainability of the insurer and fairness in risk-sharing among participants. His work also incorporated statistical techniques to model claim distributions and provided insights into optimizing premium calculations based on observed data and risk tolerances.

Meanwhile, Panjer & Willmot (1992) examined the modeling of aggregate claim distributions using a combination of Poisson and Pareto distributions. Their approach involved decomposing aggregate claims into two components: the frequency of claims, modeled by a Poisson distribution, and the severity of claims, represented by a Pareto distribution. This method provides a flexible framework for capturing the variability in both the number of claims and the magnitude of individual claims. By integrating these two distributions, Panjer & Willmot demonstrated how the resulting aggregate claim distribution could be used to analyze the impact of claim variability on premium calculations. Their findings highlighted the effectiveness of this approach in accurately estimating the required reserves and premium loadings for scenarios involving high-risk and extreme loss events, such as catastrophic insurance claims.

Collectively, these studies have significantly advanced the methodologies for determining premium loadings. By leveraging probabilistic modeling, actuarial principles, and mathematical distortions, they offer insurers tools to better anticipate and manage the financial risks associated with claim variability and extreme events. These approaches provide the foundation for further research into enhancing premium calculations and risk management strategies in modern insurance practices.

3. Materials and Methods

3.1. Materials

The data used in this study is secondary data, which includes the number of catastrophic diseases treated and the financing burden borne by the national health insurance program from 2018 to 2023. The data was obtained from the report on the monitoring and evaluation results of the National Health Insurance program for the second semester of 2023 (National Social Security Council, 2024). The number of claims indicates the amount of claims submitted by participants of the national health insurance program each year. Meanwhile, the claim amount represents the benefits paid by the national health insurance provider to participants who submitted financing claims (David, 2015).

Table 1. Data on the frequency of catastrophic diseases and the cost burden (in billion rupiah) covered by National Health Insurance

Disease	2018		2019		2020		2021		2022		2023	
	Freq.	**Cost**	**Freq.**	**Cost**	**Freq.**	**Cost**	**Freq.**	**Cost**	**Freq.**	**Cost**	**Freq.**	**Cost**
Heart Disease	11,628,273	9,389	13,041,443	10,276	11,592,990	8,296	12,934,931	8,671	15,495,646	12,144	20,037,280	17,629
Cancer	1,990,091	2,979	2,452,749	3,543	2,294,114	3,134	2,595,520	3,500	3,147,895	4,500	3,864,086	5,979
Stroke	1,914,455	2,271	2,127,609	2,549	1,789,261	2,336	1,992,014	2,363	2,536,620	3,234	3,461,563	5,209
kidney disease	1,648,667	2,116	1,763,518	2,321	1,602,059	1,922	1,417,104	1,781	1,322,798	2,355	1,501,016	2,939

Source: BPJS Kesehatan 2018-2023, kesehatan.djsn.go.id/kesehatan/doc/laporan-semester/Lapres_Sem_II_2023.pdf (diakses)

Figure 1 below illustrates the number of catastrophic disease cases covered by the National Health Insurance program, while Figure 2 shows the total costs incurred for these four catastrophic diseases.

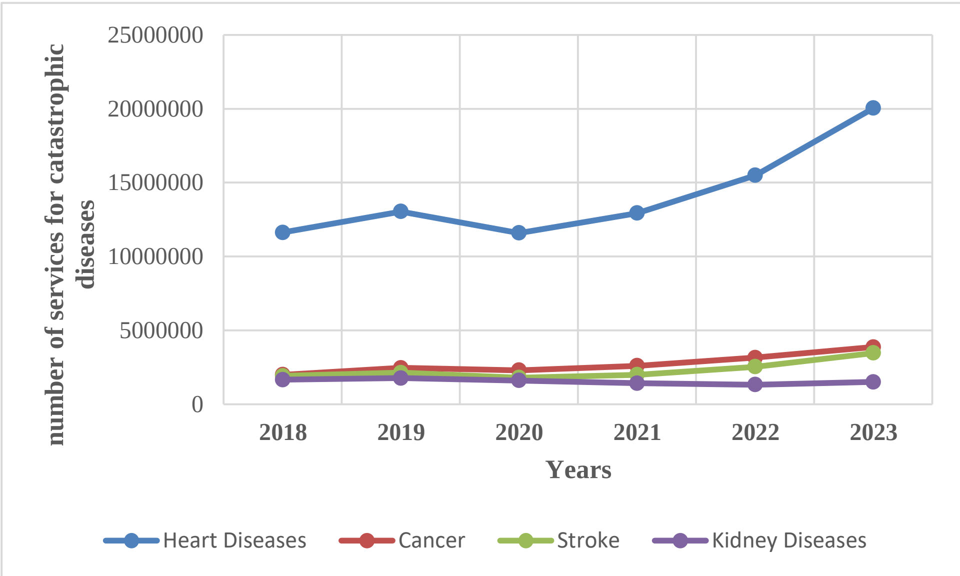


Figure 1. Number of catastrophic disease cases from 2018 to 2023

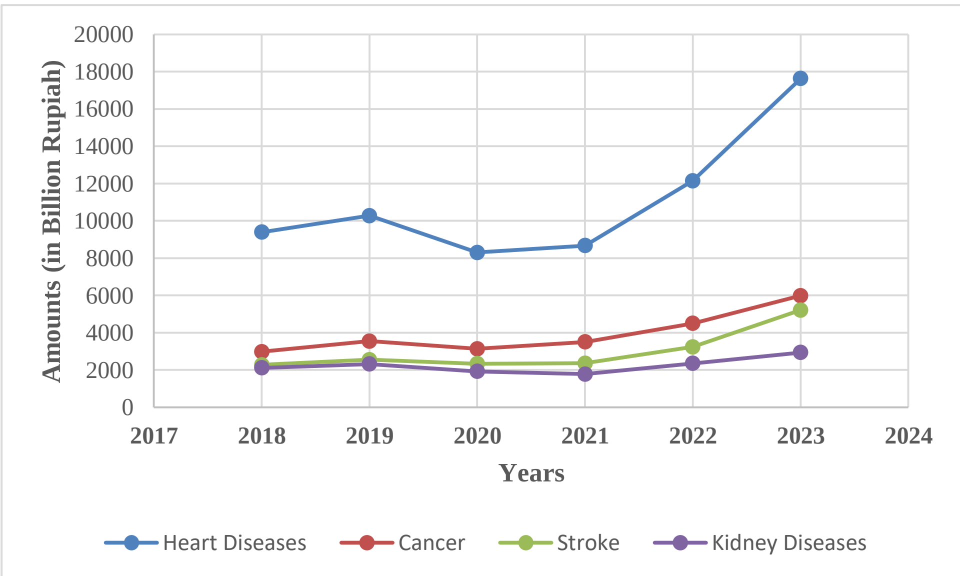


Figure 2. otal costs incurred for catastrophic disease cases from 2018 to 2023

3.2. Methods

The aggregate claim distribution model is used to represent the total claim costs that an insurance company must pay within a specific period. In this study, aggregate claims (S) are defined as the sum of all individual claims that occur within a given period, denoted as (Bowers, et al., 1997):

$$S = X_1 + X_2 + X_3 + \cdots + X_N \tag{1}$$

$$S = \sum_{i=1}^{N(t)} X_i \tag{2}$$

where N is the number of claims that occur during the period, and X_i is the amount of the i-th individual claim. This model combines two components of probability distributions: the Poisson distribution to model the number of claims (N) and the Pareto distribution to model the claim amounts (X_i).

3.2.1. Frequency Claims Distribution (Poisson)

The Poisson distribution is a discrete probability distribution used to model the number of occurrences within a specific time interval or space, where the events occur independently with a certain average rate (λ). The Poisson distribution has the following probability mass function (PMF):

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ dengan } k = 0,1,2, \dots \tag{3}$$

where,

X is the number of events,
k is a certain amount of X,
λ is the average events per time interval.

3.2.2. Distribution of Claim Cost Burden (Pareto)

The Pareto distribution is a continuous probability distribution used to model data with heavy tails, where a small portion of the population accounts for a large proportion of the total size (the 80/20 principle or Pareto Principle). The Pareto distribution has two main parameters: α as the shape parameter, which controls the heaviness of the distribution's tail, and x_m or β as the scale parameter, which determines the minimum value of the random variable. The probability density function (PDF) of the Pareto distribution is:

$$f(x) = \frac{\alpha {x_m}^{\alpha}}{x^{\alpha+1}} \text{ dengan } x \geq x_m \tag{4}$$

where,

x_m is the minimum value of random variable ($x_m > 0$).
$\alpha > 0$ is shape parameter.

3.2.3. Calculation of Premium Loading

The calculation of premium loading aims to adjust the insurance premium according to the aggregate claims that occur, considering the claim distribution that has been analyzed. This premium loading is typically calculated based on the level of risk emerging from the claims made during a specific period. One way to calculate premium loading is by using a formula based on the expectation of aggregate claims, which combines the claim frequency distribution and the claim severity distribution. The premium loading can be calculated by estimating the expected aggregate claims per period, given by the following equation (Prabowo, et al., 2019):

$$\mathbb{E}[S] = \mathbb{E}[N] \cdot \mathbb{E}[X] \tag{5}$$

where,

$\mathbb{E}[S]$ is the average of aggregate claim,
$\mathbb{E}[N]$ is the average of claim frequency,
$\mathbb{E}[X]$ is the average of amount of claim.

Premium loading is calculated by multiplying the expected aggregate claims by the desired profit margin or other risk factors. The premium loading can be computed using the following equation:

$$P = \mathbb{E}[S] \cdot (1 + \theta) \quad (6)$$

where,

P is the net premium,

θ is the risk factor or profit margin added to the base premium to cover the expected losses.

The value of θ is determined based on the security margin, which can be calculated using a confidence level of $1 - \alpha$ from the standard deviation of aggregate claims (Prabowo, et al., 2019; Sukono, et al., 2018). Specifically,

$$\theta = Z_{1-\alpha} \cdot \sigma_S \quad (7)$$

$$\theta = Z_{1-\alpha} \cdot \sqrt{Var(S)} \quad (8)$$

$$\theta = Z_{1-\alpha} \cdot \sqrt{\mathbb{E}(N) \cdot Var\,(X) + \mathbb{E}(X)^2 \cdot Var\,(N)} \quad (9)$$

where,

$Z_{1-\alpha}$ is the z score for a given level of confidence (e.g., $Z = 1.96$ for 95% of confidence),

$Var(S)$ is the variance of aggregate claim,

$Var(N)$ is the variance of claim frequency,

Var(X) is the variance of amount claim.

3.2.4. Parameter Estimation

In aggregate claim analysis, accurate estimation of the claim distribution parameters is crucial for calculating the premium loading and managing risk. There are two main methods used for parameter estimation in claim distribution models: the Method of Moments and Maximum Likelihood Estimation (MLE).

a). Method of Moments

Method of Moments is a statistical technique used to estimate the parameters of a probability distribution by equating the sample moments (i.e., the mean, variance, etc.) to the theoretical moments of the distribution (Walpole, et al., 2011). The idea is to match the first k moments of the sample with the first k moments of the distribution and solve for the unknown parameters. In the case of using the Moment Generating Function (MGF), the MGF is defined as:

$$M_X(t) = \mathbb{E}[e^{tX}] \quad (10)$$

where t is a real number and X is the random variable. The MGF is helpful in deriving moments (e.g., $E[X] = M'_X(0)$ for the first moment) and can be used to estimate parameters by equating the sample moments to the theoretical moments derived from the MGF.

b). Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is a method used to estimate the parameters of a probability distribution by maximizing the likelihood function, which represents the probability of observing the given data as a function of the parameters (Walpole, et al., 2011). The likelihood function for a sample $x_1, x_2, x_3, \ldots, x_n$ is the product of the probability density function (PDF) evaluated at each data point:

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \quad (11)$$

where $f(x_i; \theta)$ is the PDF of the distribution and θ\thetaθ is the parameter to be estimated. The log-likelihood is often used for convenience, which is the natural logarithm of the likelihood function:

$$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)\,. \quad (12)$$

MLE estimates are found by differentiating the log-likelihood function with respect to θ, setting it to zero, and solving for θ, thus obtaining the parameter estimates that maximize the likelihood of observing the given data.

4. Results and Discussion

The Pareto distribution and Poisson distribution are often chosen in insurance and risk management models due to their specific properties that align well with the characteristics of claims data, particularly in cases involving catastrophic events. The estimated parameters for the four catastrophic diseases can be seen in the table below.

Table 2. The estimated parameters for the frequency distribution and cost burden of catastrophic diseases in the national health insurance program

Diseases	Freq	Cost	
	Lamda	alpha	betha
Heart Diseases	11,068.0	45,916	661,130,000
Cancer	3,939.2	65,489	147,650,000
Stroke	2,993.7	45,447	86,068,000
Kidney Diseases	2,239.0	67,899	97,378,000

Based on equations (5) the expected aggregate claim value $E[S]$ can be formulated as follows:

$$\mathbb{E}[S] = \lambda \cdot \frac{\alpha\beta}{\alpha - 1} \tag{13}$$

The standard deviation of the aggregate claim value, based on equation (9), is defined as:

$$Var\ (S) = \lambda \cdot \frac{2\alpha\beta^2}{(\alpha - 1)^2(\alpha - 2)} + \left(\frac{\alpha\beta}{\alpha - 1}\right)^2 \cdot \lambda \tag{14}$$

$$Std\ (S) = \sqrt{Var\ (S)} \tag{15}$$

The calculation results for the aggregate claims and its variance of the four catastrophic diseases are shown in Table 3.

Tabel 3. Aggregat claims, variance and standard deviation value of four catastrophics diseases

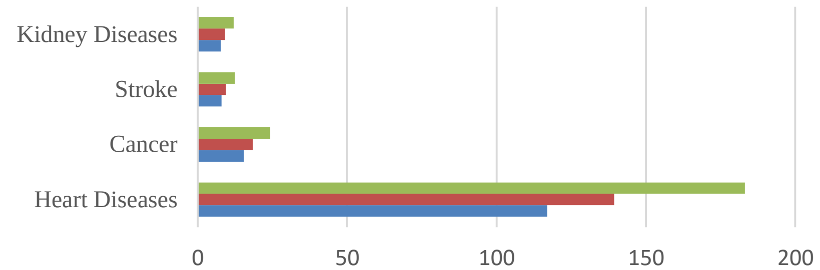
Diseases	Aggregate Claim	Variance of Aggregate Claim	Standard Deviation of Aggregate Claim
Heart Diseases	7,4803E+12	5,06057E+21	71,137,677,649
Cancer	5,9064E+11	8,86032E+19	9,412,925,259
Stroke	2,6346E+11	2,3209E+19	4,817,574,655
Kidney Diseases	2,2129E+11	2,18805E+19	4,677,660,395

The calculation of the premium loading (θ) based on equation (7) with various α is as follows:

Tabel 4. Premium loading based on the confidence level (in billion rupiah)

Diseases	Premium Loading for Confidence level of 90%	Premium Loading for Confidence level of 95%	Premium Loading for Confidence level of 99,5%
Heart Diseases	117,021	139,430	183,180
Cancer	15,484	18,450	24,238
Stroke	7,925	9,440	12,405
Kidney Diseases	7,695	9,170	12,045

Based on the estimated loss distribution parameters, the average aggregate claims for heart disease, cancer, stroke, and kidney failure are approximately 7.480,3 billion rupiah, 590,64 billion rupiah, 263,46 billion rupiah and 221,29 billion rupiah, respectively. The corresponding standard deviations are approximately 71,14 billion rupiah, 9.41 billion rupiah, 4,82 billion rupiah and 4.68 billion rupiah. To ensure the financial stability of the national health insurance program, it is recommended that the national health insurance provider prepare reserve funds for claims related to these four catastrophic diseases. The highest reserve fund for premium loading is allocated to heart disease, with a 99.5% confidence level, amounting to 183.18 billion rupiahs, while the lowest reserve fund is allocated to kidney diseases with a 90% confidence level, amounting to 7.696 billion rupiahs, as shown in Figure 3.



	Heart Diseases	Cancer	Stroke	Kidney Diseases
Premium Loading for Confidence level of 99,5%	183.18	24.238	12.405	12.045
Premium Loading for Confidence level of 95%	139.43	18.45	9.44	9.17
Premium Loading for Confidence level of 90%	117.021	15.484	7.925	7.695

Figure 3. Premium loading based on the confidence level (in billion rupiah) for four catastrophics diseases

5. Conclussion

This study demonstrates the importance of accurately estimating aggregate claims and determining premium loading for catastrophic diseases to ensure the financial stability of the national health insurance program. Using the Poisson-Pareto model, the average aggregate claims and their standard deviations for heart disease, cancer, stroke, and kidney failure were estimated. The results indicate that heart disease requires the highest reserve fund allocation, reflecting its significant financial impact, while kidney failure requires the lowest. These findings highlight the necessity of preparing adequate reserve funds, particularly for high-risk diseases, to mitigate financial risks and maintain the sustainability of the health insurance system. By considering different confidence levels for premium loading, the national health insurance provider can strategically manage its financial reserves to address uncertainties in claim costs while ensuring continued service delivery.

References

Bowers, Jr. N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). Actuarial Mathematics. Second Edition. Illinois: The Society of Actuaries.

Bühlmann, H. (1985). Premium Calculation from Top Down. ASTIN Bulletin, 15(2), 89-101.

David, M. (2015). Auto insurance premium calculation using generalized linear models. Procedia Economics and Finance, 20, 147-156.

de Zea Bermudez, P., & Kotz, S. (2010). Parameter estimation of the generalized Pareto distribution—Part I. Journal of Statistical Planning and Inference, 140(6), 1353-1373.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997): Modelling Extremal Events for Insurance and Finance, Springer-Verlag. 645 pp (1.04 kg). ISSN 0172-4568, ISBN 3-540-60931-8.

Han, M. (2020). The E-Bayesian estimation and its E-MSE of Pareto distribution parameter under different loss functions. Journal of Statistical Computation and Simulation, 90(10), 1834-1848.

Kleiber, C., & Kotz, S. (2003). Statistical size distributions in economics and actuarial sciences. Wiley.

DOI:10.1002/0471457175.

Maf'ula, F., & Mi'raj, D. A. (2022). Islamic insurance in Indonesia: Opportunities and challenges on developing the industry. Journal of Islamic Economic Laws, 5(1), 116-138.

Mukhaiyar, U., Dianpermatasari, A., Dzakiya, A., Widyani, S. B., & Syam, H. K. (2024). The Value at Risk Analysis using Heavy-Tailed Distribution on the Insurance Claims Data. JTAM (Jurnal Teori dan Aplikasi Matematika), 8(4), 1233-1248.

National Social Security Council. (2024). *Report on the results of monitoring and evaluation of the implementation of the national health insurance program for the second semester of 2023.* National Social Security Council. Retrieved from kesehatan.djsn.go.id/kesehatan/doc/laporan-semester/Lapres_Sem_II_2023.pdf

Panjer, H. H., & Willmot, G. E. (1992). Insurance risk models. Society of Actuaries.

Prabowo, A., Mamat, M. Sukono, and Taufiq, A.A. (2019). Pricing of Premium for Automibile Insurance using Bayesian Method. International Journal of Recent Technology and Engineering, Vol. 8, No. 3, 6226-6229.

Pratiwi, A. B., Setiyaningsih, H., Kok, M. O., Hoekstra, T., Mukti, A. G., & Pisani, E. (2021). Is Indonesia achieving universal health coverage? Secondary analysis of national data on insurance coverage, health spending and service availability. BMJ open, 11(10), e050565.

Sukono, Riaman, Lesmana, E., Wulandari, R., Napitupulu, H., and Supain, S. (2018). Model Estimation of Claim Risk and Premium of Motor Vehicle Insurance Using Bayesian Method. IOP Conf. Series: Material Science and Engineering, 300(2018): 012027.

Tawai, A., Afriadi, Z., & Yusuf, M. (2023). Model Implementasi Program Jaminan Kesehatan Nasional (JKN) di Dinas Kesehatan Kota Kendari. *NeoRespublica: Jurnal Ilmu Pemerintahan*, 5(1), 275-284.

Walpole, R.E., Myers, R.H., Myers, S.L., dan Ye, K. 2011. Probability and Statistics for Engineers and Scientists. Boston: Pearson Education.

Wang, S. S. (2000). A Class of Distortion Operators for Pricing Financial and Insurance Risks. Journal of Risk and Insurance, 67(1), 15-36.