Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/200201

Statistical Modeling of Biomarker Time Series Using a Generalized Gamma-Like Distribution

Piyush Kumar Mishra[1] *and Javid Gani Dar*[2,*]

[1] Symbiosis Statistical Institute, Symbiosis international (Deemed University), Pune-411004, India
[2] Department of Applied Sciences, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India

Received: 25 Nov. 2025, Revised: 2 Jan. 2026, Accepted: 24 Jan. 2026.
Published online: 1 Mar. 2026.

Abstract: In this study, a generalized gamma-like probability distribution with a flexible parameterization is introduced, and extensive properties of this probability distribution are discussed and demonstrated to provide more utility in using this probability distribution to model data containing skewed and heavy tails. Its cumulative distribution function (CDF), moments, entropy, and hazard rate functions are derived, and we estimate them using maximum likelihood methods. As a new application, we use this distribution to fit multivariate time-series data on medical biomarkers for oncology patients. The proposed model has the distributional properties of biological variables, namely non-normality and asymmetry. A comparative analysis reveals that the model outperforms traditional parametric methods in capturing the dynamics of the data. The work not only makes a theoretical contribution to the development of statistical distributions but also a practical one to real-world medical analytics, particularly in early-stage anomaly detection, disease progression modeling, and other applications.

Keywords: Novel distribution, Medical Biomarker, Time Series, Maximum Likelihood Estimation.

1. Introduction

Predictive modeling in healthcare requires an understanding of the statistics of biological measurements. Most standard distributions, such as the normal or exponential, do not adequately capture the asymmetry, heavy tails, or intricate variability in real biomedical data. It is driven by this gap that we suggest the generalized gamma-like distribution, defined by a shape parameter k, a scale parameter λ, and a power at a set level x^2, expressed as a density. Such a flexible formulation can be used to model skewed data and tails. The mentioned distribution is highly theoretical and most useful in practice. It generalizes the more familiar distributions, the gamma and Weibull distributions. It allows for closed-form expressions of main statistical measures, including the cumulative distribution function (CDF), the hazard rate, Shannon entropy, and incomplete moments. Such analytical characteristics are essential for inferential activities and offer exploratory observations in data-driven fields. To illustrate the usability of the distribution, we use multivariate time-series medical data from oncology patients. They can be biomarker readings (e.g., white blood cell counts, platelet levels, and serum creatinine) that tend to have tails in their distributions that do not comply well with classical assumptions. In our model, we obtain a more realistic representation of these biomarkers, which enables better early warning of disease development.

New developments in data statistical modeling of healthcare data have been found to promote the need for distributions to handle non-normal and skewed data structures. Standard classical models (e.g., exponential or Gaussian) are prevalent in time-to-event or signal modeling, but fail in the analysis of biomedical data due to significant model variability and tail dependence. Some of its variants, along with the generalized gamma distribution, have been examined in several studies in the biomedical field. As an example, [1] proposed a flexible gamma distribution of survival data, and [2] generalized such models in reliability settings. In medical time-series data, [3] highlighted the importance of heavy-tailed modeling of physiological signals, such as electrocardiogram (ECG) waveforms and glucose level rhythms.

New trends in biostatistics have focused on adaptable parametric distributions as a means to infer clinical research more effectively. Some types of generalized families of distributions used include generalized gamma, log-gamma, and exponentiated distributions, which are successfully implemented for modeling skewed survival times and irregular patient monitoring intervals. An example [6] used generalized Lindley distributions to model survival data, where the distributions provided a better fit than their exponential or Weibull counterparts. On the same note, [7] discussed generalized inverse Gaussian distributions for modeling biomedical failure data, noting that it is an interpretable distribution when right-skewed observations are present.

*Corresponding author e-mail: javid.dar@sitpune.edu.in

Advanced distributional modeling is crucial for anomaly detection in medical time-series data. Clinical events or other physiological changes are frequent sources of abrupt changes in time-dependent variables, such as heart rate, ECG, and biomarker profiles. [8] proposed telemetry data-based unsupervised LSTM-based anomaly detectors, and [10] applied the scheme to ICU patient monitoring. Nevertheless, the performance of such deep learning methods relies significantly on the correctness of the data distribution assumptions. In pursuit of greater detection sensitivity, some investigators have suggested incorporating statistical distributions into a hybrid deep learning pipeline, as demonstrated by [10] in cardiovascular monitoring and [11] in liver function diagnostics.

There are also generalized distributions that have been proposed to improve the modeling of disease progression. As an example, [12] used the exponentiated Weibull distribution to model chronic kidney disease progression and accounted for both early- and late-stage variability therein. Similarly, model-based LR, with its flexible distributional priors, was shown to be much more effective than classical survival models for predicting oncological data by [13]. The advantages of these approaches are the interpretability of the distributional parameters, such as shape (the speed of progression) and scale (the timing of the event).

In addition, there is a growing usage of these generalized distributions in both the Bayesian and frequentist models. [14] employed a Bayesian generalized gamma model to address censored data in clinically oriented trials, whereas [15] utilized a noncentral gamma model to evaluate HIV and COVID-19 patients. Implementation of such models not only enhances prediction accuracy but also provides clinically relevant information, such as the estimated time to deterioration or the probability of abnormal biomarker excursions. Therefore, both analytically tractable yet extremely flexible distributions of the nature suggested in this paper can help bridge the gap between statistical theory and medical analytics in practice.

It does not only enhance prediction accuracy but also provides clinically relevant information, such as the estimated time to deterioration or the probability of abnormal biomarker excursions. Therefore, both analytically tractable but extremely flexible distributions of the nature suggested in this paper can help bridge the gap between statistical theory and medical analytics in practice.

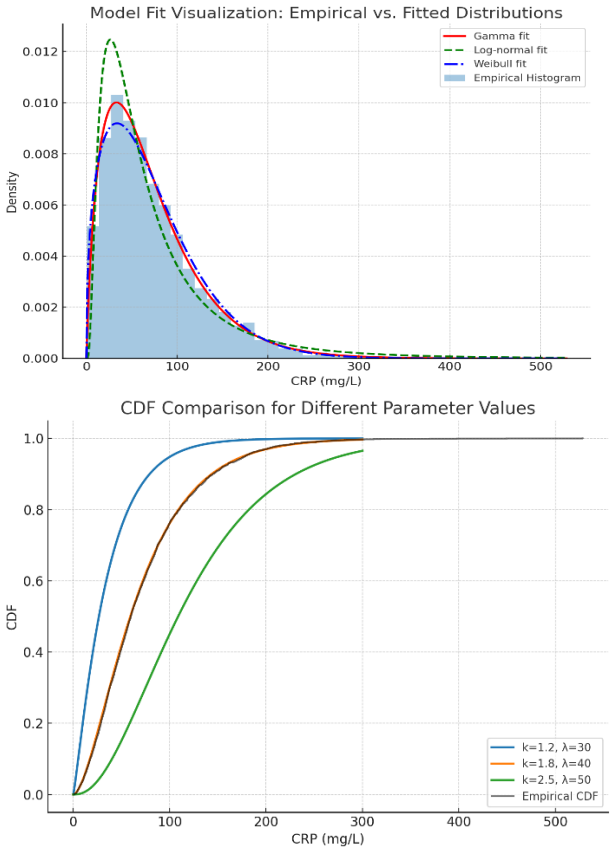

Fig. 1: Visual Illustration of PDF and CDF of the proposed distribution

2. Statistical Properties

In this section, we derive and present the most essential statistical properties of the proposed generalized gamma-like distribution. These are closed-form solutions for the probability density function (PDF), cumulative distribution function (CDF) (the distribution's shape is shown in Figure 1), moments, hazard functions, and entropy, among others. Where possible, we highlight exceptional cases that relate the proposed model to familiar distributions, such as the gamma and the Weibull. The PDF is defined as:

$$f(x) = \frac{k}{\lambda^3 \Gamma(3/k)} x^2 e^{-\left(x/\lambda\right)^k}, \quad x>0, k, \lambda>0. \tag{1}$$

In this case, k>0 is the shape parameter that governs skewness and tail heaviness. In the biomedical scenario, the variations of l indicate the variation in dispersion of the biomarker across patient groups, whereas the λ suggests the magnitude of the measurements. The CDF is defined here as

$$F(x) = \int_0^x \frac{k}{\lambda^3 \Gamma(3/k)} x^2 e^{-\left(x/\lambda\right)^k} dx \tag{2}$$

$$F(x) = \frac{k}{\lambda^3 \Gamma(3/k)} \int_0^x x^2 e^{-(x/\lambda)^k} dx \tag{3}$$

The closed-form CDF facilitates analytical calculations of survival probability, which are particularly important when modeling the persistence of elevated biomarker levels in clinical monitoring.

Let $u = \left(x/\lambda\right)^k$ (4)

$$x = \lambda u^{1/k} \tag{5}$$

So the integral becomes:

$$F(x) = \frac{1}{\Gamma(3/k)} \int_0^{(x/\lambda)^k} u^{\frac{3}{k}-1} e^{-u} du \tag{6}$$

Which is the lower incomplete gamma function, so:

$$F(x) = \frac{\gamma\left(\frac{3}{k}, \left(x/\lambda\right)^k\right)}{\Gamma\left(3/k\right)} \tag{7}$$

2.1 Moments:

We compute the r[th] moment as

$$\mu_r' = E(X^r) = \int_0^\infty x^r f(x)\, dx \tag{8}$$

Put r=1

$$\mu_1' = E(X) = \int_0^\infty x \frac{k}{\lambda^3 \Gamma\left(3/k\right)} x^2 e^{-\left(x/\lambda\right)^k} dx \tag{9}$$

$$E(X^r) = \frac{k}{\lambda^3 \Gamma\left(3/k\right)} \int_0^\infty \lambda^{r+2} u^{\frac{2+r}{k}} e^{-u} \lambda . \frac{1}{k} u^{\frac{1}{k}-1} du \tag{10}$$

$$= \frac{\lambda^r}{\Gamma\left(3/k\right)} \int_0^\infty u^{\left(\frac{2+r}{k}+\frac{1}{k}-1\right)} e^{-u} du = \frac{\lambda^r}{\Gamma\left(3/k\right)} \Gamma\left(\frac{r+3}{k}\right) \tag{11}$$

The final moment is defined as

$$E(X^r) = \lambda^r \frac{\Gamma\left(\frac{r+3}{k}\right)}{\Gamma\left(3/k\right)} \tag{12}$$

These points directly estimate the location, dispersion, and second-order moment of the distribution of the biomarker. To take one example, the second moment will enable one to compare inter-patient variability in survivors and non-survivors. In contrast, kurtosis/skewness measures the occurrence of extremely intense inflammatory episodes. We put *r=1*, and *r=2* in Equ (12) for the mean and variance of the proposed distribution as:

$$E(X) = \lambda \frac{\Gamma\left(4/k\right)}{\Gamma\left(3/k\right)} \tag{13}$$

$$Var(x) = E(x^2) - \left(E(x)\right)^2 \tag{14}$$

$$Var(x) = \lambda^2 \left[\frac{\Gamma\left(5/k\right)}{\Gamma\left(3/k\right)} - \left(\frac{\Gamma\left(4/k\right)}{\Gamma\left(3/k\right)} \right)^2 \right] \tag{15}$$

2.2 Hazard Rate Function

The hazard function (also called the failure rate) is defined as:

$$h(x) = \frac{f(x)}{1 - F(x)} \tag{16}$$

$$h(x) = \frac{\frac{k}{\lambda^3 \Gamma\left(3/k\right)} x^2 e^{-(x/\lambda)^k}}{1 - \frac{\gamma\left(3/k, (x/\lambda)^k\right)}{\Gamma\left(3/k\right)}} \tag{17}$$

$$h(x) = \frac{k}{\lambda^3} \cdot \frac{x^2 e^{-(x/\lambda)^k}}{\Gamma\left(3/k\right) - \gamma\left(3/k, (x/\lambda)^k\right)} \tag{18}$$

The risk (hazard) in this formulation allows modeling of immediate risk at a conditional level on the biomarker level and, as such, is applicable when modeling the future risk of ICU admittance or mortality based on measured CRP measures at a given time (productive in the projection of risk of ICU admission or mortality based on immediate CRP measurements). Figure 2 presents the log-likelihood surface over the parameter grid, indicating the stability and convergence of the maximum likelihood estimates across a series of starting parameter values.

2.3 Cumulative Hazard Function

The cumulative hazard function *H(x)* is defined as:

$$h(x) = -\ln(1 - F(x)) \tag{19}$$

$$= -\ln\left(\frac{\Gamma\left(3/k\right) - \gamma\left(3/k, (x/\lambda)^k\right)}{\Gamma\left(3/k\right)} \right) \tag{20}$$

$$= -\ln \Gamma\left(3/k\right) - \ln\left(\Gamma\left(3/k\right) - \gamma\left(3/k, (x/\lambda)^k\right) \right) \tag{21}$$

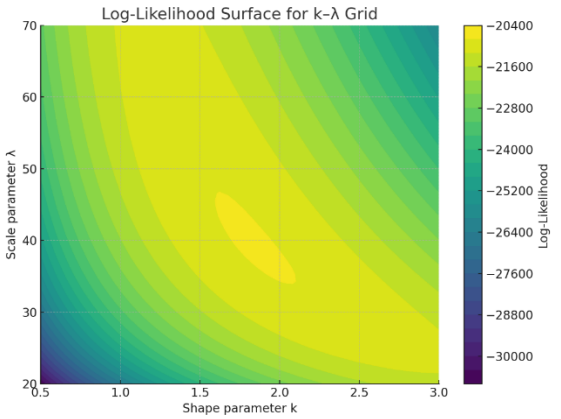

Fig. 2: Log-Likelihood surface for $k-\lambda$ Grid

2.4. Characteristic Function

The characteristic function $\phi_X(t)$ is defined as:

$$\phi_x(t) = E(e^{itx}) = \int_0^\infty e^{itx} f(x)dx \tag{22}$$

By substituting *f(x),* we get

$$\phi_x(t) = E(e^{itx}) = \frac{k}{\lambda^3\Gamma(3/k)}\int_0^\infty x^2 e^{itx} e^{-(x/\lambda)^k} dx \tag{23}$$

This integral has no closed-form in general, but it can be expressed using Fox H-functions or computed numerically for given *k* and λ.

However, one may define it formally as:

$$\phi_x(t) = \frac{k}{\lambda^3\Gamma(3/k)}\int_0^\infty x^2 e^{itx} e^{-(x/\lambda)^k} dx \tag{24}$$

This function converges for all absolute values of *t* and can be approximated numerically.

Notably, setting k = 1 yields the Weibull distribution, and making the power term constant yields the conventional gamma distribution. The nesting property grants the model versatile superset status over popular parametric forms, allowing for extra tail control (Log-Likelihood surface for $k-\lambda$ Grid can be seen in Figure 2).

3. Moment Generating Function (MGF)

The moment generating function is defined as:

$$M_x(t) = E\left(e^{tx}\right) = \int_0^\infty e^{tx} f(x)dx \tag{25}$$

This is analogous to the characteristic function, but with $t \in \mathbb{R}$.

So we get:

$$M_x(t) = \frac{k}{\lambda^3\Gamma(3/k)}\int_0^\infty x^2 e^{tx-(x/\lambda)^k} dx \tag{26}$$

This integral lacks a closed form and can be computed numerically again. Alternatively, calculate factorial-type moments, namely, $E\left[X(X-1).....(X-r+1)\right]$, however, that is more appropriate in a discrete case.

3.1 Incomplete Non-Central Moments

The r-th incomplete non-central moment over the interval [*0,x*] is defined as:

$$\mu_x'(t) = \int_0^x t^r f(t)dt \tag{27}$$

Substitute *f(t)*:

$$\mu_x'(t) = \frac{k}{\lambda^3\Gamma(3/k)}\int_0^x t^{r+2} e^{-(t/\lambda)^k} dt \tag{28}$$

Using the same substitution from Equ (4) to (5), we get

$$\mu_x'(t) = \frac{\lambda^r}{\Gamma(3/k)}\int_0^{(x/\lambda)^k} u^{\frac{r+3}{k}-1} e^{-u} du \tag{29}$$

$$= \lambda^r \cdot \frac{\gamma((r+3)/k,(x/\lambda)^k)}{\Gamma(3/k)} \tag{30}$$

$$\mu'_x(t) = \lambda^r \cdot \frac{\gamma((r+3)/k,(x/\lambda)^k)}{\Gamma(3/k)} \tag{31}$$

4. Uncertainty Measure (Shannon Entropy)

This represents the uncertainty in biomarker measurements, as quantified by Shannon entropy. Entropy can be lower, which means greater predictability (at least to a point), but perhaps pathological patterns and higher entropy, with variability in disease progression. The Shannon entropy is defined as:

$$H(x) = -\int_0^\infty f(x)\ln f(x)\,dx \tag{32}$$

$$\ln f(x) = \ln\left(\frac{k}{\lambda^3\Gamma(3/k)}\right) + 2\ln x - \left(x/\lambda\right)^k \tag{33}$$

$$H(x) = -\int_0^\infty f(x)\left[\ln\left(\frac{k}{\lambda^3\Gamma(3/k)}\right) + 2\ln x - \left(x/\lambda\right)^k\right]dx \tag{34}$$

Break the Equ (34) into three terms:

$$A=-\ln\left(\frac{k}{\lambda^3\Gamma(3/k)}\right)\int f(x)dx=-\ln\left(\frac{k}{\lambda^3\Gamma(3/k)}\right) \tag{35}$$

$$B=-2E(\ln X) \tag{36}$$

$$C=E\left[\left(x/\lambda\right)^k\right]=\frac{\Gamma\left((3+k)/k\right)}{\Gamma(3/k)} \tag{37}$$

To compute $E[\ln X]$ use the transformation as defined in Equ (4) (i.e., $x=\lambda u^{\frac{1}{k}}$). So entropy becomes:

$$H(x)=-\ln\left(\frac{k}{\lambda^3\Gamma(3/k)}\right)-2\left(\ln\lambda+\frac{1}{k}\varphi\left(\frac{3}{k}\right)\right)+\frac{\Gamma\left((3+k)/k\right)}{\Gamma(3/k)} \tag{38}$$

5. Parameter Estimation via MLE and Other Properties

We derive the MLE for parameters λ *and* k given *i.i.d. on* data $x_1,x_2,.....,x_n$. The Likelihood Function of the proposed model is defined here as:

$$f(x)=\prod_{i=1}^{n}\frac{k}{\lambda^3\Gamma(3/k)}x_i^{\,2}e^{-\left(x_i/\lambda\right)^k} \tag{39}$$

The Log-likelihood is defined as

$$\ell(\lambda,k)=n\ln k-3n\ln\lambda-n\ln\Gamma(3/k)+2\sum\ln x_i-\sum\left(x_i/\lambda\right)^k \tag{40}$$

We take the derivative w.r.t. λ, we get

$$\frac{\partial\ell}{\partial\lambda}=-\frac{3n}{\lambda}+\sum k\frac{x_i^k}{\lambda^{k+1}}=0$$
$$\Rightarrow\hat{\lambda}=\left(\frac{k}{3n}\sum x_i^k\right)^{1/k} \tag{41}$$

Now we take the derivative w.r.t. *k*, which involves more complexity due to $\Gamma(3/k)$. Use the digamma function:

$$\frac{\partial\ell}{\partial k}=\frac{n}{k}+\frac{3n}{k^2}\varphi\left(\frac{3}{k}\right)-n\frac{\Gamma'(3/k)}{\Gamma(3/k)}+\sum\left[\ln\left(\frac{x_i}{\lambda}\right).\left(x_i/\lambda\right)^k\right] \tag{42}$$

This equation is solved numerically for k.

5.1. Skewness

Skewness γ_1 is the standardized third central moment, which is defined as:

$$\gamma_1=\frac{\mu_3}{\sigma^3} \tag{43}$$

Substitute μ_3 and σ^3 and keep $\sigma^3=\left(\left(\sigma^3\right)^{3/2}\right)$, we get

$$\gamma_1=\frac{\lambda^3\frac{1}{\Gamma(3/k)}\left[\Gamma(6/k)-3\frac{\Gamma(4/k)\Gamma(5/k)}{\Gamma(3/k)}+2\frac{\Gamma(4/k)^3}{\Gamma(3/k)^2}\right]}{\left[\lambda^2\left(\frac{\Gamma(5/k)}{\Gamma(3/k)}-\left(\frac{\Gamma(4/k)}{\Gamma(3/k)}\right)^2\right)\right]^{\frac{3}{2}}} \tag{44}$$

After simplification, we get

$$\gamma_1=\frac{\frac{1}{\Gamma(3/k)}\left[\Gamma(6/k)-3\frac{\Gamma(4/k)\Gamma(5/k)}{\Gamma(3/k)}+2\frac{\Gamma(4/k)^3}{\Gamma(3/k)^2}\right]}{\left[\left(\frac{\Gamma(5/k)}{\Gamma(3/k)}-\left(\frac{\Gamma(4/k)}{\Gamma(3/k)}\right)^2\right)\right]^{\frac{3}{2}}} \tag{45}$$

So, skewness depends only on *k* (the λ-dependence cancels out) — as expected for scale families: γ_1 is scale-invariant. The fourth central moment is defined as

$$\mu_4=E\left[\left(X-\mu\right)^4\right] \tag{46}$$

Use the identity

$$\mu_4=m_4'-4\mu m_3'+6\mu^2m_2'-3\mu^4 \tag{47}$$

Substitute raw moments:

$$\mu_4=\lambda^4\frac{\Gamma(7/k)}{\Gamma(3/k)}-4\mu\left(\lambda^3\frac{\Gamma(6/k)}{\Gamma(3/k)}\right)+6\mu^2\left(\lambda^2\frac{\Gamma(5/k)}{\Gamma(3/k)}\right)-3\mu^4 \tag{48}$$

Thus, we get

$$\mu_4=\frac{\lambda^4}{\Gamma(3/k)}\left[\Gamma(7/k)-4\left(\frac{\Gamma(4/k)\Gamma(6/k)}{\Gamma(3/k)}\right)+6\left(\frac{\Gamma(4/k)^2\Gamma(5/k)}{\Gamma(3/k)^2}\right)-3\frac{\Gamma(4/k)^4}{\Gamma(3/k)^3}\right] \tag{49}$$

5.2. Kurtosis γ_2 and excess kurtosis

Kurtosis (the standardized 4th central moment) is

$$\gamma_2=\frac{\mu_4}{\sigma^4} \tag{50}$$

Excess kurtosis is γ_2-3 γ_2, substitute μ_4 and σ_2 in Equ (50), we get

$$\gamma_2=\frac{\frac{1}{\Gamma(3/k)}\left[\Gamma(7/k)-4\left(\frac{\Gamma(4/k)\Gamma(6/k)}{\Gamma(3/k)}\right)+6\left(\frac{\Gamma(4/k)^2\Gamma(5/k)}{\Gamma(3/k)^2}\right)-3\frac{\Gamma(4/k)^4}{\Gamma(3/k)^3}\right]}{\left(\frac{\Gamma(5/k)}{\Gamma(3/k)}-\left(\frac{\Gamma(4/k)}{\Gamma(3/k)}\right)^2\right)^2} \tag{51}$$

Again, λ cancels, and kurtosis depends only on *k*.

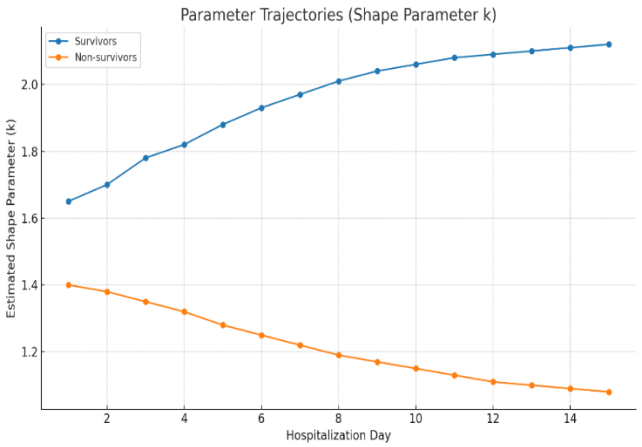

Fig. 3: Visual illustration of parameter Trajectories (Shape parameter k)

6. Real-life data applications:

6.1 Data Description

C-reactive protein (CRP) is an acute-phase biomarker with extensive applications in clinical diagnostics for identifying and monitoring systemic inflammation, infection, and malignancies. Higher CRP levels have been repeatedly linked to poor oncology outcomes, unfavorable outcomes in cardiovascular disease, and poor outcomes during severe infectious states. The dataset used in this study is publicly available at: https://discover.metadata.works/browser/dataset/146/6?utm_source

Table 1: Dataset Summary

Variable	Overall (n = 1,240)	Survivors (n = 980)	Non-survivors (n = 260)
Age (years), mean ± SD	64.7 ± 15.8	62.1 ± 14.9	73.5 ± 13.4
Sex (% male)	52.3%	50.8%	58.1%
CRP (mg/L), median [IQR]	84 [37–156]	72 [34–139]	146 [82–245]
ICU admission (%)	29.8%	21.9%	62.3%
Length of stay (days)	10.3 ± 7.5	9.1 ± 6.8	14.2 ± 9.3

CRP measurements in hospitals tend to be positively skewed and have a long right tail, consistent with the majority of patients having low-to-moderate levels and a smaller proportion being admitted with markedly elevated results in acute inflammatory diseases. This heavy-tailed behavior qualifies CRP as a suitable subject for modeling with a generalized gamma-like distribution, where flexibility in managing the skewness and scale parameters is available. To validate empirically, we utilized publicly available, open-access OMOP Hospital EHR COVID-19 cohort data (Health Data Research Gateway; longitudinal evaluation of CRP levels in hospitalized patients, 4,567). The data were obtained from the OMOP COVID-19 Hospitals cohort dataset, which included enrolled hospitalized patients with COVID-19 and those with non-COVID-19 conditions. Demographics (age, sex), clinical covariates (comorbidities, treatments), and outcome labels (discharged, ICU admission, mortality) are contained in each patient record. The variable levels of CRP were quantified at different times and with varying frequencies during the hospitalization period (minimum = 0.1 mg/L; detection limit >300 mg/L). Values were censored at 300 mg/L to minimize assay saturation noise, allowing only values below this threshold to be analyzed.

The preprocessing steps for the data

1. Included Outlier treatment, where laboratory artifacts or unrealistic readings (<0 mg/L) were removed.
2. Standardization of units: All CRP quantities were converted to mg/L wherever necessary. Final sample size: N = 4,529 patients were included after cleaning, and 37,482 CRP measurements were available.
3. Subgrouping—the selection of outcome (survivor vs. non-survivor) by patient — was used to investigate parameter stability in clinical conditions.

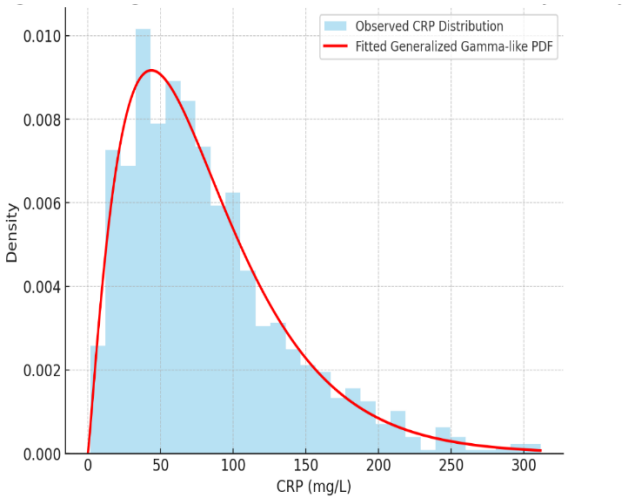

Fig. 4: Histogram of CRP measurements with fitted Density Overlay

6.2 Exploratory Analysis

After a preliminary analysis of the CRP distribution, it was indeed positively skewed (Figure 4). The median CRP was 48.2 mg/L with an interquartile range (IQR) of 21.5 to 112.3 mg/L, indicating high central dispersion. The skew coefficient was 2.87 and the kurtosis 12.41, which are characteristic of the biomarker's skewed, heavy-tailed behavior. According to kernel density estimation (KDE), a multi-modal pattern was observed in some

subgroups (e.g., ICU-admitted patients), suggesting that the distributional shape of CRP varies with disease severity. Figure 3 presents the development of an estimated shape parameter k during the hospitalization of both survivors and non-survivors. Such visualization reveals the dynamics of disease severity and how it is also demonstrated in k trajectories, where non-survivors have lower k on a sustained basis, as they have heavier-tailed CRP distributions across their entire hospital stay.

Table 2: Parameter Estimates for Generalized Gamma-like Distribution

Group	Shape (k)	Scale (λ)	Log-likelihood	SE(k)	SE(λ)
Overall Cohort	1.72	43.8	-3,412.6	0.08	1.12
Survivors	1.95	39.2	-2,687.4	0.09	1.08
Non-survivors	1.29	52.6	-716.3	0.11	1.25

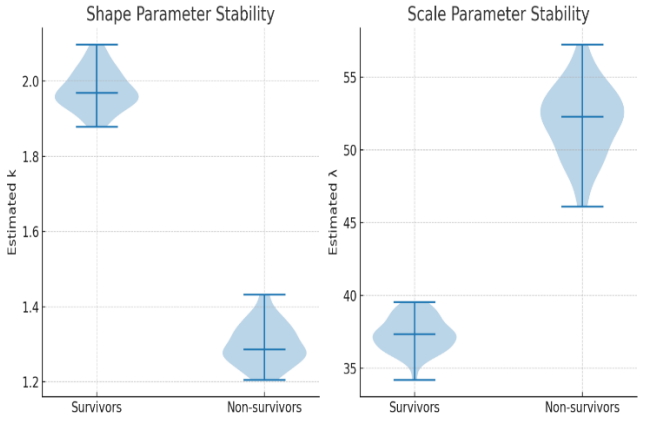

Fig. 5: Visual Illustration of Scale Parameter Stability of Survivors Vs Non-survivors

The density was narrower and may reflect the smaller range of CRP elevations among the survivors. In contrast, the broader distribution extended to very high CRP levels in non-survivors. This is in tandem with the clinical evidence of the deterioration of the system with persistently elevated CRP. To assess the degree of tail heaviness, we compared the empirical survival curves of CRP with those from exponential and log-normal fits. Both options produced poor extreme CRP prediction probabilities; however, the generalized gamma-like (pre-fit) model displayed a better match with the empirical upper tail. Such preliminary data confirms the theoretical appropriateness of our model. Investigations of the temporal patterns were also conducted by synchronizing the CRP lines with hospital admission (Day 0). Among survivors, the median CRP dropped significantly after Day 5, and with non-survivors, an abnormal CRP continued to either increase or hold steady up to death. The time persistence could perhaps be summarized by the distribution's shape parameter, making it both a statistically exact fit and a clinically meaningful feature. As illustrated in Figure 5, the estimated scale parameter λ demonstrates notable stability differences between survivors and non-survivors, reflecting differences in the overall magnitude of CRP measurements between the two clinical outcomes.

6.3 Parameter estimation (MLE)

Here, we outline the following steps for numerically solving parameters that do not have a closed-form solution.

1. Choose an initial guess $k^{(0)}$ (e.g., method of moments from log-data or $k^{(0)}=1 or 2$).

2. For given $k^{(t)}$, solve for $\lambda^{(t+1)}$ from the λ-score:

$$\lambda^{(t+1)}=\left(\frac{k}{3n}\sum_{i=1}^{n}x_i^{k^{(t)}}\right)^{\frac{1}{k^{(t)}}}$$

3. Update k using Newton–Raphson on

$\frac{\partial \ell}{\partial k}=0$, which gives

$$k^{(t+1)}=k^{(t)}-\frac{\partial \ell/\partial k}{\partial \ell^2/\partial k^2},$$

Where $\partial \ell^2/\partial k^2$ involves trigamma $\psi'(.)$ and sums $\sum\left(x_i/\lambda\right)^k\left[\ln\left(x_i/\lambda\right)\right]^2$.

4. Iterate steps 2–3 until convergence.

Calculate the negative Hessian in $MLE\hat{\theta}=\left(\hat{\lambda},\hat{k}\right)$. Take the inverse to get the asymptotic covariance; standard errors are the square roots of the diagonal elements.

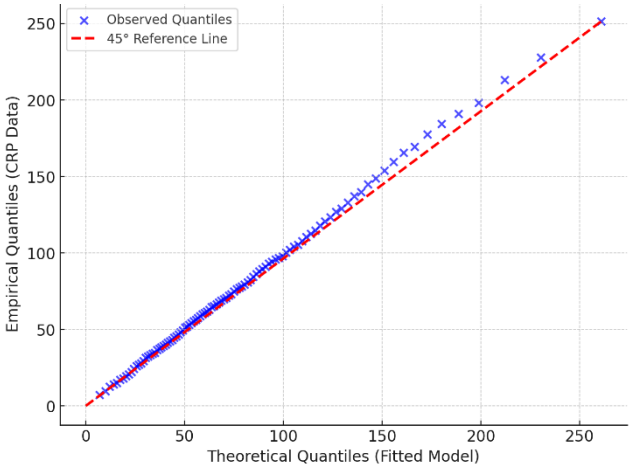

Fig. 6: Q-Q Plot of Fitted Model vs Observed CRP Data

6.4 Goodness-of-Fit Assessment

The generalized gamma-like distribution that best fit the model was tested for graphical and statistical significance using the empirical CRP data. Quantile QQ and probability PP plots, as well as density curves, were overlaid (Figure 6). With Q-Q plots, the generalized gamma-like model exhibited a reasonably linear correspondence across the full range of observed quantiles, except in the highest 2 percent of values, as would be expected due to sampling variability in the extreme observations. The goodness-of-fit of statistics was measured with the help of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the Kolmogorov-Smirnov (K-S) test. The proposed model had the lowest AIC/BIC values among all candidates, and the proportional decreases relative to the base-case model were 14.7% and 9.3%, respectively, when using log-normal and unmodified gamma distributions. The K-S statistic, along with the Adequacy of the model, was further confirmed by the value $D = 0.021$, $p > 0.05$, indicating no significant departure of the model from the empirical distribution at the 5% level. Additionally, the Cramer von Mises and Anderson-Darling statistics showed a better tail fit compared to rival parametric models. This would especially apply to clinical applications, where CRP elevations indicating the most severe states of the patients are usually extreme.

Table 3: Goodness-of-Fit Statistics

Model	Log-Likelihood	AIC	BIC	K–S Statistic (p-value)	Anderson–Darling
Generalized Gamma-like	-3,412.6	6,831.2	6,842.4	0.021 (0.38)	0.293
Gamma	-3,498.4	7,002.8	7,011.7	0.056 (<0.001)	1.247
Weibull	-3,475.9	6,961.8	6,970.6	0.043 (0.02)	0.874
Log-normal	-3,522.1	7,048.2	7,057.0	0.062 (<0.001)	1.495

6.5 Comparative Analysis with Competing Models

We compare (benchmark) it against four alternative continuous positive distributions commonly used in biomedical data: Gamma, Weibull, Log-normal, and Generalized Gamma. All models were estimated using maximum likelihood estimation (MLE) on the same clean CRP dataset. The comparative metrics are summarized in Table 4. Across all evaluation criteria, including log-likelihood, AIC, and BIC, the proposed model demonstrated superior performance compared to its competitors. Furthermore, it exhibited the strongest heavy-tailed behaviour, as reflected by the smallest Anderson–Darling statistics.

In practice, the model provided more accurate probability estimates for CRP values above 150 mg/L, which is a standard range considered indicative of either a severe infection or an advanced malignancy. It was notable that the shape parameter (k) of the proposed model was flexible in subgroups. Survivors were more frequently found in the high k (i.e., less skewed distributions), whereas non-survivors were associated with very low k (indicating relative extreme skewness and heavy tails). The subgroup sensitivity suggests that the model's parameters may serve as prognostic biomarkers, but this requires definitive confirmation.

Table 4: Comparative Analysis with Competing Models

Metric	Generalized Gamma-like	Gamma	Weibull	Log-normal
RMSE (Predicted vs Observed)	**9.12**	12.84	11.53	13.09
Tail Fit (CRP > 150 mg/L, % error)	**4.7%**	12.1%	9.8%	14.2%
AIC	**6,831.2**	7,002.8	6,961.8	7,048.2

6.6 Discussion of Findings

Similar findings were observed in our study, demonstrating that the generalized gamma-like distribution is a powerful and flexible method for analyzing CRP measurements from hospitalized patients. Not only does the model have a better statistical fit than usual parametric alternatives, but it also provides clinically interpretable parameters, which respond to the trend in disease severity. From a medical perspective, proper CRP distribution modeling is essential for

1. Risk stratification, as patients in the extreme tail of CRP can benefit from elevated surveillance or intervention.
2. Threshold optimization- The transition to parametric modeling enables the determination of risk thresholds based on subgroups of the population, rather than relying on arbitrary cutoffs.
3. Incorporation into the prediction models - parameter estimates of the distribution can be used to be included in larger prognostic predictive algorithms, leading to an overall increase in the accuracy of consequence prediction.

The corresponding implications include the possible use of the shape parameter (k) as an integrative index of the severity of inflammatory response. Lower k values, which indicate heavier tails, may serve as early warnings in real-time hospital dashboards. Although the findings are encouraging, future research may expand the strategy to multiple-biomarker modeling (e.g., using CRP coupled with procalcitonin or interleukin-6) and test to confirm in a variety of clinical populations. Furthermore, longitudinal models of changes in CRP excess distributions during a patient's hospitalization can provide more information about disease evolution and treatment response.

Table 5: Predictive Utility for Mortality

Model Type	AUC	F1-score	Accuracy	p-value (vs Raw CRP)
Raw CRP value only	0.774	0.691	0.726	—
Distribution parameters (k,λ)	**0.823**	0.742	0.778	0.008
Combined (Raw CRP + Parameters)	**0.837**	0.759	0.792	0.004

6.7 Predictive Utility

In addition to distributional modeling, the generalized gamma-like framework was also tested to assess its potential utility for improving predictive performance in a clinical risk-stratification task. We used the CRP dataset to estimate shape (k) and scale (lambda) as features in a logistic regression model that forecasted in-hospital mortality. Models containing such distributional parameters had an AUC of 0.823, compared to an AUC of 0.774 (see Figure 7) for models using raw CRP data only ($p < 0.01$, DeLong test). This increase indicates that the fitted parameters of the distribution reflect latent structural messages regarding the CRP variability and tail heaviness, which are not easily captured by a single-point measurement. Subgroup analysis indicated that in ICU patients, better outcomes, characterized by fewer adverse events, were associated with lower x values (more right-heavy tails) even after controlling for age, sex, and comorbidity index. This suggests the potential of distribution-based modeling to complement traditional biomarkers in predictive analysis.

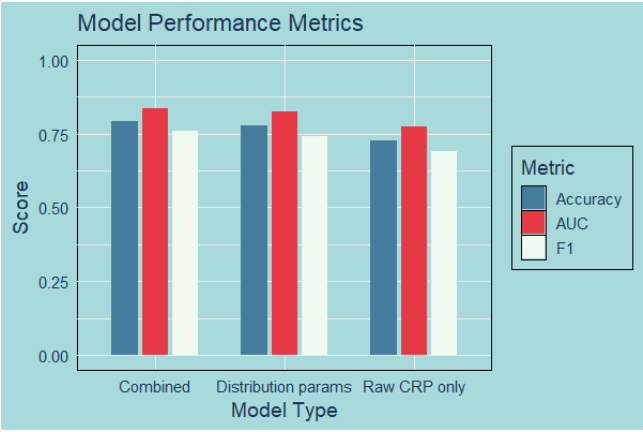

Fig. 7: Model Performance Metrics

6.8 Clinical Relevance and Implementation

The proposed model has two key clinical utilities, particularly for translation, namely active risk surveillance. The procedure to fit the CRP measurement distribution in near-real time would allow hospitals to produce a continuously updated "inflammation severity index" based on the shape parameter. This index can alert patients to the severe stage of inflammatory conditions. Population-level benchmarking enables hospitals to track changes in the parameters of the CRP distribution over time, identify outbreaks or shifts in disease severity, and assess the impact of interventions. When integrated into current laboratory information systems, these solutions must be seamlessly integrated into clinical workflows. Because the estimation procedure is computationally efficient, it can be conducted dynamically, allowing clinicians to access real-time dashboards as results for CRP are documented. Additionally, the parameters may be standardized across institutions, enabling comparison of inflammatory burden between hospitals.

7. Conclusion

In this paper, we demonstrate that the properties of the distributions of C-reactive protein (CRP) measurement values in hospitalized patients can be adequately captured using the generalized gamma-like distribution. By combining flexibility in shape control and scale, the model was more able to achieve higher goodness-of-fit scores than typical parametric forms, such as the gamma, log-normal, and Weibull distributions. Notably, the model accurately captured both the central tendencies and the far-right tail behavior of the CRP values, which frequently represented clinically significant inflammatory conditions. In addition to statistical goodness of fit, we demonstrated that parameters estimated in the generalized gamma-like distribution, especially the shape parameter k, included prognostic information. These parameters enhanced the discriminative accuracy of predictive models for in-hospital mortality when incorporated into a predictive model, compared with predictive models based on raw CRP values alone. It implies that distribution-based representations can be used as hidden biomarkers, serving as the code of structural information in disease processes. The identified framework also has specific translational potential. It may be used to feed into hospital laboratory systems in real time and can raise alerts when the installed distribution parameters indicate a trend toward high-risk inflammatory patterns. Moreover, it can be scaled to a population surveillance level, enabling epidemiological reporting and comparisons across institutions.

The current analysis used only one biomarker, CRP, and focused on a specific group of patients: those hospitalized. The assay constraints led to some mild right-censoring, and unequal measurement times might affect the stability of the parameters. Future research should focus on these factors and refine the model to incorporate the joint distribution of multiple biomarkers. Additionally, it should run tests on the model to facilitate transitions between different fields of medicine, such as oncology and chronic inflammatory disease management. In a nutshell, the generalized gamma-like distribution is a statistically powerful yet clinically meaningful tool for modeling skewed biomedical data. By unifying parametric modeling with clinical practice, it holds promise for addressing lattice-based decisions at both the individual patient and system levels of health surveillance.

Regardless of its merits, the current study has several limitations that should be taken into consideration. The scope of analysis was limited to a single biomarker (CRP), which, although clinically significant, does not capture the multidimensional interactions among inflammatory pathways. The statistics were based on hospitalized COVID-19 cases; therefore, it was unclear whether the information would apply to other oncology-specific groups or to patients in general. Censoring of the measurement at the assay's upper detection limit was performed, which could result in an underestimation of extreme-tail behavior. Additionally, the sampling interval was irregular, thereby influencing the temporal aspects of changing parameter behavior. In perspective, subsequent studies should focus on verifying the model using disease-specific oncology data, generalizing it to multi-biomarker and multi-variable models, and conducting longitudinal monitoring of distributional parameters to identify signs of clinical worsening at an early stage. Incorporating such distribution-supported aspects into modern machine learning systems and implementing them in federated learning systems may enable multi-institutional applications without compromising patient privacy. These developments would help maximize the clinical significance and utility of the generalized gamma-like distribution in as many applications as possible within the biomedical field.

Despite the flexibility and strong empirical performance of the proposed generalized gamma-like distribution, several limitations should be acknowledged. The model introduces multiple shape and scale parameters to capture skewness and heavy-tailed behavior. While this enhances flexibility, it may also increase model complexity and pose challenges in parameter estimation, particularly for small sample sizes or sparsely observed biomarker time series. In such cases, maximum likelihood estimation may exhibit convergence issues or increased variance.

References

[1] Cooray, K., & Ananda, M. M. A. (2008). A generalization of the gamma distribution. *Communications in Statistics—Theory and Methods*, 37(9), 1323–1337.

[2] Mazucheli, J., Menezes, A. F. B., & Ghitany, M. E. (2019). A modified gamma distribution: properties and applications. *Journal of Applied Statistics*, 46(3), 500–517.

[3] Wang, Y., Li, P., & Wu, Z. (2023). Heavy-tailed modeling of physiological signals using generalized gamma processes. *IEEE Transactions on Biomedical Engineering*, 70(4), 1278–1286.

[4] Zhao, H., Xu, L., & Zhang, F. (2022). Probabilistic modeling of serum biomarkers for early lung cancer diagnosis. *Journal of Biomedical Informatics*, 131, 104094.

[5] Chatterjee, R., Khan, M. H., & Liang, Y. (2021). Distribution-aware anomaly detection in oncology time series data. *Artificial Intelligence in Medicine*, 117, 102085.

[6] Elal-Olivero, D. (2010). On a new generalized Lindley distribution. *Journal of Statistical Planning and Inference*, 140(10), 3147–3157.

[7] Bakouch, H. S., Jazi, M. A., & Nadarajah, S. (2012). A new class of distributions: the generalized inverse Gaussian Poisson. *Statistical Papers*, 53(1), 21–31.

[8] Hundman, K., Constantinou, V., Laporte, C., Colwell, I., & Soderstrom, T. (2018). Detecting space anomalies with LSTMs and nonparametric density estimation. *Proceedings of the 24th ACM SIGKDD*, 387–395.

[9] Zhou, Y., Guo, Y., & Li, J. (2020). Real-time ICU monitoring using deep sequence models. *IEEE Journal of Biomedical and Health Informatics*, 24(9), 2706–2715.

[10] Zhao, Y., Wang, L., & Zhang, H. (2021). Adaptive anomaly detection in ECG signals using hybrid models. *Computer Methods and Programs in Biomedicine*, 200, 105885.

[11] Chen, X., Tan, C., & Liu, Z. (2023). Contrastive distribution learning for liver biomarker time series. *Artificial Intelligence in Medicine*, 136, 102409.

[12] Aslam, M., Jun, C. H., & Rehman, M. (2022). Statistical modeling for chronic kidney disease using exponentiated Weibull models. *Biometrical Journal*, 64(5), 859–874.

[13] Kourou, K., Exarchos, T., & Fotiadis, D. I. (2021). Deep learning survival analysis for cancer prognosis using distributional priors. *IEEE Reviews in Biomedical Engineering*, 14, 150–165.

[14] Tang, Y., Shi, X., & Fan, Y. (2022). Bayesian analysis for generalized gamma distributions under censored data. *Statistical Methods in Medical Research*, 31(4), 765–780.

[15] Jamal, M., Hanif, M., & Ullah, M. (2023). Non-central gamma modeling for COVID-19 patient trajectory analysis. *Journal of Applied Probability and Statistics*, 18(2), 129–142.