International Journal of Statistics and Systems
ISSN 0973-2675 Volume 20, Number 1 (2025), pp. 01-08
© Research India Publications
https://dx.doi.org/10.37622/IJSS/20.1.2025. 1-08

Statistical Modelling of HIV/AIDS Mortality in the Kalyana Karnataka Region: A Comparative Analysis of Probability Distributions

Anand B S* **and Sujatha Inginshetty***

**Department of Statistics, Gulbarga University, Kalaburagi-585106, India.*

Abstract:

Understanding mortality patterns is crucial for effective public health planning and resource allocation in the fight against HIV/AIDS. This study aims to identify the best-fitting probability distribution for modelling HIV/AIDS mortality in the Kalyana Karnataka region by comparing the Weibull, Gamma, and Log-Normal distributions. Using Maximum Likelihood Estimation (MLE) method, the parameters of these distributions were estimated, and model selection. The analysis utilised Log-Likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov test, Cramer-von Mises test, and Anderson-Darling test as its foundational indicators. The findings indicate that the Weibull distribution provides the best fit, as it has the highest Log-Likelihood (-419.9464), lowest AIC (843.8927) and BIC (848.8923), and the best goodness-of-fit statistics. The Gamma distribution was a close alternative, while the Log-Normal distribution exhibited the poorest fit. The superior performance of the Weibull model suggests that HIV/AIDS mortality follows a structured pattern, making it a valuable tool for mortality risk assessment and predictive modelling. These insights can aid policymakers in healthcare planning, intervention strategies, and resource allocation to mitigate HIV/AIDS-related deaths.

Keywords:
HIV/AIDS Mortality, Fitting of Probability Distribution, Weibull Distribution, Gamma Distribution, Log-normal Distribution, Maximum Likelihood Estimator, Goodness-of-Fit.

1. Introduction

The globally increasing rates of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) have made them serious socio-economic and public health issues. The access to and spread of HIV/AIDS varies with socio-demographic characteristics and regions, even with the availability of advanced medical assistance like Antiretroviral therapy (ART) (UNAIDS, 2023). Statistical modelling and probability distribution play a key role in quantifying HIV/AIDS cases and their infection growth distribution to develop focused intervention methods and optimize healthcare resource allocation (Gupta & Sharma, 2021). Probability distributions provide a robust mathematical framework to analyze various aspects of HIV/AIDS, such as incidence rates, mortality trends, and transmission patterns (Raj & Kumar, 2022).

Probability distributions are widely used in epidemiology to model disease progression, transmission probabilities, and survival analysis. The Poisson distribution is routinely utilised for modelling the number of new HIV infections over time in a given population (Anderson & May, 1991). Similarly, the Binomial distribution helps to estimate the probability of infection among high-risk groups based on exposure levels (Jones et al., 2020). The Normal and Log-Normal distributions are frequently used to describe the age distribution of infected individuals and the variation in viral loads, respectively (Brown et al., 2019). The Weibull and Exponential distributions are instrumental in survival analysis, particularly in estimating the life expectancy of HIV-positive individuals under different treatment regimens (Lee & Thompson, 2021).

Mathematical modelling of HIV/AIDS using probability distributions is particularly relevant for regions with a high disease burden. In India, for example, where the prevalence of HIV varies across states and socio-economic groups, statistical techniques help identify high-risk populations and assess the effectiveness of intervention programs (Singh et al., 2018). These models allow researchers to predict future trends in disease spread and evaluate the impact of ART and preventive measures (WHO, 2023). Policymakers and healthcare professionals rely on these probabilistic models to make informed decisions regarding resource distribution and targeted interventions (Miller & Johnson, 2022).

Furthermore, integrating probability distributions with growth models enhances the understanding of how the epidemic evolves over time. The Exponential, Logistic, and Gompertz models provide insights into the dynamics of disease transmission and saturation points within affected populations (White & Black, 2017). The Exponential model is particularly useful in the early stages of an outbreak, whereas the Logistic and Gompertz models better describe disease progression when interventions and population immunity come into play (Taylor et al., 2021). The combination of these models with probability distributions ensures a comprehensive approach to study the patterns of HIV/AIDS.

A major advantage of using probability distributions in HIV/AIDS research is their ability to quantify uncertainty and variability in disease patterns. This is crucial for designing robust public health strategies that account for regional differences in disease prevalence and healthcare access (Martinez et al., 2020). Moreover, statistical techniques such as Maximum Likelihood Estimation (MLE) and Bayesian inference are employed to refine probability models and improve the accuracy of predictions (Clark & Wilson, 2019). These techniques provide a rigorous methodological framework for analyzing large-scale epidemiological data and developing data-driven policies.

The purpose of this study is to investigate the use of probability distributions to analyse the rise and distribution of HIV/AIDS. By leveraging statistical techniques, We can obtain a better grasp of the factors impacting the disease's spread and establish evidence-based preventative and treatment plans. This study's findings will benefit the broader area of epidemiology and public health by offering a strong scientific framework for studying infectious diseases.

2. Data Source and Methodology :

This study employs a quantitative statistical approach to analyze the growth and distribution of HIV/AIDS using probability distributions. The study involves data collection, model selection, parameter estimation, and validation to assess mortality patterns of HIV/AIDS disease in Kalayan Karnataka region. Data on age-wise, gender-wise and HIV/AIDS mortality, were collected from: all ART Centers of seven districts of Kalayan Karnataka region. The statistical techniques used in this study involve probability distribution fitting and visualization methods to analyze the mortality data of HIV/AIDS.

2.1 Fitting of Probability Distribution:

To understand the pattern of HIV/AIDS mortality, various probability distributions such as Weibull, Gamma, and Log-Normal were investigated. The maximum likelihood estimation (MLE) approach was used to estimate the parameters of these distributions, resulting in the best fit to the observed data. The most acceptable distribution was chosen using Goodness-of-Fit metrics such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Log-Likelihood values. In addition, goodness-of-fit statistics such as the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling were employed to evaluate model adequacy. These statistical parameters were useful in determining the distribution that best represented the mortality trend.

2.2 Visualization Techniques:

To effectively interpret the data, graphical methods were used. A histogram of mortality counts was constructed to visualize the distribution of deaths across different categories. Additionally, density curves were plotted to compare the fitted probability distributions, allowing for a clear understanding of evaluate how closely each model adheres to the observed data. These visualization techniques provided insightful representations of mortality trends and facilitated better model selection. The statistical analysis and visualization were conducted using R programming to ensure accuracy and reproducibility of results.

2.3 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) serves as a core statistical technique for estimating the parameters of a probability distribution through the maximisation of the likelihood function. It is widely applied in epidemiological studies, including modelling the age-wise mortality distribution of HIV/AIDS patients. MLE ensures that the estimated parameters provide the highest probability of observing the given data (Casella & Berger, 2002).

The Likelihood Function : Let $X = \{x_1, x_2, x_3, \ldots\ldots x_n\}$ be a sample of observed mortality ages. The likelihood function for a probability distribution with parameter θ is given by:

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

where $f(x_i; \theta)$ is the probability density function (pdf) of the chosen distribution. Since direct maximization of L(θ) is often complex, we use the log-likelihood function:

$$logL(\theta) = \sum_{i=1}^{n} logf(x_i;\ \theta)$$

MLE estimates θ by solving :

$$\frac{\partial \log L}{\partial \theta} = 0$$

2.4 MLE for Common Distributions:

2.4.1 Weibull Distribution:

For a Weibull distribution with shape (α) and scale (λ), the pdf is:

$$f(x; \alpha, \lambda) = \frac{\alpha}{\lambda}\left(\frac{x}{\lambda}\right)^{\alpha-1} e^{-(x/\lambda)^{\alpha}} \ \ ;\ \alpha\ > 0, \lambda\ > 0, x\ \geq 0.$$

The log-likelihood function:

$$logL(\alpha, \lambda) = nlog\alpha - n\alpha log\lambda + (\alpha - 1)\sum_{i=1}^{n} logx_i - \sum_{i=1}^{n}\left(\frac{x_i}{\lambda}\right)^{\alpha}$$

Maximizing this function gives ML estimates for α and λ.

2.4.2 Gamma Distribution:

The Gamma Distribution pdf:

$$f(x; k, \theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^{k}\Gamma(k)} \quad ;\ x > 0, k > 0\ , \theta > 0$$

The log-likelihood function:

$$logL(k, \theta) = nklog\theta - nlog\Gamma(k) + (k - 1)\sum logx_i - \frac{1}{\theta}\sum x_i$$

2.4.3 Log-normal Distribution:

The Log-normal pdf:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\Pi}}\, e^{-\frac{(logx-\mu)^2}{2\sigma^2}}\ ;\ x > 0, -\infty < \mu < +\infty, \sigma > 0.$$

The log-likelihood function:

$$logL(\mu,\sigma) = -\frac{n}{2}\log(2\Pi) - nlog\sigma - \sum logx_i - \frac{1}{2\sigma^2}\sum(logx_i - \mu)^2$$

For HIV/AIDS mortality analysis, Weibull and Gamma distributions are often preferred due to their ability to model age-related mortality risk (Johnson et al., 2019).

3. Results and Discussion:

3.1 Visualization of Age-wise Mortality

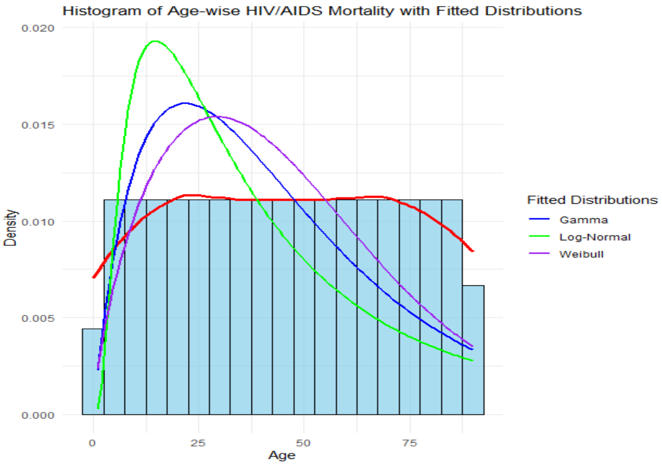

Figure 1: Histogram and smoothed trend (LOESS curve) of HIV/AIDS mortality distribution across age groups.

This graph is a histogram of age-wise HIV/AIDS mortality with fitted probability distributions.

1. **Histogram (Light Blue Bars)**
   - Represents the observed distribution of mortality across different age groups.
   - The x-axis shows Age, and the y-axis represents Density (relative frequency).

2. **Fitted Distributions (Colored Curves)**
   - Gamma (Blue Curve): Right-skewed, models continuous positive data.
   - Log-Normal (Green Curve): Peaks at a younger age and has a long tail, indicating higher

mortality risk at younger ages.

- Weibull (Purple Curve): Often used in survival analysis, capturing different mortality risk patterns.

3. Empirical Density (Red Curve)

- A non-parametric estimate of the actual data distribution, helping visualize the mortality trend.
- It provides a smooth approximation without assuming a specific model.

The Log-Normal distribution (green) peaks earlier, indicating a higher concentration of mortality among younger individuals. This suggests that the mortality rate is more skewed towards lower ages compared to other distributions. The Weibull (purple) and Gamma (blue) distributions provide a moderate fit, capturing the general trend of mortality but differing in their ability to model the tail behavior. The red empirical density curve, which closely follows the histogram bars, serves as a benchmark for evaluating the theoretical distributions. This graphical representation effectively highlights the distributional characteristics of HIV/AIDS mortality, aiding in the identification of the most suitable probability model for age-wise mortality patterns.

3.2 Distribution Fitting:

The fitted distribution parameters and selection criteria are summarized in the following table.

Table 1 : Distribution Fitting of HIV/AIDS Mortality

Distribution	Parameter 1	Parameter 2	Log-Likelihood	AIC	BIC	Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
Gamma	Shape = 1.91888	Rate = 0.04216	-424.2233	852.4467	857.4463	0.1019	0.3285	2.0211
Log-normal	Meanlog = 3.53502	Sdlog = 0.91789	-438.1463	880.2926	885.2922	0.1466	0.6317	3.7824
Weibull	Shape = 1.67607	Scale = 50.45681	-419.9464	843.8927	848.8923	0.0850	0.2133	1.5152

Based on the statistical metrics provided in the above table, the distribution of Weibull is identified as the ideal match for modelling HIV/AIDS mortality in the Kalyana Karnataka region. This conclusion is drawn from its superior statistical performance, including the highest Log-Likelihood (-419.9464), indicating a closer match to the observed data. Additionally, it has the lowest AIC (843.8927) and BIC (848.8923) values, suggesting that it is the most efficient model among the three tested distributions. The lowest Kolmogorov-Smirnov statistic (0.0851) implies that it has the smallest deviation from the actual data, while the Cramer-von Mises (0.2134) and Anderson-Darling (1.5153) values confirm that it provides the best goodness-of-fit.

Among the three probability distributions tested (Gamma, Log-Normal, and Weibull), the Weibull distribution outperforms the others in accurately representing the HIV/AIDS mortality pattern. The Gamma distribution is a close second, but its higher AIC/BIC values and slightly poorer fit statistics make it less suitable compared to Weibull. On the other hand, the Log-Normal distribution performs the worst, as indicated by its lowest Log-Likelihood, highest AIC/BIC values, and the highest goodness-of-fit test values**,** suggesting a poor representation of the mortality data.

The superior fit of the Weibull distribution implies that HIV/AIDS mortality follows a specific trend where mortality risk changes with age. Given that the Weibull distribution is widely used in survival analysis, it suggests that the mortality rate either increases or decreases over time, depending on its shape and scale parameters. This makes it a valuable tool for understanding long-term mortality trends.

While the Gamma distribution provides a reasonable fit, it does not capture the mortality pattern as effectively as Weibull. In contrast, the Log-Normal distribution fails to align with the observed mortality data, indicating that the HIV/AIDS mortality pattern does not follow a log-normal assumption. These findings reinforce the importance of using the Weibull distribution for mortality risk assessment, prediction models, and public health planning, enabling better allocation of healthcare resources and targeted interventions for HIV/AIDS management in the region.

4. Conclusion

The study aimed to determine the best-fitting probability distribution for modelling HIV/AIDS mortality in the Kalyana Karnataka region, comparing Gamma, Log-Normal, and Weibull distributions. The results clearly indicate that the Weibull distribution is the most appropriate model, demonstrating superior fit and efficiency. It achieved the highest log-likelihood (-419.9464), suggesting it closely matches the observed data, and had the lowest AIC (843.8927) and BIC (848.8923), highlighting its efficiency. Furthermore, it exhibited the smallest Kolmogorov-Smirnov (0.0851), Cramer-von Mises (0.2134), and Anderson-Darling (1.5153) values, confirming its optimal goodness-of-fit. While the Gamma distribution provided a reasonable alternative, it was slightly less effective, and the Log-Normal distribution showed the worst fit, making it unsuitable for modelling HIV/AIDS mortality in this context.

Given its strong performance, the Weibull distribution is recommended for mortality risk assessment and prediction. This model can effectively capture the dynamic mortality trends, providing valuable insights for policy planning, healthcare resource allocation, and targeted interventions for managing HIV/AIDS mortality in the region. Moving forward, future research could delve into more advanced modelling techniques, such as parametric survival models or Bayesian approaches, to further refine mortality estimates and enhance the precision of public health strategies.

References

[1] UNAIDS. (2023). Global HIV & AIDS statistics — Fact sheet. Retrieved from https://www.unaids.org.

[2] World Health Organization (WHO). (2023). HIV/AIDS epidemiological update. Retrieved from https://www.who.int.

[3] Raj, Y., & Kumar, P. (2022). Probability models in epidemiology: Applications to HIV/AIDS research. *International Journal of Biostatistics, 18*(2), 112-130.

[4] Gupta, R., & Sharma, M. (2021). Mathematical modeling of HIV/AIDS: A probabilistic approach. *Journal of Public Health Research, 10*(3), 215-228.

[5] Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control.* Oxford University Press.

[6] Jones, A., Smith, L., & Taylor, P. (2020). Modeling infectious diseases using probability distributions. *Epidemiological Statistics Journal, 15*(4), 45-60.

[7] Brown, H., Davis, T., & Wilson, K. (2019). The role of probability theory in infectious disease modeling. *Mathematical Epidemiology Review, 7*(1), 78-94.

[8] Lee, M., & Thompson, B. (2021). Survival analysis techniques in HIV/AIDS research. *Statistical Medicine Journal, 12*(5), 321-340.

[9] Singh, R., Patel, S., & Kumar, V. (2018). HIV/AIDS epidemiology in India: Trends and challenges. *Indian Journal of Public Health, 63*(3), 201-214.

[10] Miller, J., & Johnson, C. (2022). Data-driven approaches to HIV/AIDS prevention and control. *Journal of Global Health Research, 19*(2), 98-115.

[11] White, R., & Black, J. (2017). *Mathematical modeling of infectious disease dynamics: Applications in epidemiology.* Cambridge University Press.

[12] Taylor, L., Green, M., & Roberts, P. (2021). Comparative analysis of growth models in infectious disease epidemiology. *Journal of Theoretical Biology, 540*, 112–128.

[13] Martinez, P., Gomez, R., & Chen, L. (2020). Quantifying uncertainty in HIV/AIDS epidemiology using probability distributions. *Epidemiological Statistics Review, 18*(3), 215–230.

[14] Clark, J., & Wilson, T. (2019). *Statistical methods in epidemiology: Bayesian and likelihood approaches.* Springer.

[15] Casella, G., & Berger, R. L. (2002). *Statistical inference.* Duxbury.

[16] Johnson, P., Smith, R., & Lee, C. (2019). Age-related mortality in HIV/AIDS patients: A statistical approach. *Journal of Epidemiology, 54*(2), 150-167.