A Comparative Study of Slope and Informative Right Censoring Parameter Estimation under Different Distribution Functions

R. Devi[1], N. Viswanathan[2]

[1]Department of Statistics, Presidency College, Chennai, India.

[2]Associate Professor and Head (Retd), Department of Statistics, Presidency College, Chennai, India.

DOI: https://doi.org/10.51244/IJRSI.2026.1311000077

Received: 07 November 2026; Accepted: 12 November 2026; Published: 24 November 2026

ABSTRACT

Longitudinal studies aim to model the evolution of a response variable over time. But the analysis becomes more complex when participants' dropout is related to the outcome of interest, resulting in informative right censoring. In this study, the number of visits by each subject is modeled as a truncated Geometric distribution which depends on the individual's rate of change expressed through four different cumulative distribution functions, namely, Standard Normal, Standard Cauchy, Standard minimum Gumbel, and Logistic distributions. A simulation study is conducted to estimate the slope parameters at the population and individual levels, and the censoring parameters using four different estimators, namely, Unweighted, Weighted, Naive and adjusted informative right censoring under four different cumulative distribution functions. The performance of these estimators is compared based on bias, mean squared error, and goodness-of-fit criteria. The best estimator suggested by the results of simulation is used to get the slope estimates of an observed trial dataset taken from 'JM' package in R software and this dataset consists of Primary Biliary Cirrhosis patients' information with their Serum Bilirubin measure as the study variable. The statistical significance of the censoring parameter was examined using the likelihood ratio test, whereas the goodness-of-fit test is used to assess the validity of truncated geometric distribution. The simulation results provided best estimators for individual and population slopes under different cumulative distribution functions with varying values of informative right censoring parameter. In addition, the study provides estimate of informative right censoring parameter and this led to identify the best distribution for the length of follow-up and the slope and right censoring parameters for the PBC observed trial data. The combined results of simulation and that of the observed trial data leads to the informative right censoring estimate as 1.3804 and the population slope estimate as 0.1737, based on IRC2 estimator.

Keywords: Longitudinal data, Slope estimation, Informative right censoring, Truncated Geometric distribution, Mean Squared Error.

INTRODUCTION

In Longitudinal studies, one of the main focuses is to measure the change in the study variable that has been measured considerably over a long period. This measure of rate of change is extensively used in fields such as clinical research, genetics, psychology, reliability engineering, and other related health sciences. These studies allow researchers and practitioners to examine the nature of changes in outcomes, understand the progression of diseases, and evaluate the long and short-term effects of treatments or exposures on individuals. Such longitudinal studies face the problem of losing subjects in the mid-way of the study due to migration, missed visits, death and other factors. Such incomplete data are quite common in biomedical and other types of research and are often referred to as censoring.

Different statistical approaches are used to handle censoring under the assumption that it is independent of the study variable. This assumption states that, a subject censored time, does not provide any information about the individuals' survival time. In some realistic scenarios, where the mechanisms that constitutes the censoring is non-ignorable, that is, the censoring time is dependent on the study variable and this scenario is known as

informative censoring. Specifically, if subjects are lost to follow-up due to reasons related to the study variable, it is known as Informative Right Censoring (IRC). Wu and Bailey (1989) used the term IRC in the context of slope-estimation to refer the situation where the censoring probability of each individual relates to his or her underlying variable's rate of change.

Mori et al. (1994) compared four estimators under a truncated geometric distribution for number of visits, using the cumulative distribution function (CDF) of Logistic distribution to model the informative right censoring. Motivated by the work of Mori et al. (1994), the present study employs the CDFs of the Standard Normal, Standard Cauchy, and Standard Minimum Gumbel distributions. The four estimators of Mori included Unweighted, Weighted, Informative right censoring based on Naive approach (IRC1) and Informative right censoring based on adjusted method (IRC2). These estimators have been used to estimate the slopes of individuals and as well that of the entire population. IRC1 and IRC2 estimators are based on the random slope model, wherein the number of visits of each subject is treated as a random variable and is modeled as a truncated Geometric distribution with probability of success p_i of the ith subject being defined through the CDF. The performance of these estimators is compared using bias, Mean Squared Error (MSE) of individual and population slope estimators, censoring parameter estimator and goodness-of-fit criteria for the fit of truncated Geometric CDF to the number of visits by the individual.

An observed trial dataset consisting of 38 subjects from 'pbc2' dataset in the 'JM' package (Rizopoulos, 2010) of R software is considered for the analysis, with Serum Bilirubin measure as the main study variable. This data come from a Mayo Clinic trial in Primary Biliary Cirrhosis (PBC) of the liver conducted between 1974 and 1984. PBC is a progressive disease thought to be of an autoimmune origin; the subsequent inflammatory process eventually leads to cirrhosis and destruction of the liver's bile ducts and death of the patient. A description of the study along with a listing of the data set can be found in Fleming and Harrington (1991). A higher dropout rate was observed among patients with elevated serum bilirubin levels, indicating more severe liver dysfunction and a greater likelihood of death. This suggests that the missingness mechanism is informative and is associated with the study variable, serum bilirubin.

Based on the initial estimation of parameters from the observed trial dataset, simulations have been carried out in the vicinity of the right-censoring parameter of observed trial data and comparisons have been carried out to identify the best estimator of slopes and censoring parameters. This identified estimator is then applied to the observed trial data to get the improved estimators of right-censoring and slope parameters. The Likelihood Ratio Test (LRT) is used for testing the informative right censoring parameter.

REVIEW OF LITERATURE

The problem of estimation of slope of study variable in longitudinal data poses a challenging problem, particularly when observations are incomplete due to censoring. The presence of right censoring, especially when it is informative, leads to significant challenges in statistical inference, as it introduces dependence between the follow-up period and the underlying value of study variable. Traditional methods that assume non-informative censoring often leads to biased and inefficient estimates when this assumption is violated.

Mori et al. (1992) proposed a parametric empirical Bayes approach, which provides a flexible framework for incorporating both individual and group level information in the estimation process, when informative right censoring is present. A significant strength of this approach lies in the robustness compared to methods built on restrictive parametric assumptions about the censoring distribution. Moreover, in the special case where the censoring is non-informative, these estimators reduce to the classical Bayes estimators derived by Fearn (1975), thereby providing a useful theoretical connection between the two frameworks. The applicability of these methods has been demonstrated through real-world datasets, such as renal transplant studies, and their performance has been further validated through simulation studies.

As an extension of previous studies of Wu and Carroll (1988), Wu and Bailey (1989) and Mori et al. (1992), a new method was proposed by Mori et al. (1994) for slope estimation in the presence of 'Informative Right Censoring', which involves a random slope model where the number of observations is assumed to follow Geometric distribution with its mean dependent on the individual subject's rate of change. A simulation study

was carried out by the authors to evaluate the performance of the estimators such as IRC adjusted, IRC Naive, Weighted and ordinary least squares for both the individual and as well that of the population slopes. To evaluate the performance of these estimators, the authors used bias and MSE metrics as evaluation criteria and concluded that their proposed methods IRC1 and IRC2 are useful for follow-up studies in which quick intervention is needed for high-risk patients, particularly those in the early stages of bone marrow and organ transplantation.

A new likelihood-based approach had been used by Jaffa et al. (2011) in estimating the slope for informative right censored longitudinal data, wherein the likelihood function is integrated over the random effects to obtain a marginal likelihood estimation. The performance of the model was assessed using the Geometric and Poisson distributions to model the number of observations for every individual subject. Maximum likelihood estimates for the population slope were acquired by direct maximization of the marginal likelihood function and empirical Bayes estimates for the individual slopes were generated using Gaussian quadrature. The methodology was illustrated using simulated data and renal transplant cohort real data.

The slope estimation in the presence of informative right censoring using Exponential censoring structure has been proposed by Viswanathan and Ravanan (2014). Simulation studies were conducted by varying the Exponential rate parameter to generate different dropout patterns and evaluate the four estimators namely Empirical Bayes, Linear Minimum Variance Unbiased (LMVUB), Weighted and Unweighted estimators. This study compared both individual and population slopes. The authors concluded that the LMVUB estimator performs better in estimating both individual and population slopes.

The present study considers estimators based on Mori et al. (1994) with Logistic distribution replaced by three more distributions, namely, Standard Normal, Standard Cauchy, and Standard minimum Gumbel distribution for modeling the truncated Geometric distribution for the ‘number of visits’ by each individual subject. These additional distributions considered are useful in identifying a better model fit for the ‘number of visits’ by the subject close enough to the assumed truncated Geometric distribution for the PBC observed trial data used in this study. Further, the present study includes LRT for testing the censoring parameter in observed trial data.

METHODOLOGY

The analysis of repeated measurements in longitudinal study provides an important framework for understanding individual variations and population-level trends in biomedical research. Of particular interest is the estimation of the rate of change in the study variable, which helps to characterize the progression of a disease or the effect of a treatment over time. Such estimation becomes challenging when the number of observations available for each individual is influenced by the study variable. In this study, a likelihood-based framework is adopted and slope estimation is carried out using four different estimators, namely, Unweighted, Weighted, Informative right censoring based on Naive approach (IRC1) and Informative right censoring based on adjusted method (IRC2) for four different CDF’s, namely, Standard Normal, Standard Cauchy, Standard minimum Gumbel and Logistic distributions.

1. The Model

Consider a set of n independent subjects $(i = 1, 2, \dots, n)$ each measured at time points $t_{i1}, t_{i2}, \dots, t_{imi}$, where m_i is the final observed time count for i^{th} subject taking values $m_i = 2, 3, \dots, p'$, where p' is the pre-specified measurement time points termed as censoring times that need not be equally spaced and the corresponding study variable is represented by y_{ij}. In order to calculate the individual slopes, the study has considered the subjects with more than one observation. In this model, the following assumptions are made.

Model assumptions

1. $\beta_i \sim Normal(\beta, \sigma_\beta^2)$, where β_i is the ith subject slope.
2. $m_i|\beta_i \sim$ Truncated Geometric distribution with $p_i = F(\gamma_0 + \gamma_1\beta_i)$.
   where γ_0 and γ_1 are the censoring parameters, with $\gamma_1 \neq 0$ represents the informative right censoring.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

3. $b_i|m_i, \beta_i \sim Normal(\beta_i, \sigma_{b_i}^2)$, where $\sigma_{b_i}^2 = \frac{\sigma_\varepsilon^2}{\sum_{j=1}^{m_i}(t_{ij}-\bar{t}_{i.})^2}$ and $\bar{t}_{i.} = \frac{\sum_{j=1}^{m_i} t_{ij}}{m_i}$.

Probability model for m_i is

$$Pr(M = m_i) = \{1 - F(\gamma_0 + \gamma_1\beta_i)\}^{m_i-2}\{F(\gamma_0 + \gamma_1\beta_i)\}^{R_i}, \qquad m_i = 2, 3, \dots, p' \tag{1}$$

Here $F(\gamma_0 + \gamma_1\beta_i)$ is a function of individual slope β_i and it is calculated using the CDF of the following distributions.

For Standard Normal distribution,

$$F(\gamma_0 + \gamma_1\beta_i) = \int_{-\infty}^{\gamma_0+\gamma_1\beta_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \tag{2}$$

For Standard Cauchy distribution,

$$F(\gamma_0 + \gamma_1\beta_i) = \int_{-\infty}^{\gamma_0+\gamma_1\beta_i} \frac{1}{\pi(1+x^2)} dx \tag{3}$$

For Gumbel distribution,

$$F(\gamma_0 + \gamma_1\beta_i) = 1 - exp[-exp(\gamma_0 + \gamma_1\beta_i)] \tag{4}$$

For Logistic distribution,

$$F(\gamma_0 + \gamma_1\beta_i) = \frac{1}{1 + exp[-(\gamma_0 + \gamma_1\beta_i)]} \tag{5}$$

R_i in equation [1] is an indicator variable that takes 1 if there is death or withdrawal occurs in the study and 0 if the ith subject has a complete observation.

$$R_i = \begin{cases} 1, & m_i < p' \\ 0, & m_i = p' \end{cases} \tag{6}$$

Geometric distribution models the number of trials needed to achieve the first success in a series of independent Bernoulli trials. In this context, if withdrawal from the study is treated as a success, then the number of observations for each subject can be modeled as a Geometric random variable, provided that the probability of withdrawal remains constant over time. The Geometric distribution employed in this model is left-truncated since at least two observations are required to estimate a rate of change, and it is as well right-censored since the study may terminate before the withdrawal of every subject is observed.

The number of visits and their corresponding probabilities under the truncated Geometric distribution with informative right censoring based on a simulated data $(n = 38,\ t = 10,\ \beta = 0.239,\ \sigma_\beta^2 = 0.05,\ \sigma_\epsilon^2 = 2.283,\ \gamma_1 = 1.326)$ for four different CDFs are presented in Figure 1 to Figure 4. The pattern in the probabilities confirms the structure of truncated Geometric distribution.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

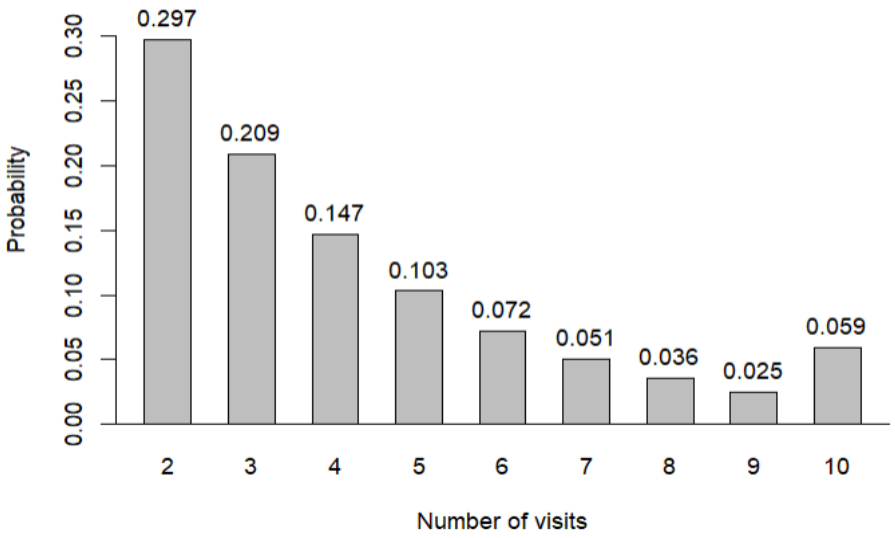

Figure 1. Truncated Geometric Distribution with Standard Normal CDF.

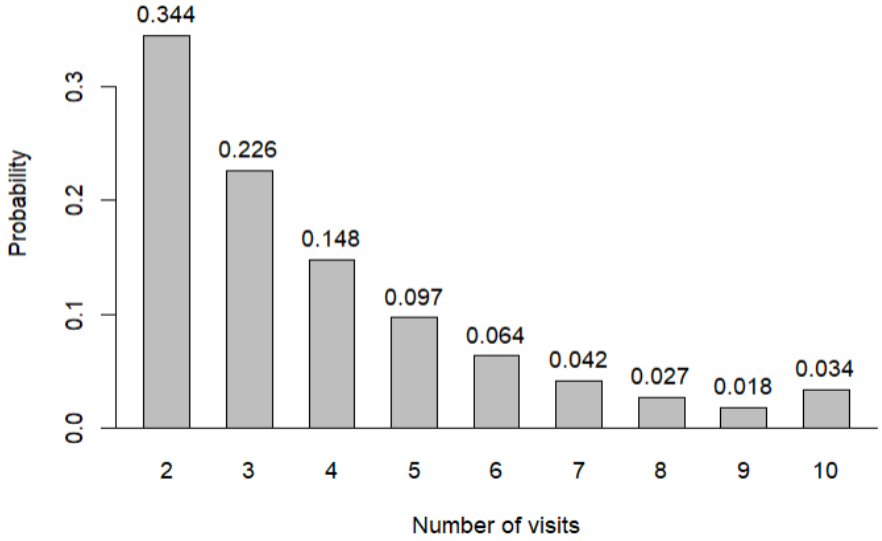

Figure 2. Truncated Geometric Distribution with Standard Cauchy CDF.

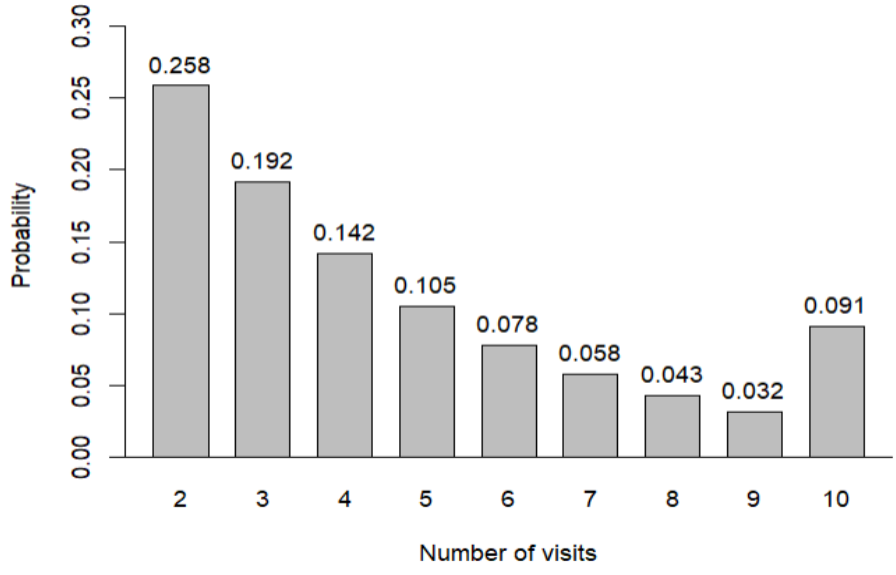

Figure 3. Truncated Geometric Distribution with Standard Minimum Gumbel CDF.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

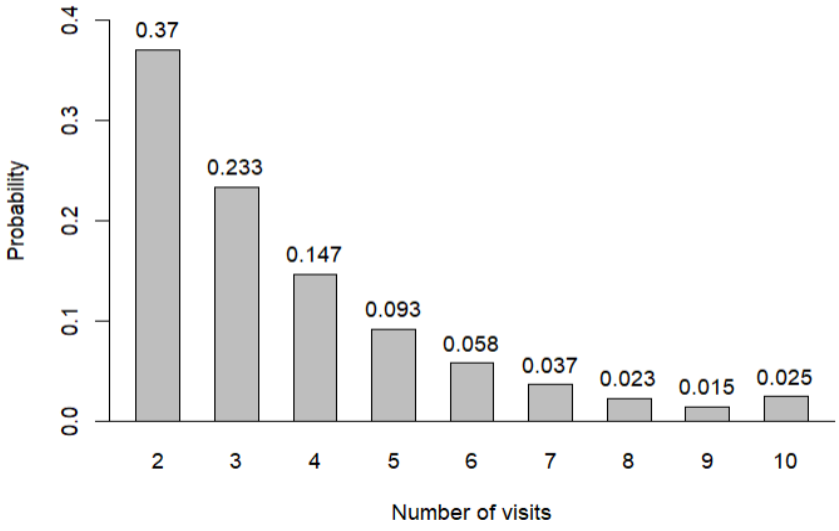

Figure 4. Truncated Geometric Distribution with Logistic CDF.

Model for study variable

The study variable y_{ij} is modeled as

$$y_{ij} = \alpha_i + \beta_i t_{ij} + \varepsilon_{ij} \qquad i = 1,2,\dots n; j = 1,2,..,m_i \tag{7}$$

Here for subject i, α_i is the intercept and β_i is the slope of the ith individual, t_{ij} are the pre-specified time points and the random error $\varepsilon_{ij} \sim \text{Normal}(0, \sigma_\varepsilon^2)$.

The ordinary least square estimators of the slope β_i and intercept α_i for i^{th} individual are

$$\hat{\beta}_i = \frac{\sum_{j=1}^{mi}(t_{ij} - \bar{t}_{i.})(y_{ij} - \bar{y}_{i.})}{\sum_{j=1}^{mi}(t_{ij} - \bar{t}_{i.})^2} \tag{8}$$

$$\hat{\alpha}_i = \bar{y}_{i.} - \hat{\beta}_i \bar{t}_{i.} \tag{9}$$

where $\bar{y}_{i.} = \frac{\sum_{j=1}^{m_i} y_{ij}}{m_i}$ are the means of the study variable of i^{th}individuals over their follow-up period.

2. Likelihood equation of the model

The likelihood involving the parameters $(\beta_i, m_i, b_i), i = 1,2,\dots,n$ is given by

$$L = \prod_{i=1}^{n} f(\beta_i, m_i, b_i) \tag{10}$$

$$= \prod_{i=1}^{n} f(\beta_i) f(m_i|\beta_i)\, f(b_i|m_i, \beta_i) \tag{11}$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma_\beta \sqrt{2\pi}} exp\left\{\frac{-1}{2}\left(\frac{\beta_i - \beta}{\sigma_\beta}\right)^2\right\} * \{1 - F(\gamma_0 + \gamma_1\beta_i)\}^{m_i - 2} * \{F(\gamma_0 + \gamma_1\beta_i)\}^{R_i} * \frac{1}{\sigma_{b_i}\sqrt{2\pi}} exp\left\{\frac{-1}{2}\left(\frac{b_i - \beta_i}{\sigma_{b_i}}\right)^2\right\} \tag{12}$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

The log-likelihood for estimating the censoring parameters γ_0 and γ_1, population slope β and the individual slope β_i is given by

$$log\, L(\gamma_0, \gamma_1, \beta, \beta_i) = -nlog(2\pi) - nlog(\sigma_\beta) - \sum_{i=1}^{n} \frac{(\beta_i - \beta)^2}{2\sigma_\beta^2} + \sum_{i=1}^{n} (m_i - 2) * log\big(1 - F(\gamma_0 + \gamma_1\beta_i)\big)$$

$$+ \sum_{i=1}^{n} R_i * log\big(F(\gamma_0 + \gamma_1\beta_i)\big) - \sum_{i=1}^{n} log(\sigma_{b_i}) - \sum_{i=1}^{n} \frac{(b_i - \beta_i)^2}{2\sigma_{b_i}^2} \quad [13]$$

$$= -nlog(2\pi) - nlog(\sigma_\beta) - \sum_{i=1}^{n} log(\sigma_{b_i}) - \sum_{i=1}^{n} \frac{(b_i - \beta)^2}{2(\sigma_{b_i}^2 + \sigma_\beta^2)}$$

$$- \sum_{i=1}^{n} \frac{\sigma_{b_i}^2 + \sigma_\beta^2}{2\sigma_{b_i}^2\sigma_\beta^2} \left(\beta_i - \frac{\sigma_{b_i}^2\beta + \sigma_\beta^2 b_i}{\sigma_{b_i}^2 + \sigma_\beta^2}\right)^2 + \sum_{i=1}^{n} (m_i - 2) * log\big(1 - F(\gamma_0 + \gamma_1\beta_i)\big)$$

$$+ \sum_{i=1}^{n} R_i * log\big(F(\gamma_0 + \gamma_1\beta_i)\big) \quad [14]$$

In the above likelihood, the consistent estimators of $\sigma_{b_i}^2$ and σ_β^2, as suggested by the third method of Henderson (Searle, 1971) were used by Mori et al. (1994). By maximizing equation [14], the estimators $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\beta}$ and $\hat{\beta}_i$ $(i = 1,2,3, \dots, n)$ are obtained, which satisfy

$$-\sum_{i=1}^{n} \lambda\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big)\big[m_i - E_i\big(M|\hat{\beta}_i\big)\big] = 0 \quad [15]$$

$$-\sum_{i=1}^{n} \lambda\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big) * \hat{\beta}_i * \big[m_i - E_i\big(M|\hat{\beta}_i\big)\big] = 0 \quad [16]$$

$$\hat{\beta}_{i,adj} = \frac{\sigma_{b_i}^2\beta + \sigma_\beta^2 b_i}{\sigma_{b_i}^2 + \sigma_\beta^2} - \left(\frac{\sigma_{b_i}^2\sigma_\beta^2}{\sigma_{b_i}^2 + \sigma_\beta^2}\right)\hat{\gamma}_1\lambda\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big)\big[m_i - E_i\big(M|\hat{\beta}_i\big)\big], \quad i = 1,2, \dots, n \quad [17]$$

where

$$E_i\big(M|\hat{\beta}_i\big) = R_i\frac{1 - F\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big)}{F\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big)} + 2, \lambda\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big) = \frac{f\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big)}{1 - F\big(\hat{\gamma}_0 + \hat{\gamma}_1\hat{\beta}_i\big)} \text{ and } f(z) = \frac{dF(z)}{dz}$$

$$\hat{\beta}_{adj} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_i \quad [18]$$

The consistent estimators of $\sigma_{b_i}^2$ and σ_β^2, provided by the third method of Henderson are given by

$$\hat{\sigma}_\beta^2 = \frac{\sum_{i=1}^{n}\big(b_i - \bar{b}_.\big)^2 \sum_{j=1}^{m_i}\big(t_{ij} - \bar{t}_{i.}\big)^2 - n\hat{\sigma}_\varepsilon^2}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\big(t_{ij} - \bar{t}_{i.}\big)^2} \quad [19]$$

$$\hat{\sigma}_{b_i}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{j=1}^{m_i}\big(t_{ij} - \bar{t}_{i.}\big)^2} \quad [20]$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

where

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(y_{ij} - \hat{\alpha}_i - b_i t_{ij}\right)^2}{\sum_{i=1}^{n} m_i - 2n} \quad [21]$$

with $\bar{b}_{.}$ representing the mean of individual $b_i^{'}s$.

3. Estimators

In this section, four different types of slope estimators under four different CDFs are defined and their efficiencies are compared through the criteria of bias and MSE.

3.1 Unweighted Estimator for Non-informative censoring

An Unweighted estimator is one in which all subjects contribute equally, without assigning different weights for each individual. The Unweighted estimator for population slope is given by

$$\hat{\beta}_{uwt} = \frac{1}{n}\sum_{i=1}^{n} b_i \quad [22]$$

Here, the ordinary least square estimator b_i is the Unweighted estimator for the individual subject's slope. The Unweighted estimators corresponding to CDFs of Standard Normal, Standard Cauchy, Standard minimum Gumbel, and Logistic distributions are denoted by UWN, UWC, UWG, and UWL, respectively.

3.2 Weighted Estimator for Non-informative censoring

Weighted estimator, instead of treating all the individuals equally, gives different weights to different subjects. It is obtained when the right-censoring process is not informative. In other words, the Weighted estimator for individual subject's slope is obtained, when $\gamma_1 = 0$ in equation [17] and is given by

$$\hat{\beta}_{i,wt} = \frac{\hat{\sigma}_{b_i}^2\hat{\beta} + \hat{\sigma}_{\beta}^2 b_i}{\hat{\sigma}_{b_i}^2 + \hat{\sigma}_{\beta}^2} = \left(\frac{w_{1i}}{w_{1i} + w_2}\right) b_i + \left(\frac{w_2}{w_{1i} + w_2}\right)\hat{\beta}_{wt} \quad [23]$$

where $w_{1i} = \frac{1}{\hat{\sigma}_{b_i}^2}$, $w_2 = \frac{1}{\hat{\sigma}_{\beta}^2}$ and $\hat{\beta}_{wt}$ is the Weighted estimator of the population slope given by

$$\hat{\beta}_{wt} = \left(\sum_{i=1}^{n}\frac{w_{1i}}{w_{1i} + w_2}\right)^{-1}\sum_{i=1}^{n}\left(\frac{w_{1i}}{w_{1i} + w_2}\right) b_i \quad [24]$$

The Weighted estimators corresponding to four different CDFs are denoted by WN, WC, WG, and WL, respectively.

3.3 Naive Estimator for Informative Right Censoring

Mori et al. (1994) proposed a method to find the population and individual subject's slope referred to as Naive method. This method reduces the computational difficulties and is described in the following steps.

Step1: Estimate $\hat{\gamma}_0$ and $\hat{\gamma}_1$ by maximizing the truncated Geometric component of the log-likelihood function given in equation [15] and equation [16]. Here $\hat{\beta}_i$ is replaced by the observed individual subject's slope b_i.

Step2: With the estimated values of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ from step1, $\hat{\beta}_{i,nav}$ is computed with

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

$$\hat{\beta}_{i,nav} = \frac{\hat{\sigma}_{b_i}^2 \hat{\beta} + \hat{\sigma}_{\beta}^2 b_i}{\hat{\sigma}_{b_i}^2 + \hat{\sigma}_{\beta}^2} - \left(\frac{\hat{\sigma}_{b_i}^2 \hat{\sigma}_{\beta}^2}{\hat{\sigma}_{b_i}^2 + \hat{\sigma}_{\beta}^2}\right) \hat{\gamma}_1 \lambda(\hat{\gamma}_0 + \hat{\gamma}_1 b_i)[m_i - E_i(M|b_i)] \quad [25]$$

where $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} b_i$.

The above IRC1 estimator is derived using the CDF of Logistic distribution and is denoted by IRC1L. Similarly, the estimators corresponding to the CDFs of Standard Normal, Standard Cauchy and Standard minimum Gumbel distributions are denoted by IRC1N, IRC1C, and IRC1G, respectively.

3.4 Adjusted Estimator for Informative Right Censoring

An adjusted method proposed by Stefanski and Carroll (1985) was adopted by Mori et al. (1994), to derive the estimators $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\beta}_{adj}$ and $\hat{\beta}_{i,adj}$ $(i = 1,2,3, \dots, n)$ by maximizing the log likelihood in equation [14]. The steps are given below.

Step1: From equation [15] and equation [16], find the estimators $\hat{\gamma}_0$ and $\hat{\gamma}_1$ by maximizing the truncated Geometric component of the log-likelihood function. Here, $\hat{\beta}_i$ is replaced by the weighted individual subject's slope $\hat{\beta}_{i,wt}$ given in equation [23].

Step2: With the calculated $\hat{\gamma}_0$ and $\hat{\gamma}_1$ from step1, compute $\hat{b}_{i,adj}$ given by

$$\hat{b}_{i,adj} = \hat{\beta}_{i,wt} - \left(\frac{\hat{\sigma}_{b_i}^2 \hat{\sigma}_{\beta}^2}{\hat{\sigma}_{b_i}^2 + \hat{\sigma}_{\beta}^2}\right) \hat{\gamma}_1 \lambda\left(\hat{\gamma}_0 + \hat{\gamma}_1 \hat{\beta}_{i,wt}\right)\left[m_i - E_i\left(M|\hat{\beta}_{i,wt}\right)\right] \quad [26]$$

Step3: Estimate $\hat{\gamma}_0$ and $\hat{\gamma}_1$ again by maximizing the truncated Geometric component of the log-likelihood function given in equation [15] and equation [16]. Now, replace $\hat{\beta}_i$ with the adjusted observed individual subject slope $\hat{b}_{i,adj}$ obtained from equation [26].

Step4: Finally, the individual subject's slope $\hat{\beta}_{i,adj}$ using informative right censoring adjusted method is obtained by substituting these new estimates $\hat{\gamma}_0$, $\hat{\gamma}_1$ and $\hat{b}_{i,adj}$ in equation [17],

where $\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \hat{b}_{i,adj}$.

Here, the IRC2 estimator is derived using four different CDFs and they are denoted by IRC2N, IRC2C, IRC2G, and IRC2L, respectively.

4. Comparison Metrics

The bias of population slope estimator is the difference between the estimated and the true value of the parameter.

$$Bias = \frac{\sum_{s=1}^{r}(\hat{\beta}_s - \beta)}{r} \quad [27]$$

where *r* is the total number of replications.

Mean Squared Error measures the average squared difference between the actual and estimated parameter values. MSE of population slope and that of individuals are respectively given by

$$MSE(a) = \frac{\sum_{s=1}^{r}(\hat{\beta}_s - \beta)^2}{r} \quad [28]$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

$$MSE(b) = \frac{\sum_{s=1}^{r}\sum_{i=1}^{n}(\hat{\beta}_{is} - \beta_{is})^2}{nr} \quad [29]$$

where n is the number of subjects in the data. Furthermore, MSE(a) measures the accuracy of the estimator of the population slope, whereas MSE(b) measures the accuracy of the estimators of the individual slopes.

A goodness-of-fit test was carried out to assess whether the distribution of the number of visits follows a truncated Geometric distribution. Here, the observed and expected frequencies are compared using Pearson's Chi-square statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \quad [30]$$

where O_i and E_i are the observed and expected frequencies, respectively.

APPLICATION TO OBSERVED TRIAL DATA

Primary Biliary Cirrhosis is a chronic autoimmune liver disease, in which the body's immune system mistakenly attacks and damages the small bile ducts within the liver. This gradually decreases the flow of bile from the liver to the small intestine. The buildup of bile damages and destroys the liver tissues. If a patient is left untreated, PBC can lead to scarring of tissues in the liver and liver failure. PBC is a relatively rare disease, affecting up to one in 3,000–4,000 people. It most often affects middle-aged and older women, with a sex ratio of at least 9:1 female to male (Poupon, 2010).

Serum bilirubin is one of the most critical prognostic indicators in PBC. It is a 'blood test' that measures the amount of bilirubin, a yellowish pigment produced during the normal breakdown of red blood cells circulating in the blood. Under normal conditions, bilirubin is transported to the liver, where it is processed and subsequently excreted through bile. An elevated serum bilirubin levels above 2.5 mg/dL may reflect impaired bile excretion and indicate the progression of PBC. Because of the high degree of skewness present in Serum Bilirubin measure, a log transformation of the value is used in the analysis (Therneau and Grambsch, 2000).

The pbc2 dataset used in this study is taken from 'JM' package in R language. It contains information on patients diagnosed with PBC, a rare autoimmune liver disease, based on data collected at the Mayo Clinic during the period between 1974 and 1984. The dataset comprises 312 patients with repeated follow-up visits, resulting in a total of 1,945 observations. Serum Bilirubin measure is considered as the study variable and in addition to it, survival information such as time-to-death or censoring and event indicators are also available in the dataset. Several covariates, including treatment group, age, sex, ascites, hepatomegaly, albumin, prothrombin, and status of the event are recorded for each patient. This study is restricted to a subset of the pbc2 dataset with 38 patients having atleast two visits and platelets value below 200. The parameter values chosen for the simulation were estimates obtained from this subset. In all, the study considers those with two to ten visits and more than 10 visits put together in a single category.

SIMULATION STUDY

A simulation study is conducted to investigate the performance of four different estimators, namely, Unweighted, Weighted, Naive, and adjusted informative right censoring with four different CDFs based on Standard Normal, Standard Cauchy, Standard minimum Gumbel, and Logistic distributions. The performance of these estimators is evaluated using bias and MSE criteria for the individual and population slopes, and for censoring parameters. Chi-square goodness-of-fit test is used to test the distribution of 'number of visits' by each individual. The required data are simulated using R program of version 4.5.3 (2026-03-11 ucrt) for varying parameters of informative right censoring and for selected parameters of population slope and variance parameters derived from the crude estimates of PBC data.

The abbreviations used to denote the Unweighted, Weighted, Naive Informative Right Censoring, and Adjusted Informative Right Censoring estimators under the four different CDFs are presented in Table 1.

Table 1. Notations used for the given estimators under four different CDFs.

CDF	Unweighted	Weighted	IRC1	IRC2
Standard Normal	UWN	WN	IRC1N	IRC2N
Standard Cauchy	UWC	WC	IRC1C	IRC2C
Standard Minimum Gumbel	UWG	WG	IRC1G	IRC2G
Logistic	UWL	WL	IRC1L	IRC2L

Ten thousand datasets were generated using simulation with each dataset containing 38 observations and having 2, 3, 4, 5, 6, 7, 8, 9 and 10 visits. The time points (t_{ij}) at which the ten measurements were made correspond to visits at 0 (initial visit), 6, 12, 24, 36, 48, 60, 72, 84, and 96 months. The parameters used in the simulation, namely, $\beta = 0.239, \gamma_0 = -1.1093, \gamma_1 = 1.326, \sigma_\beta^2 = 0.05$ and $\sigma_\varepsilon^2 = 2.283$ were selected using the crude estimates obtained from the pbc2 dataset. In addition, without loss of generality, a common intercept of $\alpha = 3.91$ was assumed, based on the mean of individual intercepts.

The study considers variations in the informative right censoring parameter with $\gamma_1 = 0$ indicating the non-informative right censoring and $\gamma_1 = 0.7, 1.326, 2.652, 3.978$ and 5.304 representing different scenarios under IRC. These values are taken around the crude estimates arrived from observed trial data and these values have some common regions across the four different CDFs and the two different IRC estimators. These parametric values indicate the degree of dependence between the censoring process and the rate of change of the study variable at varying levels.

RESULTS AND DISCUSSION

Simulation Results

The simulation results provide insights into the performance of individual and population slope estimators and censoring parameters under the four different CDFs, namely, Standard Normal, Standard Cauchy, Standard minimum Gumbel, and Logistic distributions. Furthermore, the performance of the estimators is compared in terms of bias for the population slope and MSE for both individual and that of population slope. These results are presented in Table 2.

CDF of Standard Normal distribution

When the Standard Normal CDF is used in the truncated Geometric distribution, the values from Table 2 indicate that the minimum bias is observed in unweighted estimators for almost all values of γ_1 except for 0.7 for which IRC2 estimator has minimum bias. When the individual slope estimators are concerned IRC1 performs better for low and moderate degree ($\gamma_1 = 0.7, 1.326,$ and 2.652) of informative right censoring in terms of MSE. IRC2 estimator performs better for high degree ($\gamma_1 = 3.978$ and 5.304) of informative right censoring. For the population slope, when informative right censoring appears low ($\gamma_1 = 0, 0.7$), weighted estimator provides minimum MSE. When a moderate ($\gamma_1 = 1.326$) and high degree ($\gamma_1 = 5.304$) of informative right censoring is present, IRC1 and IRC2 estimators perform better, respectively. In rest of the cases unweighted estimator performs better.

CDF of Standard Cauchy distribution

When the Standard Cauchy CDF is used in the truncated Geometric distribution, the values from Table 2 indicate that the minimum bias is observed in IRC1 estimator for moderate degree ($\gamma_1 = 1.326, 2.652,$ and 3.978)of informative right censoring and IRC2 for low degree ($\gamma_1 = 0, 0.7$) of informative right censoring. When the degree of informative right censoring is high ($\gamma_1 = 5.304$) unweighted estimator has minimum bias. When the individual slope estimators are concerned IRC2 performs better for high degree ($\gamma_1 = 5.304$) of informative right censoring, followed by IRC1 estimator for moderate informative right censoring ($\gamma_1 = 1.326, 2.652,$ and 3.978) and Weighted estimator for no ($\gamma_1 = 0$) or low ($\gamma_1 = 0.7$) informative right

censoring. When the population slope estimator is concerned, weighted estimators perform better for no ($\gamma_1 = 0$) or low degree ($\gamma_1 = 0.7$ and 1.326) of informative right censoring. For high degree ($\gamma_1 = 3.978$ and 5.304) of informative right censoring, unweighted slope estimator performs better, while IRC1 performs better for $\gamma_1 = 2.652$

CDF of Standard Minimum Gumbel distribution

When the Standard Minimum Gumbel CDF is used in the truncated Geometric distribution, the values from Table 2 indicate that the minimum bias is observed in unweighted estimators for moderate to high degree ($\gamma_1 = 1.326, 2.652, 3.978,$ and 5.304) of informative right censoring. When the degree of informative right censoring is low ($\gamma_1 = 0.7$) IRC2 estimator performs better and when no ($\gamma_1 = 0$) informative right censoring is present weighted estimator performs better. When the individual slope estimators are concerned IRC1 performs better for moderate degree ($\gamma_1 = 1.326$ and 2.652) of informative right censoring, followed by IRC2 estimator for high degree ($\gamma_1 = 3.978$ and 5.304) of informative right censoring and weighted estimator for no ($\gamma_1 = 0$) or low ($\gamma_1 = 0.7$) informative right censoring. When the population slope estimator is concerned weighted estimators perform better for no ($\gamma_1 = 0$) or low ($\gamma_1 = 0.7$) degree of informative right censoring. For moderate to high degree ($\gamma_1 = 2.652, 3.978,$ and 5.304) of informative right censoring unweighted slope estimator performs better, while IRC1 performs better for $\gamma_1 = 1.326$.

Table 2. Comparisons of slope estimators: Unweighted, Weighted and IRC estimators under Standard Normal, Standard Cauchy, Standard minimum Gumbel and Logistic distributions: Bias and Mean Squared Error[a]

γ_1	Metric	Standard Normal				Standard Cauchy				Standard Minimum Gumbel				Logistic			
		UWN	WN	IRC1N	IRC2N	UWC	WC	IRC1C	IRC2C	UWG	WG	IRC1G	IRC2G	UWL	WL	IRC1L	IRC2L
0	Beta Bias	-0.0004	-0.0005	-0.0004	-0.0004	0.0008	-0.0002	0.0005	0.0002	-0.0008	-0.0007	-0.0008	-0.0008	0.0003	0.0004	0.0004	0.0004
	MSE(a)*100	0.1880	0.1572	0.1646	0.1649	0.2208	0.1771	0.1953	0.1878	0.2436	0.1899	0.2125	0.2113	0.2344	0.1858	0.2025	0.2017
	MSE(b)	0.0219	0.0085	0.0088	0.0088	0.0361	0.0138	0.0143	0.0142	0.0436	0.0165	0.0170	0.0171	0.0388	0.0147	0.0151	0.0151
0.7	Beta Bias	-0.0007	-0.0114	-0.0021	0.0000	0.0015	-0.0072	0.0006	0.0001	-0.0009	-0.0116	-0.0020	-0.0008	-0.0002	-0.0084	-0.0014	-0.0002
	MSE(a)*100	0.2033	0.1767	0.1771	0.1783	0.2283	0.1853	0.2006	0.1981	0.2600	0.2170	0.2313	0.2337	0.2473	0.1999	0.2159	0.2180
	MSE(b)	0.0283	0.0110	0.0110	0.0110	0.0404	0.0154	0.0158	0.0157	0.0495	0.0189	0.0191	0.0193	0.0436	0.0166	0.0169	0.0170
1.326	Beta Bias	-0.0003	-0.0215	-0.0022	0.0029	0.0019	-0.0150	0.0000	0.0004	-0.0008	-0.0214	-0.0025	0.0009	-0.0006	-0.0164	-0.0023	0.0010
	MSE(a)*100	0.2202	0.2196	0.1975	0.2036	0.2488	0.2192	0.2221	0.2321	0.2767	0.2641	0.2558	0.2726	0.2557	0.2276	0.2285	0.2369
	MSE(b)	0.0348	0.0141	0.0132	0.0133	0.0451	0.0177	0.0176	0.0179	0.0554	0.0215	0.0212	0.0216	0.0483	0.0185	0.0184	0.0186
2.652	Beta Bias	-0.0007	-0.0394	0.0024	0.0131	0.0021	-0.0313	-0.0011	0.0059	0.0000	-0.0368	0.0011	0.0114	-0.0009	-0.0299	-0.0015	0.0071
	MSE(a)*100	0.2619	0.3744	0.2696	0.2762	0.2748	0.3196	0.2605	0.2969	0.3075	0.4061	0.3179	0.3611	0.2797	0.3196	0.2745	0.2996
	MSE(b)	0.0502	0.0211	0.0173	0.0173	0.0557	0.0224	0.0208	0.0214	0.0690	0.0279	0.0250	0.0257	0.0577	0.0229	0.0214	0.0219
3.978	Beta Bias	0.0004	-0.0478	0.0145	0.0170	0.0037	-0.0389	0.0034	0.0136	-0.0000	-0.0435	0.0090	0.0189	-0.0001	-0.0370	0.0040	0.0157
	MSE(a)*100	0.2977	0.5125	0.3672	0.3050	0.3061	0.4269	0.3237	0.3354	0.3353	0.5199	0.3809	0.3789	0.3094	0.4136	0.3305	0.3506
	MSE(b)	0.0639	0.0268	0.0204	0.0194	0.0657	0.0260	0.0230	0.0233	0.0800	0.0329	0.0274	0.0271	0.0668	0.0271	0.0238	0.0239
5.304	Beta Bias	-0.0005	-0.0524	0.0245	0.0105	0.0065	-0.0409	0.0110	0.0169	-0.0016	-0.0466	0.0160	0.0158	-0.0001	-0.0410	0.0106	0.0199
	MSE(a)*100	0.3272	0.6135	0.4709	0.3069	0.3226	0.4799	0.3762	0.3370	0.3544	0.6029	0.4432	0.3690	0.3220	0.4851	0.3783	0.3681
	MSE(b)	0.0738	0.0306	0.0229	0.0219	0.0736	0.0285	0.0249	0.0249	0.0872	0.0363	0.0287	0.0283	0.0744	0.0306	0.0255	0.0251

[a] *10,000 simulated data sets, each consisting of 38 observations, are generated under the following parameter settings: measurement error variance* $\sigma_\varepsilon^2 = 2.283$; *slope variance* $\sigma_\beta^2 = 0.05$; *the expected value of slope* $\beta = 0.239$; *a common initial value* $\alpha = 3.91$; *the censoring parameter* $\gamma_0 = -1.109$.
[b] Bias of population slope.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

CDF of Logistic distribution

When Logistic CDF is used in the truncated Geometric distribution, the values from Table 2 indicate that the minimum bias is observed in unweighted estimators for moderate to high ($\gamma_1 = 1.326, 2.652, 3.978,$ and 5.304) degree of informative right censoring. When the degree of informative right censoring is low ($\gamma_1 = 0.7$), IRC2 estimator performs better and when no ($\gamma_1 = 0$) informative right censoring is present unweighted estimator performs better. When the individual slope estimators are concerned IRC1 performs better for moderate degree ($\gamma_1 = 1.326, 2.652,$ and 3.978) of informative right censoring, followed by IRC2 estimator for high degree ($\gamma_1 = 5.304$) of informative right censoring and weighted estimator for no ($\gamma_1 = 0$) or low ($\gamma_1 = 0.7$) informative right censoring. When the population slope estimator is concerned weighted estimators perform better for no ($\gamma_1 = 0$) or low ($\gamma_1 = 0.7$ and1.326) degree of informative right censoring. For high degree ($\gamma_1 = 3.978$ and 5.304) of informative right censoring, unweighted slope estimator performs better, while IRC1 performs better for $\gamma_1 = 2.652$.

By comparing the results from all the four CDFs, it is inferred that, in case of bias, the unweighted and weighted estimators perform better in most of the cases for Standard Normal, Standard Minimum Gumbel and Logistic distributions. But, in case of Standard Cauchy, IRC1 and IRC2 estimators perform better. For individual slope estimation, IRC1 and IRC2 estimators perform better when moderate to high informative right censoring is present. In case of population slope estimation, the unweighted and weighted slope estimators have minimum MSE in most of the cases and IRC1 in some cases where moderate informative right censoring is present.

Table 3. Comparisons of informative right censoring parameter estimators under Standard Normal, Standard Cauchy, Standard minimum Gumbel and Logistic distributions: Bias and Mean Squared Error[a]

γ_1	Metric	Standard Normal		Standard Cauchy		Standard Minimum Gumbel		Logistic	
		IRC1N	IRC2N	IRC1C	IRC2C	IRC1G	IRC2G	IRC1L	IRC2L
0	Bias	0.00524	0.00654	0.06038	0.04225	-0.00713	-0.00685	-0.00423	-0.00468
	MSE	0.32775	0.47156	2.48689	1.72046	0.67072	1.17091	0.85287	1.59006
0.7	Bias	-0.12078	0.10019	0.08931	0.08978	-0.16797	0.08352	-0.17808	0.08657
	MSE	0.32243	0.54824	1.74325	1.67830	0.61937	1.29379	0.84650	1.84331
1.326	Bias	-0.30582	0.22049	-0.0219	0.18188	-0.37158	0.23575	-0.36428	0.26736
	MSE	0.39398	0.75351	1.23767	2.13115	0.67749	1.78121	0.92216	2.31572
2.652	Bias	-0.89868	0.69143	-0.53991	0.96762	-0.98702	0.90029	-0.91608	0.72547
	MSE	1.17736	1.93055	1.28869	6.60069	1.49044	5.07310	1.66056	4.84819
3.978	Bias	-1.65901	1.32096	-1.16851	2.10566	-1.73500	1.79312	-1.59390	1.34159
	MSE	3.25851	4.29650	2.72374	14.03021	3.65882	16.87673	3.51186	9.51861
5.304	Bias	-2.57427	1.66614	-1.75355	3.18467	-2.56492	2.20808	-2.32736	1.96621
	MSE	7.32929	6.50522	5.09033	25.12359	7.48681	13.66557	6.64045	13.09836

[a]The conditions of the simulations are explained in [a] of Table 2.

Table 3 compares the IRC estimators for the selected six values of informative right censoring parameter γ_1 across the CDF of four selected distributions in terms of their bias and MSE. The values of MSE from Table 3 indicate that IRC1 provides better estimator in case of Standard minimum Gumbel and Logistic distributions for all values of γ_1 under consideration. Further, it serves as a best estimator for most of the γ_1 values in Standard Normal and Standard Cauchy distributions.

IRC2 estimator works well for higher informative right censoring ($\gamma_1 = 5.304$) in Standard Normal distribution and for non-informative ($\gamma_1 = 0$) and low informative right censoring ($\gamma_1 = 0.7$) in Standard Cauchy distribution. Considering the bias values from Table 3, it is observed that IRC2 estimator performs better in Standard Normal, Standard minimum Gumbel, and Logistic distributions for most of the γ_1 values. In case of Standard Cauchy distributions, it is noticed that IRC1 estimator performs better in terms of bias.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

PBC Observed Trial Data Results

In this section, the model requirements of PBC data in terms of informative right censoring and goodness-of-fit test for number of visits, are discussed. Further, the best choice of the parameter estimates, in the light of simulation results, has been provided. LRT is used to test the significance of the estimated parameters.

The average log transformed Serum bilirubin measure is calculated at each follow-up month for groups of subjects classified according to their respective number of visits. From Figure 5, it is observed that patients with fewer visits generally exhibit higher average Serum bilirubin levels, whereas those with a greater number of visits tend to have relatively lower levels. Furthermore, some groups with fewer visits show considerable fluctuations and increasing trends in Serum bilirubin over time.

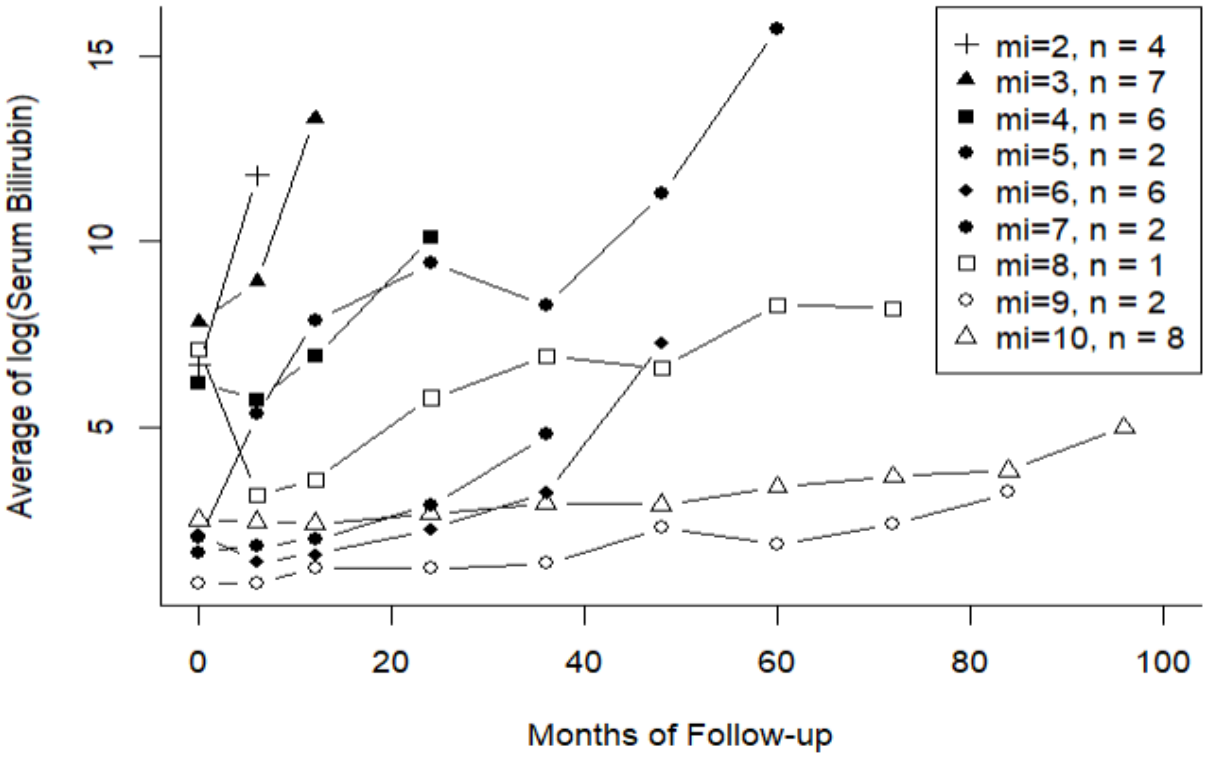

Figure 5. Mean of log transformation of Serum Bilirubin vs months following the initial visit by the number of patients indicated in the legend.

The correlation between the ordinary least square estimates of the individual slopes and the number of visits is -0.39 (p-value = 0.014), indicating a moderate and statistically significant negative association. It is further observed that the slope of Serum Bilirubin varies according to the dropout pattern. Subjects with consistently higher (positive) slope of Serum Bilirubin level tend to have fewer follow-up visits and, consequently, shorter follow-up periods. Thus, the number of follow-up visits appears to be associated with the individual longitudinal trajectories of serum bilirubin. These findings suggest that the missingness mechanism is unlikely to be completely random and may be informative.

The appropriateness of the distribution of 'number of visits' for four different CDFs for PBC observed trial data is tested using Chi-square test for goodness of fit and the results are presented in Table 4.

The p-values from Table 4 indicate that all the four distributions have statistically significant fit under both IRC1 and IRC2 estimators. Among these four distributions, the Standard minimum Gumbel provides the best fit under IRC1 with the maximum p-value (0.359). Thus, the goodness-of-fit results suggests that the Standard Minimum Gumbel distribution provides a more suitable framework for estimating both the slope and informative right censoring parameters compared with the other CDFs.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue XI November 2026

Table 4. PBC Observed Trial data - Comparison of observed and expected frequencies under truncated Geometric distribution for Standard Normal, Standard Cauchy, Standard Minimum Gumbel, and Logistic CDFs.

No. of visits	Observed Freq	Standard Normal		Standard Cauchy		Standard Minimum Gumbel		Logistic	
		IRC1N	IRC2N	IRC1C	IRC2C	IRC1G	IRC2G	IRC1L	IRC2L
2	4	7.026	7.632	7.629	9.000	6.822	7.324	7.109	7.757
3	7	5.506	5.498	5.389	5.488	5.474	5.488	5.513	5.523
4	6	4.418	4.323	4.255	3.990	4.443	4.379	4.409	4.334
5	2	3.592	3.478	3.433	3.092	3.634	3.563	3.582	3.477
6	6	2.943	2.826	2.798	2.486	2.988	2.921	2.934	2.818
7	2	2.423	2.312	2.298	2.045	2.466	2.403	2.416	2.301
8	1	2.002	1.903	1.901	1.707	2.040	1.983	1.995	1.890
9	2	1.658	1.574	1.581	1.438	1.690	1.640	1.652	1.562
10	8	8.432	8.454	8.717	8.754	8.443	8.300	8.391	8.339
χ^2 Test Statistic[c]		6.824	7.593	7.822	10.138	6.607	7.102	6.896	7.671
p-value[d]		0.337	0.269	0.251	0.118	0.359	0.312	0.331	0.263

[c]The goodness-of-fit test uses the statistic: $\sum_{i=1}^{9}(O_i - E_i)^2/E_i$.

[d]The p-value is obtained by comparing the test statistic with a χ^2 distribution having six degrees of freedom.

Table 5. PBC Observed Trial Data - Estimation of the Population Slope and Informative Right Censoring Parameters

Parameter	Methods	Standard Normal	Standard Cauchy	Standard Minimum Gumbel	Logistic
Population Slope	Unweighted	0.2389	0.2389	0.2389	0.2389
	Weighted	0.1653	0.1653	0.1653	0.1653
	IRC1	0.1846	0.1847	0.1829	0.1847
	IRC2	0.1774	0.1817	0.1737	0.1776
Censoring Parameter IRC1	Gamma0	-1.0575	-2.2554	-1.802	-1.7732
	Gamma1	0.7645	3.471	0.8385	1.3954
	LRT[e]	8.0127	10.261	5.8661	8.1055
	p-value	0.0046	0.0014	0.0154	0.0044
Censoring Parameter IRC2	Gamma0	-1.1093	-2.4714	-1.8456	-1.8593
	Gamma1	1.3257	5.4811	1.3804	2.3786
	LRT[e]	10.3443	14.5236	7.2335	10.08
	p-value	0.0013	0.0001	0.0072	0.0015

[e] LRT test for $H_0: \gamma_1 = 0$

Table 5 provides the population and individual slope estimates with their associated estimates of informative right censoring parameter under IRC1 and IRC2. Going by the goodness-of-fit test for the number of visits, it is observed that, from Table 4, Standard minimum Gumbel CDF under IRC1 provides the best fit for the considered observed trial data. Thus, the focus is restricted to the Standard minimum Gumbel distribution in Table 5, though estimates from other distributions are also provided for comparison. Since the observed trial data from Table 4 confirms considerable evidence of informative right censoring, using the information from simulation results in Table 2, the estimate of informative right censoring parameter is taken as 1.3804, that falls under Standard minimum Gumbel with IRC2 estimate.

If IRC1 is taken, then the results from the simulation table indicate a weighted estimate which contradicts the significant results of the goodness-of-fit. Hence, the associated best estimate for the population slope is 0.1737. For individual slope estimates, it is recommended that IRC2 under Standard minimum Gumbel CDF could be used.

CONCLUSION

In longitudinal studies it is of main interest to know how certain important study variables change over time. Estimating this rate of change or slope becomes difficult when participants drop out for reasons related to the study variable, in particular the rate of change of the study variable. Identification of certain patterns in the rate of change, particularly at the earlier stage of observations helps to device quicker intervention to change the course of the patient in reducing the impact of the disease. In this study, the number of visits by each subject is modeled as a truncated Geometric distribution which depends on the individual's rate of change in the study variable expressed through the Cumulative Distribution Function of Standard Normal, Standard Cauchy, Standard minimum Gumbel, and Logistic distributions. Simulated and PBC observed trial datasets are used to find out appropriate distribution for the follow-up and better estimators of slope and right censoring parameters. The fit metrics included bias, MSE and goodness of fit.

In this study, apart from the existing four methods compared by Mori et.al (1994) with Logistic CDF, the other three CDFs mentioned above are also used. The four methods of Mori included Unweighted, Weighted, Informative Right Censoring based on Naive approach (IRC1) and Informative Right Censoring based on adjusted method (IRC2). The simulation results indicate that for all the four CDFs, in case of bias, the unweighted and weighted estimators perform better in most of the informative right censoring parameter values for Standard Normal, Standard Minimum Gumbel and Logistic distributions. But, in case of Standard Cauchy, IRC1 and IRC2 estimators perform better. For individual slope estimation, IRC1 and IRC2 estimators perform better when moderate to high informative right censoring is present. In case of population slope estimation, the unweighted and weighted slope estimators have minimum MSE in most of the cases and IRC1 in some cases where moderate informative right censoring is present.

While estimating the informative right censoring parameter using IRC1 and IRC2 estimators for the four CDFs, IRC1 provides better estimator in case of MSE for Standard minimum Gumbel and Logistic distributions for all values of the informative right censoring parameter under consideration. Further, it serves as a best estimator for most of the informative right censoring parameter values in Standard Normal and Standard Cauchy distributions. IRC2 estimator works well for higher informative right censoring in Standard Normal distribution and for no and low informative right censoring in Standard Cauchy distribution. In terms of bias, IRC2 estimator performs better in Standard Normal, Standard minimum Gumbel, and Logistic distributions for most of the informative right censoring parameter values. In case of Standard Cauchy distributions, it is noticed that IRC1 estimator performs better in terms of bias.

For the observed trial PBC data, the association between the slope of Serum Bilirubin and the length of follow-up is well established with statistical significance, indicating the presence of informative right censoring in the data. The distribution of 'number of visits' is better approximated by Standard minimum Gumbel under IRC1. The combined results of simulation and that of the observed trial data leads to the informative right censoring estimate 1.3804 and the population slope estimate as 0.1737, based on IRC2 estimator. The results further suggest that IRC2 under Standard minimum Gumbel CDF could be used for estimating the individual slopes of the patients.

LIMITATIONS

The models included in this study may not adequately capture more complex longitudinal data structures in which subjects appear intermittently. Apart from the study variable, if information on additional correlated variables is available, the accuracy of the slope estimation could be improved further. Moreover, in this study the number of visits of the individuals is modeled using the truncated Geometric distribution, whereas alternative discrete distributions, such as Poisson or Negative Binomial could also be explored.

SCOPE FOR FUTURE WORK

Future work may focus on dropout of patients with the intermittent patterns and with additional correlated covariates of the study variable. In addition, other distributions such as Poisson or Negative Binomial could be used to accommodate different types of dropout patterns in the patients.

ACKNOWLEDGEMENT

The authors sincerely thank the editors and the reviewers for their constructive and critical comments and valuable suggestions, which greatly helped to improve the content and the presentation of this manuscript.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

DATA AVAILABILITY STATEMENT

The observed trial data used in this study, namely PBC (Primary Biliary Cirrhosis) is taken from the R package "JM" (Joint Modeling of Longitudinal and Survival Data, July21,2025). The details of the data are available in https://cran.r-project.org/web/packages/JM/index.html

REFERENCE

1. Fearn, T. (1975), "A Bayesian approach to growth curves", Biometrika, 62, 89-100. DOI: https://doi.org/10.2307/2334490.
2. Fleming, T. and Harrington, D. (1991), "Counting Processes and Survival Analysis", Wiley, New York.
3. Jaffa, M. A., Lipsitz, S. R. and Woolson, R. F. (2011), "Slope estimation for informatively right censored longitudinal data modelling the number of observations using Geometric and Poisson distributions: application to renal transplant cohort", Statistical Methods in Medical Research, 24(6), 819-835. DOI: https://doi.org/10.1177/0962280211430681.
4. Mori, M., Woodworth, G. G. and Woolson R. F. (1992), "Application of Empirical Bayes Inference to Estimation of Rate of Change in the Presence of Informative Right Censoring", Statistics in Medicine, 11, 621-631. DOI: https://doi.org/10.1002/sim.4780110507.
5. Mori, M., Woolson R. F. and Woodworth, G. G. (1994), "Slope Estimation in the Presence of Informative Right Censoring: Modeling the Number of Observations as a Geometric Random Variable", Biometrics, 50, 39-50. DOI: https://doi.org/10.2307/2533195.
6. Poupon, R. (2010), "Primary biliary cirrhosis: a 2010 update", Journal of Hepatology. 52(5), 745-758. DOI:10.1016/j.jhep.2009.11.027.
7. Rizopoulos, D. (2010), "JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data", Journal of Statistical Software, 35(9), 1–33. DOI: https://doi.org/10.18637/jss.v035.i09.
8. Searle, S. R. (1971), "Linear Models", Wiley, New York.
9. Stefanski, L. A. and Carroll, R. J. (1985), "Covariate measurement error in logistic regression", The Annals of Statistics, 13, 1335-1351. DOI: https://www.jstor.org/stable/2241358.
10. Therneau, T. and Grambsch, P. (2000), "Modeling Survival Data: Extending the Cox Model", Springer Verlag, New York.
11. Viswanathan, N. and Ravanan, R. (2014), "Linear Minimum Variance Unbiased Estimation of Individual and Population slopes in the presence of Informative Right Censoring with Exponential Censoring Structure", International Journal of Current Medical and Applied Sciences, 4(1), 41-46.
12. Wu, M. C. and Carroll, R. J. (1988), "Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process", Biometrics, 44, 175-188.
13. Wu, M. C. and Bailey, K. (1989), "Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model", Biometrics, 45, 939-955.