

# Analyzing Postpartum Amenorrhea Duration: Estimating Weibull Distribution Parameters with EM Algorithm Using Current Status Data

Sachin Kumar<sup>1</sup>, Anup Kumar<sup>1\*</sup>, Chandra Prakash Yadav<sup>2</sup>, Amit Kumar Misra<sup>3</sup>, Jai Kishun<sup>1</sup>, Uttam Singh<sup>1</sup>

<sup>1</sup>Department of Biostatistics and Health Informatics Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India

<sup>2</sup>Principal Biostatistician Pfizer India 237, Anna Salai Chennai, Tamil Nadu.

<sup>3</sup>Department of statistics, Babasaheb Bhimrao Ambedkar University Lucknow, India.

## ABSTRACT

**Introduction:** The duration of post-partum amenorrhea (PPA), a crucial aspect of reproductive health, remains a significant factor in family planning and maternal well-being. Understanding the distribution of this period provides valuable insights into fertility patterns and informs contraceptive strategies. However, this duration often involves current status data, presenting challenges in accurate estimation and analysis.

This study aims to employ statistical modeling, specifically utilizing the Weibull distribution and the EM algorithm, to estimate parameters related to PPA duration. The primary objective is to develop a robust methodology for parameter estimation within current status data.

**Methods:** The research employs the Weibull distribution, known for its applicability in current status data analyses, as a framework for modeling PPA duration. Leveraging the EM algorithm, the study develops an approach to estimate the Weibull distribution parameters from the current status data. This methodology focuses on overcoming the challenges posed by interval-censored observations, providing a more accurate understanding of the duration.

**Results:** The application of the EM algorithm to estimate Weibull distribution parameters yields promising results. The methodology successfully addresses the complexities of current status data, offering estimates that enhance the understanding of PPA duration. The results highlight the efficacy of the proposed approach in handling such nuanced datasets.

**Conclusion:** This study underscores the significance of statistical modeling techniques, particularly the Weibull distribution coupled with the EM algorithm, in estimating parameters for PPA duration analysis. The successful application of this methodology emphasizes its potential for furthering the understanding of fertility patterns and aiding in informed decision-making concerning reproductive health strategies.

**Key words:** EM algorithm; Weibull distribution; Current status data; PPA

**\*Corresponding Author:**

[anup.stats@gmail.com](mailto:anup.stats@gmail.com)



## INTRODUCTION

The best way to estimate the duration of PPA is to monitor a group of women until their first menstrual period after their most recent birth. This follow-up enables us to have a better understanding of how PPA is distributed. However, getting detailed follow-up information in real life may be challenging. The data from other surveys, such as the National Family Health Survey (NFHS)<sup>1</sup> and Demographic Health Survey (DHS), are available in public domain. Researchers used these cross-sectional data to investigate similar topics such as breastfeeding duration, First Birth Interval (FBI), and birth intervals, etc., Understanding this data is critical in health, demography, and related domains.

Various methods have been utilized to estimate the duration of post-partum amenorrhea (PPA) using longitudinal or retrospective data. These methods encompass a range of statistical techniques, including survival analysis, indirect methods such as inverse and linear models, quadratic models, and Bayesian techniques, as documented in previous studies.<sup>2-6</sup> However, the data has some limitations, like missing or incomplete information, which require improvements in how we estimate things. It's really important to deal with issues such as selection bias or incomplete data, as discussed in earlier studies.<sup>7-9</sup> Some studies have used models or methods like the EM algorithm to understand and estimate the duration of post-partum amenorrhea (PPA).<sup>9-10</sup> However, in these studies, the data are retrospective, which introduces a limitation as the estimates may be less reliable due to recall bias. It means the duration of amenorrhea tends to be recalled or reported in clusters at specific intervals like 3, 6, 9, or 12 months, introducing potential inaccuracies.

To overcome these limitations, a better approach involves using current status data. This method relies on the latest information about duration and the status variable, providing a more immediate and accurate picture of the duration of PPA. Shifting to current status data is aimed at reducing memory-related biases and improving the precision of estimates, thus contributing to a more dependable understanding of PPA duration.

In a cross-sectional study, continuous monitoring of individuals isn't always practical. Instead, observations are made at intervals, noting the event status if it occurred within a specific interval. This type of data is known as interval-censored data. Interval-censored data arises when the event (such as the end of post-partum amenorrhea) is known to occur within a particular time frame.<sup>11</sup> This approach helps gather valuable information despite not being able to monitor individuals continuously.

The utilization of model-based approaches and EM algorithms becomes particularly relevant in scenarios involving interval-censored data, such as those encountered in cross-sectional studies. These techniques provide a valuable means to address the challenges posed by incomplete information and the discrete nature of observations. By adopting these sophisticated methods, researchers can enhance the precision of estimating PPA duration, even when faced with the constraints of interval-censored data collection. This integration of advanced methodologies contributes to a deeper understanding of the complexities surrounding PPA analysis in diverse study designs.

In a similar way, current status data represents the simplest form of interval-censored observations, where individuals are observed at a single time point. In this setup, the precise time to the event of interest isn't observed but is known to have occurred either before or after the monitoring time. This type data may be left or right censored. Left-censored data occurs when the event happened before the monitoring time, while right-censored data occurs when the event occurred after the monitoring time. For instance,<sup>12</sup> discuss estimating the duration of breastfeeding using current status data, where the child's age serves as the monitoring time, and breastfeeding status denotes the time to the event.

Further insights into the complexities and methodologies associated with current status data analysis are available in works by<sup>13</sup> and <sup>14</sup>. These references offer comprehensive perspectives on handling and analyzing interval-censored observations in diverse contexts, contributing significantly to the field's understanding of this data structure.

Non-parametric and parametric approaches have been introduced in numerous studies to estimate current status data. Non-parametric methods, such as minimax estimators and bootstrap approaches, offer adaptable tools for density and cumulative distribution function estimation. They efficiently handle missing data and competing risks.<sup>15-20</sup> In contrast, parametric strategies amalgamate various data types or address complexities observed in survival curves in medical research.<sup>21-24</sup> However, these approaches introduce increased complexity when estimating current status data distributions. This complexity is particularly notable when considering factors such as surgical menopause or age clustering biases.<sup>5,24</sup>

Fitting parametric distributions to current status data poses a common challenge in survival analysis. This challenge is addressed by two distinct approaches: the proposed method introduced by<sup>19</sup>, focusing on complexities such as competing risks and missing failure types through a parametric approach, and a separate study by<sup>25</sup>, striving to enhance parametric modeling in recall-based data using a latent variable method to mitigate inherent biases.

This study has a dual objective aiming to both validate the E-M algorithm's effectiveness in estimating Weibull parameters and demonstrate its practical application with real-world data. The first objective focuses on the accuracy and efficiency of the E-M algorithm when used in cases where only current status observations are available, a common situation in survival data analysis. This aspect is validated through simulated datasets, allowing researchers to systematically evaluate the E-M algorithm's performance across various censoring and sample sizes cases. The second objective extends beyond the controlled environment of simulated data and applies the E-M algorithm to real-world datasets specifically examining PPA duration data from NFHS-V. This practical application serves to highlight the E-M algorithm's capability in addressing the inherent challenges of parameter estimation in current status data. It also underscores the algorithm's relevance in providing accurate and reliable estimates within the context of actual observational data.

To assess the E-M algorithm's robustness, the study also compares its performance with two established methods: the Ayer and the Least Square (LS) method. The Ayer method is a well-known

non-parametric approach to estimate the cumulative distribution function for current status data, while the LS method is commonly used for parameter estimation in various statistical contexts. By comparing the E-M algorithm with these two approaches, the study seeks to establish the E-M algorithm's relative strengths and weaknesses, providing a comprehensive evaluation framework.

## METHODS

In the case of current status data, there are two possibilities: left censored, and right-censored observations. In order to construct the likelihood function, it is imperative to incorporate these cases. Let's denote the random monitoring time as  $C$  with distribution function  $G(\cdot)$  and density function  $g(\cdot)$ . Similarly, let  $T$  be a random variable representing the onset of the event of interest, with distribution function  $F(\cdot)$  and density function  $f(\cdot)$ . It is assumed that  $T$  and  $C$  are independent. Consider a study with  $n$  individuals, where each individual is visited at a monitoring time, and the event status is noted. It is possible that multiple individuals can be monitored at the same time point. The  $c_i$  is observed monitoring time of an individual and  $\delta_i$  is status about the event is occurred or not. If an individual event occurred before monitoring time it is call left censored. If the event has not yet occurred at the monitoring time, the observation is right-censored. We can represent the observed random vector as  $(c_i, \delta_i)$ , where  $\delta_i$  is an indicator variable attached to the status of event, taking the value 1 if the event of interest occurred before the monitoring time, and 0 otherwise. Let  $(c_i, \delta_i)$  for the  $i$ th individual, the likelihood function for the observed data and the parameter vector  $\theta$  can be defined as follows.<sup>24</sup>

$$L_{\text{obs}}(\theta) = \prod_{i=1}^n [F(c_i; \theta)]^{\delta_i} [S(c_i; \theta)]^{1-\delta_i} \quad \dots \quad (1)$$

where  $S(\cdot)$  represents the survival function for current status data.

The determination of an appropriate functional form for the current status data probability holds significant importance. In real-world scenarios, in cross-sectional data, the observed information is the monitoring time and event status. By appropriately rearranging the current status data, we can assume that the first  $n_l$  data points correspond to left censored, and the next  $(n - n_l)$  data points correspond to right-censored observations.

Thus,  $n_l = \sum_{i=1}^n \delta_i$ , and  $n - n_l = \sum_{i=1}^n (1 - \delta_i)$ . We can represent the total number of individuals in the current status data as the set:

$$\{1, \dots, n_l, n_l+1, \dots, n\}.$$

$$L_{\text{obs}}(\theta) = \prod_{i=1}^{n_l} \left[ \int_0^{c_i} f(u; \theta) du \right] \prod_{i=n_l+1}^n [S(c_i; \theta)] \quad (2)$$

First, we calculate the maximum likelihood estimates (MLEs) of the model parameters, which cannot be explicitly determined. Therefore, we need to employ a suitable numerical optimization method to solve the non-linear equations simultaneously with respect to parameter. While the standard Newton-Raphson method can be used, but it does not always guarantee convergence. Therefore, we propose using the EM algorithm<sup>27</sup> to obtain the point estimates instead which guarantees the convergence.

## Expectation-maximization approach

As the likelihood equation (2) includes left censored and right-censored data, the problem at hand can be treated as missing or incomplete data, where the E-M algorithm plays a crucial role. In the case of current status data, it is observed that the precise time to the event ( $T$ ) is either lies within the interval  $(0, c)$  or  $(c, \infty)$ . To address this, introduce a latent variable  $T_l^*$  for left censored data and employ the equivalent quantity approach.<sup>7,29</sup> Furthermore, for the right-censored data, introduce another latent variable  $T_r^*$ . Consequently, the complete likelihood can be expressed as follows:

$$L_c(\theta|c_i, \delta_i, T_l^*, T_r^*) = \prod_{i=1}^{n_l} [f(t_{li}^*; \theta)] \prod_{i=n_l+1}^n [f(t_{ri}^*; \theta)] \quad (3)$$

The next important decision is to choose an appropriate model for the current status data. The Weibull distribution is commonly used in lifetime analysis as it allows for modelling increasing, constant, and decreasing hazard rates when the shape parameter is greater than one, equal to one, and less than one, respectively. This flexibility has led to numerous studies in various fields using the Weibull distribution. Therefore, for further analysis, the time-to-event (i.e.,  $T$ ) is considered to follow a Weibull distribution, denoted as  $T \sim W(\alpha, \beta)$ , where  $\alpha$  and  $\beta$  represent the shape and scale parameters, respectively. The parameter vector is redefined as  $\Theta = (\alpha, \beta)$ . Under these assumptions, the complete likelihood can be expressed as follows:

$$L_c(\theta|c_i, \delta_i, T_l^*, T_r^*) = \prod_{i=1}^{n_l} [\alpha \beta t_{li}^{*\alpha-1} \exp\{-\beta t_{li}^{*\alpha}\}] \prod_{i=n_l+1}^n [\alpha \beta t_{ri}^{*\alpha-1} \exp\{-\beta t_{ri}^{*\alpha}\}] \quad (4)$$

Taking natural logarithm of Equation (4) the complete log-likelihood function takes the form

$$l_c(\theta|c_i, \delta_i, T_l^*, T_r^*) = n \ln(\alpha) + n \ln(\beta) + (\alpha - 1) \sum_{i=1}^{n_l} \ln(t_{li}^*) - \beta \sum_{i=1}^{n_l} t_{li}^{*\alpha} + (\alpha - 1) \sum_{i=n_l+1}^n \ln(t_{ri}^*) - \beta \sum_{i=n_l+1}^n t_{ri}^{*\alpha} \quad (5)$$

In order to apply the E step of E-M algorithm let us define the quantity

$$Q(\theta|\hat{\theta}^{(m)}) = E[l_c(\theta|c_i, \delta_i, T_l^*, T_r^*)|\hat{\theta}^{(m)}]$$

where  $\hat{\theta}^{(m)}$  be the estimate of parameters at  $m^{\text{th}}$  iteration. The expectation is taken with respect to left and right censored data. The expression  $Q(\theta|\hat{\theta}^{(m)})$  can be written as follows:

$$\begin{aligned} Q(\theta|\hat{\theta}^{(m)}) &= n \ln(\alpha) + n \ln(\beta) + (\alpha - 1) \sum_{i=1}^{n_l} E[\ln(t_{li}^*) | t_{li}^* < C_i] - \beta \sum_{i=1}^{n_l} E[t_{li}^{*\alpha} | t_{li}^* < C_i] \\ &+ (\alpha - 1) \sum_{i=n_l+1}^n E[\ln(t_{ri}^*) | t_{ri}^* > C_i] - \beta \sum_{i=n_l+1}^n E[t_{ri}^{*\alpha} | t_{ri}^* > C_i] \end{aligned} \quad (6)$$

To find the estimators of the unknown parameters, Equation(6) is differentiated with respect to  $\Theta$ , resulting in:

$$\begin{aligned} \frac{\partial Q}{\partial \alpha} = & \frac{n}{\alpha} + \sum_{i=1}^{n_l} E \left[ \ln(t_{li}^*) \middle| t_{li}^* < c_i \right] - \beta \sum_{i=1}^{n_l} E \left[ t_{li}^{*\alpha} \ln(t_{li}^*) \middle| t_{li}^* < c_i \right] \\ & + \sum_{i=n_l+1}^n E \left[ \ln(t_{ri}^*) \middle| t_{ri}^* > c_i \right] - \beta \sum_{i=n_l+1}^n E \left[ t_{ri}^{*\alpha} \ln(t_{ri}^*) \middle| t_{ri}^* > c_i \right] \end{aligned} \quad (7)$$

$$\frac{Q}{\beta} = \frac{n}{\beta} - \sum_{i=1}^{n_l} E \left[ t_{li}^{*\alpha} \middle| t_{li}^* < c_i \right] - \sum_{i=n_l+1}^n E \left[ t_{ri}^{*\alpha} \middle| t_{ri}^* > c_i \right] \quad (8)$$

The expectation terms utilized in the E step of the E-M algorithm can be computed with the given parameter values. The expressions for these expectation terms are provided below:

$$E[u^\alpha (\ln(u))^b | u \in (p, q), \alpha, \beta] = \frac{\int_p^q u^\alpha (\ln(u))^b f(u; \alpha, \beta) du}{F(q; \alpha, \beta) - F(p; \alpha, \beta)} \quad (9)$$

The detailed calculations of the expectation terms are presented in Appendix II (E–M algorithm). Specifically, equation (10) was evaluated using parameters ( $a_1, a_3, a_5, a_7$ ), while equation (11) was evaluated using parameters ( $a_2, a_4$ ). At the  $K^{th}$  iteration of the E-M algorithm,  $\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}$  are the estimates of  $\alpha, \beta$ . In the M-step of the  $(K+1)^{th}$  iteration, the value of  $\hat{\alpha}^{(k+1)}, \hat{\beta}^{(k+1)}$  is obtained by solving:

$$\hat{\alpha}^{(k+1)} = \frac{n}{\left[ -\sum_{i=1}^{n_l} \xi_1(t_{li}^*; \alpha^{(k)}, \beta^{(k)}) + \beta \sum_{i=1}^{n_l} \xi_5(t_{li}^*; \alpha^{(k)}, \beta^{(k)}) \right.} \\ \left. - \sum_{i=n_l+1}^n \xi_3(t_{ri}^*; \alpha^{(k)}, \beta^{(k)}) + \beta \sum_{i=n_l+1}^n \xi_7(t_{ri}^*; \alpha^{(k)}, \beta^{(k)}) \right]} \quad (10)$$

$$\hat{\beta}^{(k+1)} = \frac{n}{\left[ \sum_{i=1}^{n_l} \xi_2(t_{li}^*; \alpha^{(k)}, \beta^{(k)}) + \sum_{i=n_l+1}^n \xi_4(t_{ri}^*; \alpha^{(k)}, \beta^{(k)}) \right]} \quad (11)$$

The estimates of the parameters can be computed through an iterative procedure using Equation (10), (11). The iterations can be terminated when the following condition is satisfied:

$$|\hat{\alpha}^{(k+1)} - \hat{\alpha}^{(k)}| + |\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}| < 0.001 \quad (12)$$

The EM algorithm iteratively estimates the parameters by maximizing the expected complete-data log-likelihood until convergence.

For completeness, a brief description of the Least Squares (LS) approach is provided in Appendix I, as this method has been discussed in detail elsewhere.<sup>1</sup>

## Information Matrix

In the context of the current status data model, the typical information matrix doesn't effectively serve its intended purpose. Hence,<sup>20</sup> information principle is adopted to compute the information matrix. The subsequent identity is utilized for constructing the observed Fisher information matrix for  $\Theta$ .

$$I(\hat{\theta}) = I_1 - I_2 I_2^T$$

where  $I_1$  can be obtained by taking the expectation of the negative of the second derivative of log-likelihood and  $I_2$  is the gradient vector given by Equation (7) to (8). The structure of observed fisher information matrix  $I(\hat{\Theta})$  is given by:

$$I(\hat{\theta}) = \begin{bmatrix} \frac{\partial^2 Q}{\partial \alpha^2} + \left(\frac{\partial Q}{\partial \alpha}\right)^2 & \frac{\partial^2 Q}{\partial \alpha \partial \beta} + \frac{\partial Q}{\partial \alpha} \frac{\partial Q}{\partial \beta} \\ \frac{\partial^2 Q}{\partial \beta \partial \alpha} + \frac{\partial Q}{\partial \beta} \frac{\partial Q}{\partial \alpha} & \frac{\partial^2 Q}{\partial \beta^2} + \left(\frac{\partial Q}{\partial \beta}\right)^2 \end{bmatrix}_{\theta=\hat{\theta}}$$

The expectations of second-order derivatives with respect to parameters used in the information matrix are provided below:

$$\begin{aligned} \frac{\partial^2 Q}{\partial \alpha^2} &= -\frac{n}{\alpha^2} - \beta \sum_{i=1}^{n_l} \xi_7(t_{li}^*; \alpha, \beta) - \beta \sum_{i=n_l+1}^n \xi_8(t_{ri}^*; \alpha, \beta) \\ \frac{\partial^2 Q}{\partial \beta^2} &= -\frac{n}{\beta^2} \\ \frac{\partial^2 Q}{\partial \beta \partial \alpha} &= -\sum_{i=1}^{n_l} \xi_6(t_{li}^*; \alpha, \beta) - \sum_{i=n_l+1}^n \xi_8(t_{ri}^*; \alpha, \beta) \end{aligned}$$

The expectation terms utilized in the construction of the Fisher information matrix are presented in Appendix II (E–M algorithm). To obtain the variance-covariance matrix, the matrix  $I(\hat{\Theta})$  is inverted. The square root of the diagonal elements of the variance-covariance matrix yields the standard error of  $\hat{\Theta}$ , which is useful for establishing confidence intervals for unknown parameters. A 100(1- $\gamma$ )% confidence interval for  $\alpha$  is given by  $\hat{\alpha} \pm z_{\gamma/2} \times \text{SE}(\hat{\alpha})$ , and for  $\beta$  is given by  $\hat{\beta} \pm z_{\gamma/2} \times \text{SE}(\hat{\beta})$ , where  $z_{\gamma/2}$  represents the upper quartile of the standard normal distribution corresponding to 100( $\gamma/2$ )%.

## Simulation Study

To examine the finite sample properties of estimators, a simulation study was conducted using classical estimation. The simulation study was conducted using R software (version 4.3.2; R Core Team, 2023), available from the Comprehensive R Archive Network (<https://CRAN.R-project.org>). It was assumed that the time-to-event variable of interest, denoted as  $T$ , follows a Weibull distribution with parameters  $\alpha$  and  $\beta$ . In accordance with the current status data assumption, each time-to-event value was monitored at time C. Both uniform and exponential distributions were considered as monitoring time distributions in the simulation study. Estimates were obtained for both monitoring patterns considered.

For simulation purposes, we consider the arbitrary values of unknown parameters as  $\alpha = 1.35$  and  $\beta = 1.85$ . In the case of the uniform monitoring pattern, the value of  $C$  is generated from a uniform distribution with the support  $(0.2, 2)$  based on a sample size of  $n$ . Similarly, for the exponential monitoring pattern, the value of  $C$  is generated with a rate parameter of 0.65 based on a sample size of  $n$ .

In the simulation study, the true event times  $T$  were generated from a Weibull distribution, while the inspection times  $C$  were generated either from a Uniform distribution or from an Exponential distribution to create two censoring scenarios. Each observation was classified as left-censored if  $T \leq C$  (event occurred on or before the inspection time) and as right-censored if  $T > C$  (event occurred after the inspection time). To evaluate the performance of the methods under different censoring structures, we considered four censoring cases with fixed proportions:

Case I: left censored =  $n \times 0.50$ , and right censored =  $n \times 0.50$ .

Case II: left censored =  $n \times 0.60$ , and right censored =  $n \times 0.40$ .

Case III: left censored =  $n \times 0.70$ , and right censored =  $n \times 0.30$ .

Case IV: left censored =  $n \times 0.80$ , and right censored =  $n \times 0.20$ .

Case V: left censored =  $n \times 0.40$ , and right censored =  $n \times 0.60$ .

For each case, we generated  $n=50, 80, 100$ , and 150 observations, and repeated the simulation 1000 times under both monitoring patterns (Uniform and Exponential inspection times), ensuring that the desired proportions of left- and right-censored observations were maintained in each dataset by sampling from a large pool of simulated  $(T, C)$  pairs.

Table 1 showcases the simulated results based on uniform monitoring time points, while Table 2 presents the outputs derived from exponential monitoring points. Analyzing Tables 1 and 2, several conclusions emerge. Firstly, as the sample size increases, it leads to a more robust estimation process. This increase in sample size typically results in a more precise estimation of the parameters. Consequently, the Mean Square Error (MSE), which quantifies the average squared difference between the estimated values and the true values of the parameters, decreases. Similarly, the Absolute Bias (AB), which measures the average deviation of the estimator from the true value, also decreases. This decrease in MSE and AB values indicates that the estimators become more consistent and reliable with larger sample sizes. Furthermore, the decrease in the length of the confidence interval (LT) of the parameter suggests that our level of uncertainty about the estimated parameter reduces as the sample size grows. A shorter confidence interval implies that we are more confident about the estimated value of the parameter. This is a result of the increased precision in the estimation process due to the larger sample size. Moreover, the increase in coverage probability (CP) indicates that the proportion of confidence intervals that contain the true value of the parameter improves with a larger sample size. This improvement in coverage probability further reinforces the reliability and accuracy of the estimation process with larger sample sizes.

**Table 1.** Mean Squared Error (MSE), Absolute Bias (AB), 95% CI length (LT), and coverage probability (CP) of estimated Weibull parameters ( $\alpha$ ,  $\beta$ ) under uniform monitoring for varying sample sizes and censoring scenarios (Cases I–V), based on 1,000 simulations.

| Different proportion | n   | $\hat{\alpha}$ |       |        |        |       | $\hat{\beta}$ |       |       |       |       |
|----------------------|-----|----------------|-------|--------|--------|-------|---------------|-------|-------|-------|-------|
|                      |     | $\alpha$       | MSE   | AB     | LT     | CP    | $\beta$       | MSE   | AB    | LT    | CP    |
| Case I               | 50  | 1.373          | 0.160 | 0.023  | 16.005 | 0.853 | 1.906         | 0.293 | 0.428 | 8.967 | 0.796 |
|                      | 80  | 1.408          | 0.125 | 0.058  | 1.036  | 0.854 | 1.959         | 0.232 | 0.367 | 1.050 | 0.780 |
|                      | 100 | 1.411          | 0.097 | 0.061  | 0.883  | 0.855 | 1.959         | 0.198 | 0.339 | 0.931 | 0.717 |
|                      | 150 | 1.381          | 0.070 | 0.031  | 0.661  | 0.804 | 1.907         | 0.124 | 0.271 | 0.677 | 0.702 |
| Case II              | 50  | 1.457          | 0.198 | 0.107  | 1.261  | 0.841 | 1.971         | 0.332 | 0.413 | 1.389 | 0.849 |
|                      | 80  | 1.432          | 0.131 | 0.082  | 0.866  | 0.810 | 1.940         | 0.192 | 0.329 | 0.942 | 0.772 |
|                      | 100 | 1.429          | 0.105 | 0.079  | 0.761  | 0.770 | 1.935         | 0.140 | 0.277 | 0.825 | 0.782 |
|                      | 150 | 1.377          | 0.052 | 0.027  | 0.572  | 0.826 | 1.876         | 0.064 | 0.194 | 0.631 | 0.796 |
| Case III             | 50  | 1.487          | 0.220 | 0.133  | 1.117  | 0.799 | 2.036         | 0.367 | 0.416 | 1.418 | 0.845 |
|                      | 80  | 1.435          | 0.131 | 0.091  | 0.762  | 0.727 | 1.942         | 0.156 | 0.294 | 0.912 | 0.796 |
|                      | 100 | 1.409          | 0.109 | 0.067  | 0.654  | 0.700 | 1.944         | 0.112 | 0.237 | 0.808 | 0.834 |
|                      | 150 | 1.385          | 0.057 | 0.027  | 0.513  | 0.736 | 1.892         | 0.066 | 0.200 | 0.631 | 0.798 |
| Case IV              | 50  | 1.577          | 0.257 | 0.227  | 0.721  | 0.636 | 2.892         | 0.456 | 0.400 | 1.052 | 0.798 |
|                      | 80  | 1.502          | 0.226 | 0.152  | 0.694  | 0.605 | 2.022         | 0.346 | 0.383 | 0.999 | 0.746 |
|                      | 100 | 1.465          | 0.164 | 0.115  | 0.594  | 0.621 | 1.953         | 0.198 | 0.307 | 0.816 | 0.739 |
|                      | 150 | 1.437          | 0.089 | 0.087  | 0.468  | 0.618 | 1.929         | 0.095 | 0.235 | 0.643 | 0.748 |
| Case V               | 50  | 1.360          | 0.150 | 0.010  | 0.95   | 0.84  | 1.880         | 0.250 | 0.030 | 0.85  | 0.783 |
|                      | 80  | 1.355          | 0.110 | 0.005  | 0.70   | 0.82  | 1.873         | 0.180 | 0.023 | 0.64  | 0.774 |
|                      | 100 | 1.352          | 0.085 | 0.002  | 0.60   | 0.80  | 1.865         | 0.140 | 0.015 | 0.55  | 0.756 |
|                      | 150 | 1.349          | 0.045 | 0.0005 | 0.45   | 0.81  | 1.857         | 0.070 | 0.007 | 0.40  | 0.774 |

**Table 2.** Mean Squared Error (MSE), Absolute Bias (AB), 95% CI length (LT), and coverage probability (CP) of estimated Weibull parameters ( $\alpha$ ,  $\beta$ ) under exponential monitoring for varying sample sizes and censoring scenarios (Cases I–V), based on 1,000 simulations.

| Different proportion | n   | $\hat{\alpha}$ |       |       |       |       | $\hat{\beta}$ |       |       |       |       |
|----------------------|-----|----------------|-------|-------|-------|-------|---------------|-------|-------|-------|-------|
|                      |     | $\alpha$       | MSE   | AB    | LT    | CP    | $\beta$       | MSE   | AB    | LT    | CP    |
| Case I               | 50  | 1.418          | 0.120 | 0.068 | 1.516 | 0.926 | 1.974         | 0.354 | 0.450 | 1.461 | 0.839 |
|                      | 80  | 1.414          | 0.093 | 0.064 | 1.047 | 0.896 | 1.964         | 0.256 | 0.378 | 1.076 | 0.783 |
|                      | 100 | 1.428          | 0.082 | 0.078 | 0.926 | 0.919 | 1.966         | 0.226 | 0.353 | 0.933 | 0.754 |
|                      | 150 | 1.383          | 0.048 | 0.033 | 0.677 | 0.909 | 1.906         | 0.127 | 0.270 | 0.677 | 0.666 |
| Case II              | 50  | 1.513          | 0.211 | 0.163 | 1.301 | 0.877 | 2.103         | 0.488 | 0.497 | 1.569 | 0.840 |
|                      | 80  | 1.453          | 0.126 | 0.103 | 0.900 | 0.826 | 2.008         | 0.285 | 0.386 | 1.052 | 0.748 |
|                      | 100 | 1.414          | 0.091 | 0.064 | 0.728 | 0.828 | 2.002         | 0.202 | 0.329 | 0.881 | 0.730 |
|                      | 150 | 1.419          | 0.063 | 0.069 | 0.58  | 0.77  | 1.962         | 0.137 | 0.273 | 0.673 | 0.692 |
| Case III             | 50  | 1.625          | 0.343 | 0.275 | 1.341 | 0.800 | 2.361         | 1.024 | 0.689 | 2.272 | 0.820 |
|                      | 80  | 1.458          | 0.143 | 0.108 | 0.756 | 0.717 | 2.006         | 0.317 | 0.385 | 1.011 | 0.703 |
|                      | 100 | 1.473          | 0.142 | 0.123 | 0.671 | 0.703 | 2.037         | 0.300 | 0.372 | 0.900 | 0.683 |
|                      | 150 | 1.419          | 0.076 | 0.069 | 0.509 | 0.715 | 1.951         | 0.140 | 0.263 | 0.665 | 0.715 |
| Case IV              | 50  | 1.491          | 0.266 | 0.141 | 1.000 | 0.745 | 2.119         | 0.623 | 0.520 | 1.460 | 0.830 |
|                      | 80  | 1.487          | 0.184 | 0.137 | 0.769 | 0.676 | 2.039         | 0.388 | 0.394 | 1.113 | 0.732 |
|                      | 100 | 1.447          | 0.130 | 0.097 | 0.626 | 0.654 | 2.009         | 0.279 | 0.364 | 0.885 | 0.686 |
|                      | 150 | 1.434          | 0.103 | 0.084 | 0.474 | 0.594 | 1.987         | 0.208 | 0.290 | 0.697 | 0.682 |
| Case V               | 50  | 1.360          | 0.150 | 0.010 | 0.950 | 0.840 | 1.880         | 0.250 | 0.030 | 0.850 | 0.782 |
|                      | 80  | 1.355          | 0.110 | 0.005 | 0.700 | 0.820 | 1.873         | 0.180 | 0.023 | 0.640 | 0.771 |
|                      | 100 | 1.352          | 0.085 | 0.002 | 0.600 | 0.800 | 1.865         | 0.140 | 0.015 | 0.550 | 0.758 |
|                      | 150 | 1.349          | 0.045 | 0.001 | 0.450 | 0.810 | 1.857         | 0.070 | 0.007 | 0.400 | 0.773 |

We varied the proportions of left- and right-censored observations to represent different event-time patterns under current status data. This variation directly influences the estimation process and the quality of the estimators. When the proportion of censored observations decreases, meaning that more complete data points are available, the estimation becomes more straightforward and the resulting estimates more reliable. However, despite the potential benefits of having less censoring, there are certain consequences that can affect the accuracy of the estimators. As the level of censoring decreases, the Mean Square Error (MSE) and Absolute Bias (AB) values for the parameters tend to increase. This is because the presence of censoring introduces additional uncertainty into the estimation process. Censored observations contain partial information about the underlying distribution, which can lead to less precise parameter estimates compared to complete observations. Therefore, when there are fewer censored observations in the dataset, the estimators may not accurately capture the true characteristics of the population, resulting in higher MSE and AB values. Furthermore, the decrease in the length of the confidence interval (LT) of the parameter suggests that our level of uncertainty about the estimated parameter reduces as the censoring decreases. A shorter confidence interval implies that we are more confident about the estimated value of the parameter. This is a result of the increased precision in the estimation process due to the reducing censoring. Moreover, the increase in coverage probability (CP) indicates that the proportion of confidence intervals that contain the true value of the parameter improves with reducing censoring. This improvement in coverage probability further reinforces the reliability and accuracy of the estimation process with reducing censoring. Additionally, the rise in MSE and AB values further supports this observation, indicating a decrease in the accuracy and consistency of the estimators as the level of censoring decreases. Estimating parameters from data with minimal censoring becomes more challenging due to the limited information available, leading to larger discrepancies between the estimated and true parameter values.

## RESULTS

In this section, we focus on the duration of PPA, collected by National Family Health Survey-V<sup>1</sup>(NFHS-v) 2020-21 for four states – Uttar Pradesh, West Bengal, Maharashtra, and Kerala (Government of India, 2021).<sup>5</sup> The original dataset from NFHS-V contains a substantial amount of information, encompassing a wide range of data points. For the analysis, a subset of this data is utilized by considering the four states.

In the NFHS-V dataset, there is question asked about current status of PPA after most recent birth. Those women who had not experienced the event at time of interview are considered as right censored observations. On the other hand, the women who had experienced PPA are considered as left censored observations. Hence, the data satisfies of basic assumptions our analysis framework. The above method is applied to the data available from NFHS data. Only those women whose difference of date of interview and date of last birth i.e., open birth interval (OBI) is less than 35 months and status of PPA is noted are considered.

**Table 3.** Descriptive summary of postpartum amenorrhea (PPA) duration (months) in four Indian states, showing quartiles (Q1, median, Q3), mean, and standard deviation (SD) based on current status data.

| States        | Quartile(Q1) | Quartile(Q2)/Median | Quartile(Q3) | Mean | SD   |
|---------------|--------------|---------------------|--------------|------|------|
| Uttar Pradesh | 1            | 3                   | 8            | 5.24 | 2.25 |
| West Bengal   | 1            | 3                   | 8            | 4.00 | 2.03 |
| Maharashtra   | 1            | 5                   | 9            | 6.84 | 2.46 |
| Kerala        | 1            | 4                   | 9            | 5.18 | 2.55 |

The Table 3 provides an overview of the duration of PPA across four states: Uttar Pradesh, West Bengal, Maharashtra, and Kerala. In all states, the duration of PPA ranges from 1 to 9 months, with median values (Q2) spanning from 3 to 5 months. The mean duration exhibits slight variations across states, with Uttar Pradesh having the longest average duration of PPA 5.24 months, followed by Kerala 5.18 months, Maharashtra 6.84 months, and West Bengal 4.00 months. The standard deviations (S.D.) also vary, indicating the extent of variability around the mean duration within each state.

For the algorithm implementation process, we utilized the R software. The initial value  $\theta_0$  was assigned based on the result obtained by the least squares method. The iterative process stopped when the difference between the  $k$ th and  $(k + 1)$ th iteration, as defined by Equation (12), was less than 0.001. The estimated weibull distribution parameters, chi-square ( $\chi^2$ ) values, and degrees of freedom are presented in Tables A3.1, A3.2, A3.3, and A3.4. Among these, Table 4 provides parameter estimates for the weibull distribution in four states. Notably, the mean duration of PPA, as determined by the weibull distribution, closely approximates the mean duration obtained directly from raw data. In assessing the goodness of fit between the observed data and the expected observations based on the weibull distribution, employ the  $\chi^2$  statistics.

Furthermore, our analysis revealed that the difference between the expected and observed mean duration of PPA is statistically insignificant in all four states. This robust agreement underscores the suitability of the weibull distribution for modelling the duration of PPA. To enhance the reliability of our estimates, initial parameters for the EM algorithm calculations are derived from the parameters estimated by<sup>2</sup> using the least square method.

**Table 4.** Estimated Weibull parameters ( $\alpha$ : shape,  $\beta$ : scale) for postpartum amenorrhea (PPA, months) across selected Indian states using LS and EM methods, with 95% CIs (EM) and corresponding mean and standard deviation (SD) for comparison

| States        | Least Square Method |               |      |      | E.M Method     |               |               |               |       |       |
|---------------|---------------------|---------------|------|------|----------------|---------------|---------------|---------------|-------|-------|
|               | Shape $\alpha$      | Scale $\beta$ | Mean | SD   | Shape $\alpha$ | CI            | scale $\beta$ | CI            | Mean  | SD    |
| Uttar Pradesh | 0.604               | 0.43          | 6.02 | 2.14 | 0.624          | (0.597,0.677) | 0.412         | (0.390,0.456) | 5.694 | 2.28  |
| West Bengal   | 0.715               | 0.346         | 5.46 | 2.54 | 0.750          | (0.721,0.781) | 0.388         | (0.343,0.430) | 4.19  | 2.95  |
| Maharashtra   | 0.924               | 0.16          | 6.82 | 2.11 | 0.904          | (0.880,0.998) | 0.203         | (0.145,0.262) | 6.11  | 2.34  |
| Kerala        | 1.017               | 0.161         | 5.94 | 2.04 | 1.012          | (0.975,1.078) | 0.205         | (0.189,0.224) | 4.38  | 2.177 |

Additionally, Table 5 presents the notation and terminology used throughout the subsequent tables

to facilitate interpretation of the results. Tables 6, 7, 8, and 9 showcase the empirical and fitted distributions of the duration of postpartum amenorrhoea (PPA) in four states, as presented in Appendix

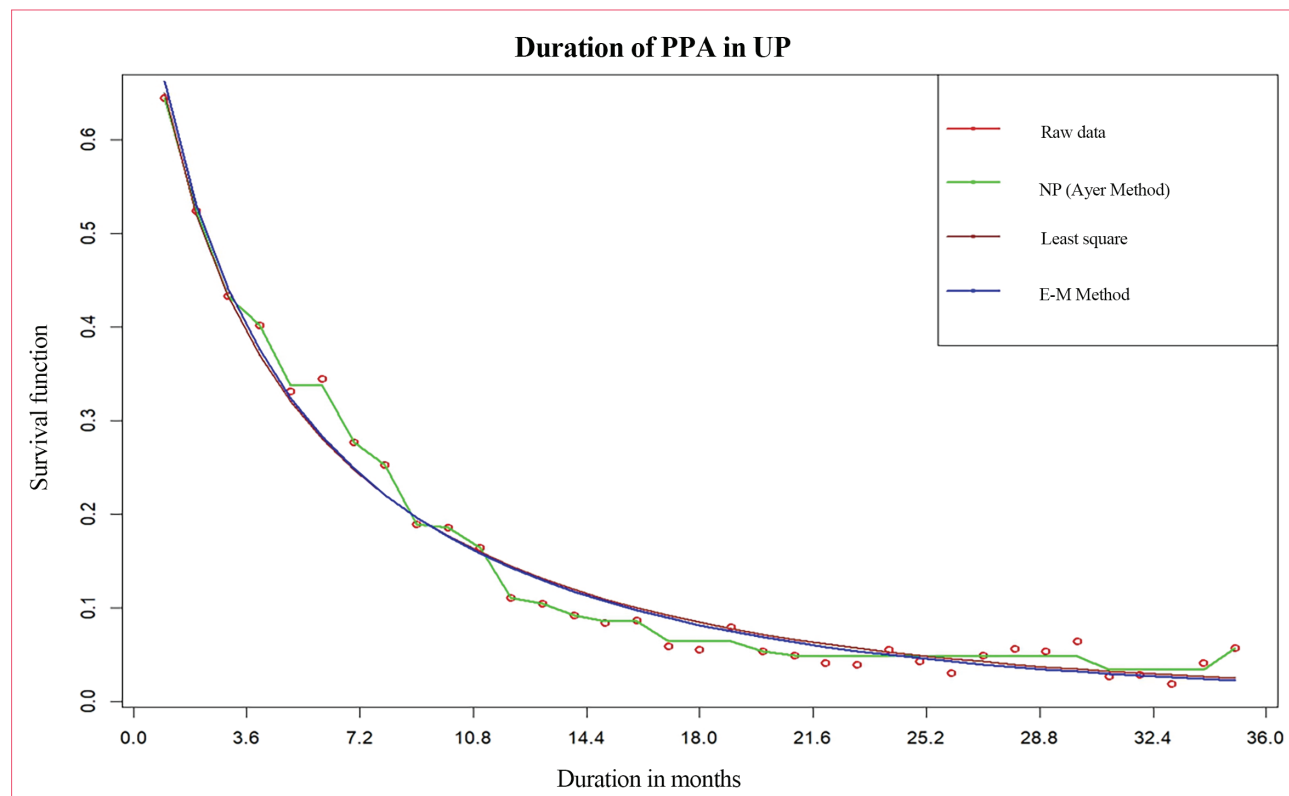


Figure 1. Estimates of survival function duration of PPA by EM method and Non-Parametric method in UP

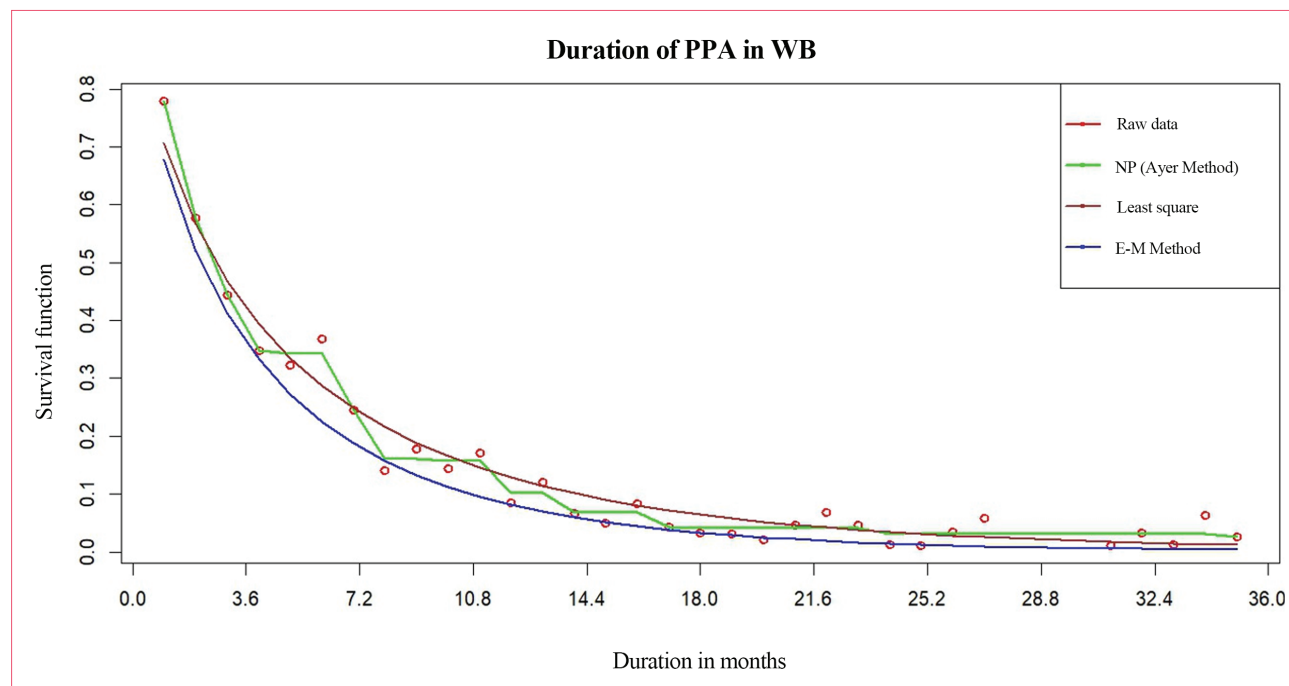


Figure 2. Estimates of survival function duration of PPA by EM method and Non-Parametric method in West- Bengal

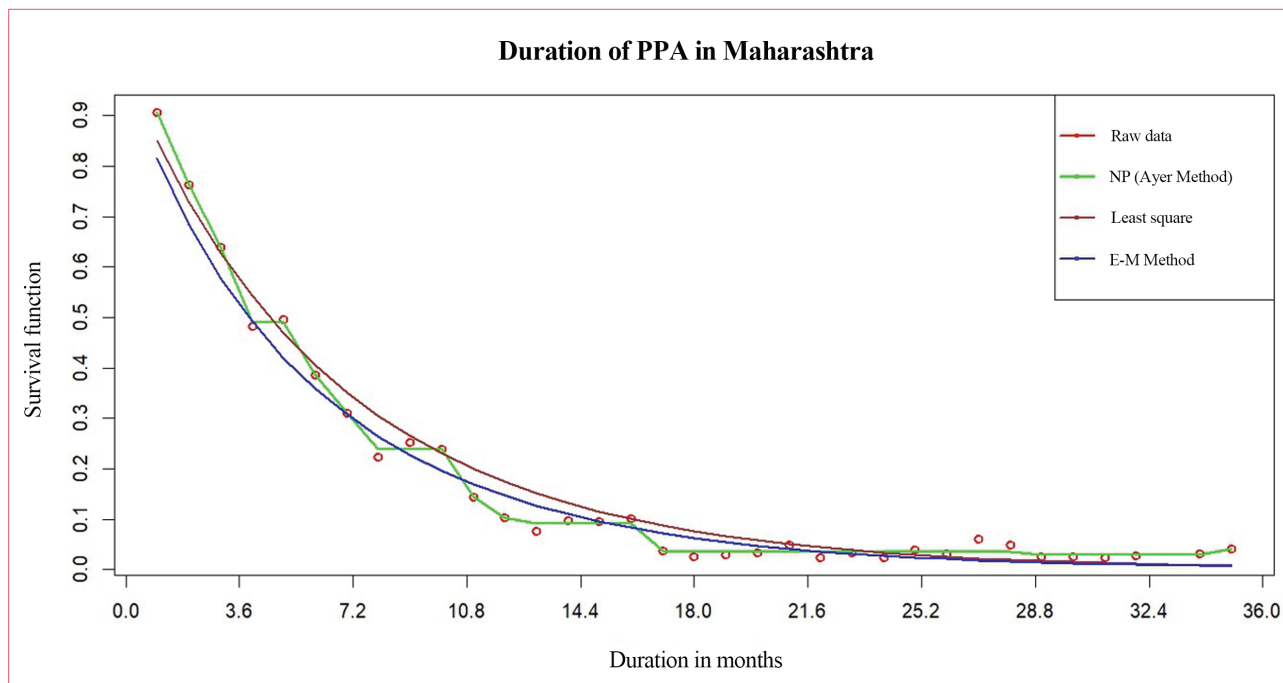


Figure 3. Estimates of survival function duration of PPA by EM method and Non-Parametric method in Maharashtra

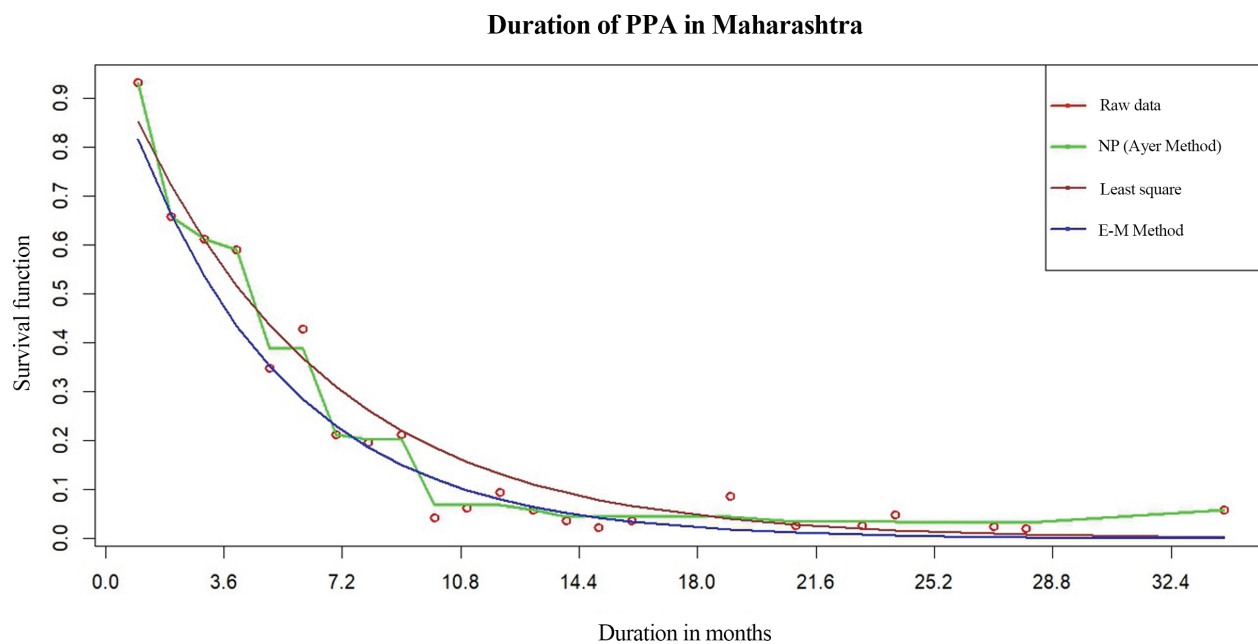


Figure 4. Estimates of survival function duration of PPA by EM method and Ayer method in Kerala

II. Furthermore, Figures 1, 2, 3, and 4 provide the survival curves for Uttar Pradesh, West Bengal, Maharashtra, and Kerala, respectively, using the various methods described above. These visual representations offer valuable insights into the survival characteristics of PPA durations across the four states and facilitate comparison of the performance of different estimation methods.

## DISCUSSION

This study conducted a thorough analysis of Postpartum Amenorrhea (PPA) duration in selected Indian states, This study conducted a thorough analysis of Postpartum Amenorrhea (PPA) duration in selected Indian states, employing a Weibull distribution model based on current status data from the NFHS-V (2020-21) survey. The analysis confirmed the suitability of the Weibull distribution in accurately capturing the distribution of PPA duration across most states.

The research contributes to existing knowledge by documenting and applying a parametric approach to estimate the distribution of PPA duration. Specific estimations, such as the average duration of PPA in Uttar Pradesh (5.24 months) with a standard deviation of 2.25 months, provide valuable insights into the variability of PPA duration across different states.<sup>30</sup> These findings, detailed in Table 4, offer a deeper understanding of PPA duration dynamics and its variations among Indian states.

Additionally, the study highlights the potential of Bayesian analysis as a promising direction for refining parameter estimations of the Weibull distribution. This advanced methodological approach could enhance the precision and depth of our understanding of PPA duration, particularly in the context of current status data.

Table 5. Notation and Terminology used in the table

|          |                                                                                             |
|----------|---------------------------------------------------------------------------------------------|
| $n_j$    | the number of women having OBI in the interval $(t_{j-p}, t_j]$ months                      |
| $y_j$    | the number of women with OBI in the interval $(t_{j-p}, t_j]$                               |
|          | months who is currently amenorrhoea (i.e., the duration of PPA than $t_j$ months)           |
| $t_j$    | the Open Birth Interval                                                                     |
| $S(t_j)$ | Proportion of women with duration of PPA greater than $t_j$ months for $j = 1, 2, \dots, k$ |

## CONCLUSION

In summary, this study provides a comprehensive assessment of the distribution of PPA duration across Indian states, demonstrating the applicability of the Weibull model for analyzing current status data. The EM algorithm showed superior precision and accuracy compared to the Least Squares (LS) method, as validated against the non-parametric Ayer estimator, which served as the benchmark. These findings highlight the robustness of the EM-based parametric approach for estimating event-time distributions under incomplete observation schemes.

From a public health perspective, the study emphasizes notable state-level differences in PPA duration, underscoring the need for region-specific reproductive health strategies. Future research should focus on identifying the socio-demographic and biological factors influencing both short and prolonged durations of PPA, thereby contributing to a deeper understanding of postpartum health and fertility dynamics across India.

## Conflict of Interest

The authors declare that there is no conflict of interest.

## ACKNOWLEDGMENT

The authors sincerely acknowledge the guidance, encouragement, and valuable support provided by Dr. Anup Kumar, Department of Biostatistics, Sanjay Gandhi Postgraduate Institute of Medical Sciences (SGPGIMS), Lucknow, throughout the course of this research. The authors are also grateful to Dr. Amit Kumar Misra for his valuable suggestions, critical inputs, and support during the development of this work.

The authors extend their appreciation to the Department of Biostatistics, SGPGIMS, Lucknow, for providing the academic environment and resources that facilitated this research. The authors also acknowledge the Demographic and Health Surveys (DHS) Program for providing access to the National Family Health Survey (NFHS) data used in this study.

The authors thank all colleagues and individuals who contributed directly or indirectly to the successful completion of this research.

## REFERENCES

1. Tika Ram Aryal. Estimation of the duration of post-partum amenorrhea among nepalese mothers: some indirect techniques. *Nepal Journal of Science and Technology*, 10:213–218, 2009.
2. Tika Ram Aryal. Weibull distribution to describe the patterns of the duration of post-partum amenorrhea. 31(1), 2011.
3. TR Aryal. Retrospective reporting of the duration of post-partum amenorrhea: a survival analysis. *Kath- mandu University medical journal*, 4(2):211–217, 2006.
4. Elodie Brunel and Fabienne Comte. Cumulative distribution function estimation under interval censoring case 1. 2009.
5. Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977.
6. David B Dunson and Donna D Baird. A flexible parametric model for combining current status and age at first diagnosis data. *Biometrics*, 57(2):396–403, 2001.
7. Sam Efromovich. Efficient nonparametric estimation of distribution for current status censoring.

- Electronic Journal of Statistics, 16(1):998–1057, 2022.
8. Tsai-Hung Fan, Yi-Fu Wang, and She-Kai Ju. A competing risks model with multiply censored reliability data under multivariate weibull distributions. *IEEE Transactions on Reliability*, 68(2):462–475, 2019.
  9. Luís Andrade FerreirA and José Luís SiLvA. Parameter estimation for weibull distribution with right cen- sored data using em algorithm. *Eksploatacja i Niezawodnos´c´*, 19(2):310–315, 2017.
  10. Kathleen Ford and Young Kim. Distributions of postpartum amenorrhea: some new evidence. *Demography*, pages 413–430, 1987.
  11. Antonio Eduardo Gomes and Cibele Queiroz DA-SILVA. Nonparametric maximum likelihood estimation for current status data with misclassification. *Rev. Bras. Biom*, 29(1):102–120, 2011.
  12. Piet Groeneboom and Kim Hendrickx. The nonparametric bootstrap for the current status model. 2017.
  13. International Institute for Population Sciences (IIPS) and ICF.National Family Health Survey (NFHS-5). [https://dhsprogram.com/publications/ publication-fr347-dhs-final-reports.cfm](https://dhsprogram.com/publications/publication-fr347-dhs-final-reports.cfm), 2020. Accessed: [date accessed].
  14. Nicholas P Jewell and Mark van der Laan. Current status data: Review, recent developments and open problems. *Handbook of statistics*, 23:625–642, 2003.
  15. Mimi Y Kim, Victor G De Gruttola, and Stephen W Lagakos. Analyzing doubly censored data with covari- ates, with application to aids. *Biometrics*, pages 13–22, 1993.
  16. Tamalika Koley and Anup Dewanji. Current status data with two competing risks and missing failure types: a parametric approach. *Journal of Applied Statistics*, 49(7):1769–1783, 2022.
  17. MD Krailo and MC Pike. Estimation of the distribution of age at natural menopause from prevalence data. *American journal of epidemiology*, 117(3):356–361, 1983.
  18. Anup Kumar, Abhishek Bharti, and RC Yadava. An adjustment of truncation and selection effect for estimating conception rate from first birth interval data. *Journal of Data Science*, 17(4):756–767, 2019.
  19. Anup Kumar, Sachin Kumar, Jai Kishun, Uttam Singh, and Pushpraj Pushpraj. Estimation of the distribution of duration of breastfeeding from cross-sectional data: Some methodological issue. *Journal of Biostatistics and Epidemiology*, 6(4):290–304, 2020.
  20. Thomas A Louis. Finding the observed information matrix when using the em algorithm. *Journal*

- of the Royal Statistical Society Series B: Statistical Methodology, 44(2):226–233, 1982.
21. Ruchi Mishra, Kaushalendra K Singh, and P Singh Brijesh. Duration of post-partum amenorrhoea: A model- based approach. *Indian Journal of Population and Development*, 1(1):41–50, 2021.
  22. Dilip C Nath, Atanu Bhattacharjee, and HB Singh. An estimation of PPA duration for manipuri women: A bayesian approach in zero-inflated longitudinal data analysis.
  23. MS Panwar, Vikas Barnwal, and CP Yadav. A latent variable approach for modeling recall-based time-to- event data with weibull distribution. *Computational Statistics*, pages 1–32, 2024.
  24. Sedigheh Mirzaei Salehabadi, Debasis Sengupta, and Rituparna Das. Parametric estimation of menarcheal age distribution based on recall data. *Scandinavian Journal of Statistics*, 42(1):290–305, 2015.
  25. Girish Singh. Estimation of the duration of postpartum amenorrhea in the presence of some censored data. *The Journal of Obstetrics and Gynecology of India*, 57(1):53–57, 2007.
  26. Pragya Singh, Kaushalendra K Singh, and Brijesh P Singh. An EM algorithm approach to estimate parameters of fluctuating nature of PPA. *Indian Journal Population and Development*, 1(2):285–294, 2021.
  27. SN Singh, BN Bhattacharya, and RC Yadava. An adjustment of a selection bias in postpartum amenorrhea from follow-up studies. *Journal of the American Statistical Association*, 74(368):916–920, 1979.
  28. Jianguo Sun and John D. Kalbfleisch. The analysis of current status data on point processes. *Journal of the American Statistical Association*, 88(424):1449–1454, 1993.
  29. Zhibin Tan. A new approach to mle of weibull distribution with interval data. *Reliability Engineering & System Safety*, 94(2):394–403, 2009.
  30. Bert van Es and Catharina Elisabeth Graafland. Nonparametric kernel density estimation for univariate current status data. *arXiv preprint arXiv:1707.00544*, 2017.
  31. Kumar, S., Kumar, A., Misra, A. K., Kishun, J., & Singh, U. (2024). Adjustment of Truncation Effect in First Birth Interval using Current Status Data Technique. *Journal of Biostatistics and Epidemiology*.
  32. Qian, C., Srivastava, D. K., Pan, J., Hudson, M. M., & Rai, S. N. (2024). Estimating transition intensity rate on interval-censored data using semi-parametric with EM algorithm approach. *Communications in Statistics-Theory and Methods*, 53(17), 6038-6054.

## Supplementary

Table A3.1–A3.4 present the observed and expected distributions of postpartum amenorrhea (PPA) duration (in months) among women from four Indian states, estimated from current status data using three approaches: Ayer's Nonparametric Maximum Likelihood Estimation (NPMLE), Least Squares (L.S.), and Expectation–Maximization (EM). Each table includes the observed number of events ( $y_j$ ), total examined women ( $n_j$ ), empirical survival estimates  $S(t_j)$ , and fitted survival probabilities  $\hat{S}(t_j)$  with corresponding expected frequencies  $E(y_j)$  under each method. Estimated Weibull parameters ( $\hat{\alpha}$ ,  $\hat{\beta}$ ) and  $\chi^2$  goodness-of-fit statistics are also reported for model comparison.

**Table A3.1.** Observed and expected distribution of postpartum amenorrhea (PPA, months) among women in Uttar Pradesh using Ayer's NPMLE, L.S., and EM methods, with fitted survival probabilities, expected frequencies, and Weibull parameter estimates.

| $t_j$ | $y_j$ | $n_j$ | $S(t_j)$ | Ayer Method    | L.S Method     | EM Method |                |          |
|-------|-------|-------|----------|----------------|----------------|-----------|----------------|----------|
|       |       |       |          | $\hat{S}(t_j)$ | $\hat{S}(t_j)$ | $E(y_j)$  | $\hat{S}(t_j)$ | $E(y_j)$ |
| 1     | 422   | 655   | 0.644    | 0.644          | 0.650          | 425.743   | 0.662          | 433.742  |
| 2     | 359   | 685   | 0.524    | 0.524          | 0.519          | 355.810   | 0.530          | 362.805  |
| 3     | 293   | 677   | 0.433    | 0.433          | 0.433          | 293.156   | 0.441          | 298.541  |
| 4     | 297   | 739   | 0.402    | 0.402          | 0.369          | 272.967   | 0.375          | 277.364  |
| 5     | 242   | 732   | 0.331    | 0.337          | 0.320          | 234.159   | 0.324          | 237.276  |
| 6     | 236   | 685   | 0.345    | 0.337          | 0.280          | 191.872   | 0.283          | 193.825  |
| 7     | 160   | 579   | 0.276    | 0.276          | 0.247          | 143.225   | 0.249          | 144.202  |
| 8     | 127   | 503   | 0.252    | 0.252          | 0.220          | 110.639   | 0.221          | 111.005  |
| 9     | 93    | 492   | 0.189    | 0.189          | 0.197          | 96.776    | 0.197          | 96.746   |
| 10    | 86    | 464   | 0.185    | 0.185          | 0.177          | 82.012    | 0.176          | 81.682   |
| 11    | 78    | 475   | 0.164    | 0.164          | 0.159          | 75.756    | 0.158          | 75.166   |
| 12    | 65    | 586   | 0.111    | 0.111          | 0.144          | 84.637    | 0.143          | 83.656   |
| 13    | 61    | 582   | 0.105    | 0.105          | 0.131          | 76.370    | 0.129          | 75.193   |
| 14    | 56    | 612   | 0.092    | 0.092          | 0.120          | 73.171    | 0.117          | 71.762   |
| 15    | 49    | 582   | 0.084    | 0.086          | 0.109          | 63.566    | 0.107          | 62.096   |
| 16    | 53    | 610   | 0.087    | 0.086          | 0.100          | 61.006    | 0.097          | 59.359   |
| 17    | 35    | 598   | 0.059    | 0.064          | 0.092          | 54.880    | 0.089          | 53.187   |
| 18    | 32    | 583   | 0.055    | 0.064          | 0.084          | 49.194    | 0.081          | 47.486   |
| 19    | 43    | 541   | 0.079    | 0.064          | 0.078          | 42.050    | 0.075          | 40.428   |
| 20    | 24    | 451   | 0.053    | 0.053          | 0.072          | 32.345    | 0.069          | 30.973   |
| 21    | 19    | 387   | 0.049    | 0.049          | 0.066          | 25.650    | 0.063          | 24.463   |
| 22    | 16    | 390   | 0.041    | 0.048          | 0.061          | 23.923    | 0.058          | 22.725   |
| 23    | 16    | 408   | 0.039    | 0.048          | 0.057          | 23.195    | 0.054          | 21.945   |
| 24    | 25    | 450   | 0.056    | 0.048          | 0.053          | 23.741    | 0.050          | 22.371   |
| 25    | 19    | 442   | 0.043    | 0.048          | 0.049          | 21.666    | 0.046          | 20.334   |
| 26    | 13    | 432   | 0.030    | 0.048          | 0.046          | 19.698    | 0.043          | 18.413   |
| 27    | 23    | 471   | 0.049    | 0.048          | 0.042          | 19.999    | 0.040          | 18.619   |
| 28    | 27    | 479   | 0.056    | 0.048          | 0.040          | 18.960    | 0.037          | 17.581   |
| 29    | 25    | 465   | 0.054    | 0.048          | 0.037          | 17.174    | 0.034          | 15.861   |
| 30    | 27    | 422   | 0.064    | 0.048          | 0.034          | 14.557    | 0.032          | 13.390   |

Table A3.1 (continued).

| $t_j$ | $y_j$ | $n_j$ | $S(t_j)$ | Ayer Method    | L.S Method     |                | EM Method      |          |
|-------|-------|-------|----------|----------------|----------------|----------------|----------------|----------|
|       |       |       |          | $\hat{S}(t_j)$ | $\hat{S}(t_j)$ | $E(y_j)$       | $\hat{S}(t_j)$ | $E(y_j)$ |
| 31    | 12    | 341   | 0.027    | 0.034          | 0.032          | 14.221         | 0.030          | 13.029   |
| 32    | 10    | 352   | 0.028    | 0.034          | 0.030          | 10.620         | 0.028          | 9.691    |
| 33    | 6     | 321   | 0.019    | 0.034          | 0.028          | 9.069          | 0.026          | 8.242    |
| 34    | 13    | 320   | 0.041    | 0.034          | 0.026          | 8.472          | 0.024          | 7.669    |
| 35    | 19    | 331   | 0.057    | 0.057          | 0.025          | 8.218          | 0.022          | 7.410    |
| Total | 3081  | 17942 |          |                |                | 3078.496       |                | 3078.238 |
|       |       |       |          |                | $\chi^2$       | 44.606         |                | 42.472   |
|       |       |       |          |                | df             | 34.000         |                | 34.000   |
|       |       |       |          |                | Parameter      | $\hat{\alpha}$ | 0.604          | 0.624    |
|       |       |       |          |                | "              | $\hat{\beta}$  | 0.430          | 0.422    |

$X^2_{(0.05,34)} = 48.60$

**Table A3.2.** Observed and expected distribution of postpartum amenorrhea (PPA, months) among women in West Bengal using Ayer's NPMLE, L.S., and EM methods, with fitted survival probabilities, expected frequencies, and Weibull parameter estimates.

| $t_j$                     | $y_j$ | $n_j$ | $S(t_j)$ | Ayer Method    | L.S Method     |          | EM Method      |          |
|---------------------------|-------|-------|----------|----------------|----------------|----------|----------------|----------|
|                           |       |       |          | $\hat{S}(t_j)$ | $\hat{S}(t_j)$ | $E(y_j)$ | $\hat{S}(t_j)$ | $E(y_j)$ |
| 1                         | 74    | 95    | 0.779    | 0.779          | 0.707          | 67.195   | 0.678          | 64.439   |
| 2                         | 56    | 97    | 0.577    | 0.577          | 0.566          | 54.935   | 0.52           | 50.482   |
| 3                         | 36    | 81    | 0.444    | 0.444          | 0.468          | 37.885   | 0.413          | 33.415   |
| 4                         | 24    | 69    | 0.348    | 0.348          | 0.393          | 27.128   | 0.333          | 22.994   |
| 5                         | 31    | 96    | 0.323    | 0.343          | 0.335          | 32.113   | 0.273          | 26.182   |
| 6                         | 28    | 76    | 0.368    | 0.343          | 0.287          | 21.825   | 0.225          | 17.131   |
| 7                         | 22    | 90    | 0.244    | 0.244          | 0.248          | 22.347   | 0.188          | 16.898   |
| 8                         | 15    | 106   | 0.142    | 0.161          | 0.216          | 22.89    | 0.157          | 16.684   |
| 9                         | 22    | 124   | 0.177    | 0.161          | 0.189          | 23.401   | 0.133          | 16.451   |
| 10                        | 14    | 97    | 0.144    | 0.157          | 0.166          | 16.066   | 0.112          | 10.898   |
| 11                        | 15    | 88    | 0.17     | 0.157          | 0.146          | 12.838   | 0.096          | 8.407    |
| 12                        | 8     | 94    | 0.085    | 0.103          | 0.129          | 12.119   | 0.082          | 7.664    |
| 13                        | 11    | 91    | 0.121    | 0.103          | 0.114          | 10.398   | 0.07           | 6.353    |
| 14                        | 6     | 89    | 0.067    | 0.068          | 0.102          | 9.037    | 0.06           | 5.335    |
| 15                        | 4     | 80    | 0.05     | 0.068          | 0.09           | 7.235    | 0.052          | 4.129    |
| 16                        | 8     | 96    | 0.083    | 0.068          | 0.081          | 7.75     | 0.045          | 4.277    |
| 17                        | 3     | 71    | 0.042    | 0.042          | 0.072          | 5.127    | 0.039          | 2.737    |
| 18                        | 3     | 91    | 0.033    | 0.042          | 0.065          | 5.889    | 0.033          | 3.041    |
| 19                        | 3     | 97    | 0.031    | 0.042          | 0.058          | 5.634    | 0.029          | 2.816    |
| 20                        | 2     | 92    | 0.022    | 0.042          | 0.052          | 4.805    | 0.025          | 2.324    |
| 21                        | 4     | 87    | 0.046    | 0.042          | 0.047          | 4.091    | 0.022          | 1.916    |
| 22                        | 7     | 102   | 0.069    | 0.042          | 0.042          | 4.325    | 0.019          | 1.962    |
| 23                        | 5     | 108   | 0.046    | 0.042          | 0.038          | 4.135    | 0.017          | 1.817    |
| 24                        | 1     | 78    | 0.013    | 0.031          | 0.035          | 2.7      | 0.015          | 1.149    |
| 25                        | 1     | 88    | 0.011    | 0.031          | 0.031          | 2.757    | 0.013          | 1.137    |
| 26                        | 3     | 87    | 0.034    | 0.031          | 0.028          | 2.47     | 0.011          | 0.987    |
| 27                        | 5     | 87    | 0.057    | 0.031          | 0.026          | 2.241    | 0.01           | 0.868    |
| 31                        | 1     | 89    | 0.011    | 0.031          | 0.018          | 1.567    | 0.006          | 0.537    |
| 32                        | 3     | 90    | 0.033    | 0.031          | 0.016          | 1.445    | 0.005          | 0.48     |
| 33                        | 1     | 82    | 0.012    | 0.031          | 0.015          | 1.2      | 0.005          | 0.387    |
| 34                        | 7     | 111   | 0.063    | 0.031          | 0.013          | 1.483    | 0.004          | 0.464    |
| 35                        | 2     | 78    | 0.026    | 0.026          | 0.012          | 0.952    | 0.004          | 0.289    |
| Total                     | 425   | 2907  |          |                |                | 435.983  |                | 334.65   |
|                           |       |       |          |                | $\chi^2$       | 17.156   |                | 18.104   |
|                           |       |       |          |                | df             | 31       |                | 31       |
| Parameter                 |       |       |          |                | $\hat{\alpha}$ | 0.715    |                | 0.750    |
| "                         |       |       |          |                | $\hat{\beta}$  | 0.346    |                | 0.388    |
| $X^2_{(0.05,34)} = 43.77$ |       |       |          |                |                |          |                |          |

**Table A3.3.** Observed and expected distribution of postpartum amenorrhea (PPA, months) among women in Maharashtra using Ayer's NPMLE, L.S., and EM methods, with fitted survival probabilities, expected frequencies, and Weibull parameter estimates.

| $t_j$                     | $y_j$ | $n_j$ | $S(t_j)$ | Ayer Method    | L.S Method     |          | EM Method      |          |
|---------------------------|-------|-------|----------|----------------|----------------|----------|----------------|----------|
|                           |       |       |          | $\hat{S}(t_j)$ | $\hat{S}(t_j)$ | $E(y_j)$ | $\hat{S}(t_j)$ | $E(y_j)$ |
| 1                         | 135   | 149   | 0.906    | 0.906          | 0.849          | 126.567  | 0.816          | 121.625  |
| 2                         | 93    | 122   | 0.762    | 0.762          | 0.729          | 88.931   | 0.684          | 83.443   |
| 3                         | 83    | 130   | 0.638    | 0.638          | 0.628          | 81.614   | 0.578          | 75.150   |
| 4                         | 55    | 114   | 0.482    | 0.490          | 0.542          | 61.781   | 0.491          | 56.002   |
| 5                         | 67    | 135   | 0.496    | 0.490          | 0.469          | 63.263   | 0.419          | 56.576   |
| 6                         | 67    | 174   | 0.385    | 0.385          | 0.406          | 70.600   | 0.359          | 62.398   |
| 7                         | 44    | 142   | 0.310    | 0.310          | 0.352          | 49.942   | 0.308          | 43.683   |
| 8                         | 30    | 135   | 0.222    | 0.239          | 0.305          | 41.194   | 0.264          | 35.700   |
| 9                         | 40    | 159   | 0.252    | 0.239          | 0.265          | 42.129   | 0.228          | 36.210   |
| 10                        | 45    | 188   | 0.239    | 0.239          | 0.230          | 43.284   | 0.196          | 36.930   |
| 11                        | 24    | 167   | 0.144    | 0.144          | 0.200          | 33.431   | 0.170          | 28.337   |
| 12                        | 17    | 165   | 0.103    | 0.103          | 0.174          | 28.737   | 0.147          | 24.214   |
| 13                        | 12    | 158   | 0.076    | 0.092          | 0.152          | 23.953   | 0.127          | 20.077   |
| 14                        | 14    | 145   | 0.097    | 0.092          | 0.132          | 19.144   | 0.110          | 15.971   |
| 15                        | 14    | 148   | 0.095    | 0.092          | 0.115          | 17.025   | 0.096          | 14.144   |
| 16                        | 15    | 148   | 0.101    | 0.092          | 0.100          | 14.840   | 0.083          | 12.283   |
| 17                        | 6     | 159   | 0.038    | 0.038          | 0.087          | 13.902   | 0.072          | 11.470   |
| 18                        | 3     | 120   | 0.025    | 0.036          | 0.076          | 9.152    | 0.063          | 7.530    |
| 19                        | 4     | 134   | 0.030    | 0.036          | 0.067          | 8.918    | 0.055          | 7.320    |
| 20                        | 4     | 120   | 0.033    | 0.036          | 0.058          | 6.971    | 0.048          | 5.710    |
| 21                        | 7     | 145   | 0.048    | 0.036          | 0.051          | 7.355    | 0.041          | 6.014    |
| 22                        | 4     | 164   | 0.024    | 0.036          | 0.044          | 7.266    | 0.036          | 5.933    |
| 23                        | 5     | 148   | 0.034    | 0.036          | 0.039          | 5.729    | 0.032          | 4.673    |
| 24                        | 4     | 167   | 0.024    | 0.036          | 0.034          | 5.649    | 0.028          | 4.605    |
| 25                        | 6     | 154   | 0.039    | 0.036          | 0.030          | 4.554    | 0.024          | 3.710    |
| 26                        | 4     | 126   | 0.032    | 0.036          | 0.026          | 3.258    | 0.021          | 2.654    |
| 27                        | 7     | 116   | 0.060    | 0.036          | 0.023          | 2.623    | 0.018          | 2.137    |
| 28                        | 6     | 122   | 0.049    | 0.036          | 0.020          | 2.413    | 0.016          | 1.966    |
| 29                        | 3     | 114   | 0.026    | 0.029          | 0.017          | 1.973    | 0.014          | 1.608    |
| 30                        | 3     | 117   | 0.026    | 0.029          | 0.015          | 1.772    | 0.012          | 1.446    |
| 31                        | 3     | 125   | 0.024    | 0.029          | 0.013          | 1.657    | 0.011          | 1.353    |
| 32                        | 3     | 110   | 0.027    | 0.029          | 0.012          | 1.277    | 0.009          | 1.044    |
| 34                        | 4     | 127   | 0.031    | 0.029          | 0.009          | 1.130    | 0.007          | 0.927    |
| 35                        | 5     | 119   | 0.042    | 0.042          | 0.008          | 0.928    | 0.006          | 0.762    |
| Total                     | 836   | 4766  |          |                |                | 892.966  |                | 793.607  |
|                           |       |       |          |                | $\chi^2$       | 31.040   |                | 25.560   |
|                           |       |       |          |                | df             | 33.000   |                | 33.000   |
| Parameter                 |       |       |          |                | $\hat{\alpha}$ | 0.924    |                | 0.904    |
| "                         |       |       |          |                | $\hat{\beta}$  | 0.160    |                | 0.203    |
| $X^2_{(0.05,34)} = 46.19$ |       |       |          |                |                |          |                |          |

Table A3.4. Observed and expected distribution of postpartum amenorrhea (PPA, months) among women in Kerala using Ayer's NPMLE, L.S., and EM methods, with fitted survival probabilities, expected frequencies, and Weibull parameter estimates.

| $t_j$     | $y_j$ | $n_j$ | $S(t_j)$ | Ayer Method    | L.S Method     |          | EM Method      |          |
|-----------|-------|-------|----------|----------------|----------------|----------|----------------|----------|
|           |       |       |          | $\hat{S}(t_j)$ | $\hat{S}(t_j)$ | $E(y_j)$ | $\hat{S}(t_j)$ | $E(y_j)$ |
| 1         | 27    | 29    | 0.931    | 0.931          | 0.851          | 24.673   | 0.815          | 23.625   |
| 2         | 25    | 38    | 0.658    | 0.658          | 0.721          | 27.395   | 0.661          | 25.132   |
| 3         | 22    | 36    | 0.611    | 0.611          | 0.610          | 21.957   | 0.536          | 19.305   |
| 4         | 36    | 61    | 0.590    | 0.590          | 0.515          | 31.444   | 0.434          | 26.499   |
| 5         | 15    | 43    | 0.349    | 0.388          | 0.435          | 18.719   | 0.352          | 15.123   |
| 6         | 18    | 42    | 0.429    | 0.388          | 0.367          | 15.432   | 0.285          | 11.952   |
| 7         | 11    | 52    | 0.212    | 0.212          | 0.310          | 16.117   | 0.230          | 11.969   |
| 8         | 9     | 46    | 0.196    | 0.203          | 0.261          | 12.022   | 0.186          | 8.561    |
| 9         | 7     | 33    | 0.212    | 0.203          | 0.220          | 7.269    | 0.150          | 4.964    |
| 10        | 2     | 47    | 0.043    | 0.068          | 0.186          | 8.723    | 0.122          | 5.713    |
| 11        | 3     | 48    | 0.063    | 0.068          | 0.156          | 7.504    | 0.098          | 4.713    |
| 12        | 5     | 53    | 0.094    | 0.068          | 0.132          | 6.977    | 0.079          | 4.203    |
| 13        | 3     | 51    | 0.059    | 0.059          | 0.111          | 5.652    | 0.064          | 3.266    |
| 14        | 2     | 54    | 0.037    | 0.043          | 0.093          | 5.037    | 0.052          | 2.792    |
| 15        | 1     | 45    | 0.022    | 0.043          | 0.078          | 3.532    | 0.042          | 1.878    |
| 16        | 1     | 28    | 0.036    | 0.043          | 0.066          | 1.849    | 0.034          | 0.943    |
| 19        | 3     | 35    | 0.086    | 0.043          | 0.039          | 1.375    | 0.018          | 0.619    |
| 21        | 1     | 38    | 0.026    | 0.034          | 0.028          | 1.055    | 0.012          | 0.437    |
| 23        | 1     | 38    | 0.026    | 0.034          | 0.020          | 0.745    | 0.007          | 0.284    |
| 24        | 2     | 41    | 0.049    | 0.034          | 0.016          | 0.675    | 0.006          | 0.247    |
| 27        | 1     | 41    | 0.024    | 0.032          | 0.010          | 0.400    | 0.003          | 0.129    |
| 28        | 1     | 49    | 0.020    | 0.032          | 0.008          | 0.402    | 0.003          | 0.125    |
| 34        | 2     | 34    | 0.059    | 0.059          | 0.003          | 0.098    | 0.001          | 0.024    |
| Total     | 198   | 982   |          |                |                | 219.053  |                | 172.503  |
|           |       |       |          |                | $\chi^2$       | 15.312   |                | 11.248   |
|           |       |       |          |                | df             | 22       |                | 22.000   |
| Parameter |       |       |          |                | $\hat{\alpha}$ | 1.017    |                | 1.012    |
| "         |       |       |          |                | $\hat{\beta}$  | 0.161    |                | 0.205    |

$$X^2_{(0.05,22)} = 33.92$$