IJSME

International Journal of Science and Mathematics Education

E-ISSN: 3062-7214
P-ISSN: 3062-7206

*Research Article*

# Estimating the Survival Function of Viral Hepatitis Patients in A Sample of Baghdad Hospitals

**Osama Abdulaziz Kadhim Al-Quraishi**

Ministry of Education Iraq **:** ousamastat@gmail.com
* Corresponding Author: Osama Abdulaziz Kadhim Al-Quraishi

**Abstract** Most people are exposed to sudden death as a result of viral hepatitis infection, especially at an advanced age, so we need to know the continuation of the disease until death in a sample of hospitals to know the differences in the services provided by those hospitals. The research aims to estimate a function for the Rayleigh distribution (RD) using several methods (maximum likelihood function, White, s, weighted least square) and choose the best method for estimation using the mean square error (MSE) and compare the survival rates of patients admitted to (Al-Kadhimiya, Al-Karkh, and Yarmouk) hospitals, The results of the research showed in the experimental aspect that the maximum likelihood method for all sample sizes is the best, followed by the weighted least squares method and finally the White method. In the applied aspect, on a sample consisting of (63) patients (liver cirrhosis + liver cancer) for the first quarter of 2025, it was found that their survival rate until death is (33%) for a time average of (2) months and (15) days.

**Keywords :** Hepatitis Virus Patients, Patients, Survival

## 1. Introduction

The estimation process for any community under study is done by taking a sample from that community that represents the same or similar characteristics of the community. The estimation process is one of the basic pillars of statistical inference, and through it the process of inferences about the study community is made on the basis of the results obtained from the sample selected from the community, Estimating the reliability function of censored data requires clarifying some types of data, as the examination and testing of the device or product (machine) requires monitoring and recording the failed units (data) according to the type of data. The first type of censored occurs when the time of the experiment is determined in advance, the failure cases are random, and the experiment ends at the end of the specified fixed time. The second type of censored occurs when the number of failure cases is determined in advance, and the experiment is stopped after failure occurs for all previously determined failure cases, and the time of the experiment is random, But there are some life test experiments in which some units are withdrawn or excluded from the experiment while they are still alive, and here it is not possible to obtain the survival or work times until failure for all the experiment components, An example of this is the treatment of a group of patients with kidney dialysis in a specific hospital. When a specific patient dies, some of the living patients leave the hospital for treatment in another hospital. When the second patient dies, another number of living patients leave the

**Received:** May 15, 2025
**Revised:** May 30, 2025
**Accepted:** June 24, 2025
**Published:** June 26, 2025
**Curr. Ver.:** June 26, 2025

**Copyright:** © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

hospital. Thus, the process continues until a predetermined number of failure cases is recorded, and ($m < n$). This model of censored is called type II progressive censored.

Statistical models play a pivotal role in modeling and statistically analyzing population age data. Among for the best distribution are the minimum number of parameters, the density function, and the most flexible hazard rate. One such distribution is the Rayleigh distribution.

The research problem is summarized in the difference in the quality of the estimated statistical models, some of which are not suitable for use because they do not have the characteristics of good estimators, which leads to a lack of reliability in their predictive or estimation accuracy, and that most people are exposed to sudden death as a result of infection with viral hepatitis, especially in advanced age, so we need to know the persistence of the disease until death in a sample of hospitals to know the differences in the services provided by those hospitals, The aim of the research is to estimate a function for the Rayleigh distribution (RD) using several methods ( Maximum likelihood function , White ›s , Weighted Least square )& Choosing the best estimation method by using the mean squared error ( MSE)& Comparison of survival rates for patients admitted to (Al-Kadhimiya, Al-Karkh, and Al-Yarmouk) hospitals ,

## 2. The theoretical aspect :

### Survival Function :[2] [4]

The survival function is defined as the probability of an organism remaining alive after the passage of time (t) and is symbolized by the symbol s(t)and can be expressed mathematically as follows:

$$S(t) = 1 - F(t) \quad \dots (1)$$

where :

F(t): represents the cumulative density function of the random variable.

(t):represents the time an organism remains alive

### Rayleigh distribution :[5] [7]

The Rayleigh distribution (RD) belongs to the family of continuous distributions, which was first presented by the English physicist Lord Raleigh, The Raleigh distribution (RD) is a special case of the two-parameter Weibull distribution when the shape parameter is equal to (2), The Raleigh distribution has been widely used in reliability theory and survival analysis, as its failure rate is a linear function of time. This distribution plays an important role in real-life applications because it is related to a number of distributions such as the generalized extreme value, weibull, and chi-square distributions.

The probability density function (p.d.f.) of the random variable T is distributed as follows:

$$f(t:\theta) = 2\theta\, t_i\, e^{-\theta {t_i}^2} \quad \dots(2) \quad ; \ \theta : scal\ parameter\ \ , \theta > 0 \,, t > 0$$

The distribution function (cumulative) for the following distribution is in the following form:

$$F(t) = 1 - \ e^{-\theta {t_i}^2} \dots (3)$$

### Statistical description of type II progressive censored data:[3] [10 ]

Assume (n) a random sample that follows the Raleigh distribution $(X_1, \ldots, X_n)$in a life test experiment , When

the first failure occurs $(X_{1:m:n}$ ، $R_1)$ *,Non-failing units from the remaining (n-1) correct units are randomly excluded from the testing process, Here* $(R_1)$ *is a random variable and its value is bounded by* $\{0,1,\ldots,n-m\}$.

the second failure occurs $(X_{2:m:n}$ ، $R_2)$ *, Non-failing units from the remaining* $(n-R_1-2)$ *correct units are randomly excluded from the testing process , Here* $(R_2)$ *is a random variable and its value is bounded by* $\{0,1,\ldots,n-m-R_1\}$.

*the third failure occurs* $(X_{3:m:n}$ ، $R_3)$ *, Non-failing units from the remaining* $(n-R_1-R_2-3)$ *correct units are randomly excluded from the testing process, Here* $(R_3)$ *is a random variable and its value is bounded by* $\{0,1,\ldots,n-m-R_1-R_2\}$.

*And so on until failure (m-1th)*

*where:*

$R_i: Random\ variables\ with\ a\ limited\ range\ of \{0,1,\ldots,n-m-\sum_{j=1}^{i-1} R_j\}$ ; ، $i = 1,2,\ldots,m-1$

*Finally, when failure m is observed,* $R_m$*equals* $n-m-\sum_{j=1}^{m-1} R_j$*, This means that all remaining healthy units after the (m th) failure are removed from the experiment; therefore, the removable set of* $R_i$ *is:*

$$\xi_{n,m}^{m} = \left\{(r_1, r_2, \ldots, r_m) \in N_0^{n,m} | \sum_{j=1}^{m} r_j = n-m\right\}$$

Where : $r_i \in \{0,1,\ldots,n-m-\sum_{j=1}^{i-1} r_j\}$& $N_0^{n,m} = \{0,1,\ldots,n-m\}$

Thus, the progressive randomized censored design of the second type is in the form of random variables $R_m = (R_1, R_1, \ldots, R_m)$

The maximum likelihood function of type II progressive censored data is:

$$f(x,a,b,p) = k \prod_{i=1}^{m} f(xi)[1-F(xi)]^{ri} \ldots.(4)$$

Where:

$t_i: is\ used\ instead\ of\ t_{i:m:n,}\ 0 < t_1 < t_2 < \cdots < t_m < \infty$

$f(t_i, \theta)$: is the pdf of $T$

$F(t_i)$: is the cdf of T

$r_i$: Excluded units ; r$\geq 0$ ; i=1,2,…,m

$k = \prod_{i=1}^{m} ri$ where $ri = \sum_{j=1}^{m}(r_j + 1)$

$n, m \in N$ , $1 \leq i \leq m$

**Estimation methods:**

- Maximum likelihood function:[3][10]

If the random variable (T) has a probability density function as in equation (23-2), then the maximum likelihood function of the independent random variables $T_1, T_2, \dots \dots . T_n$ is:

$$L(T_1, T_2, \dots, T_n, \theta) = f\left(t_{1,}, \theta\right). \mathrm{f}(t_2, \theta) \dots \mathrm{f}(t_n, \theta) \dots (5)$$

$$\therefore L = \prod_{i=1}^{n} f(t_i, \beta, \theta) \quad \dots (6)$$

Since the maximum likelihood function of type II progressive censored data is:

$$L = k2^m\ \theta^m \prod_{i=1}^{m} t_i\ \ e^{-\theta \sum_{i=1}^{m}(r+1){t_i}^2} \quad \dots (7)$$

$$L = k \prod_{i=1}^{m} 2\theta\ t_i\, e^{-\theta {t_i}^2} \left[ e^{-\theta {t_i}^2} \right]^{ri} \quad \dots (8)$$

$$L = k2^m\ \theta^m \prod_{i=1}^{m} t_i\ \ e^{-\theta \sum_{i=1}^{m}(r+1){t_i}^2} \quad \dots (9)$$

$$L = k2^m\ \theta^m \prod_{i=1}^{m} t_i\ \ e^{(-\theta z)} \quad \dots (10)$$

Where :

$$Z = \sum_{i=1}^{m}(r+1){t_i}^2$$

Taking the natural logarithm of both sides of the equation, we get:

$$LOG(L) = logk + mlog2 + mlog\theta + \sum_{i=1}^{m} logt_i\ \ - \theta z \quad \dots (11)$$

By partially deriving the above equation with respect to the measurement parameter (θ) and equating the derivative to zero, we obtain the maximum likelihood estimators for them as follows:

$$\frac{\partial LogL}{\partial \theta} = \frac{m}{\theta} - \mathrm{z} \quad \dots (12)$$

$$\hat{\theta}_{MLE} = m/z \quad \dots (13)$$

- **Weighted Least square:[6][9]**

The weighted least squares method uses the same technique as the conventional least squares method, except for the use of weights.

In the case of a general linear model, the system of equations can be written in weighted least squares form using matrices, as follows:

$$p^{-1}y = p^{-1}\, x\, p + P^{-1}u \quad \dots (14)$$

$or$

$$\begin{bmatrix}\sqrt{w_1}y_1\\ \sqrt{w_2}y_2\\ .\\ .\\ .\\ \sqrt{w_n}y_n\end{bmatrix}=\begin{bmatrix}\sqrt{w_1} & \sqrt{w_1}x_{11}\ ....\sqrt{w_1}\ x_{1k}\\ \sqrt{w_2} & \sqrt{w_2}x_{21}\ ....\sqrt{w_2}\ x_{2k}\\ & .\\ & .\\ & .\\ & .\\ \sqrt{w_n} & \sqrt{w_n}x_{n1}\ ....\sqrt{w_n}\ x_{nk}\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ .\\ .\\ .\\ .\\ \beta_k\end{bmatrix}+\begin{bmatrix}\sqrt{w_1}\ U_1\\ \sqrt{w_2}\ U_2\\ .\\ .\\ .\\ .\\ \sqrt{w_n}\ U_K\end{bmatrix}\ldots(15)$$

To verify the basic assumptions necessary to apply the least squares method as follows:

$$E(u\,.\,u')=E[(p^{-1}u)(p^{-1}u)']=\sigma^2\,p^{-1}\,p\,p'\,p^{|-1}=\sigma^2\,I_p\quad\ldots(16)$$

This result confirms the assumptions of homogeneity of error variance and the absence of autocorrelation (covariance). Therefore, the basic assumptions of the regression model are met. The least squares method can then be used to obtain the estimators of the model's parameter vector. The following weighting is used:

$$W_i=\frac{(n+1)^2(n+2)}{i(n-i+1)}\qquad\ldots(17)$$

Therefore, the sum of squared error equation can be formulated as follows:

$$Q=\sum_{i=1}^{n}W_i\left[F(x_i)-\frac{i}{n+1}\right]^2\qquad\ldots(18)$$

$$Q=\sum_{i=1}^{n}W_i\left[1-e^{-\theta x^2}-\frac{i}{n+1}\right]^2\quad\ldots(19)$$

$$Q=\left\{\sum_{i=1}^{n}\frac{(n+1)^2(n+2)}{i(n-i+1)}\left[1-e^{-\theta x^2}-\frac{i}{n+1}\right]^2\right\}\quad\ldots(20)$$

By deriving the previous value with respect to θ, the natural equations will be as follows:

$$\frac{dQ}{d\theta}=2\sum_{i=1}^{n}\frac{e^{-x^2\theta}(1+n)^2(2+n)\,(1-e^{-x^2\theta}-\frac{i}{1+n})x^2}{i(1-i+n)}\quad\ldots\quad(21)$$

*The previous equation is nonlinear and cannot be solved using conventional methods except by using numerical methods in order to obtain the maximum likelihood estimator (θ ̂) for the unknown parameter θ*

- **White ›s :[*8*][11]**

*The basic idea of this method is based on the survival function (reliability) of the probability distribution function in formulating a simple linear regression model as follows:*

*S(t)*= $e^{-\theta x^2}$ …(22)

The inverse of the survival function is given by the following formula:

$$s_{(t)}{}^{-1}=\frac{1}{e^{-\theta x^2}}\quad\ldots(23)$$

$$\ln\left[\ln\left(\frac{1}{s_{(t)}}\right)\right]=\ln\theta\ +2\ln x$$

Thus, a linear regression model was obtained:

$$Y_i=a+b\,T_i\qquad\ldots(24)$$

$where:$

$Y_i = \ln\left[\ln\left(\frac{1}{s_{(t)}}\right)\right]$ … (25)

a = ln $\theta$

$T_i = \ln x$

b = 2

Using the ordinary least squares method,

$\hat{a} = \bar{y} - \hat{b}\bar{T}$ … (26)

Ln $\theta = \bar{y} - \hat{b}\bar{T}$ … (27)

$\hat{\theta} = e^{\bar{y}-2\bar{T}}$ …(28)

## 3. Experimental aspect

Simulaion:[1]

- Random number generation :The Monte-Carlo method is used to generate observations for most known statistical distributions. It is one of the most important and common simulation methods. The mechanism of the Monte Carlo method is: Generating random numbers that follow a regular distribution over the interval (0,1) using the clustering density function, By using a statistical mathematical method to convert the generated regular random number to obtain a random variable that desWLSibes the model under experiment, as shown in the two equations below:

  $y = F(X)$ … (29)

  $x = F^{-1}(y)$ … (30)

- Set default values for the distribution parameters (θ) as defined below:

  θ= 20 , 40

- Generating samples of size (25, 50, 100) with imposed values for the excluded units (m=1, 2, 3), Three models can be WLSeated to represent the sample size, distribution parameter, and excluded units (m).

model (A)

Default values for the distribution parameter, sample size, and excluded units

| Model | n | θ | m |
|---|---|---|---|
| 1 | 25 | 20 | 1 |
| 2 | 25 | 40 | 2 |
| 3 | 25 | 60 | 3 |

model (b)

Default values for the distribution parameter, sample size, and excluded units

| Model | n | θ | m |
|---|---|---|---|
| 1 | 50 | 20 | 1 |
| 2 | 50 | 40 | 2 |
| 3 | 50 | 60 | 3 |

model (c)

Default values for the distribution parameter, sample size, and excluded units

| Model | n | θ | m |
|---|---|---|---|
| 1 | 100 | 20 | 1 |
| 2 | 100 | 40 | 2 |
| 3 | 100 | 60 | 3 |

- The experiment was repeated (1000) times to obtain greater homogeneity, Calculating the estimated values of the reliability function and all the estimation methods mentioned in the theoretical aspect after calculating the distribution parameter estimate for each method, The statistical measure of mean square error (MSE) is adopted to compare the estimated values of the reliability function, knowing that MSE can be found using the formula below:

$$MSE\left[\hat{R}(t_i)\right] = \frac{\sum_{i=1}^{R}[R(t_i)-\hat{R}(t_i)]^2}{R} \quad \dots (31)$$

Where : R: Repeat the experiment , $\hat{R}(t_i)$: *Estimated reliability function*

**Simulation results**:

The analysis results were obtained using the Matlab language program. The following are the results shown in the tables that will be analyzed according to the sequence of the tables:

Table (1) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=25 , m=1 , θ = 20 | $Ml$ | 0.01505 |
| | WLS | 0.024275 |
| | WH | 0.189575 |

Table (2) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=25 , m=2 , θ = 20 | $ML$ | 0.00908 |
| | WLS | 0.02128 |
| | WH | 0.19957 |

Table (3) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=25 , m=3 , θ = 20 | $ML$ | 0.00102 |
| | WLS | 0.01137 |
| | WH | 0.18662 |

By looking at tables (1, 2, 3) we see that the maximum likelihood method is the best because it has the lowest mean (MSE), followed by the (Weighted Least square) method and finally the (White,s) method.

Table (4) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=25 , m=1 , θ = 40 | $ML$ | 0.01308 |
| | WLS | 0.01522 |
| | WH | 0.13544 |

Table (5) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=25 , m=2 , θ = 40 | $ML$ | 0.01098 |
| | WLS | 0.02048 |
| | WH | 0.17131 |

Table (6) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=25 , m=3 , $\theta = 40$ | $ML$ | 0.01004 |
| | WLS | 0.02010 |
| | WH | 0.2013 |

By looking at tables (4, 5, 6) we see that the maximum likelihood method is the best because it has the lowest mean (MSE), followed by the (Weighted Least square) method and finally the .(White,s) method

Table (7) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=50 , m=1 , $\theta = 20$ | $ML$ | 0.11234 |
| | WLS | 0.12128 |
| | WH | 0.17957 |

Table (8) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=50 , m=2 , $\theta = 20$ | $ML$ | 0.10255 |
| | WLS | 0.13001 |
| | WH | 0.16671 |

Table (9) Mean (MSE) of the reliability function for each estimation metho

| MODEL | Method | average mse |
|---|---|---|
| n=50 , m=3 , $\theta = 20$ | $ML$ | 0.01601 |
| | WLS | 0.12564 |
| | WH | 0.13359 |

By looking at tables (7, 8, 9) we see that the maximum likelihood method is the best because it has the lowest mean (MSE), followed by the (Weighted Least square) method and finally the (White,s) method.

Table (10) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=50 , m=1 , $\theta = 40$ | $ML$ | 0.02034 |
| | WLS | 0.02883 |
| | WH | 0.10307 |

Table (11) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=50 , m=2 , $\theta = 40$ | $ML$ | 0.10902 |
| | WLS | 0.13124 |
| | WH | 0.18334 |

Table (12) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=50 , m=2 , $\theta = 40$ | $ML$ | 0.00908 |
| | WLS | 0.02128 |
| | WH | 0.19957 |

By looking at tables (10, 11, 12) we see that the maximum likelihood method is the best because it has the lowest mean (MSE), followed by the (Weighted Least square) method and .finally the (White,s) method

Table (13) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=100 , m=1 , $\theta = 40$ | $ML$ | 0.03010 |
| | WLS | 0.13012 |
| | WH | 0.13901 |

Table (14) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=100 , m=2 , $\theta = 20$ | $ML$ | 0.00908 |
| | WLS | 0.14126 |
| | WH | 0.29778 |

Table (15) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=100 , m=3 , $\theta = 20$ | $ML$ | 0.00537 |
| | WLS | 0.12158 |
| | WH | 0.29779 |

By looking at tables (13, 14, 15) we see that the maximum likelihood method is the best because it has the lowest mean (MSE), followed by the (Weighted Least square) method and finally the .(White,s) method

Table (16) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=100 , m=1 , $\theta = 40$ | $ML$ | 0.00802 |
| | WLS | 0.04124 |
| | WH | 0.28987 |

Table (17) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=100 , m=2 , $\theta = 40$ | $ML$ | 0.00708 |
| | WLS | 0.09129 |
| | WH | 0.29876 |

Table (18) Mean (MSE) of the reliability function for each estimation method

| MODEL | Method | average mse |
|---|---|---|
| n=100 , m=3 , $\theta = 40$ | $ML$ | 0.00711 |
| | WLS | 0.12321 |
| | WH | 0.29782 |

By looking at tables (14, 15, 16) we see that the maximum likelihood method is the best because it has the lowest mean (MSE), followed by the (Weighted Least square) method and finally the (White,s) method

The applied aspect:

sum data:

Data related to the study were collected for a number of viral hepatitis patients from the records of Yarmouk, Karkh, and Kadhimiya hospitals, amounting to (19, 23, and 21), respectively. The data represented the duration of the patients' stay in months under observation and treatment until death, for the time period, the first quarter of 2025, as shown in Table No. (19):

survival function calculation:

After verifying that the data are subject to Rayleigh distribution by using chi-square, the maximum likelihood method was used to estimate the distribution coefficient and then calculate the survival function after determining it, which is the best estimation method in the experimental aspect as shown in Table (19).

**Table (19) represents the survival times and function of patients with hepatitis in months (liver cirrhosis + liver cancer)**

| Yarmouk Hospital(ri=9) | | | | Al-Karkh Hospital(ri=4) | | | | Al-Kadhimiya Hospital(ri=1) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| i | ($t_i$) | $s_{(t)}$ | | i | ($t_i$) | $s_{(t)}$ | | i | ($t_i$) | $s_{(t)}$ | |
| 1 | 0.4 | 0.95 | | 1 | 0.3 | 0.97 | | 1 | 0.7 | 0.86 | |
| 2 | 0.7 | 0.86 | | 2 | 0.6 | 0.90 | | 2 | 1 | 0.74 | |
| 3 | 0.9 | 0.78 | | 3 | 0.9 | 0.78 | | 3 | 1.4 | 0.56 | |
| 4 | 1 | 0.74 | | 4 | 0.9 | 0.78 | | 4 | 1.4 | 0.56 | |
| 5 | 1.1 | 0.70 | | 5 | 1 | 0.74 | | 5 | 1.6 | 0.46 | |
| 6 | 1.3 | 0.60 | | 6 | 1 | 0.74 | | 6 | 1.8 | 0.38 | |
| 7 | 1.3 | 0.60 | | 7 | 1.5 | 0.51 | | 7 | 1.8 | 0.38 | |
| 8 | 1.8 | 0.38 | | 8 | 1.5 | 0.51 | | 8 | 2 | 0.30 | |
| 9 | 2.2 | 0.23 | | 9 | 1.7 | 0.42 | | 9 | 2 | 0.30 | |
| 10 | 2.2 | 0.23 | | 10 | 1.9 | 0.34 | | 10 | 2.1 | 0.27 | |
| 11 | 2.5 | 0.15 | | 11 | 1.9 | 0.34 | | 11 | 2.1 | 0.27 | |
| 12 | 2.6 | 0.13 | | 12 | 2 | 0.30 | | 12 | 2.4 | 0.18 | |
| 13 | 2.9 | 0.08 | | 13 | 2 | 0.30 | | 13 | 2.4 | 0.18 | |
| 14 | 3 | 0.07 | | 14 | 2 | 0.30 | | 14 | 2.8 | 0.10 | |
| 15 | 3.1 | 0.06 | | 15 | 2.2 | 0.23 | | 15 | 2.8 | 0.10 | |
| 16 | 3.1 | 0.06 | | 16 | 3.1 | 0.06 | | 16 | 2.8 | 0.10 | |
| 17 | 3.2 | 0.05 | | 17 | 3.3 | 0.04 | | 17 | 2.9 | 0.08 | |
| 18 | 3.4 | 0.03 | | 18 | 3.3 | 0.04 | | 18 | 3.1 | 0.06 | |
| 19 | 3.4 | 0.03 | | 19 | 3.5 | 0.03 | | 19 | 3.3 | 0.04 | |
| | | | | 20 | 3.5 | 0.03 | | 20 | 3.3 | 0.04 | |
| | | | | 21 | 3.8 | 0.01 | | 21 | 3.4 | 0.03 | |
| | | | | 22 | 3.9 | 0.01 | | | | | |
| | | | | 23 | 4 | 0.01 | | | | | |
| mean | 2.1 | 0.35 | | | 2.3 | 0.37 | | | 2.24 | 0.26 | |

Where :

i: Patient sequence.

$t_i$: Time the patient remained alive until death is recorded in months.

$s_{(t)}: survival\ function$

From the table above, we see that the survival rate of patients in Al-Karkh Hospital until death reached (37%) for every two months and nine days, meaning that (73%) of patients lose their lives within 2 months and 9 days, which indicates the seriousness of hepatitis (liver cirrhosis - liver cancer). The survival rate of patients in Al-Yarmouk Hospital was close to the survival rate of patients in Al-Karkh Hospital, as it appeared equal to (35%) for every 2 months and 3 days, while the survival rate of patients in Al-Kadhimiya Hospital appeared lower than the survival rate of patients in both Al-Karkh Hospital and Al-Yarmouk Hospital, as the survival rate appeared equal to (26%) for every 2 months and 7 days. This clear difference in the survival rate may indicate that the services provided in Al-Kadhimiya Hospital are less than the medical services provided in other hospitals& it was found that their survival rate until death is (33%) for a time average of (2) months and (15) days.

## 4. Conclusions:

- The maximum likelihood method appeared in the experimental aspect as the best estimation method for different sample sizes because it has the lowest average Mse
- There is an inverse relationship between the survival function and the time to death. The longer the survival time, the smaller the survival function.
- The average survival time of patients until death in Yarmouk Hospital was (%35) per (2) month and (3)day
- The average survival time of patients until death in Al-Karkh Hospital was (37%) per (2) month and (9)day
- The average survival time of patients until death in Al-Kadhimiya Hospital was (26%) per (2) month and (7)day
- It was found that Al-Kadhimiya Hospital had the lowest survival rate compared to Al-Karkh and Yarmouk Hospitals.

Recommendations:

- Use and propose new distributions to study the survival function and its indices, and compare them with the Rayleigh distribution.
- The time patients spend until death in Al-Kadhimiya Hospital is shorter than the time patients spend until death in Al-Karkh and Yarmouk Hospitals.
- The relevant authorities should pay attention to the medical services provided to liver patients, as their time in the hospital until death is short.

## Referenc:

[1] A. H. H. Al-Taie and O. A. K. Al-Quraishi, "Comparing Between the Method of Moments and the Method of Maximum Likelihood for Estimating the Reliability Function of Weibull Distribution Under Observation Data Using Simulation," *J. Kerbala Univ.*, vol. 17, no. 1, 2019.

[2] S. K. Bhattacharya and R. K. Tyagi, "Bayesian Survival Analysis Based on the Rayleigh Model," *Trab. Estad.*, vol. 5, pp. 81–92, 1990.

[3] A. C. Cohen, "Progressively Censored Samples in Life Testing," *Technometrics*, vol. 5, no. 3, pp. 327–339, 1963.

[4] M. J. Crowder, A. C. Kimber, R. L. Smith, and T. J. Sweeting, *Statistical Analysis of Reliability Data*. Great Britain: Chapman and Hall, 1991.

[5] E. Elviana and J. Purwadi, "Parameters Estimation of Rayleigh Distribution in Survival Analysis on Type II Censored Data Using the Bayesian Method," in *J. Phys.: Conf. Ser.*, vol. 1503, no. 1, p. 012004, IOP Publishing, 2020.

[6] M. Galrinho, C. Rojas, and H. Hjalmarsson, "A Weighted Least-Squares Method for Parameter Estimation in Structured Models," in *53rd IEEE Conf. on Decision and Control*, 2014, pp. 3322–3327.

[7] I. Gupta, "Bayesian and E-Bayesian Method of Estimation of Parameter of Rayleigh Distribution—A Bayesian Approach under Linex Loss Function," *Int. J. Stat. Syst.*, vol. 12, no. 4, pp. 791–796, 2017.

[8] H. L. Harter and A. H. Moore, "Maximum-Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and from Censored Samples," *Technometrics*, vol. 7, no. 4, pp. 639–643, 1965.

[9] H. L. Lu and S. H. Tao, "The Estimation of Pareto Distribution by a Weighted Least Square Method," *Qual. Quant.*, vol. 41, pp. 913–926, 2007.

[10] A. Pathak, M. Kumar, S. K. Singh, and U. Singh, "Bayesian Inference: Weibull Poisson Model for Censored Data Using the Expectation–Maximization Algorithm and Its Application to Bladder Cancer Data," *J. Appl. Stat.*, vol. 49, no. 4, pp. 926–948, 2022.

[11] S. K. Sinha and B. K. Kate, *Life Testing and Reliability*. Wiley Eastern Limited, 1980.

[12] O. A. K. Al-Quraishi, "Estimating the Survival Function of Viral Hepatitis Patients in a Sample of Baghdad Hospitals," Unpublished manuscript, 2025.

[13] Ministry of Education Iraq, "Hepatitis Statistics and Healthcare Services in Baghdad," Internal Report, 2024.

[14] World Health Organization, "Hepatitis B and C," WHO Fact Sheet, Mar. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b

[15] R. D. Smith, "Survival Analysis and Medical Decision Making," *Stat. Med.*, vol. 32, no. 15, pp. 2507–2516, 2013.