UAR Publisher

UAR Journal of Multidisciplinary Studies (UARJMS)

ISSN: 3049-4346 (Online) | Published by UAR Publisher

NORMALIZED SPACING PROCEDURE FOR ESTIMATING STAGGERED PATIENTS ENTRY TIMES FOR CENSORED VALUES IN CLINICAL TRIAL

Eric Boahen

Department of Statistics School of Mathematical Sciences CK Tedan University of Technology and Applied Sciences

*Corresponding Author: **Eric Boahen**

Abstract

The study was carried out to provide a general procedure to estimate censored data. Parameters in censored data are not estimated in the same way that those in uncensored data are. Problems in estimating survival time events are compounded due to inappropriate choice of starting time in survival studies. Normalized spacingprocedure for handling censored data are investigated by using the proposed generalized model. Survival times are arranged to start from time zero and thereby eliminating truncations in the algorithm. Recursive partitioning for placing of the survival time are used to normalize spacing of intervals so that at least an event time (failure time) is located in every interval for easy possible parameter estimation. Simulations showed that the generalized algorithm developed was better and adequate.

INTRODUCTION

One of the branches of statistics that has been popular for research and used for analyzing duration of time until an event happens is survival analysis. In biological organisms, the event is death and in mechanical systems the event is failure. The clarification of an event in certain phenomena is treated in many research areas with a variety of names. For this reason, the topic of survival analysis is known as reliability theory in engineering, duration analysis in economics and event history in sociology (Chen 2013). The purpose of survival Analysis goes beyond description of an event; it focuses on the time a certain population survives a certain time scale and also concerns the rate with which it dies or fails. Apart from these purposes of survival

Article DNA

Article Type:
Original research article

DOI:
10.5281/zenodo.18329308

Article History:
Received: 08-01-2026
Accepted: 14-01-2026
Published: 21-01-2026

Keywords:
Censored, recursive, normalized spacing, algorithm, patients

How to Cite

Eric Boahen. (2026). NORMALIZED SPACING PROCEDURE FOR ESTIMATING STAGGERED PATIENTS ENTRY TIMES FOR CENSORED VALUESIN CLINICAL TRIAL. *UAR Journal of Multidisciplinary Studies (UARJMS)*, 2(1), 1–14. 10.5281/zenodo.18329308

***Related declarations are provided in the final section of this article.*

Analysis, it is also concerned with whether or not there are multiple deaths or failures in a population. To respond to these areas in survival analysis, the term lifetime of an event often poses a problem and needs to be defined in the context of the research area. In biological survival context: "death" is unambiguous and the time to death is discrete, but system failure can be partial and the degree of biological failure such as kidney and heart attack are localized in time. For this reason, time to an event in survival analysis has been a problem for many researchers; (Miyakoshi, Y. 2021) some use the time failure begins, others use the time a system has completely failed while others use the average time of failure. In survival analysis, death or failure is taken as an event and traditionally, it is assumed that only one event occurs for an individual subject after which the organism is dead. Repeated events relax this assumption. Repeated events also known as recurring events are basically predominant in system reliability analysis (Azur, 2011). Not all subjects have the event during the observation time and such subjects are described as censored (Stene, 2009). These subjects in the population of observation are censored because nothing is observed or known about them after the time of censoring. For this reason, censored subjects do not carry information needed for survival analysis but are yet part of the population. The presence of censored subjects in a population or data poses problems in survival analysis and decrease precision in estimations. Censoring occurs in different forms. If an event is expected to happen at a specified time and the subject has been observed beyond the specified time, the subject is said to be right censored. Left censoring is when a subject experiences an event before its specified time of occurrence. There are situations in survival analysis where nature determines an occurrence of an event without a specified time for observations from the researcher; this kind of censoring is known as random censoring (Singh, 2013). Right censoring is of two forms: Singly Type 1 and Type 2. Singly Type 1 censoring occurs when the researcher fixes the censoring time to terminate an observation (Wu, 2010). The researcher has control over the experiment with the censoring limit. Type 2 censoring on the other hand is when the researcher determines the time to terminate the experiment after an expected satisfactory event has occurred in the process (Mroz, 1987). Censoring in survival analysis makes estimation uncomfortable.

Research Questions

The incomplete information on censored events in censored data makes it difficult to estimate parameters. For this reason many researchers have come out with several robust models to deal with the inconsistencies in estimation (Chib, 1992). Most of the models developed to solve censored estimation problems are not robust and efficient enough to handle censored data and the problem is compounded when those existing models in survival literature are applied to staggered censored data, because different entry points of the subjects create continuous interval censoring that demand interval estimation. The questions that the paper seeks to answer are as follows:

1. Can individual staggered patients' entry with different starting points be made to start at common starting points?

2. Can patients entry points censored data be converted to interval censoring to make the censoring model general?

Normalized spacing Algorithm

In a study of disease condition in survival analysis, it is not possible to admit all patients to the study sample at the same time because of paucity of patients. It is therefore reasonable to accept the patients as they enter into the study for treatment. Patients are then followed to death. Unfortunately, a portion of patients may be lost to follow-up because of one or two reasons. For these reasons, it is possible to end up with a censored data that contains variation in interval occurrences. This situation makes it difficult to come out with appropriate estimations on censored data. Figure 3.1 is the general representation of staggered entry point.

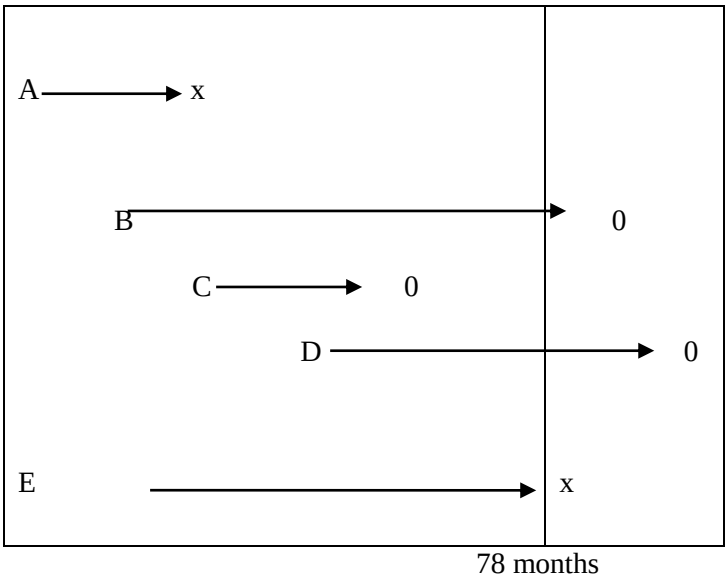


Figure 3.1: Patients staggered entry into the study.

In Figure 3.1, patients enter into the study at different times. This is the time patients report for treatment and the study is terminated at fixed censoring time of 78 months. This type of censoring is called Random Censoring. Random censoring is one where there is a single censoring limit for all the patients but the entry times vary across the patients. The letters A, B, C, D, E are the symbols which indicate individual patients. From Figure 3.1, patient A started the treatment and was admitted to the study but died in the process (indicated by symbol x) without surviving to censoring limit of 78 months. Patient C also entered the study and was censored in the process of the study. Patient C either stopped reporting for treatment or relocated to a different treatment center. For this reason Patient C is considered missing from the study. Patients B and D entered the study and survived beyond the censoring limit but later died. These two patients were still alive at the time of terminating the study at the censoring time of 78 months. Patient E

also entered the study but was alive at the time of terminating the study at the censoring time of 78 months. Data made up of these types of entries are considered in the work. Censored figures carry hidden information because the stochastic value of the figures is masked which makes estimations in censored data difficult. Many researchers have proposed several methods to estimate censored data that seeks to minimize bias and also increase precision but the data with several censoring emanating from staggered entry described in figure 3.1 always pose a problem.

To overcome this difficulty in calculation, it is reasonable to put all staggered observations at the same beginning time. This technique means that all values are allowed to be observational at a common starting point at time zero. The choice of time zero is very useful to begin all patients so as to make meaningful analyses. This is because; common time zero makes no difference in substantial coefficient estimates and fit models. This is to say that, all patients are given an equal platform for comparison as indicated in Figure 3.2. The technique of using a common platform for all values in censored data is a contribution that seeks to make this work unique from other methods.

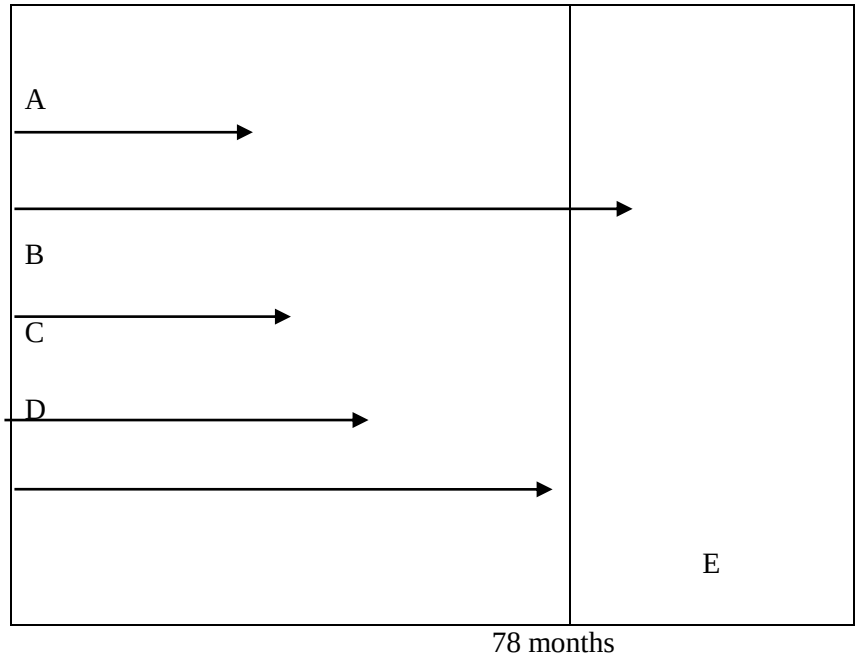


Figure 3.2: A common start for all the patients

Figure 3.2 shows the arrangement of patients to start at common starting point. This makes analyses very simple since patient's length of survival is compared with the censoring limit or time. From Figure 3.2, we face a situation where a censored observation may be in any interval. To overcome this challenge, let $t_1, t_2, t_3, \ldots\ t_n$ be n patients' survival times out of which r patients fail and $n - r$ are censored

To do an analysis at time t, it is appropriate to put the observations in recursive partitioning. This recursive partitioning is a statistical algorithm for multivariable analysis that seeks to create a decision tree that correctly classifies individual patients in the study population by splitting the population into sub-populations based on censored and uncensored independent observations. The idea of using recursive partitioning here is to ensure that each sub-population created contains at least both censored and uncensored observations. This process is recursive because each sub-population created in the patient's survival time may in turn be split in an indefinite number of times until the splitting process terminates after a particular stopping criterion is reached.

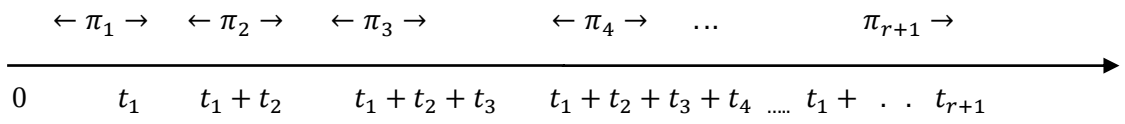


Figure 3.3:Recursive partitioning

$\pi_i$is the recursive interval in a Poisson process. This means that π_1 is the total observed survival time to the first failure, π_2 is the total observed survival time between the second and the first failure time, π_r is the total observed time between the r^{th} and the $(r-1)th$ failure.π_{r+1}corresponds to the slop over time in the process without observing an event of failure. From Figure 3.3, we observe that an event at $t_1$and $t_1 + t_2$, so t_2 is an exponential distribution. To do an analysis at time t, it is appropriate to put the observations in normalized spacing. The idea of using normalized spacing is important because infinite intervals are terminated. Let the i^{th}interval be denoted as $I_i = (t_{(i-1)}, t_i]$.

Also denote the time for the censored observation to be ${t_1}^i, {t_2}^i, {t_3}^i, \ldots, {t_{ci}}^i$ with order values ${t_1}^i \leq {t_2}^i \leq {t_3}^i \leq \ldots, \leq {t_{ci}}^i$.

Suppose the probability density function of an observation believed to be drawn at random from an interval is $q(t)$ and its survival density function is $Q(t) = P\{T > t\}$. In the i^{th} interval, there would be a failure at the right hand side of the interval of observations with probability $q(t_i)dt_i$ and c_i which represent censored observations with probability$\prod_{j=1}^{c_i} Q(t_j^{(i)})$. The contribution of the i^{th} interval to the likelihood is $q(t_i)dt_i \prod_{j=1}^{c_i} Q(t_j^{(i)})$. From Figure 3.3, we have r such intervals with failures at the right hand end point and $\boldsymbol{r+1}$ infinite with failure. $\boldsymbol{r+1}$which are defined as those patients who survived beyond the censoring limit. The likelihood function is

$$\prod_{i=1}^{r} q(t_i)dt_i \prod_{j=1}^{c_i} Q(t_j^{(i)}) \prod_{j=1}^{c_{r+1}} Q(t_j^{(r+1)}))$$

To derive the probability density function, it is useful to obtain the constant of proportionality. If all the censored observations are in the last interval, the constant of proportionality is $\frac{n!}{(n-r)!}$, since at the first

failure any of the n can fail, at the second interval any of the $n-1$ can fail, at the third interval any of the $n-2$ can fail, and so on. The constant equals $n(n-1),\ .\ .\ .\ ,n-r+1$. Using the same idea above, n and c_1 are censored in the first interval, so there are $n-c_1=n_1$ ways that the first failure can occur. In the second interval c_2 are censored in the interval and based on this second failure rate, there are $n-c_1-c_2-1=n_2-1$ ways it can occur. If there is to be a failure at interval 3 and m_3 observations are censored in the third interval, then there are $n-c_1-c_2-c_3-2=n_3-2$ ways this can happen. At this point the constant term is

$$n_1(n_2-1)(n_3-2)(n_4-3),\ .\ .\ .(n_r-r+1)=K$$

Denote $\pi_i$to be the total number of patients observed in the i^{th} interval. Then

$$\pi_1=(n-c_1)t_1+S_1=n_1t_1+S_1$$

$$\pi_2=(n-c_1-c_2-1)(t_2-t_1)+S_2-c_2t_1,\qquad \pi_1=(n_2-1)(t_2-t_1)+S_2-c_2t_1$$

$$\pi_i=(n_i-i+1)t_i-t_{(i-1)}+[S_i-c_it_{(i-1)}]$$

For i= 1, . . , r..

$$\pi_{r+1}=S_{r+1}-c_{r+1}t_r$$

The sum of $\pi_i=V$ ie $\sum_{i=1}^{r+1}\pi_i=V$, where V is the sum of the patients' time in all the intervals

At this point we need to estimate the joint distribution of π_i. This is obvious because each $\pi_i$is the sum of the survival time observed in the i^{th} interval and V is the sum of the survival time observed in all the intervals.

We propose the distribution of π_i. When ordering of data is not important, the joint distribution of π_i is

$f(\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_{r+1}\ )$. Define the transformation

$\pi_1=u_1(t_1,t_2,t_3\ .\ .\ t_r),\ \pi_2=u_2(t_1,t_2,t_3\ .\ .\ t_r),\pi_r=u_r(t_1,t_2,t_3\ .\ .\ t_r)$ which maps H_t onto H_π, where H_t is the disjoint spaces $H_1,H_2,\ .\ .\ H_r$ such that the transformation H_t onto H_y is one-to-one for all $i=1,2,3,\ .\ .\ r$. Since patients' times in individual intervals are not ordered, the intervals are independent and identical. For this reason, there exist for i, a unique inverse transformation (Gibbson,1956). Again denote $t_1=z_{1i}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r),\ t_2=z_{2i}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r),$ $t_3=z_{3i}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r),t_r=z_{ri}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)$

The Jacobian$J_i((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)=J_i\left(\pi^{(r)}\right)=\frac{\partial(z_{1i},z_{2i},z_{3i}\ .\ .\ z_{ri})}{\partial((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)}=det\frac{\partial z_{ri}}{\partial\pi_r}$ exist,

where$det(a_{ij})$ is the determinant of the $r\times r$ matrix. The joint pdf for the interval

$(\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)$where$\pi_i=u_i(t_1,t_2,t_3\ .\ .\ t_r)$ is

$$\sum_{i=1}^{r}J_i((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)|f(t^r)[(z_{1i}(\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)],z_{2i}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r),$$

$z_{3i}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)\ \ .\ .\ .\ .z_{ri}((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)$for all $((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)$

The Jacobian of the inverse transformation is the reciprocal of the Jacobian of the direct transformation $\frac{\partial(z_{1i},z_{2i},z_{3i}\ .\ .\ z_{ri})}{\partial((\pi_1,\pi_2,\pi_3,\ .\ .\ .\ \pi_r)} = \left[\frac{\partial(u_1,u_2,u_3\ ,\ .\ .u_r}{\partial t_1,t_2,t_3\ .\ .\ t_r}\right]^{-1}$

$$J_i(\pi^r) = [J_i(t^r)^{-1}$$

$$f(\pi_i) = [\mu \left| \frac{\Delta\pi}{\Delta t} \right|^{-1}] \qquad 3.1$$

whereμ is the joint probability density function of patients' survival time, $\Delta\pi$ is the total number of patients observed in the intervals and Δt represents the change in survival time of patients.

Equation 3.1 is the proposed general model when ordering of data is necessary for estimation.
The joint distribution of the π_i order statistics for the i^{th} interval is not the same for an independent distribution since order statistics are not independent or identically distributed. This is because; order statistics depends on the strength of the chain. For this reason the joint distribution of π_i is easily obtained by the method of Jacobian transformations. The r^{th} order statistics for $\pi_i$is obtained by the transformation

$$\pi_1 = \text{smallest of } (t_1,t_2,t_3\ .\ .\ t_{r+1})$$

$$\pi_2 = \text{second smallest of } (t_1,t_2,t_3\ .\ .\ t_{r+1})$$

$$\vdots$$

$$\pi_3 = r^{th}\text{Smallest of } (t_1,t_2,t_3\ .\ .\ t_{r+1})$$

This transformation is not one-to-one because there are a total of $r!$ possible arrangements of the

patients' survival times in increasing order of magnitude. For these reasons, it is reasonable to think of $r!$ inverses to the transformation.
$\pi_i = N_1(N_2-1)(N_3-2)(N_4-3),\ .\ .\ .(N_r-r+1)\lambda^r e^{-\lambda\sum_{i=1}^{r+1}.\pi_i}\left|\frac{\partial(t_1,t_2,\ .\ .\ .t_r)}{\partial(\pi_1,\pi_2,.\ .\ .\pi_r)}\right|.$
(3.2)

To get the Jacobian, it is appropriate to get reciprocal $|\frac{\partial\pi}{\partial t}|$ and reciprocate that. The matrix of $|\frac{\partial\pi}{\partial t}|$ whose $(i,j)th$ element is

$$\begin{array}{c} \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ . \\ . \\ . \\ \pi_r \end{array}\begin{array}{c} t_1\ t_2 t_3 t_r \\ \left[\begin{array}{cccc} N_1 & 0 & 0 & 0 \\ -(N_2-1)-M_2 N_2-1 & 0 & 0 \\ 0 & 0 & N_3-2 & 0 \\ . & . & 0 & . \\ . & . & . & . \\ . & . & . & . \\ 0 & 0 & 0 & N_r-r+1 \end{array}\right. \end{array} \quad \ldots \quad \left.\begin{array}{c} \\ 0 \\ 0 \\ . \\ . \\ . \end{array}\right]$$

$\frac{\partial \pi}{\partial t}$is a triangular matrix so its determinant equals the product of the diagonal terms

$$N_1(N_2-1)..\ (N_r-r+1)=N_i-i+1. \qquad (3.3.)$$

Thus

$$|\frac{\partial \pi}{\partial t}|=N_1(N_2-1)(N_3-2)(N_4-3),\ \ldots(N_r-r+1)=K, \qquad (3.4)$$

where$K=\ ^{N_i!}/_{(N_i-r)!}$, therefore

$$\frac{\partial t}{\partial \pi}=\ ^1/_K. \qquad (3.5)$$

$$f(\pi)=K\lambda^r e^{-\lambda\sum_{i=1}^{r+1}.\pi_i}\ ^1/_K. \qquad (3.6)$$

$$f(\pi)=\ ^{N_i!}/_{(N_i-r)!}\lambda^r e^{-\lambda\sum_{i=1}^{r+1}.\pi_i}\ (N_i-r)!/_{N_i!} \qquad (3.7)$$

$$f(\pi)=\lambda^r e^{-\lambda\sum_{i=1}^{r+1}.\pi_i} \qquad (3.8)$$

$$f(\pi)=\prod_{i=1}^{r+1}\lambda e^{-\lambda\pi_i} \qquad (3.9)$$

It is obvious that all π_i are independently and identically distributed and follow an exponential distribution given by

$$f(\pi_i)=\lambda^r e^{-\lambda V}$$

Where V=$\sum_{i=1}^{r+1}\pi_i$ is the total number of patients observed in the study. Clearly, V contains all the information needed to estimate λ.

The aim is to estimate λ . From the equation

$$\log f(\pi_1,\pi_2,\pi_3,\ \ldots\ \pi_{r+1})=r log\lambda-\lambda V+C$$

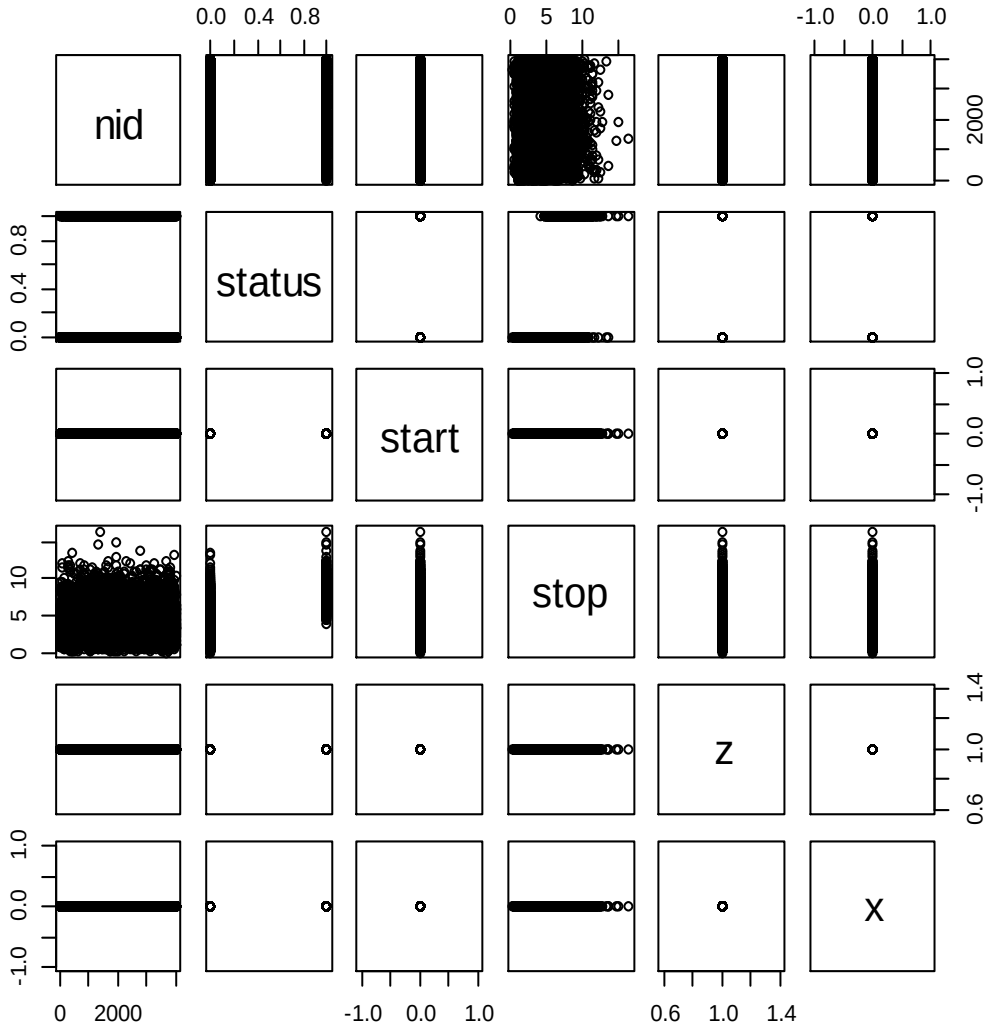
$$\frac{dL}{d\lambda}=\frac{r}{\lambda}-V.\ \text{At}\ \ \frac{dL}{d\lambda}=0,$$

$$\hat{\lambda}=\frac{r}{V}$$

$$\hat{\lambda}=\frac{r}{V}$$

which is the optimal approach for estimating the hazard, this is because V is the total number of both observed values and censored values.

Simulation of the normalized spacing algorithm



Model simulations and Comparisons

In this section, simulation of the optimal model is undertaken using clinical censored data to ascertain the adequacy of the model. The software R is used to simulate the function $f(\pi_i) = \lambda^r e^{-\lambda V}$ and a symmetrical plot of simulation is generated in Figure 3.4. The distribution of the model plot is symmetrical and therefore normally distributed.

nid represents a vector of unique identifier that identifies the individual patients who participated in the study. This means that, individual identity is maintained in the simulation process just as it is present in the original data. For this reason, the model accounts for the censored values even though their stochastic realizations are masked. This means that simulation does not clump up patients' survival time but each patient's survival time is simulated according to the specified number of observations. The plot of the model captured patients survival status which represent a

logical value indicating whether the survival time corresponds to the event (status = 1) or the survival time is censored (status = 0). From the plot, statuses are separated from each other. Censored status (status = 0) is shown in the upper corner and the event status is at the bottom of the status box. The status box is located at the right hand side of the status caption horizontally. Since clinical data are mostly multiple censored data the general proposed model $f(\pi_i) = \lambda^r e^{-\lambda V}$ is adequate for all types of censoring data. From the simulation plot, since all patients are allowed to start from a common point at time zero, the starting time is indicated on the plot. All patients are arranged to start at time zero to ensure normalized spacing t_i of patients' survival times. On the plot, it is obvious that patients' survival starting time in the simulation process is still maintained at zero. This is seen at the starting boxes horizontal to the start box. For this reason, the model accounted for reduction of bias likely to emanate from staggered entry that makes statistical analysis difficulty. The time the study is terminated is captured in the plot of the simulation. This corresponds to the censoring time. Censoring is the time the study is terminated. From the plot, the first box from the right on the horizontal stop box indicated black at time zero but begins to fade to the terminating. From the simulation plot, symbol *Z* indicates whether there is an individual heterogeneity. Heterogeneity is the variation in individual survival times among the patients. The model does not give room for heterogeneity. This is indicated by the horizontal dark line. The symbol *X* represents the covariate variable. No covariate was added to the original data before simulation and this means that the simulated data contains no covariate as indicated by the dark horizontal line which passes through zero on the plot.

Applications

Table 1.Total life in the I-th intervals.

I_i	m_i	A	R	B	C	I_i	m_i	A	r	B	C
1	1	1	0	0	1	47	5	235	2	94	392
2	14	28	2	4	32	48	11	528	2	96	624
3	19	57	5	15	72	49	13	637	2	98	735
4	20	80	3	12	92	50	4	200	3	150	350
5	18	90	5	25	115	51	6	306	2	102	408
6	25	150	2	12	162	52	4	208	1	52	260
7	27	189	5	35	224	53	5	265	3	159	424
8	29	232	12	96	328	54	3	162	1	54	216
9	26	234	5	45	279	55	4	220	1	55	275
10	23	230	7	70	300	56	4	224	1	56	280

11	26	286	8	88	374	57	1	57	1	57	114
12	21	252	4	48	300	58	5	290	2	116	406
13	18	234	9	117	351	59	5	295	1	59	354
14	20	280	3	42	322	60	6	360	1	60	420
15	20	300	5	75	375	61	5	305	5	305	610
16	26	416	3	48	468	62	4	248	3	186	434
17	14	238	5	85	323	63	2	126	3	189	315
18	17	306	2	36	342	64	1	64	1	64	128
19	14	266	1	19	285	65	6	390	3	195	585
20	25	500	5	100	600	66	1	66	1	66	132
21	7	147	5	105	252	67	3	201	1	67	268
22	9	198	1	22	220	68	2	136	1	68	204
23	8	184	4	92	276	69	3	207	2	138	345
24	15	360	2	48	408	70	4	280	1	70	350
25	11	275	1	25	300	71	1	71	1	71	142
26	13	338	3	78	416	72	3	216	1	72	288
27	12	324	2	54	378	73	2	146	1	73	219
28	15	420	4	112	532	74	5	370	3	222	590
26	17	493	2	58	551	75	2	150	1	75	225
30	11	330	3	90	420	76	3	228	1	76	304
31	10	310	5	155	465	77	3	231	1	77	308
32	11	352	4	128	480	78	1	78	1	78	156
33	10	330	1	33	363	79	2	158	0	0	158
34	3	102	1	34	136	80	3	240	0	0	0
35	6	210	1	35	245	81	1	81	0	0	81
36	11	396	4	144	540	82	3	246	0	0	246
37	11	407	1	37	444	83	2	166	0	0	166
38	7	266	1	38	304	84	0	0	0	0	0
39	6	234	2	78	312	85	3	255	0	0	255
40	6	240	3	120	360	86	1	86	0	0	86
41	6	246	1	41	287	85	3	255	0	0	255
42	11	462	2	84	546	86	1	86	0	0	86
43	6	258	1	43	301	87	0	0	0	0	0
44	11	484	1	44	528	88	0	0	0	0	0
45	10	450	2	90	540	108	1	108	0	0	108
46	5	230	2	92	322	Total	807		204		26783

The table (Table 1) was constructed by arranging patients' survival times in such a way that they are given the opportunity to start at time zero. This is to avoid possible truncations of patients' survival times that may arise due to staggered entry of patients into the study. On the table $I_i = [t_{i-1} + t_i]$ represents recursive intervals where $i = 1, 2, \ . \ . \ r + 1$. The $r + 1$ interval is the interval that contains survival time after the study has been terminated at 78 months. From

the table a total of 17 patients survived beyond the termination point. $m_i$represents the number of censored values in I_i and the count of the censored values in the i^{th} interval is denoted j as seen in page 59. $\sum m_i$Represents the sum of the values in the i^{th} interval. From the table,$\sum m_i = 807$ and $\sum r$ which represents the total number of failure values (event times) in I_i is 204. The sum of both censored and the event time is denoted by π_i. From the table, $\pi_i = \sum m_i + r = 1011$ which is the total sample size obtained on breast cancer patients from Korle–Bu Teaching Hospital in Ghana.

The likelihood function

$$f(\pi) = \frac{1011!}{(1011-204)!}\boldsymbol{\lambda^r e^{-\lambda V}}\frac{(1011-204)!}{1011!} = \boldsymbol{\lambda^r e^{-\lambda V}}$$

$L(\pi) = \boldsymbol{\lambda^r e^{-\lambda V}}$)

$$\hat{\mu} = \frac{V}{r} = \frac{26783}{204} = 131.73$$

This is the mean time to death for cancer patients at Kole-Bu.

E($\hat{\mu}$) $= \mu$, $\hat{\mu}$ is an unbiased estimator of μ

Log of $L(\pi) = rlog\lambda - \lambda \mathrm{V} + \mathrm{C}$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = \frac{r}{\boldsymbol{\lambda}} - V$$

Setting $\frac{\partial L}{\partial \boldsymbol{\lambda}} = 0$, $\frac{r}{\lambda} - V = 0 \Rightarrow \boldsymbol{\lambda} = \frac{\boldsymbol{r}}{\boldsymbol{V}}$

$\hat{\lambda} = \frac{204}{26783} = 0.007616$/month

There is 0.007616 hazard rate of cancer death per month amongst the patients who reported for treatment at Kole-Bu Teaching Hospital

Var $(\hat{\boldsymbol{\lambda}}) = \frac{\boldsymbol{r}}{\hat{\boldsymbol{\lambda}}^{\boldsymbol{2}}} = \frac{204}{5.80034\times10^{-05}} = 3517035$ (4.1)

At $\hat{\boldsymbol{\lambda}} = \frac{\boldsymbol{r}}{\boldsymbol{V}} = (\frac{V}{r})^{\boldsymbol{-1}}$this is the mean time to death

Maximum likelihood estimates are invariant under to one-to-one

$$\hat{\mu} = (0.007616)^{-1} = 131.730$$

The minimum variance unbiased estimate of μ is $\breve{\mu}$

$$\breve{\mu} = \sum_{i=1}^{r} a_i \frac{\boldsymbol{V}}{\boldsymbol{r}}$$

$\sum_{i=1}^{r} a_i = 1$for this reason, $a_i = \frac{\mathbf{1}}{\boldsymbol{r}} = \frac{\mathbf{1}}{\mathbf{204}} = 0.0049$ for all i

For this reason, $\breve{\mu} = \frac{\boldsymbol{V}}{\boldsymbol{r}} = 131.73$

Discussion

The hazard rate is given by 0.007616 as seen in Equation 4.10. This means that the death rate among the patients who reported for cancer treatment at Korle-Bu Teaching Hospital is 0.007616 per month. This death rate has been influenced by 80% censored observations. This affects the reliability and unbiased estimates of the hazard rate. The effect on censored values contributes to unequal interval hazard rate. As a result of this, there is an interval non constant variation. This shows in a large variance of 3,517,035 as shown in Equation 4.11. The estimate for this natural parameter (λ) may not be the one that makes inner product work nicely; there is therefore the need for an estimate of sufficient statistics. $\hat{\mu} = (0.007616)^{-1} = 131.730$given by Equation 4.12 is the sum of the times to death of patients who die on the study. It was against this discrepancy that (Kubler, 1994) used the mean time to death $\hat{\mu}$ vis-a-vis the exponential survival distribution to obtain $\hat{\mu} = \hat{\lambda}^{-1}$.

Article Publication Details

This article is published in the **UAR Journal of Multidisciplinary Studies (UARJMS)**, ISSN 3049-4346 (Online). In Volume 2 (2026), Issue 1 (January)

The journal is published and managed by **UAR Publisher.**

References

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40–49. https://doi.org/10.1002/mpr.329

Boahen, E. (2014). *Logistic model in reliability of TB drugs*. LAP Lambert Academic Publishing.

Chen, H. (2013). A Bayesian multiple imputation method for handling longitudinal data. *Environmetrics*, *24*(2), 132–142. https://doi.org/10.1002/env.2195

Chib, S. (1992). Bayes regression with autoregressive errors: A Gibbs sampling approach. *Journal of Econometrics, 51*(1-2), 79–102. https://doi.org/10.1016/0304-4076(92)90032-W

Čížek, P. (2002). Robust estimation with discrete explanatory variables. In J. Antoch (Ed.), *COMPSTAT: Proceedings in Computational Statistics* (pp. 509–514). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-57489-4_81

Coster, D. (2017, June 13). Driver who caused man's death is placed into dementia care. *Stuff*. https://www.stuff.co.nz/

Gibson, J. H. (1956). *Transformation of variables in multivariate analysis*. [Cited in text regarding Jacobian transformations].

Gospodinov, N., & Komunjer, I. (2013). *Misspecification-robust standard errors for the continuous variables*. Imperial College London. https://workspace.imperial.ac.uk/business-school/Public/Misspecification-robust.pdf

Kubler, J. (1994). *The exponential survival distribution and mean time to death*. [Cited in text regarding $\mu = \lambda^{-1}$ derivation].

Miyakoshi, Y. (2012). A missing value imputation method using a Bayesian network. *Electronics and Communications in Japan, 95*(2), 14–22. https://doi.org/10.1002/ecj.10391

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica, 55*(4), 765–799. https://doi.org/10.2307/1911029

Prinja, S., Gupta, N., & Verma, R. (2010). Censoring in clinical trials: Review of survival analysis techniques. *Indian Journal of Community Medicine, 35*(2), 217–221. https://doi.org/10.4103/0970-0218.66859

Singh, R., & Mukhopadhyay, K. (2011). Survival analysis in clinical trials. *Indian Journal of Pharmaceutical Sciences, 73*(3), 346–353. https://doi.org/10.4103/0250-474X.93514

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ, 338*, b2393. https://doi.org/10.1136/bmj.b2393

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag.

Wu, L. (2010). *Mixed effects models for complex data*. CRC Press.