

CENTRAL ASIAN JOURNAL OF MEDICAL AND NATURAL SCIENCES
https://cajmns.centralasianstudies.org/index.php/CAJMNS
Volume: 08 Issue: 01 | January 2026 ISSN: 2660-4159

Article

Nonparametric Estimation of Residual Extropy for Censored Data: An Applied Study at the Cancer Center for Lung Cancer Patients in Iraq 2025

Abbas Salem Kazim[1]

1. Department of Statistics, College of Administration and Economics, University of Sumer, Dhi-Qar, Iraq
* Correspondence: abbassalmm@gmail.com

Abstract: The research aims at filling a critical gap in the field of applied statistics, which is to come up with a measure of uncertainty in censored data, especially in survival analysis. The main aim of the research is to develop a non-parametric approach to estimating the "Residual Extropy". It is a new measure that acts as a complement to entropy, with a specific focus on the uncertainty in the tail part of the distribution, but considering censored data. Kaplan-Meier estimation approach was used to estimate the survival function, which helped to effectively deal with censored data. Following this, the kernel estimator was used to smooth the survival function and to estimate the probability density function. It is an important step that sets the stage for estimation of the extropy. The findings revealed a significant connection between the survival function, hazard (mortality) density, and residual uncertainty. We found that the level of uncertainty reaches its highest during the periods with the highest mortality rates (the peak of the mortality density), This proves that extropy is a sensitive and effective measure for capturing hazard dynamics. It was discovered that residual extropy estimation is a valuable technique when used in the non-parametric setting, giving medical practitioners and others in the field an ability to make better decisions through a more profound comprehension of uncertainty in survival data.

Keywords: Residual Extropy, Nonparametric Estimation, Censored Data, Survival Analysis, Kaplan–Meier Estimator, Kernel Estimator

Citation: Kazim, A. S. Nonparametric Estimation of Residual Extropy for Censored Data: An Applied Study at the Cancer Center for Lung Cancer Patients in Iraq 2025. Central Asian Journal of Medical and Natural Science 2027, 8(1), 122-133.

Received: 10th Aug 2026
Revised: 11th Seo 2026
Accepted: 24th Oct 2026
Published: 19th Nov 2026

Copyright: © 2027 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

In many critical fields such as medical research and reliability engineering life-or-death decisions rely on an accurate understanding of survival times. However, due to the complex nature of censored data, the fate of a large portion of the sample remains unknown, which hinders the ability to effectively measure the residual uncertainty in the system. Despite the fact that there is a new and accurate measure for this called residual extropy, the absence of proper methods to deal with such data is an important research problem. This shortage of appropriate tools can lead to inaccurate and misleading estimates of uncertainty. This may negatively affect clinical decision-making. Therefore, there is an urgent need to develop and evaluate nonparametric methodologies that can provide reliable and objective estimates of residual extropy, in order to fill this gap and offer practitioners a fundamental tool for making better decisions in the face of uncertainty. Measuring uncertainty in random variables is a fundamental pillar in many

research studies, such as engineering, statistical sciences, and medicine. The concept of entropy, introduced by Shannon, has been the most established measure of uncertainty in the context of statistics and information theory. Entropy is based on measuring the "expected average amount of information" and is defined as an indicator of how widely probabilities are spread within a distribution, where a higher value indicates greater uncertainty[1]. However, in recent years, the concept of extropy has emerged as a complementary measure to entropy for quantifying uncertainty. Introduced by Lad et al. the fundamental difference between the two lies in the fact that extropy does not measure uncertainty based on the spread of probabilities, but rather on the distance between the probability density function and the cumulative distribution function. Therefore, extropy demonstrates greater sensitivity to changes in the tails of the distribution, which makes it particularly important in survival studies, where the focus is on long-term time horizons and late-occurring deaths [2], [3].

Many researchers have shown interest in this area; for example, in 2012 S. M. Sunoj and M. N. Linu introduced a generalized measure known as the Cumulative Residual Renyi's Entropy and studied its properties. It was applied to various practical problems, such as weighted models and equilibrium models and they further extended this measure to the bivariate set-up [4]. In 2015 Frank Lad et al. investigated and identified the duality between entropy and extropy focusing on coarse and fine partitions. They also analyzed the extropy function in the context of probability densities, demonstrating that relative extropy constitutes a dual counterpart to the widely recognized Kullback–Leibler divergence, which is a measure of continuous entropy. Furthermore, they applied this concept statistically to the scoring of sequential forecast distributions which led to the discovery of extropy [5]. In 2021 Mehdi Bibak et al. compared the biota-sediment accumulation factor (BSAF) with the Empirical Cumulative Entropy (ECE), which is based on comparing the elemental information in algae with that in sediments across three species of algae and their associated sediments on the northern margin of the Persian Gulf. Algae samples were collected along a 5-kilometer stretch of coastline at each station and surface sediment samples were simultaneously taken from the low and middle intertidal zones. Their results indicated that the ECE index serves as a good alternative to BSAF[6]. In 2021 Nastaran Marzban Vaselabadi et al. conducted a study presenting results related to the dispersion measure of extropy for random variables. They focused on investigating several properties of this measure and discussed the concept of varextropy for residual and past lifetimes, order statistics, record values, and proportional hazard rate models. They also studied conditional varextropy and some of its properties. Furthermore, another stochastic ordering technique known as "varextropy ordering" has been developed along with some of its features [7]. The nonparametric estimates for extropy measures from length biased data have been proposed by 2025 R. Dhanya Nair and E. I. Abdul Sathar in 2025. It has been proven that the estimates are consistent and asymptotically normal. The efficiency of these estimates has been studied on simulated and empirical data sets with an application of marine pollution [8]. Consequently, the current research will use extropy as a statistical tool to deal with censored data, which may pose as a problem when dealing with survival times. Censored data can be explained as those which have been observed until a certain time and it is known for sure that the person is alive at this moment, but it is unknown where their life ends. In this study, the measure of residual extropy will illustrate certain times where there is a lot of uncertainty which coincide with the times when there is a peak mortality rate interval.The aim of this study is to develop a nonparametric methodology for estimating the residual extropy function in the presence of censored data, relying on the survival function estimated by the Kaplan–Meier estimator. Kernel smoothing is then applied to refine the survival function results and estimate the probability density function, which is a fundamental step for calculating residual extropy. The suggested methodological approach was implemented for real data (lung cancer data). The analysis of the results was done in order to clarify the

interconnection between the survival rate, mortality density, and the degree of residual uncertainty [9].

1-1**Incomplete data**

Incomplete data represent a major challenge in statistical analysis and arise in various situations. Among the most prominent types are: [10]

1-2 **Censored Data**

This occurs in time-to-event studies, particularly in survival analysis, where only a portion of the complete event time is observed [11]. For instance, the study might be terminated before the event (such as death of the patient) occurs, thus making the observation incomplete [12].

2. Materials and Methods

Survival Function for Censored Data

Let X be a random variable with survival time X, and let S(x) = P(X>x) be the probability that someone will survive more than time x. The Kaplan–Meier estimate is used to measure the survival probability in cases when there is censored data. This is done by accounting for both the time at which events and censoring occurs.

2-Probability Density Function f(x)

The probability density function f(x) is defined as the derivative of the cumulative distribution function F(x). In the context of censored data, it is estimated using the kernel estimator, which smooths the stepwise survival function produced by the Kaplan–Meier estimator, resulting in a smooth estimate of f(x). The kernel estimator is particularly used for calculating extropy, which fundamentally relies on the integral of the square of the density function.

Survival Function for Censored Data

Let X be a random variable representing survival time, and let the survival function $S(x) = P(X > x)$ denote the probability that an individual survives beyond time xxx. To estimate S(x) in the presence of censored data, the Kaplan–Meier estimator is used. This estimator accounts for both event times and censoring times, producing a step function that serves as the basis for subsequent analyses **Probability Density Function f(x)**

The probability density function f(x) is known to be a derivative of the cumulative distribution function F(x). In case of censored data , f(x) is estimated via the kernel estimator that smoothes the stepwise survival function calculated via the Kaplan-Meier estimator. This method is applied specifically for the calculation of extropy since the latter depends on the integral of the squared density function.

Extropy

Extropy is a measure of uncertainty in a probability distribution and serves as the dual concept to entropy. While entropy expresses the amount of information provided by a random variable, extropy can be interpreted as a measure of the information that is lost or consumed when observing the value of the random variable. The extropy J(X) for a continuous random variable X with probability density function f(x) is defined by the following expression:

$$J(X) = -\int f(x)^2 dx \qquad (1)$$

Residual Extropy J(X;t)

Residual extropy at time t is a measure of the remaining uncertainty in the lifetime of a system beyond time t. It is defined by the expression:

$$J(X;t) = \frac{1}{2}\int_0^{\infty} f_t^2(x)dx \qquad (2)$$

where $f_t(x) = \frac{f(x+t)}{S(t)}$ is the conditional probability density function of the residual lifetime. This value reflects the degree of remaining uncertainty, and is practically computed using the estimated survival function and density function.

-Nonparametric Estimation Methods

Nonparametric estimation methods aim to estimate the characteristics of a probability distribution without the need to assume any specific form for it, which makes them flexible and suitable for handling complex or irregular data.

- Kernel Estimator

The kernel estimator is used to estimate the probability density function f(x) in a nonparametric manner. This estimator relies on defining a "window" around each data point and uses a kernel function to assign greater weight to observations that are closer to the point at which the estimate is made. The kernel density estimator $\widehat{f_n}(x)$ for the probability density function f(x) is given by the following expression [13]:

$$\widehat{f_n}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \qquad (3)$$

where:

n is the sample size.

h is the bandwidth, which controls the smoothness of the estimate.

K is the kernel function (e.g., the Gaussian kernel).

This estimator can be used to estimate extropy by substituting f(x) in the extropy formula with $\widehat{f_n}(x)$.

2-9 Kaplan–Meier Estimator

The Kaplan–Meier estimator (also known as the product-limit estimator) is one of the leading methods for estimating the survival function in the presence of censored data. It estimates the probability of survival beyond a given time point by multiplying the conditional survival probabilities at each time an event occurs. The Kaplan–Meier estimator for the survival function S(t) is given by the following expression

$$\hat{S}(t) = \prod_{t_i \le t} (1 - \frac{d_i}{n_i}) \qquad (4)$$

where:

t_i are the times at which deaths or events occur.

d_i is the number of individuals who experienced the event at time t_i.

n_i is the number of individuals at risk at time t_i

3- Data

Data were obtained from the Cancer Center in Dhi Qar Governorate / Nasiriyah, with a sample size of 83 observations. The dataset included two key variables for analysis: "Time," which represents the patient's survival time in days, and "Status," which indicates whether the patient had died or whether the data were censored.

The analysis and extraction of results were conducted using the MATLAB programming language. The Kaplan–Meier estimator was applied to handle the censored data, and the kernel estimator was used to accurately estimate the probability density function.

Finally, these consistent estimates will be used to calculate the residual extropy, where this approach contributes to performing precise and meaningful analysis of censored data within the framework of nonparametric estimation methods. This, in turn, supports making well-informed decisions based on the derived results.

3. Results

4Kaplan–Meier Survival Estimate Results

As seen from the graph, the probability of patient survival gradually declines with time. It should be noted that the findings can be described in several stages.Interpretation of Kaplan-Meier Survival Estimate Results[12]

Table 1 demonstrates how the chances of survival decline with time and accurately captures the key points in the course of the disease. The findings can be categorized into several stages.:

1. Early Phase (Day 3 to Day 14)

Initial Decline: The survival function starts at a value of 1 (100% of patients alive) at baseline but shows a rapid decrease in the first few days, with the daily decline reaching approximately 2.4% on days 3 and 4.

Critical Period: The critical phase begins on day 10, where a sharp drop in survival probability is observed, reaching 4.8% within just two days, indicating a critical period of increased mortality rate[13].

Table 1. Kaplan-Meier Survival Estimate

Time	S(t)	ΔS	Time Interval	Clinical Interpretation
3	1.0000	-	Baseline	Start of Observation – All Patients Alive
3	0.9759	-0.0241	Early Drop	First Death (2.4% Decrease)
4	0.9518	-0.0241	Early Phase	Steady Decline in Survival
5	0.9398	-0.0120		Slowdown in Mortality Rate
8	0.9277	-0.0121		
10	0.8795	-0.0482	Critical Phase	Sharp Decline (4.8% in Two Days)
11	0.8313	-0.0482		Continued Sharp Decline
12	0.7831	-0.0482		
13	0.7711	-0.0120		Temporary Slowdown
14	0.7590	-0.0121		
15	0.6978	-0.0612	Crisis Point	Largest Daily Decline (6.12%)
16	0.6733	-0.0245		
17	0.6611	-0.0122		
18	0.6366	-0.0245		
19	0.6244	-0.0122		
20	0.5994	-0.0250	Milestone	Survival Drops Below 60%
21	0.5744	-0.0250		
22	0.5619	-0.0125		
23	0.4995	-0.0624	Second Crisis	Second Largest Daily Decline (6.24%)
24	0.4870	-0.0125		
25	0.4620	-0.0250		
26	0.4371	-0.0249		
27	0.4246	-0.0125		
28	0.4121	-0.0125		
30	0.3871	-0.0250		
31	0.3621	-0.0250		
33	0.3496	-0.0125		
35	0.3372	-0.0124		
39	0.3247	-0.0125		
40	0.3122	-0.0125	Late Phase	Survival Drops Below 35%
41	0.2997	-0.0125		
42	0.2872	-0.0125		
43	0.2747	-0.0125		
48	0.2622	-0.0125		
67	0.2491	-0.0131		
82	0.2345	-0.0146		
84	0.2198	-0.0147		
100	0.2052	-0.0146	Long-term	20% Survival After 100 Days
103	0.1905	-0.0147		
110	0.1612	-0.0293		
118	0.1451	-0.0161		
139	0.1290	-0.0161		
153	0.1128	-0.0162		
155	0.0967	-0.0161		
190	0.0806	-0.0161	Very Late Phase	8% Survival — Occurred 8 Times
314	0.0604	-0.0202	End Phase	
326	0.0403	-0.0201	Final	Lowest Recorded Survival (4.03%)

2. Crisis Points (Days 15 and 23):

• First Critical Point (Day 15): It can be seen from the table that the maximum daily fall is registered on day 15, when there is a dramatic fall of 6.12%. This critical point is the maximum daily death rate in the group.

• Second Critical Point (Day 23): After a temporary halt, a second sharp fall is registered, which is the second maximum daily fall of 6.24%[14].

3. Late Stage (Day 40 and Beyond):

• Reduced mortality rate: From about the third week onwards, the rate of decrease in survival rate gets stabilized and reduces in intensity as compared to the critical period [15].

• Long-term survival: It can be seen from the table that some patients live for long periods. Around 20% of patients survive for over 100 days.

It can be seen from the Kaplan-Meier survival estimate that the progression of the disease has a critical period in its early days (specifically between day 10 and 25), where there is a drastic reduction in survival rate. Post the critical period, the survival rate starts to fall, although at a slower rate [16].

4-**2**Commentary on the Results of Kernel Density Estimation

The table shows how the probability density of death f(t) varies over time, giving more insight into those periods when mortality rate is the highest. The results may be viewed from different stages:

1. Early Stage of Risk (Day 3 – Day 17):

Start of Risk: The risk density starts at zero point; thereafter, it grows quickly starting from Day 3 due to the emergence of the first death cases.

Slow Growth: During the next days, there is gradual growth of the density, with consistently positive daily changes showing that the mortality rate grows slowly.

2. High Stage of Risk (Day 18 - Day 30):

Beyond the Risk Threshold: On Day 18, the value of f(t) is higher than 1, which shows the beginning of the "high-risk" period.

Risk Maximum: The density achieves its maximum on Day 23, when there is a highest daily change of +12.41%, thus, marking the highest mortality rate period. It corresponds to the "crisis point" found in the Kaplan-Meier survival function analysis [17].

3. Absolute Peak Phase (from Day 31 to Day 43):

Late Maximum: The density continues growing in an oscillatory way until it reaches its temporary "maximum peak" value of 1.8748 on Day 40.

Second Absolute Maximum: The density achieves its maximum value on Day 43, which equals 2.0497. This second peak denotes the highest mortality rate of the whole period of observation.

4. Final Phase (From Day 44 Till the End of the Study):

Gradual Decline: After reaching its peak on Day 43, the value of f(t)f(t)f(t) tends to decrease gradually, signaling the decreasing trend in mortality rates.

Irregular Spikes: Irregular spikes in density occur later on, for example, on Day 326. These irregular spikes are due to occasional death events that happen after a long time without any deaths, which could impact the density values at those points [18].

Table 2. Kernel Density

Time	f(t)	Trend Compared to Previous Stage	Risk Level	Epidemiological Interpretation
0	0.0000	-	There is no risk	Start of the Study – No Deaths Recorded
1	0.0000	0%		
2	0.0000	0%		
3	0.6006	+∞	Early risk	First Deaths Recorded
4	0.6234	+3.80%		
5	0.6387	+2.45%		
6	0.6457	+1.10%		
7	0.6523	+1.02%		
8	0.6671	+2.27%		
9	0.6729	+0.87%		
10	0.7156	+6.35%	Increasing risk	
11	0.7626	+6.57%		
12	0.8150	+6.87%		
13	0.8326	+2.16%		
14	0.8503	+2.13%		
15	0.9292	+9.28%		
16	0.9667	+4.04%		
17	0.9877	+2.17%		
18	1.0283	+4.11%	High risk	Exceeding the Hazard Threshold of 1.0
19	1.0504	+2.15%		
20	1.0954	+4.28%		
21	1.1436	+4.40%		
22	1.1687	+2.20%		
23	1.3137	+12.41%	Peak risk	Highest Mortality Rate
24	1.3453	+2.40%		
25	1.4148	+5.17%		
26	1.4914	+5.41%		
27	1.5298	+2.57%		
28	1.5696	+2.60%		
29	1.5620	-0.48%		
30	1.6536	+5.86%		
31	1.7568	+6.24%		
32	1.7450	-0.67%		
33	1.7939	+2.80%		
34	1.7796	-0.80%		
35	1.8295	+2.80%		
36	1.8126	-0.92%		
37	1.7948	-0.98%		
38	1.7760	-1.05%		
39	1.8239	+2.70%		
40	1.8748	+2.79%	Maximum peak	Highest Hazard Density Value (1.8748)
41	1.9290	+2.89%		
42	1.9871	+3.01%		
43	2.0497	+3.15%	Absolute maximum value	Second Peak (2.0497)
44	2.0211	-1.40%		
45	1.9919	-1.44%		
46	1.9620	-1.50%		
47	1.9315	-1.55%		
48	1.9909	+3.08%		
49	1.9579	-1.66%		
50	1.9244	-1.71%		
...	...	...	...	All values inclusive
326	0.5413	+47.24%	Late increase	All values inclusive
...	...	...	...	
411	0.0228	-96.32%	End of the study	Start of the study – no deaths

4-3Commentary of the results Extropy Results J(Xt)

The remaining extropy J(Xt) represents a measure of the residual uncertainty about the system's lifetime after elapsed time t. The greater the absolute value of J(Xt) (i.e., the more negative it is), the higher the uncertainty. This table demonstrates how uncertainty varies over time, demonstrating periods of certainty and uncertainty.

1. Phase of Onset and Rapid Rise of Uncertainty (Days 10 – Days 50)

Onset of Uncertainty: The value begins with NaN (undefined) on Day 0, then appears at a relatively low value on Day 10, which marks the onset of uncertainty.

Acceleration: Uncertainty grows rapidly and sharply. It can be seen from the table that there is a doubling of uncertainty values between Days 10 and 20, and rapid growth between Days 20 and 30.

Peak Value: Uncertainty becomes very high around Days 40 and 50, thus this period was the most ambiguous during the first phase of the experiment [19].

2. Phase of Fluctuations and Multiple Peaks (From Day 60 to Day 200):

Unexpected Variability: After the first peak, the uncertainty level becomes stable and starts to fluctuate in an unpredictable manner, implying that the system faces a period of instability.

Multiple Peaks: Three peaks in uncertainty occur during this phase:

Second Peak (Day 160): Demonstrates a significant increase of 50% from the previous level.

Third Peak (Day 190): Demonstrates a further increase of 30%.

3. Period of Maximum Peak and Anomaly (Days 320 to 330):

Maximum Value Ever Recorded: The data reveal that there is an unusual and extreme peak in uncertainty, the absolute value of which reaches the maximum ever recorded on Day 320 (-1192.3874).

An Extreme Anomaly: On Day 330, an extremely anomalous value arises (-2059.9406). This extreme value might be because of the sharp decline in the population of surviving individuals, wherein the death of even one individual results in the extreme increase in uncertainty.

4. Final Stage (from Day 340 until the end of the experiment):

A Sudden Fall in Uncertainty: After the unexpected spike, there is a sudden and consistent fall in the level of uncertainty. The positive change in percentage values is very high, even 95%, signaling the rapid dissipation of uncertainty [20].

Final Equilibrium: At the end of the experiment, the final value of extropy becomes close to zero (–0.1719), implying that uncertainty about the future of the survivors becomes very small.

Table 3. Residual Extropy J(X;t) at Selected Time Points

Time	$J(X;t)$	Relative Change (%)	State of uncertainty	Statistical Remarks
0	NaN	-	Undefined	Start of the Study – No Data
10	-149.4194	-	Low	Beginning of uncertainty emergence
20	-309.9527	-107.45%	Medium	Doubling of uncertainty
30	-677.5621	-118.60%	High	Sharp acceleration in uncertainty
40	-877.7713	-29.55%	Very high	Approaching the first peak
50	-961.0958	-9.49%	Critical condition	Close to the maximum values
60	-737.0843	+23.30%	High	Significant decrease in uncertainty
70	-650.4147	+11.76%	High	
80	-540.8046	+16.85%	Medium-to-high	
90	-592.6732	-9.59%	High	Sudden increase
100	-601.7408	-1.53%	High	
110	-865.9180	-43.90%	Very high	Large jump in uncertainty
120	-926.3972	-6.99%	Critical condition	
130	-802.3073	+13.39%	Very high	
140	-892.4356	-11.24%	Critical condition	
150	-779.3224	+12.67%	Very high	
160	-1169.8218	-50.11%	Maximum uncertainty	Second peak
170	-983.8309	+15.90%	Critical condition	
180	-857.8019	+12.81%	Very high	
190	-1116.4365	-30.16%	Maximum uncertainty	Third peak
200	-1014.3098	+9.15%	Critical condition	
210	-956.0782	+5.74%	Very high	
220	-924.8188	+3.27%	Very high	
230	-908.6217	+1.75%	Very high	
240	-899.8492	+0.97%	Very high	
250	-893.9745	+0.65%	Very high	
260	-888.1493	+0.65%	Very high	
270	-879.7756	+0.94%	Very high	
280	-865.5627	+1.62%	Very high	
290	-841.6650	+2.76%	High	
300	-805.1819	+4.33%	High	
310	-756.3423	+6.07%	Medium-to-high	
320	-1192.3874	-57.65%	Exceptional	Highest value ever recorded
330	-2059.9406	-72.76%	Abnormal	Anomalous value
340	-1283.2608	+37.71%	Exceptional	
350	-709.7676	+44.69%	High	
360	-345.8303	+51.27%	Medium	
370	-147.3702	+57.38%	Low–medium	
380	-54.3846	+63.09%	Low	
390	-16.9561	+68.82%	Very low	
400	-4.0149	+76.33%	Stable	
410	-0.1719	+95.72%	Very stable	End of the study

Figure 1. Integrated Analysis of Patient Survival Dynamics: Survival Function, Mortality Density, and Remaining Uncertainty

Figure 1 presents an integrated system of interconnected functions based on the analysis of lung cancer. Each function in the diagram represents a certain aspect of survival analysis of patients.

The first plot (Kaplan-Meier survival estimate) starts with a steep fall within the first days of the experiment, which indicates that most deaths occurred during the first period of the study. After about 50 days, the line becomes stable, which means that those patients who survive the initial period have a much longer life expectancy , [21]. The middle panel (kernel density function) illustrates when deaths occurred at a high rate. The peak that is clearly visible in this graph from Days 20 through 50 coincides exactly with the sharp drop noted in the survival function in the upper graph. This proves that mortality was at its highest rate at the initial stage of the illness.

The third graph measures the degree of uncertainty – the remaining extropy. Sharp spikes in the level of uncertainty (sharp decrease in the graph) correspond exactly to the peaks in the graph of the kernel density function. As mortality was high – as seen in the range between Days 20 and 50 – there was much uncertainty associated with the further development of the disease.

More specifically, the abnormal spike in the level of uncertainty around Day 320, which corresponds to an upward spike in the kernel density function, indicates that there was a rare event, namely the death after long periods of stability, that has greatly increased the level of uncertainty due to the unpredictable further life prospects of a small number of patients left [22].

To conclude, all these three graphs illustrate the point that the levels of uncertainty were highest in cases when mortality rate was highest in the lung cancer dataset.

4. Discussion

The findings bring together three views of the observed survival experience. The Kaplan–Meier curve shows the main decline in survival early in follow-up, while the kernel density estimate places its reported largest peak at day 43. Residual extropy is strongly negative around days 40–50 and shows a further pronounced excursion around days 320–330. Taken together, these patterns suggest that estimated residual uncertainty is elevated during periods of concentrated mortality and can also change sharply when very few patients remain under observation. The timing is not identical across the summaries, so the results are best interpreted as related patterns rather than exact point-by-point agreement.

The late follow-up results require particular caution. The sample contains 83 patients, and the manuscript reports that only about one fifth survived beyond 100 days. When the number at risk is small, individual events can strongly affect smoothed density and residual-extropy estimates. The extreme values around days 320–330 may therefore reflect sparse tail data as well as the underlying survival pattern; they should not be read as stable estimates of patient-level risk or as direct evidence for treatment decisions.

This analysis is descriptive and based on one center's dataset. The manuscript does not report confidence intervals, a sensitivity analysis for the kernel bandwidth, or validation in an independent cohort, which limits assessment of precision and reproducibility. Future work could quantify uncertainty with resampling, examine the sensitivity of the estimates to smoothing choices, account for relevant patient characteristics, and compare the patterns in additional cohorts. These steps would help determine whether residual extropy adds information beyond the survival and density summaries for clinical research.

5. Conclusion

It can be summarized that the following main conclusions have been drawn from the analysis of the results: Concentration of the Risk at the Initial Stage: It is revealed that the maximum rate of deaths among the patients with the diagnosed pulmonary cancer occurs at the initial stage of the analysis (around Days 20 to 50). Such a conclusion is based on the rapid fall of the Kaplan–Meier survivor function as well as the first significant peak of the kernel density function within the same period. Direct Relationship between the Mortality Density and the Level of Uncertainty: There is a direct relationship between the mortality density function and the uncertainty level as it is revealed in the negative values of remaining extropy. Thus, extropy is the good and appropriate metric for revealing the dynamics of risks in censored data. Efficacy of the Nonparametric Approach: The results allow to conclude that the nonparametric approach, which is implemented using such techniques as Kaplan–Meier estimator and the kernel density estimator, is efficient enough for the analysis of censored data. Practical Applications: Such results represent an important instrument for decision making in applied fields. By identifying periods of higher risk and greater uncertainty, practitioners can design better treatment strategies or interventions during those critical phases. We would like to express our thanks to the staff of Cancer Oncology Center / Nasiriyah /Dhi Qar Governorate /Iraq for their effort for

providing data access. We declare that we have no conflicts of interest to disclose regarding this manuscript. This manuscript received no specific grant from any funding agency in the public or commercial sectors.

REFERENCES

[1]. M. Bibak, S. Tahmasebi, M. Sattari, R. Kafaei, and B. Ramavandi, "Empirical cumulative entropy as a new trace elements indicator to determine the relationship between algae-sediment pollution in the Persian Gulf, southern Iran," Environ. Sci. Pollut. Res., vol. 28, no. 4, pp. 4634–4644, 2021, doi: 10.1007/s11356-020-10838-5.

[2]. F. Lad, G. Sanfilippo, and G. Agrò, "Extropy: Complementary dual of entropy," Stat. Sci., vol. 30, no. 1, pp. 40–58, 2015, doi: 10.1214/14-STS430.

[3]. R. D. Nair and E. I. Abdul Sathar, "Nonparametric Estimation of Extropy-Related Measures with Length-Biased Data," Stat. Appl., vol. 23, no. 1, pp. 321–334, 2025.

[4]. S. M. Sunoj and M. N. Linu, "Dynamic cumulative residual Renyi's entropy," Statistics (Ber)., vol. 46, no. 1, pp. 41–56, 2012, doi: 10.1080/02331888.2010.494730.

[5]. N. M. Vaselabadi, S. Tahmasebi, M. R. Kazemi, and F. Buono, "Results on varextropy measure of random variables," Entropy, vol. 23, no. 3, pp. 1–14, 2021, doi: 10.3390/e23030356.

[6]. S. Tahmasebi and A. Toomaj, "Communications in Statistics - Theory and Methods On negative cumulative extropy with applications," 2020.

[7]. M. J. H. Rawa, D. W. P. Thomas, and M. Sumner, "Simulation of non-linear loads for harmonic studies," Proceeding Int. Conf. Electr. Power Qual. Util. EPQU, vol. 00037, pp. 102–107, 2011, doi: 10.1109/EPQU.2011.6128915.

[8]. D. A. N. Njomen and J. W. Ngatchou, "Consistency of the Kaplan-Meier Estimator of the Survival Function in Competiting Risks," Open Stat. Probab. J., vol. 9, no. 1, pp. 1–17, 2018, doi: 10.2174/1876527001809010001.

[9]. E. L. Kaplan and P. Meier, "Nonparametric Estimation from Incomplete Observations," J. Am. Stat. Assoc., vol. 53, no. 282, p. 457, 1958, doi: 10.2307/2281868.

[10]. D. Centre, "Smoothing Parameter Selection for Nonparametric Density Estimation for Length-biased Data : A Bayesian Perspective," 2015.

[11]. Y. P. Chaubey, "On An Entropy Estimator Based On a Non-parametric Density Estimator For Non-negative Data," no. 2, 2020.

[12]. A. K. Nanda, P. G. Sankaran, and S. M. Sunoj, "Rényi's residual entropy: A quantile approach," Stat. Probab. Lett., vol. 85, no. 1, pp. 114–121, 2014, doi: 10.1016/j.spl.2013.11.016.

[13]. M. Askari, F. Yousefzadeh, and S. Jomhoori, "GENERALIZED EXTROPY OF k -RECORDS: PROPERTIES AND APPLICATIONS," vol. 15, no. 1, pp. 89–110, 2026.

[14]. I. Gijbels, "Kaplan-Meier Estimator," Int. Encycl. Stat. Sci., pp. 709–710, 2011, doi: 10.1007/978-3-642-04898-2_322.

[15]. Y. C. Chen, "A tutorial on kernel density estimation and recent advances," Biostat. Epidemiol., vol. 1, no. 1, pp. 161–187, 2017, doi: 10.1080/24709360.2017.1396742.

[16]. "Cancer Oncology Center / Nasiriyah /Dhi Qar Governorate /Iraq"

[17]. F. Buono, O. Kamari, and M. Longobardi, "Interval extropy and weighted interval extropy," Ric. di Mat., vol. 72, no. 1, pp. 283–298, 2023, doi: 10.1007/s11587-021-00678-x.

[18]. A. S. Krishnan, S. M. Sunoj, and P. G. Sankaran, "Some reliability properties of extropy and its related measures using quantile function," Statistica, vol. 80, no. 4, pp. 413–437, 2020, doi: 10.6092/issn.1973-2201/9887.

[19]. H. H. Qubbaj, H. A. Bayoud, and H. M. Hilow, "Extropy and entropy estimation based on progressive Type-I interval censoring," Stat. Transit. New Ser., vol. 25, no. 3, pp. 83–102, 2024, doi: 10.59170/STATTRANS-2024-029.

[20]. S. M. Sunoj and P. G. Sankaran, "Quantile based entropy function," Stat. Probab. Lett., vol. 82, no. 6, pp. 1049–1053, 2012, doi: 10.1016/j.spl.2012.02.005.

[21]. M. Goel, P. Khanna, and J. Kishore, "Understanding survival analysis: Kaplan-Meier estimate," Int. J. Ayurveda Res., vol. 1, no. 4, p. 274, 2010, doi: 10.4103/0974-7788.76794.

[22]. I. Etikan, "The Kaplan Meier Estimate in Survival Analysis," Biometrics Biostat. Int. J., vol. 5, no. 2, pp. 55–59, 2017, doi: 10.15406/bbij.2017.05.00128.