Original Article
A modern approach of survival analysis of patients with pancreatic cancer

Aditya Chakraborty, Chris P Tsokos

Department of Mathematics & Statistics, University of South Florida, Florida 33620, USA

Abstract: Pancreatic cancer is one of the deadliest diseases and becoming an increasingly common cause of cancer mortality. It continues to give rise to massive challenges to clinicians and cancer researchers. One of the main goals of our present study is to determine if there exists any statistically significant difference in the survival probabilities of male and female pancreatic cancer patients in different cancer stages and irrespective of stages. Another goal is to investigate if there exists any parametric probability distribution function that best fits the male and female patient survival times in different stages of cancer, irrespective of stages, and compare the survival probabilities with the non-parametric Kaplan-Meier (KM) method. We employed both parametric and non-parametric statistical approaches to examine the survival probabilities of 10,000 patients diagnosed with pancreatic cancer and showed that there is no significant difference in male and female survival times at any stage except stage IV. We also found no evidence of a statistically significant difference in overall mean survival durations between male and female pancreatic cancer patients, regardless of stage. We used parametric survival analysis and identified the Generalized Pareto (GP) probability distribution as the best fit to the overall survival data for pancreatic cancer patients. Also, we identified the appropriate probability distributions for patients in different cancer stages. We then estimated the overall survival probabilities and compared them with the frequently used non-parametric Kaplan-Meier (KM) survival method, which is not as powerful as our parametric analysis. An assessment of the survival probability estimates generated by the two procedures found that the parametric method produced a better survival probability estimate than the Kaplan-Meier approach. We further compared the median survival times of patients using descriptive, parametric, and non-parametric techniques of analysis and found that the results were relatively consistent. We found that parametric survival analysis is more reliable and efficient than non-parametric Kaplan-Meier estimates since it is based on a well-defined parametric probability distribution.

Keywords: Pancreatic cancer, parametric survival functions, Generalized Pareto (GP) probability distribution, Probability-Weighted Moments (PWM) estimates

Introduction

Pancreatic cancer is still one of the most lethal diseases affecting human mortality, and it remains a serious and intractable health issue at the start of the twenty-first century. It is estimated that this disease kills approximately 30,000 people in the United States each year [1]. Researchers discovered that pancreatic cancer is the fourth leading cause of mortality from cancer in the United States and accounts for an estimated 227,000 fatalities worldwide each year. The incidence and mortality rates from pancreatic tumors have gradually increased, although the incidence and mortality rates from other prevalent cancers have

decreased. "In spite of the developments in the detection and management of pancreatic cancer, it is estimated, approximately 4% of patients will live five years after diagnosis" [2]. The pancreas is made up of digestive enzyme-secreting acinar cells, bicarbonate-secreting ductal cells, Centro acinar cells (the geographical transition between acinar and ductal cells), hormone-secreting endocrine islets, and largely dormant stellate cells. Adenocarcinomas are the most lethal pancreatic neoplasms. "Rare pancreatic neoplasms include neuroendocrine tumors (responsible for the secretion of hormones like insulin or glucagon) and acinar carcinomas (which can release digestive enzymes into the circulation). Precisely, ductal adenocarcinoma

is the most common malignancy of the pancreas; this tumor (commonly referred to as pancreatic cancer) presents a substantial health problem, with an estimated 367,000 new cases diagnosed worldwide in 2015 and an associated 359,000 deaths in the same year" [3, 4]. After the detection of pancreatic cancer, doctors usually perform some additional tests to understand better if cancer has been spread or the locations of spreading areas of cancer. Imaging tests, such as a PET scan, assist doctors by identifying the presence of cancerous growths. With these tests, doctors try to establish the cancer stage of a given patient with pancreatic cancer. Staging helps explicate the advancement of cancer. It also assists doctors in deciding treatment options. Once a diagnosis has been made, the doctor assigns a stage to the patient depending on the results of the following tests:

• Stage I: Tumors exist solely in the pancreas.

• Stage II: Tumors have spread to adjacent abdominal tissues or lymph nodes.

• Stage III: Cancer has spread to major blood vessels and lymph nodes.

• Stage IV: Tumors have spread to other organs, such as the liver, lung, bone, etc.

Despite pancreatic cancer remains incurable in most situations, most researchers studying this kind of cancer have concentrated on how to enhance the survival periods of individuals diagnosed with pancreatic cancer at various stages. Recently, clinical researchers are using the Kaplan-Meier (KM) method extensively in order to analyze the clinical data. Based on the log-rank test, this method is often used in health sciences to compare the survival differences of two or more groups of patients. Our work gives a parametric and non-parametric survival analysis of Pancreatic Cancer patients' survival periods. Identifying the unique probability distribution that characterizes the probabilistic behavior of the survival times is vital for any real-world phenomenon. For our case, we can proceed to find the analytical form of the survival function of the data, driven by the specific probability distribution. This sort of parametric analysis is typically more potent than non-parametric analysis. "Feigl and Zelen [5] have shown that the assumption of exponential

distribution works well for studying some of the survival of cancer-related studies" [6-8]. Assuming such a probability distribution without rationale, on the other hand, may result in misleading results. As a result, it is critical to determine the accurate probability distribution of patient survival times based on gender, race, and other factors. In this study, we identify the probability distribution that best fits the survival times, and then we proceed to acquire the survival functions of male and female patients in four cancer stages. We also made a comparison of our parametric probability estimates with the frequently used Kaplan-Meier (KM) method. The structural organization of the paper is as follows: In Section 2.1, we provide the data discussion and perform the non-parametric Wilcoxon test to investigate if there exists any significant difference between the male and female patients at any individual stage. In section 2.2, we discuss the stage-based descriptive analysis with graphical representation. In section 3, we discuss specific details about the parametric survival analysis of pancreatic cancer patients at different stages. In section 4, we investigate the significant difference in overall survival times of male and female patients by log-rank test [9, 10] and discuss in detail the overall parametric survival analysis of patients irrespective of stages. We also describe the parameter estimation procedure of GP probability distribution in Section 4.3 elaborately. Section 5 describes the KM estimate and compares patients' median survival times using descriptive, parametric, and non-parametric techniques. Section 6 compares the Generalized Pareto (GP) probability distribution with non-parametric KM estimates for patient survival probability estimates. Sections 7 and 8 provide results & discussion, and conclusion, respectively.

Methodology

Brief description of data

The study data has been collected from the Surveillance, Epidemiology, and End Results (SEER) database. It includes information on the survival times of patients with pancreatic adenocarcinoma. We are interested in each patient's survival time (in months) and cause-specific mortality (deaths owing to pancreatic cancer). The patient survival time is one of the
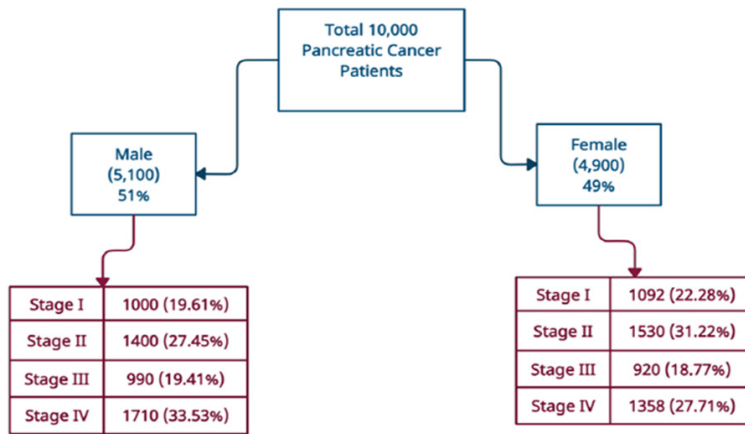
Figure 1. Pancreatic cancer data sorted by gender and stages.

Table 1. Wilcoxon test results for different stages, classified by gender

Stages	P-Values	Result
I	0.75	Difference does not Exist
II	0.25	Difference does not Exist
III	0.84	Difference does not Exist
IV	0.001	Difference Exists

male and female survival times in stage I. That is, $\mu_M \neq \mu_F$.

After we analyze the data for male and female patients in each stage, we proceed to perform the combined analysis for all stages, classified by gender. **Table 1** illustrates the test results along with the p-values in different stages for male and female pancreatic cancer patients.

As the results of **Table 1** suggest, there does not exist a significant difference between the male and female pancreatic cancer patient survival times at stage I, stage II, and stage III. However, at stage IV, the difference is significant. In the next section, we proceed to identify the parametric probability distributions and survival functions of the survival times of patients, along with some important descriptive statistics.

Descriptive analysis of pancreatic cancer patients in different stages-A gender-based classification

Descriptive statistics provide a basic picture of different cancer stages before performing any sophisticated statistical methods on the data.

Table 2 illustrates the different descriptive statistics for male and female patients in four different stages.

We now proceed to identify the most appropriate probability distributions that drive the survival times of patients in different stages (I, II, III, and IV), classified by gender. We came to know from the last section that there does not exist any significant difference between male and female survival times in stages I, II, and III. However, we found a significant difference in survival times of male and female patients in stage IV. We have obtained the best fits for each stage and estimated their individual parameter estimates. Identification of the most suitable probability distribution is crucial since it gives better survival probability estimates for both male and female patients in each of the stages that are driven by the specific probabili-

most important components considered in research relating to cancer. It is vital to assess the extent of cancer, which aids in determining the prognosis and determining the best treatment techniques. We considered a random sample of 10,000 patients diagnosed with pancreatic cancer, including males and females. A schematic data diagram used in this study with necessary attributes is shown in **Figure 1**. As the following figure describes, in our dataset, we have information on survival times regarding 5,100 male and 4,900 female patients diagnosed with pancreatic cancer.

Before we begin the parametric analysis of patient survival times, we must first determine whether there is a statistically significant difference in the true survival times of genders, i.e., male and female patients in different stages of cancer. We employ the two-sample Wilcoxon test with the following hypothesis for this purpose.

H_0: There is no significant difference between the true mean survival times of male (μ_M) and female (μ_F) patients at stage $i.i$ = 1, 2, 3, 4. That is, $\mu_M = \mu_F$. Vs. H_1: Differences exist between

Table 2. Descriptive statistics of survival time (in months) of pancreatic cancer patients classified by gender in different stages

Gender	Mean	Median	Std. Dev.	Skewness	Kurtosis	Std. Error
Combined (Stage I)	30.6	20	31.5	1.33	1.14	0.76
Combined (Stage II)	21.44	14	23.50	2.14	5.10	.33
Combined (Stage III)	16.92	8	14.71	3.73	20.01	.37
Male (Stage IV)	6.7	3	12.73	4.78	30.44	.18
Female (Stage IV)	7.50	3.11	13.67	4.63	27.80	.20

ty distribution. Once we obtain the parametric estimates from the probability distributions at each of the stages, for male and female patients, we can derive the probability density functions (pdfs), cumulative distribution functions (cdfs), and parametric survival function driven by the specific probability distribution.

Parametric analysis of pancreatic cancer survival time for different stages

Johnson (1949) [28] proposed systems of different frequency curves based on transformations of the following form:

$$z = \gamma + \delta\, f\left(\frac{x - \zeta}{\lambda}\right)$$

where z is a standard Normal variable, f is a function taking different forms S_L, S_B, and S_U. Our data in stage I follows Johnson S_B Probability distribution with parameters γ (shape parameter), δ (shape parameter), ζ (location parameter), and λ (scale parameter). In stage II, and stage III, the data follows a generalized extreme value (GEV) probability distribution. Chakraborty & Tsokos [27] describes in detail the parameter estimation procedure of acute myeloid leukemia cancer data modeled by GEV probability distribution using probability-weighted moment (pwm). In stage IV, the data follows a generalized Pareto (GP) probability distribution. In section 4.3, we discuss in detail the parameter estimation procedure of generalized pareto (GP) probability distribution for overall survival times of the patients. We now proceed to discuss the parameter estimation process of the Johnson S_B, distribution in stage I. SFIEKIERS [29] has given a brief summary about the parameter estimation procedure of Johnson S_B probability distribution using moments of transformed values of a random Variable. Let T be a random variable denoting the survival times of

patients in stage I. Then, the p.d.f of T is given by:

$$f(t) = \frac{\delta}{\sqrt{2\pi}} \frac{\lambda}{(t-\zeta)(\lambda+\zeta-t)} exp\left[-\frac{1}{2}\left(\gamma + \delta ln\left(\frac{t-\zeta}{\lambda+\zeta-t}\right)\right)^2\right],$$
$$\zeta < t < \zeta + \lambda,\ ^-\infty < \zeta < \infty,\ \lambda > 0,\ ^-\infty < \gamma < \infty,\ \delta > 0$$

From the data, the extreme order statistics t_{min} and t_{max} are determined. In our case, at stage I, t_{min} = 0, and t_{max} = 155. Since, ζ, and λ are the location and spread parameters respectively, $\hat{\zeta}$ = t_{min} = 0 (since the minimum value of the survival time t is 0), and $\hat{\lambda}$ = $(t_{max}-t_{min})$ = (155-0) = 155 = t_{max}. Given the estimated values, $\hat{\zeta}$ and $\hat{\lambda}$, we proceed with the following transformation, that is, the values of t_i are transformed to:

$$f_i = ln\left(\frac{t_i - \hat{\zeta}}{\hat{\lambda} - t_i + \hat{\zeta}}\right)$$

The estimates of the other parameters $\hat{\gamma}$, and $\hat{\delta}$ take the following form:

$$\hat{\gamma} = -\frac{\bar{f}}{s_f} \text{ and } \hat{\delta} = \frac{1}{s_f}$$

Where $\bar{f} = \frac{\Sigma_i f_i}{n}$, $S_f = \left[\frac{\Sigma_i(f_i - \bar{f})^2}{n}\right]^{\frac{1}{2}}$.

The validity of the study data following different probability distributions has been justified using the goodness of fit tests. Soukissian [30] fitted a Johnson SB probability distribution to the wind speed data and used Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests to justify the goodness of fit assumptions. We followed the same approach using Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Cramér-von Mises (CVM) goodness of fit tests. **Table 3** provides the goodness of fit test results along with the *p*-values for all probability distributions in the four different stages.

Table 3. Goodness of tests for four stages

Stages	Gender	Prob. Distribution	GOF Tests	p-Values
I	Combined	Johnson S_B	A-D	.11
			K-S	.13
II	Combined	GEV	A-D	.27
			K-S	.21
III	Combined	GEV	A-D	.09
			K-S	.1
IV	Male	GPD	CVM	.22
			K-S	.18
IV	Female	GPD	CVM	.19
			K-S	.17

Table 4. Probability distributions and and parameter estimates of survival times of pancreatic cancer patients in different stages

Stages	Gender	Probability Distributions	Parameter Estimates
I	Combined	4-Parm. Johnson S_B	$\hat{\gamma}=1.2$, $\hat{\delta}=.62$, $\hat{\lambda}=155$, $\hat{\xi}=0$
II	Combined	Gen. Extreme Value (GEV)	$\hat{\mu}=10.18$, $\hat{\sigma}=10.83$, $\hat{k}=.32$
III	Combined	Gen. Extreme Value (GEV)	$\hat{\mu}=5.54$, $\hat{\sigma}=6.07$, $\hat{k}=.37$
IV	Male	Gen Pareto (GP)	$\hat{\mu}=0$, $\hat{\sigma}=4.12$, $\hat{k}=.25$
IV	Female	Gen Pareto (GP)	$\hat{\mu}=0$, $\hat{\sigma}=4.63$, $\hat{k}=.41$

As the *p*-values shown in **Table 3** of the given data, we fail to reject the fact that the observations (survival times) follow the specified probability distributions in each of the four stages. **Table 4** provides the specific probability distributions in each stage and their individual parameter estimates (approximate), classified by gender.

Table 5 illustrates the analytical forms of the probability density functions of male and female patients for the different stages, with their parametric estimates.

After we estimate the parameters for the specific probability distributions, we can present the exact analytical forms of the cancer survival times in each of four different stages.

Table 5 illustrates the analytical forms of the probability density functions of male and female patients for different cancer stages.

Figures 2-**4** illustrate the probability density function (pdf) and cumulative distribution function (cdf) of the patients at stage I, stage II, and stage III, respectively. **Figures 5** and **6** show the histogram, pdf, and cdf of male and female survival time at stage IV, respectively.

Parametric survival analysis for different stages

Once we have the analytical structures of the survival times of patients in different stages, driven by different parametric probability distributions, we can express the survival function S(t) analytically as a function of the cumulative distribution function (cdf). Now we proceed to express the analytical forms of the survival functions for the four different stages, with respect to **Table 4**. The estimate of the parametric survival function of patients diagnosed with pancreatic cancer in stage I is given by:

$$\hat{S}_I(t;\hat{\xi},\hat{\lambda},\hat{\gamma},\hat{\delta}) = 1 - \hat{F}_I(t;\hat{\xi},\hat{\lambda},\hat{\gamma},\hat{\delta})$$

$$= 1 - \Phi\left[\hat{\gamma} + \hat{\delta}ln\left(\frac{t-\hat{\xi}}{\hat{\lambda}-t+\hat{\xi}}\right)\right] \quad (1)$$

$$= 1 - \Phi\left[1.2 + .62ln\left(\frac{t}{155-t}\right)\right], t \geq 0$$

Where $\Phi(\cdot)$ is the cdf of a standard normal probability distribution. $\hat{F}_I(t;\hat{\xi},\hat{\lambda},\hat{\gamma},\hat{\delta})$ is the cdf of Johnson S_B Probability distribution. The survival function $S(\cdot)$ can be used to estimate the probability that a patient diagnosed with pancreatic cancer would survive beyond time *t*, which is

Table 5. Analytical forms of the probability density functions of patient survival times in different cancer stages

Gender	Analytical Forms
Combined (I)	$f(t) = \dfrac{.62}{\sqrt{2\pi}} \dfrac{155}{t(155-t)} \exp\left[-\dfrac{1}{2}\left(1.2 + .62\ln\left(\dfrac{t}{155-t}\right)\right)^2\right]$
Combined (II)	$f(t) = \dfrac{1}{10.83} \exp\left[-\left(1 - .32\ln\left(\dfrac{t-10.18}{10.83}\right)\right)^{3.125}\right]\left(1 - .32\ln\left(\dfrac{t-10.18}{10.83}\right)\right)^{2.125}$
Combined (III)	$f(t) = \dfrac{1}{6.07} \exp\left[-\left(1 - .37\ln\left(\dfrac{t-5.54}{6.07}\right)\right)^{2.7}\right]\left(1 - .37\ln\left(\dfrac{t-5.54}{6.07}\right)\right)^{1.7}$
Male (IV)	$f(t) = \dfrac{1}{4.12}\left[1 - .25\left(\dfrac{t}{4.12}\right)^3\right]$
Female (IV)	$f(t) = \dfrac{1}{4.63}\left[1 - .34\left(\dfrac{t}{4.63}\right)^{1.44}\right]$

denoted by $P(T \geq t)$. For example, we can compute the probability that a male patient diagnosed with pancreatic cancer would survive beyond 30 months. For example, for $t=40$ in Equation (1), we estimate the probability is 0.29 approximately. Thus, we can infer that a randomly chosen patient classified at stage I with pancreatic cancer has a 29% chance of survival beyond 40 months, as shown by **Figure 7**.

Similarly, the estimate of the parametric survival function of patients, driven by GEV probability distribution function diagnosed with pancreatic cancer in stage II is given by:

$$\hat{S}_{II}(t; \hat{\mu}, \hat{\sigma}, \hat{k}) = 1 - \hat{F}_{II}(t; \hat{\mu}, \hat{\sigma}, \hat{k})$$
$$= 1 - \exp\left[-\left(1 - k\left(\dfrac{t-\hat{\mu}}{\hat{\sigma}}\right)\right)^{\frac{1}{k}}\right] \quad (2)$$
$$= 1 - \exp\left[-\left(1 - .32\left(\dfrac{t-10.18}{10.83}\right)\right)^{\frac{1}{32}}\right], t \geq 10.18$$

As the following survival plot (**Figure 8**) for stage II patients illustrates, patients in stage II have a comparatively lower survival probability than stage I patients, which is quite natural. With reference to the last example, we can predict the survival probability as 13% for a stage II patient, surviving beyond $t=40$ months.

Now we proceed to express the GEV in analytical form for the stage III patients in a similar manner. The survival function at stage III can be given by:

$$\hat{S}_{III}(t; \hat{\mu}, \hat{\sigma}, \hat{k}) = 1 - \hat{F}_{III}(t; \hat{\mu}, \hat{\sigma}, \hat{k})$$
$$= 1 - \exp\left[-\left(1 - k\left(\dfrac{t-\hat{\mu}}{\hat{\sigma}}\right)\right)^{\frac{1}{k}}\right] \quad (3)$$
$$= 1 - \exp\left[-\left(1 - .37\left(\dfrac{t-5.54}{6.07}\right)\right)^{\frac{1}{37}}\right], t \geq 5.54$$

From **Figure 9** illustrates that the survival probability is decreasing, and it is approximately 5% for a randomly chosen patient who will survive beyond $t=40$ months after the patient is diagnosed with pancreatic cancer, stage III.

Results from **Table 1** suggested that there is a significant difference between the true mean survival times of stage IV patients, classified by gender. Thus, we now proceed to express the analytical forms of the survival times for male and female patients separately at stage IV. The parametric survival function, driven by GPD, at stage IV male patients is expressed as:

$$\hat{S}_{IV}(t; \hat{\mu}, \hat{\sigma}, \hat{k}) = 1 - \hat{F}_{IV}(t; \hat{\mu}, \hat{\sigma}, \hat{k})$$
$$= 1 - \left[1 - \left(\left[1 + \hat{k}\left(\dfrac{t-\hat{\mu}}{\hat{\sigma}}\right)\right]\right)^{\frac{1}{k}}\right] \quad (4)$$
$$= \left(\left[1 + .25\left(\dfrac{t}{4.12}\right)\right]^{\frac{1}{25}}\right), t \geq 0$$

Similarly, the parametric survival function, driven by GPD, at stage IV female patients is given by:

$$\hat{S}_{IV}(t; \hat{\mu}, \hat{\sigma}, \hat{k}) = 1 - \hat{F}_{IV}(t; \hat{\mu}, \hat{\sigma}, \hat{k})$$
$$= 1 - \left[1 - \left(\left[1 + \hat{k}\left(\dfrac{t-\hat{\mu}}{\hat{\sigma}}\right)\right]\right)^{\frac{1}{k}}\right] \quad (5)$$
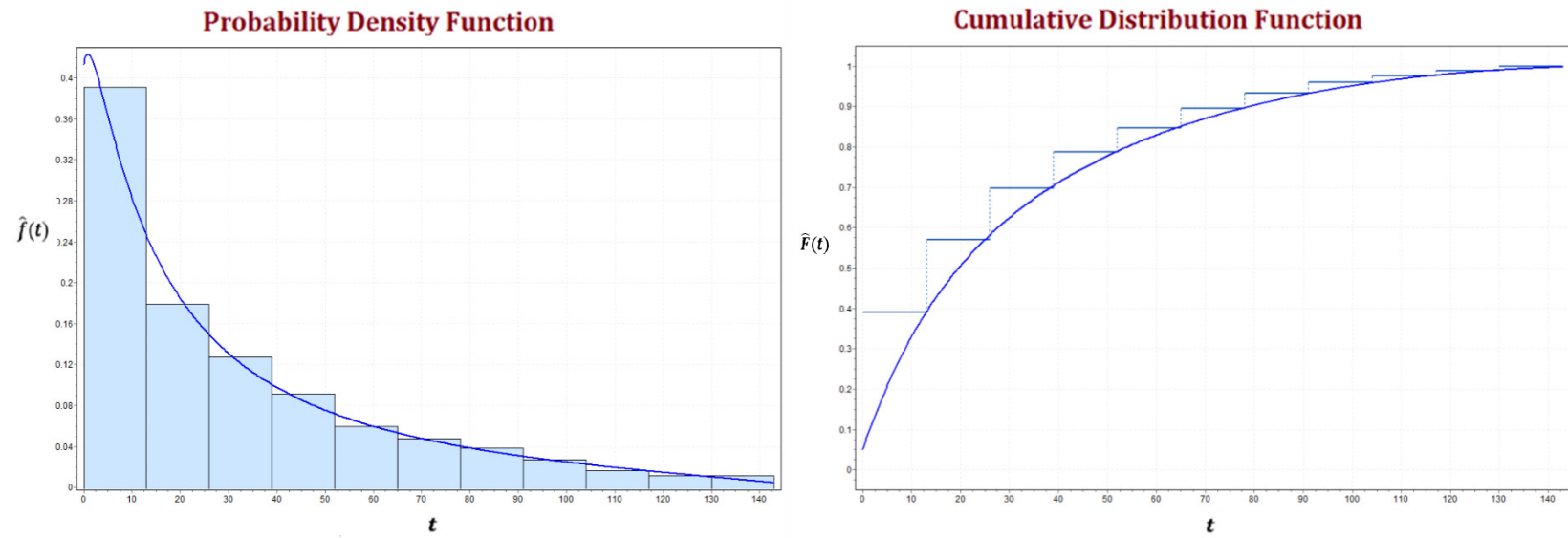$$= \left(\left[1 + .41\left(\dfrac{t}{4.63}\right)\right]^{\frac{1}{41}}\right), t \geq 0$$

Figure 2. Showing the probability density function (pdf), and cumulative distribution function of survival times of patients at stage I.
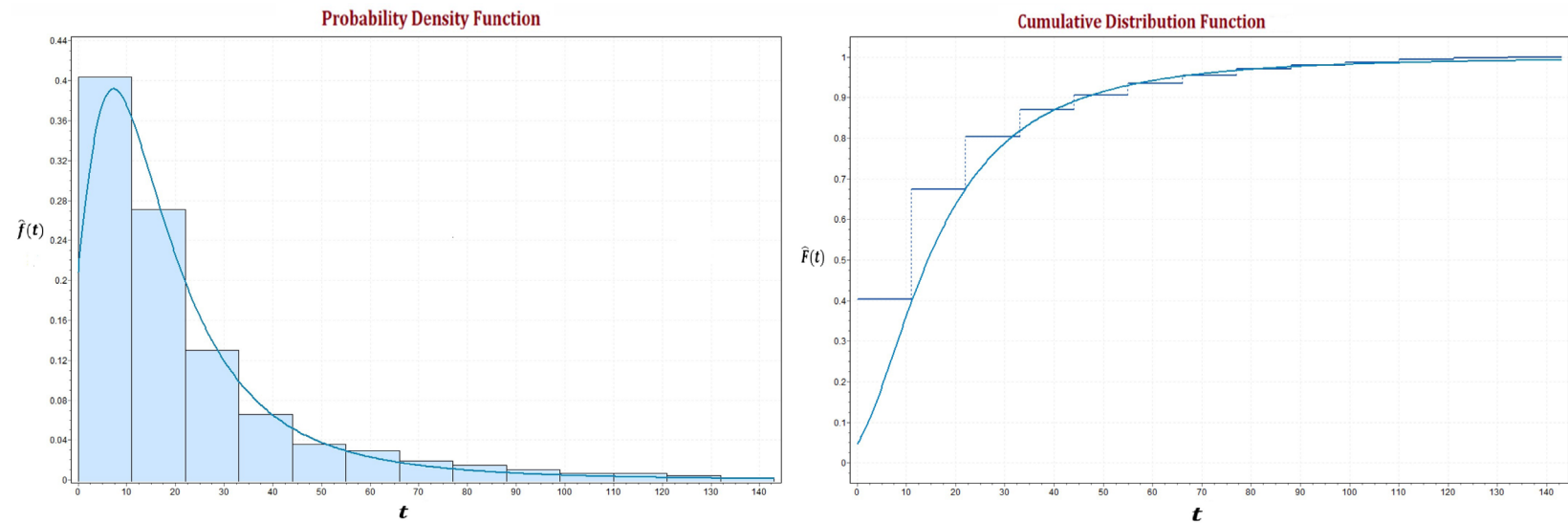

Figure 3. Showing the probability density function (pdf), and cumulative distribution function of survival times of patients at stage II.
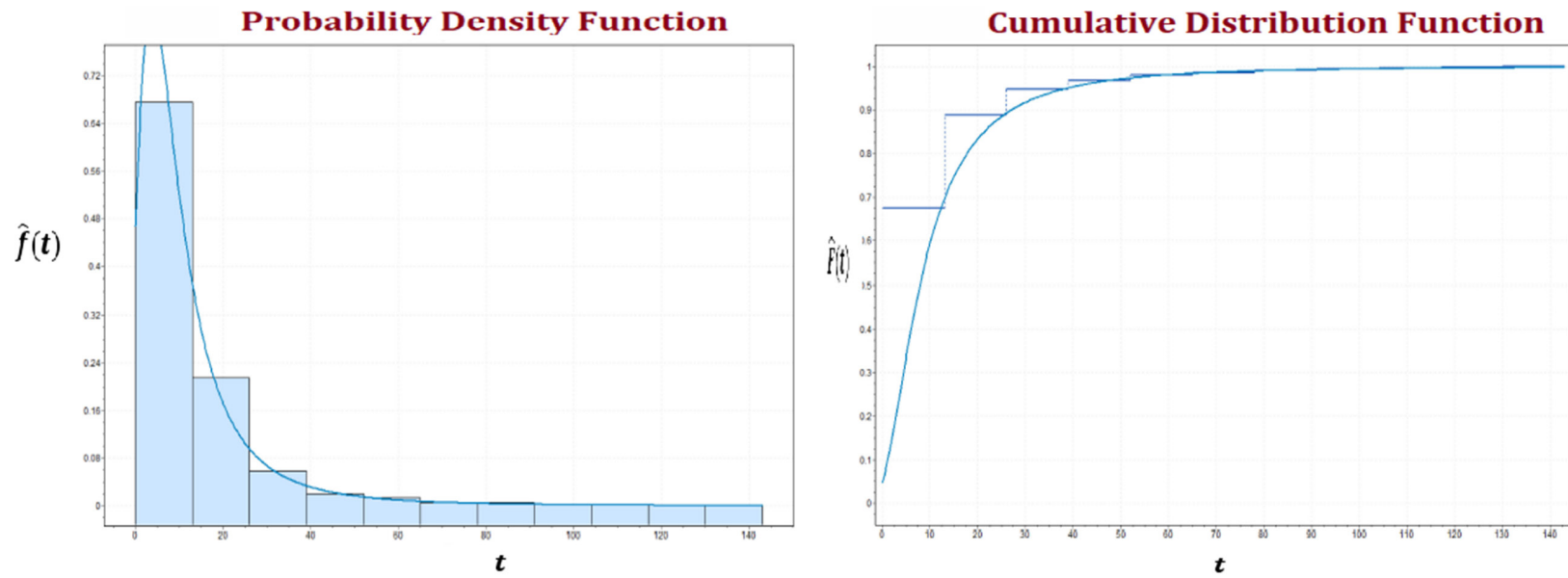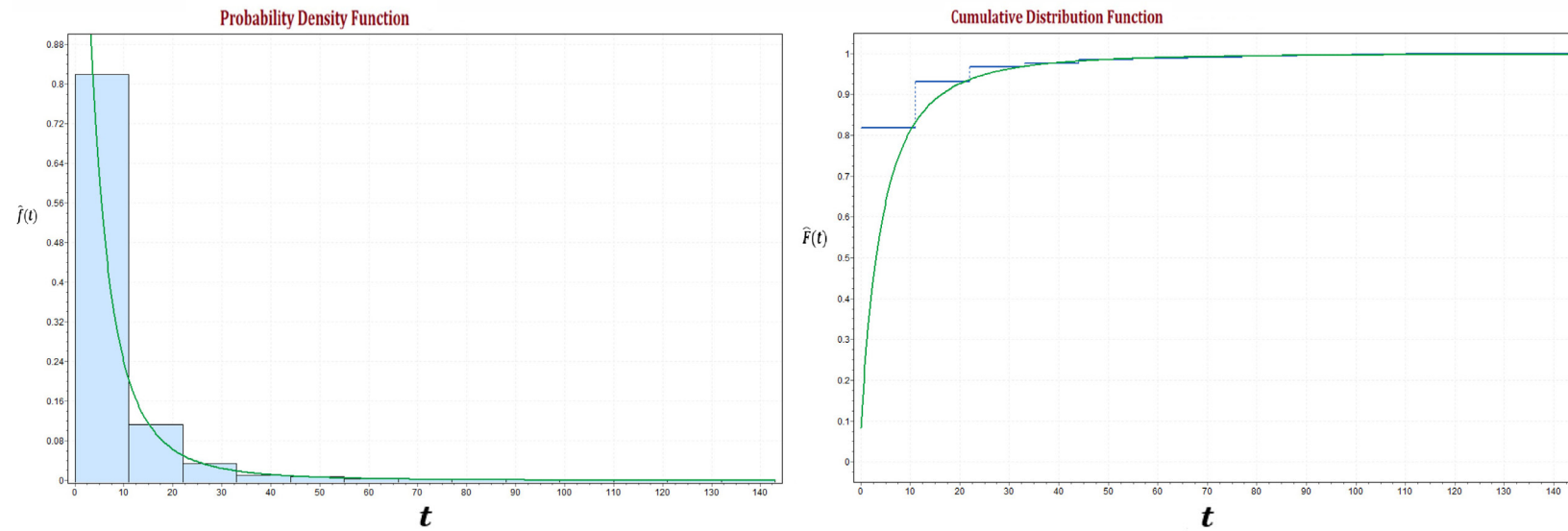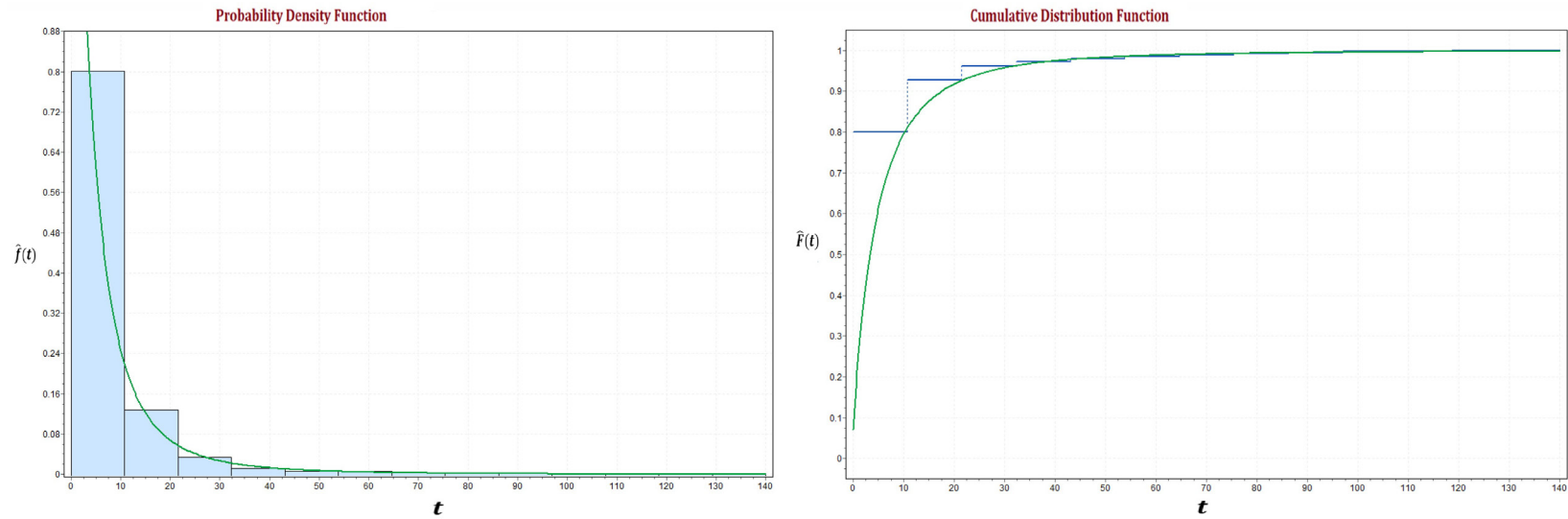
Figure 4. Showing the probability density function (pdf), and cumulative distribution function of survival times of patients at stage III.

Figure 5. Showing the probability density function (pdf), and cumulative distribution function of survival times of male patients at stage IV.

Figure 6. Showing the probability density function (pdf), and cumulative distribution function of survival times of female patients at stage IV.
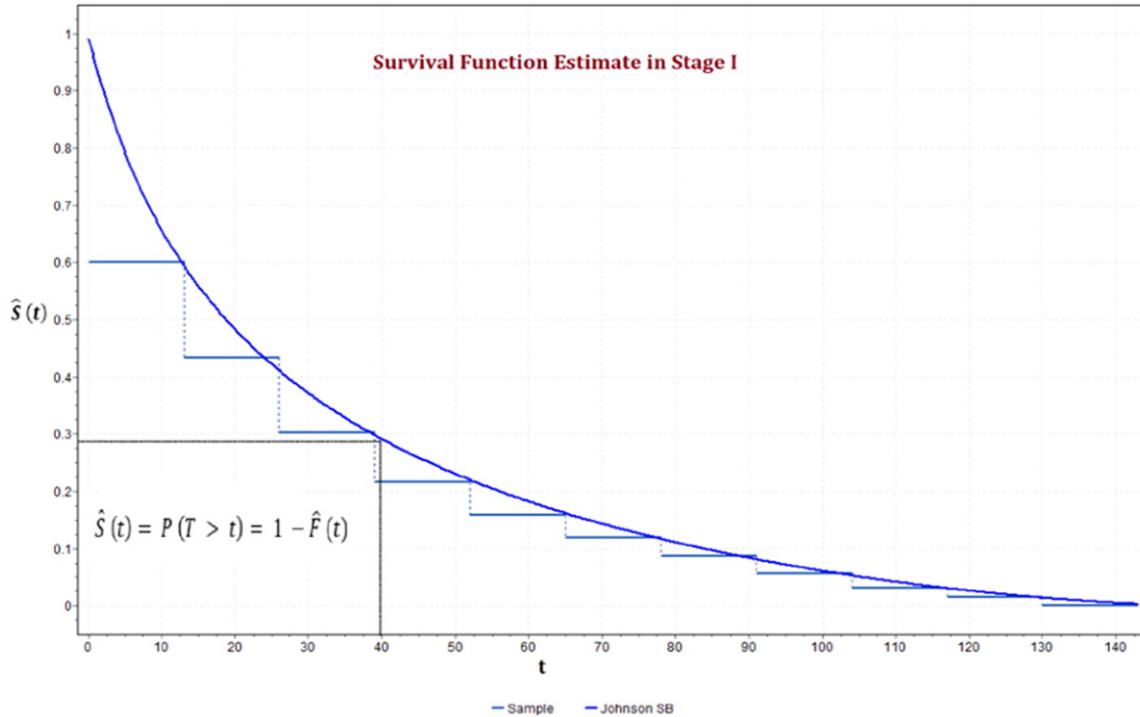
Figure 7. Parametric survival plot of pancreatic cancer patients at stage I.

As **Figures 10** and **11** indicate, the survival probabilities are extremely low (2% for male patients and 3% for female patients) for surviving beyond t=40 months after the diagnosis at stage IV.

Parametric analysis of the survival times of patients with pancreatic cancer-a combined analysis

So far, we have discussed the parametric analytical forms of the survival times of patients in different stages. We also computed the survival functions of patients in different stages. We found no significant difference in the true mean of the survival times of male and female patients except stage IV. We now proceed to do the same for the combined data, irrespective of stage. At first, we will check if there exists a significant difference between the true mean survival times of male and female pancreatic cancer patients. For this purpose, we use the log-rank test and found that there is insufficient sample evidence to reject the hypothesis that the distribution of mean survival times between the Male and Female patients diagnosed with pancreatic cancer is the same. **Figure 12** illustrates the behavior of overall survival curves of

male and female patients. As **Figure 12** illustrates, the survival curve of males (sky-blue) and the survival curve of females (red) are almost identical, which implies that they exhibit similar characteristics.

Descriptive statistics of the survival times of pancreatic cancer patients

In this section, we proceed to analyze the combined survival data descriptively. We represent the histogram and probability density function (pdf) to investigate the probability distribution of the survival times of pancreatic cancer patients, as shown in **Figure 13**. The figure follows that the probability distribution of the overall survival time is right-skewed. **Table 6** displays the descriptive statistics of the overall survival times for pancreatic cancer patients. We see that the mean (average) survival time for patients diagnosed with pancreatic cancer is 10.87 months. It implies that a randomly chosen patient diagnosed with pancreatic cancer is expected to survive for 10.87 months on average. Also, the median survival time is six months, which implies that the probability/chance of survival of a male or female patient beyond six months is approximately 50%. The
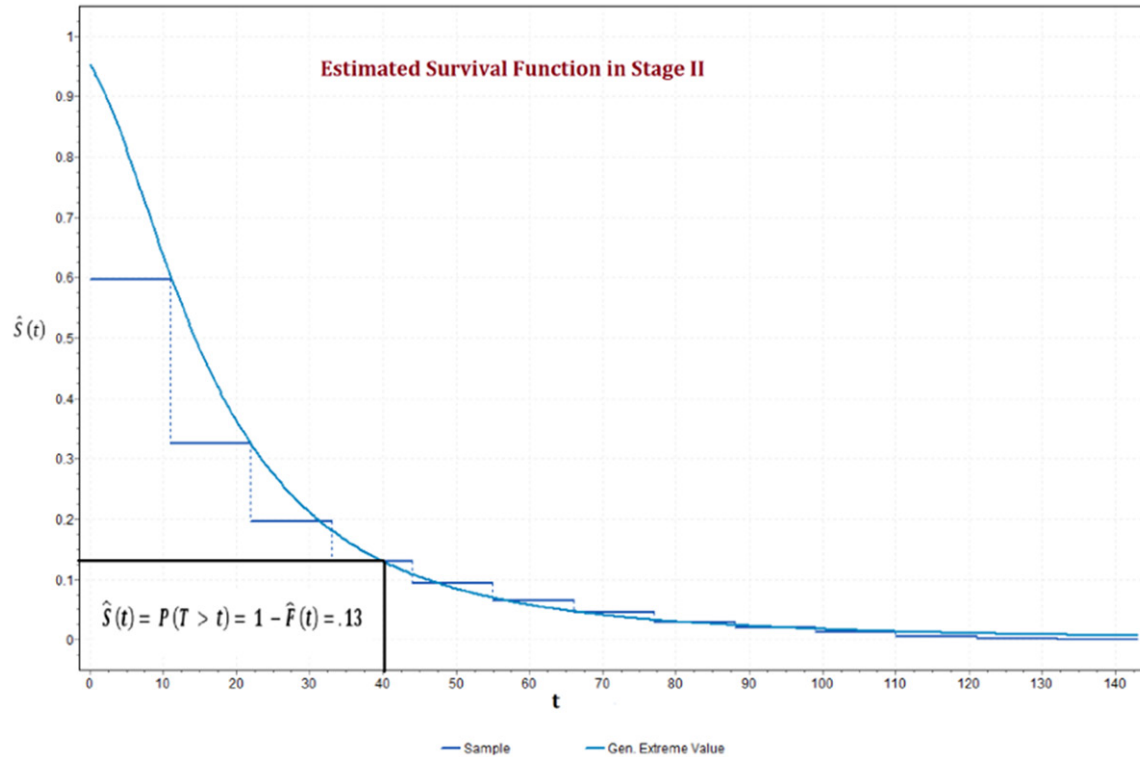
Figure 8. Parametric survival plot of pancreatic cancer patients at stage II.

positive skewed value of 3.07, as shown in **Table 6**, for patients diagnosed with pancreatic cancer, is further evidence to support the right-skewed behavior of the data, as shown in **Figure 13**, and the kurtosis value of 12.67 in **Table 6** attests to the leptokurtic behavior of the survival data. **Table 6** illustrates the different descriptive statistics for survival times of all patients combined diagnosed with pancreatic cancer.

Some literature reviews on the three parameter Generalized Pareto (GP) probability distribution and estimation

We will now conduct a parametric analysis of the survival times of patients diagnosed with pancreatic cancer to find the underlying probability distribution that best describes the probabilistic behavior of patient survival times (both genders). In order to obtain the best-fitted probability distribution, a number of classical distributions were tested to fit the data. We used the famous Anderson-Darling test [12] and Cramér-von Mises test [11] to find the best-fitted probability distribution function that describes the probabilistic pattern of the patients' survival

times. Also, we estimate the expected survival times and median survival times that are driven by the best-fitted probability distribution. The best-fitted probability distribution that characterizes the probabilistic behavior of the survival times of the male and female patients accurately is the three-parameter (3-P) Generalized Pareto (GP) probability distribution. **Table 7** provides the goodness of fit (GOF) results of the 3-P GPD distribution.

The findings of the GOF test demonstrate that we are unable to reject the null hypothesis that the subject data (survival times for males and females) follow a GP probability distribution. In this section, we define the probability density function (pdf) of the Generalized Pareto distribution and the statistical methods for computing the approximate parameter estimates. The Generalized Pareto distribution (GPD) is a family of continuous probability distributions developed on the basis of extreme value theory in the field of probability theory and statistics [13]. The GPD is a generalization of the Pareto distribution (PD). "The PD was studied extensively by Arnold (1983), and the problem of estimation in the PD was considered by Arnold and
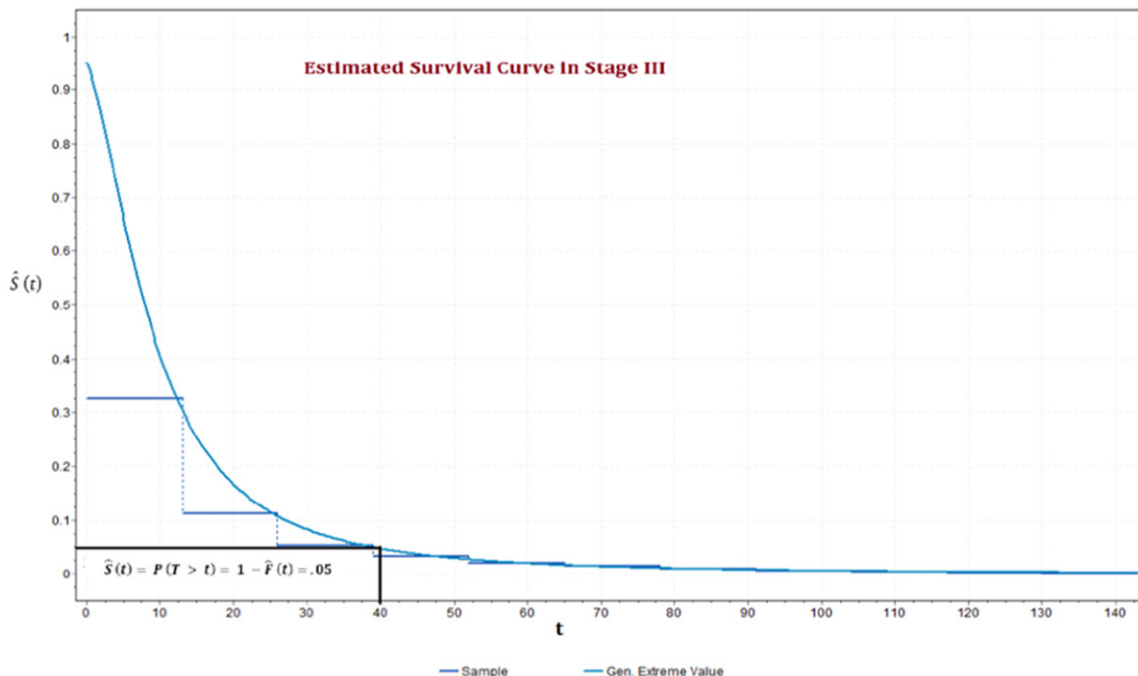
Figure 9. Parametric survival plot of pancreatic cancer patients at stage III.

Press (1989)" [14]. It has been used broadly by several researchers to model data arising from several fields. "Hosking and Wallis used the GPD to model the annual maximum flood of the River Nidd at Hunsingore, England" [15]. "Grimshaw used it to model tensile strength data from a random sample of nylon carpet fibers" [16]. Other estimation procedures and applications of the GPD in extreme value analysis using numerical optimization have been illustrated by Castillo and Daoudi [17]. Let T be a random variable following GPD with location parameter μ, scale parameter $\sigma > 0$ and shape parameter k. That is, $T \sim GDP (\mu, \sigma, k)$ with the domain $\mu \le t \le \mu - \frac{\sigma}{k}$, for $k<0$ and $\mu \le t < \infty$, for $k \ge 0$. Then, the probability density function (pdf) of T is given as follows:

$$f_{GDP}(t;\mu,\sigma,k) = \begin{cases} \frac{1}{\sigma}\left(\left[1 + k\left(\frac{t-\mu}{\sigma}\right)\right]^{\frac{1}{k}-1}\right), k \neq 0 \\ \frac{1}{\sigma}exp\left(-\frac{(t-\mu)}{\sigma}\right) \quad , k = 0 \end{cases} \quad (6)$$

The associated cumulative distribution function (cdf) is shown below:

$$F_{GDP}(t;\mu,\sigma,k) = \begin{cases} 1 - \left(\left[1 + k\left(\frac{t-\mu}{\sigma}\right)\right]^{\frac{1}{k}}\right), k \neq 0 \\ 1 - exp\left(-\frac{(t-\mu)}{\sigma}\right) \quad , k = 0 \end{cases} \quad (7)$$

There are different procedures to estimate the parameters μ, σ, and k of the GP probability distribution. Some of these methods include the elemental percentile method (EPM) proposed by Castillo and Hadi [18]. "Grimshaw proposed an algorithm for computing the maximum likelihood estimation (MLE) of the parameters of the GPD" [16]. Hosking & Wallis [15] derived a parameter and quantile estimation mechanism based on Probability-weighted moments (PWM). Zhang proposed an improved maximum likelihood estimation using the empirical Bayesian method to overcome the non-existence problem of the PWM estimator [19]. Castillo and Hadi proposed a more efficient optimization algorithm for estimators of the GPD parameters where the proposed estimators are defined for all possible values of the parameters [20]. The performance of the estimators was found to be better than the method of moments (MOM) and Probability-Weighted Moments (PWM) estimates. Pham, Tsokos, & Choi proposed a GP parameter estimation method for censored data and validated their results using a sensitivity and specificity test [21]. Singh & Gao developed a parameter estimation method using the principle of maximum entropy (POME) for 3-P GPD [22]. Since we have enough data to analyze, we can choose any well-known method for our parameter esti-
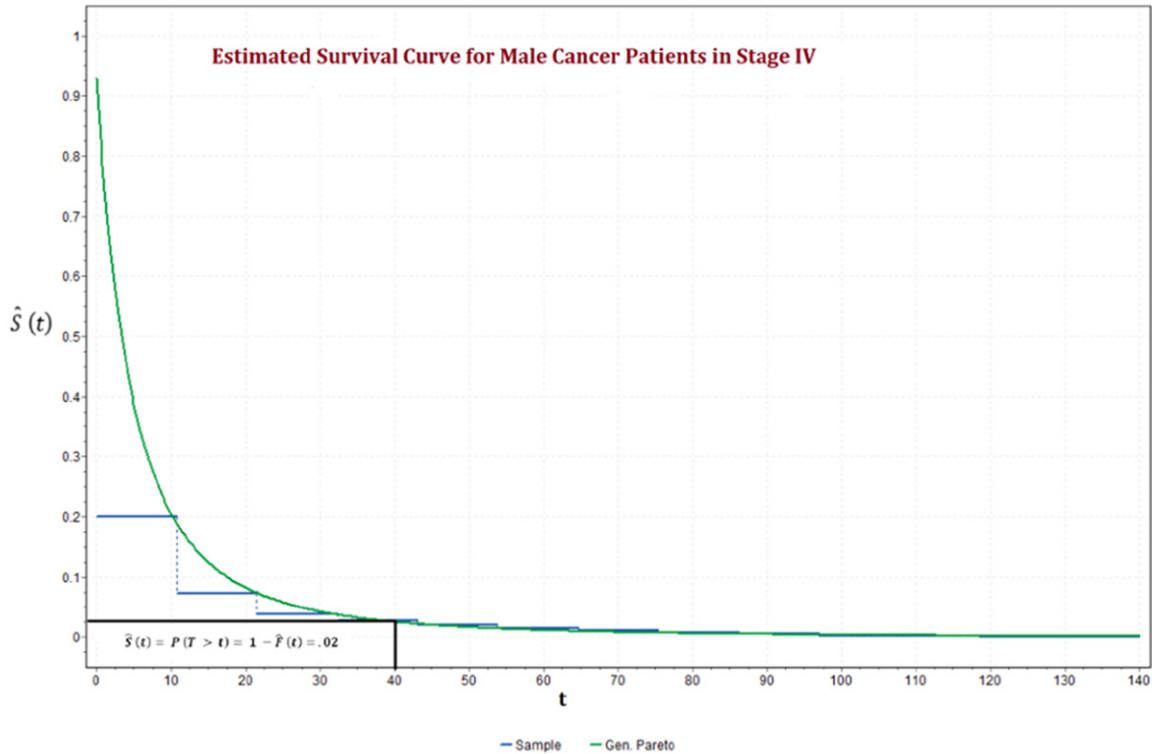
Figure 10. Parametric survival plot of male pancreatic cancer patients at stage IV.

mation purpose. In the next subsection, we discuss the parameter estimation procedure of 3-P GPD briefly by the PWM method.

Estimating the parameters of 3-P GP probability distribution using the probability-weighted moments (PWM) method

The probability-weighted moments (PWM) of a random variable T with cumulative distribution function $F(t) = P(T \leq t)$ is given by:

$$M_{p,r,s} = E[T^p\{F(t)\}^r\{1-F(t)\}^s] \qquad (8)$$

Where p, r, and s are real numbers. Probability-weighted moments can be expressed as a function of the inverse distribution function $F^{-1}(t) = t(F)$ in closed form by:

$$M_{p,r,s} = \int_0^1 \{t(F)\}^p F^r\{1-F\}^s] \qquad (9)$$

The two special cases of $M_{p,r,s}$ which are commonly used are:

$$\alpha_s = M_{1,0,s} = E[T\{1-F(t)\}^s], (s = 0, 1, 2, \ldots)$$
and $\qquad (10)$
$$\beta_r = M_{1,r,0} = E[T\{F(t)\}^r], (r = 0, 1, 2, \ldots)$$

Where T inside the $E[\cdot]$ is the inverse distribution of T, denoted by $t(F)$ To estimate the parameters of GPD, we use $\alpha_s = M_{1,0,r} = E[T\{1-F(t)\}^s]$ according to the approach used by Singh & Gao [22].

From (7), we can solve for T to obtain the inverse cdf, $t(F)$. The inverse distribution function is given by:

$$t(F) = \begin{cases} \mu + \dfrac{\sigma}{k}\{1 - (1 - F^k)\} & \text{if } k \neq 0 \\ \mu - \sigma log(1 - F) & \text{if } k = 0 \end{cases} \qquad (11)$$

The analytical form of α_s for the 3-P GPD is given as follows. Using expressions (10) and (11).

From (10), we have:

$$\alpha_s = M_{1,0,s}$$
$$= \int_0^1 \left[\mu + \frac{\sigma}{k}\{1 - (1 - F^k)\}\right][1 - F^s] dF \qquad (12)$$
$$= \frac{1}{s+1}\left(\mu + \frac{\sigma}{k}\right) - \frac{\sigma}{k}\left(\frac{1}{k+s+1}\right), (s = 0, 1, 2, \ldots)$$

Thus, for $k \neq 0$; the probability-weighted moments (PWM) of the 3-P GP distribution is given by (12). In equation (12), substituting $s = 0$, $r =$
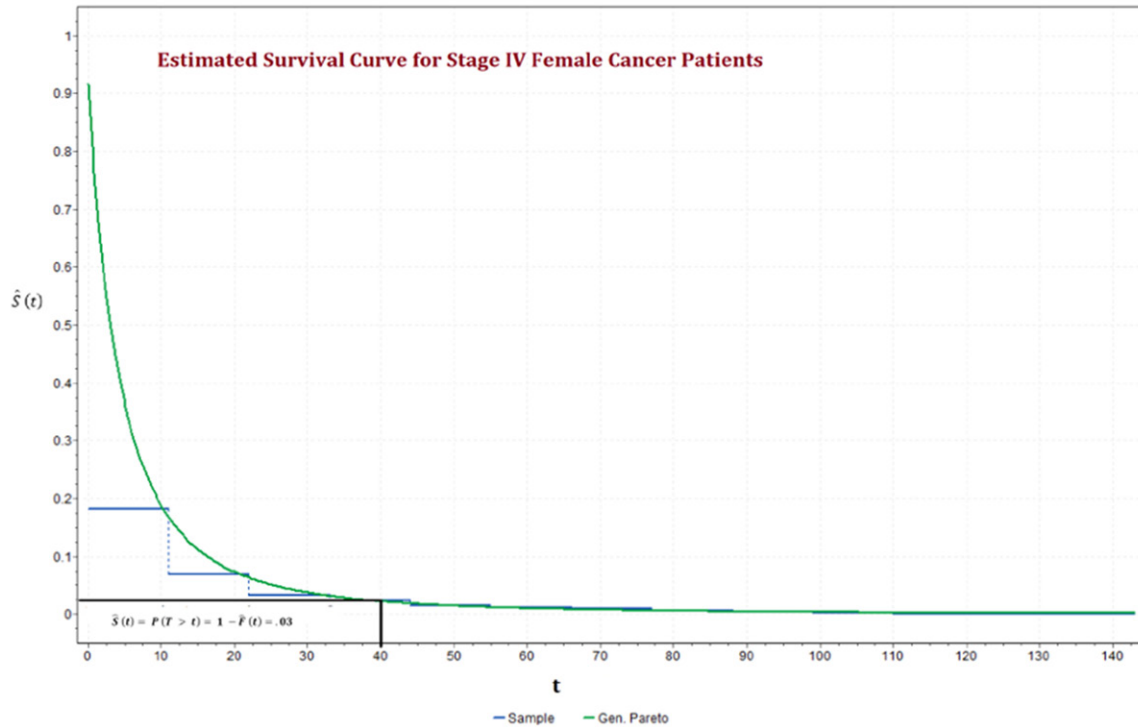
Figure 11. Parametric survival plot of female pancreatic cancer patients at stage IV.

1 and $r = 2$ we can obtain explicit expressions of α_0, α_1, and α_2 in terms of μ, σ, and k. That is:

$$\alpha_0 = \left(\mu + \frac{\sigma}{k}\right) - \frac{\sigma}{k}\left(\frac{1}{k+1}\right) \qquad (13)$$

$$\alpha_1 = \frac{1}{2}\left(\mu + \frac{\sigma}{k}\right) - \frac{\sigma}{k}\left(\frac{1}{k+2}\right) \qquad (14)$$

and

$$\alpha_3 = \frac{1}{3}\left(\mu + \frac{\sigma}{k}\right) - \frac{\sigma}{k}\left(\frac{1}{k+3}\right) \qquad (15)$$

The PWM estimates of the parameters $(\hat{\mu}, \hat{\sigma}, \hat{k})$ can be obtained by solving the equations (13), (14), and (15) for μ, σ, and k. After solving the above three equations, we obtain the explicit expressions of the PWM estimates [22] as follow:

$$\hat{k} = \frac{\alpha_0 - 8\alpha_1 - 9\alpha_2}{-\alpha_0 + 4\alpha_1 - 3\alpha_2} \qquad (16)$$

$$\hat{\sigma} = \frac{(\alpha_0 - 2\alpha_1)(\alpha_0 - 3\alpha_2)(-4\alpha_1 + 6\alpha_2)}{(-\alpha_0 + 4\alpha_1 - 3\alpha_2)^2} \qquad (17)$$

and

$$\hat{\mu} = \frac{2\alpha_0\alpha_1 - 6\alpha_0\alpha_2 + 6\alpha_1\alpha_2}{-\alpha_0 + 4\alpha_1 - 3\alpha_2} \qquad (18)$$

Table 8 shows the approximate parameter estimates of survival times driven by a 3-P GP probability distribution.

Now substituting the parameter estimates of μ, σ, and k in (6), we obtain the analytical form of the probability density function (pdf) of patients' survival times. The analytical form of the GP probability density function (pdf) for combined pancreatic cancer survival time is given by:

$$f_{Combined}(t) = \frac{1}{8.9}\left[1 + .22\left(\frac{t - .22}{8.9}\right)\right]^{-5.54}, t \geq .22 \qquad (19)$$

The probabilistic behavior of the overall survival times of male and female patients with pancreatic cancer is characterized by the above probability density function. Now, we will compute the expected survival times of patients driven by GP probability distribution. Using estimates given in **Table 8**, we can find the expectations and median survival times for the patients that follow *GDP* (.65, 8.9, .22).

The expected value of a random variable *T* following *GDP* (μ, σ, K) is given by:

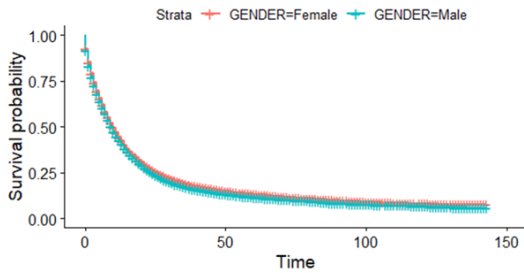$$E(T) = \hat{\mu} + \frac{\hat{\sigma}}{1 - \hat{k}}, \hat{k} < 1 \qquad (20)$$

Figure 12. Log-rank test showing in difference in survival times.

Using Equation (20), the expected survival time for pancreatic cancer patients following *GDP* (.65, 8.9, .22) is:

$$E(T) = .65 + \frac{8.9}{(1 - .22)} = 12.06 \text{ months.}$$

The median survival time *T* of *GDP* (*μ, σ, k*) is given by:

$$Med_{GDP}(t;\mu,\sigma,k) = \hat{\mu} + \frac{\hat{\sigma}(2^{\hat{k}} - 1)}{\hat{k}} \quad (21)$$

The overall median survival periods of male and female pancreatic patients can be computed from Equation (21):

$$Med_{GDP}[T] = .65 + \frac{8.9(2^{.22} - 1)}{.22} = 7.31 \text{ months.}$$

We can compute the cumulative distribution functions (cdf) of the random variable *T* after we have the analytical forms of the pdf. The analytical form of the GPD is given by:

$$F_{Combined}(t) = 1 - \left[1 + .22\left(\frac{t - .65}{8.9}\right)^{-4.45}\right], t \geq .65 \quad (22)$$

Figure 14 illustrates the cdf plot of the overall survival times. As **Figure 14** illustrates, the cdf plot is highly useful for estimating the chances that a given male or female patient diagnosed with pancreatic cancer would survive until a certain point in time. For example, the probability that a randomly diagnosed patient will survive up to time months is approximately 91.5%. The parametric survival analysis of pancreatic cancer patients' overall survival times, which is one of the essential components of this study, will be discussed in the next section.

Parametric survival analysis

The process of estimating the parametric survival function is used to evaluate the survival probabilities of pancreatic cancer patients (male or female) as a function of time. In Equation (22), we calculated the cdf of survival periods for patients diagnosed with pancreatic cancer. We now proceed with estimating the analytical form of the GP parametric survival function of patients, irrespective of stages, which is given by:

$$\hat{S}(t;\hat{\mu},\hat{\sigma},\hat{k}) = 1 - \hat{F}(t;\hat{\mu},\hat{\sigma},\hat{k})$$
$$= \left[1 + .22\left(\frac{t - .65}{8.9}\right)^{-4.45}\right], t \geq .65 \quad (23)$$

The survival function $\hat{S}(\cdot)$ can be used to estimate the probability that a randomly selected patient diagnosed with pancreatic cancer would survive beyond time *t*, which is denoted by *P* (*T≥t*). For example, we can compute the probability that a patient diagnosed with pancreatic cancer would survive beyond 30 months. That is, in Equation (23), we estimate the probability as 0.09. As a result, we can conclude that a randomly selected pancreatic cancer patient has a 9% chance of surviving beyond 30 months. The GP parametric survival curve for pancreatic cancer patients is depicted in **Figure 15**. The non-parametric Kaplan-Meier Survival function for pancreatic cancer is discussed briefly in the next section.

Kaplan-Meier estimation of survival probability of the survival times of patients with pancreatic cancer

Many clinical scientists use the Gaussian probability distribution to model the time-to-event phenomena or take the logarithmic transformation of the data to fit a probability distribution to the data parametrically. Such assumptions about the data and transformations should be justified. However, non-parametric methods are often used to estimate the survival probabilities in some cases where the appropriate parametric structure of the data is unknown or difficult to estimate. The Kaplan-Meier (KM) estimator [24, 25], a.k.a, the product-limit estimator, is a non-parametric statistic used to estimate the survival function from data. In health sciences, it is used to calculate the percentage
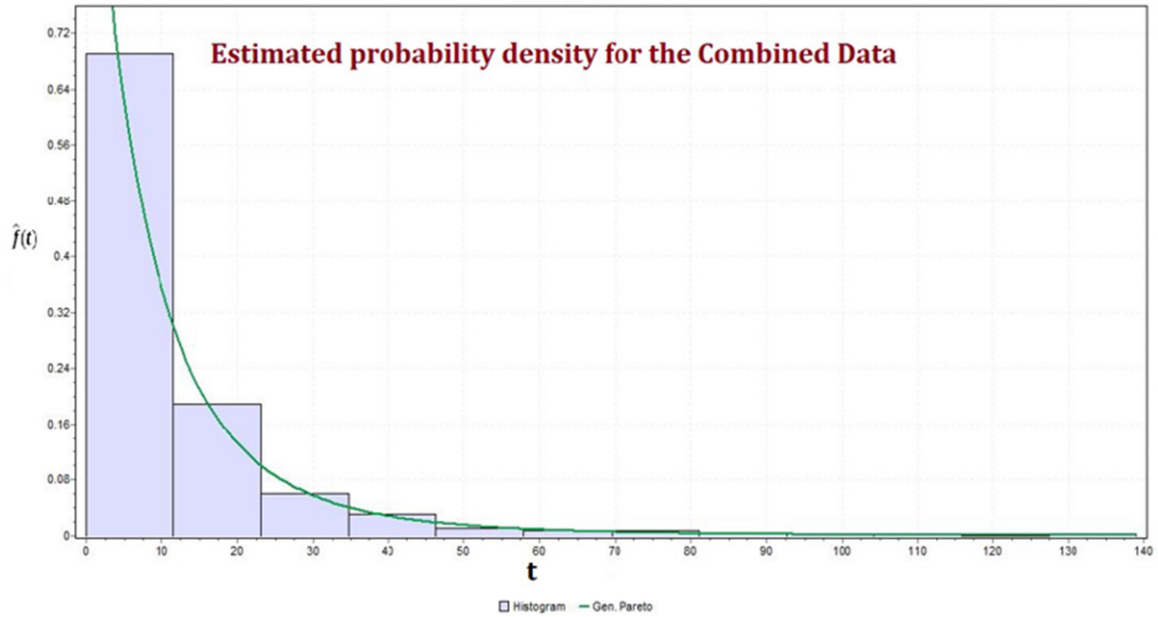
Figure 13. The probability distribution for the combined data.

Table 6. Descriptive statistics of survival times (in months) of overall pancreatic cancer patients

Descriptive Statistic	Measures
Mean	10.87
Median	6
Std. Dev.	14.63
Skewness	3.07
Kurtosis	12.67
Std. Error	.24

Table 7. Goodness-of-fit test of the GPD of the survival times of male and female

Statistical Tests	P-Values Male	P-Values Female
Kolmogorov-Smirnov	0.27	.38
Cramér-von Mises	0.22	.18

of patients who survive for a period of time after receiving any specific therapy. Edward L. Kaplan and Paul Meier proposed it in 1958, which is defined as the product of the conditional probabilities of surviving to the next failure time multiplied by the failure times. Theoretically, the estimate is defined as:

$$\hat{S}(t) = \prod_{t \leq t} (1 - \hat{q}_i) = \prod_{t \leq t} (1 - \frac{d_i}{n_i}) \qquad (24)$$

Table 8. Parameter Estimates of 3-P GP probability distribution

Parm. Estimates	Values
Location ($\hat{\mu}$)	.65
Sacale ($\hat{\sigma}$)	8.9
Shape ($\hat{k}$)	.22

where n_i represents the number of patients at risk at the time t_i, and d_i denotes the number of individuals who fail (die) at a specific time.

Figure 16 demonstrates the overall non-parametric survival curve for patients diagnosed with pancreatic cancer.

Median survival using KM estimate

Median survival time describes how long a group of patients survives with an ailment in general or after a specific treatment has been implemented. Median survival time is when half the patients who are susceptible to a specific ailment are expected to be alive. It is defined as the probability of surviving beyond a specific time t is 50%. Generally, the median survival time [26] is defined as, $\hat{t}_{med} = inf\{t; \hat{S}(t) \leq 0.5\}$.

It means that it is the smallest t such that the estimated survival function S(t) is less than or
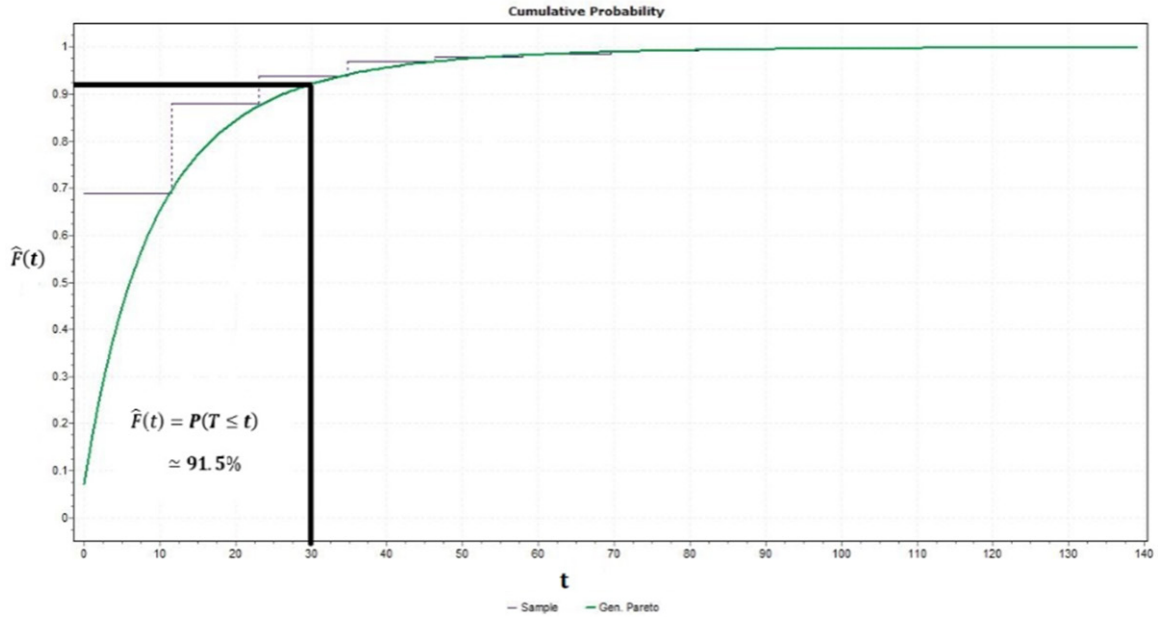
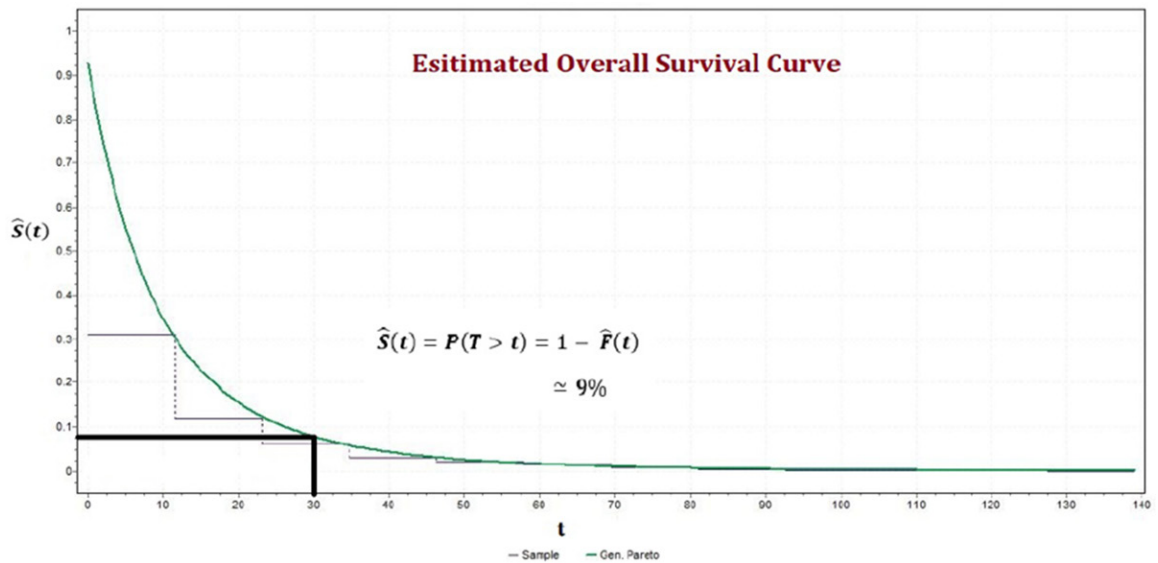**Figure 14.** cdf plot for the combined data.

Figure 15. Parametric survival plot of overall survival times.

equal to 0.5. The median survival times computed using a non-parametric KM estimator, for the pancreatic cancer patients are given as six, which is evident from **Figure 16**. It is very interesting to note that the median survival time we obtained by the descriptive method (**Table 5**) is exactly the same as what we obtained from the non-parametric method. However, the median survival times we obtained using the parametric method (implementing the GPD) are significantly higher than the descriptive and non-

parametric methods. **Table 9** compares the median survival times for all patients diagnosed with pancreatic cancer, computed using the three methods.

Comparison of GP probability distribution with the Kaplan-Meier (KM) estimation of the survival function

In the parametric analysis (section 4.2), we found that patients' survival times (both male
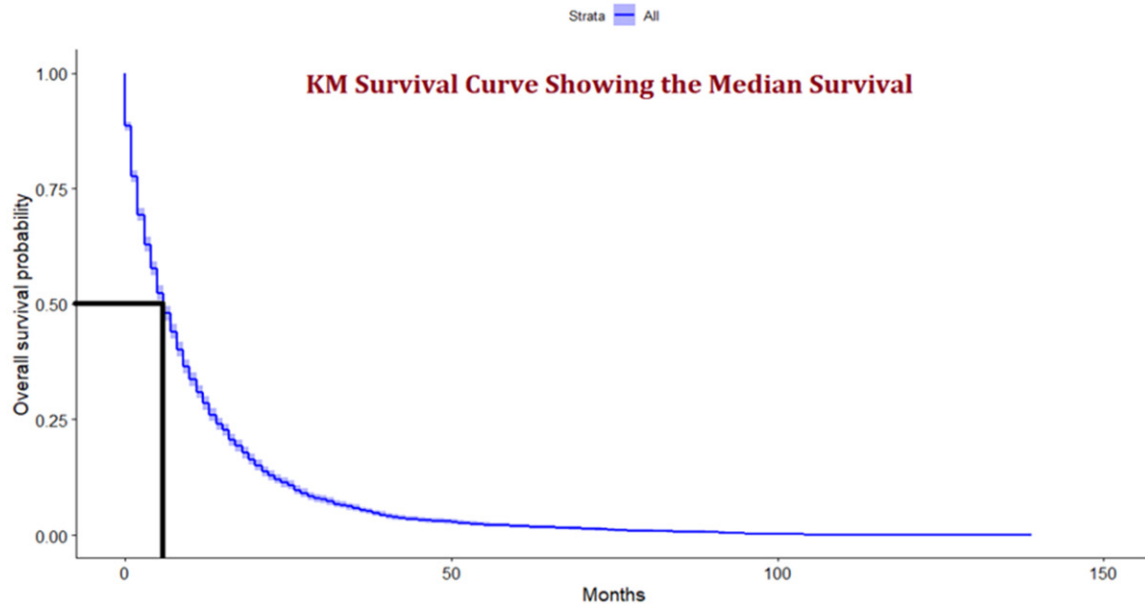
Figure 16. Overall KM survival plot for pancreatic survival times.

Table 9. Table of comparison of the median survival times for all pancreatic cancer patients

Methods	Median Survival Time
Descriptive	6
Parametric	7.31
Non-Parametric	6

Table 10. Table of comparison of survival probabilities of pancreatic cancer patients computed using parametric and non-parametric methods

t	$\hat{S}_P(t)$	$\hat{S}_{KM}(t)$
0	.96	.88
1	.87	.77
2	.81	.69
3	.77	.62
4	.7	.57
5	.63	.52
6	.57	.47
7	.51	.44
8	.47	.4
9	.43	.36
10	.39	.33

and female) with pancreatic cancer follow a Generalized Pareto (GP) distribution. In section 5, we performed a non-parametric analysis using the Kaplan-Meier to estimate a randomly selected patients' survival probability. We now compare the survival probability estimates obtained from GP probability distribution with the non-parametric Kaplan-Meier survival estimates of the survival times of the pancreatic cancer patients. The survival function of the two techniques is important in estimating the chance of survival of a patient diagnosed with pancreatic cancer beyond a particular time. **Table 10** compares the survival probabilities associated with different time periods (in months). We observe that the probability estimates computed by the GP survival function are significantly higher than that of Kaplan-Meier probability estimates. Because parametric methods are more powerful, robust, and efficient than non-parametric approaches, parametric survival estimates must be utilized as the most accurate estimates, provided that

the appropriate probability distribution has been appropriately established.

In **Table 10**, $\hat{S}_P(t)$, is the parametric survival probability estimates for pancreatic cancer patients using GP probability distribution. $\hat{S}_{KM}(t)$ is the non-parametric survival probability estimates for pancreatic cancer patients using the non-parametric KM estimate.

Results and discussions

Given the increased incidence of pancreatic cancer in recent years, it is critical to investi-

gate the prognosis to improve pancreatic cancer therapeutic/treatment strategies. The primary treatment for most types of pancreatic cancer is chemotherapy and targeted therapy drugs in some cases. A stem cell transplant can be thought of as a treatment option. However, surgery and radiation therapy do not fall under crucial treatments for pancreatic cancer; they may be employed in exceptional circumstances. Furthermore, the therapeutic approach for children with pancreatic cancer may differ from that employed for adults. Several research strategies and methodologies have been designed to treat pancreatic cancer patients in order to increase their chances of survival. Chakraborty & Tsokos [27] performed data-driven research on Acute Myeloid Leukemia (AML) by doing some parametric and non-parametric analyses to improve the survival probabilities of patients of different gender groups. In our present study:

• We analyzed a total of 10,000 patient information and have shown that there was no significant difference between the overall survival times of male and female pancreatic cancer patients.

• We identified a well-defined probability distribution that characterized the survival times of a total of 10,000 patients (5,100 male and 4,900 female) diagnosed with pancreatic cancer and used the information to estimate the parametric survival function driven by generalized Pareto (GP) probability distribution.

• We have tested if there is any significant difference between the mean survival times of male and female patients in each of the four stages.

• We have identified the probability distributions of male and female survival times in four different cancer stages and derived their analytical forms. Also, we derived the parametric survival functions in each stage, driven by different parametric probability distributions.

• We compared the median survival times of patients using descriptive, parametric, and non-parametric methods and obtained very consistent results.

• We calculated the overall survival probabilities utilizing the frequently used non-paramet-

ric Kaplan-Meier (KM) cancer survivorship analysis method and compared those estimates with the parametric probability estimates obtained from the GP probability distribution.

Conclusion

We have determined the survival probabilities of pancreatic cancer patients using different statistical methods; the parametric Generalized Pareto (GP) distribution and the non-parametric Kaplan-Meier (KM) method and obtained better parametric estimates of the survival probabilities by implementing the parametric method in contrast to the non-parametric KM method. The intrinsic distributional assumption regarding the survival times under study is one of the difficulties in the parametric survival analysis. However, if one could justify the distributional assumptions, it is possible to obtain a better estimate from the parametric analysis, which has greater statistical power in contrast to the non-parametric method. Based on our data analysis and study results relating to pancreatic cancer patients, we impart the following important suggestions.

• If we have information on the survival times of male and female cancer patients, we must first evaluate whether there is a statistically significant difference between the true mean survival times of male and female cancer patients. If a statistically significant difference is found, it is more practical to conduct different analyses for the two gender groups. In our current study, we discovered that there is no statistically significant difference in overall survival durations between male and female pancreatic cancer patients.

• After identifying the appropriate probability distributions of male and female cancer patients, if we have further data available regarding the different stages, it is essential to identify the analytical forms of the probability distributions that drive the survival data in each of the four individual stages.

• If we have information available, then the stage-by-stage analysis most appropriately reflects the survival probability of patients in individual stages.

• If the data simply includes patient survival time, then evaluating the survival probability

parametrically usually yields more precise, robust, and reliable findings than the commonly used non-parametric Kaplan-Meier survival estimate.

• If no unique or well-defined parametric probability distribution is identified, we propose using the kernel density estimate or the Kaplan-Meier (KM) technique to estimate survival probabilities.

Although in certain circumstances, the use of non-parametric Kaplan-Meier survival analysis may result in a very close or higher probability estimate of the survival rate (if we include the censored observations in our study), the parametric analysis continues to be more powerful, robust, and efficient when there is no information about the censored individuals. As a result, it is reasonable to proceed with parametric analysis at the beginning of any particular cancer survivorship data study. By evaluating cancer survivorship data, this study provides a more effective and realistic approach for estimating survival probability in order to improve the therapeutic/treatment process of pancreatic cancer.

This research is protected by the University of South Florida TOT.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Chris P Tsokos, Department of Mathematics & Statistics, University of South Florida, 4202 E. Fowler Ave., CMC 366 Tampa, Florida 33620, USA. Tel: (813) 974-9734; Fax: (813) 974-2700; E-mail: ctsokos@usf.edu

References

[1] Li D, Xie K, Wolff R and Abbruzzese JL. Pancreatic cancer. Lancet 2004; 363: 1049-1057.
[2] Vincent A, Herman J, Schulick R, Hruban RH and Goggins M. Pancreatic cancer. Lancet 2011; 378: 607-620.
[3] Kleeff J, Korc M, Apte M, La Vecchia C, Johnson CD, Biankin AV, Neale RE, Tempero M, Tuveson DA, Hruban RH and Neoptolemos JP. Pancreatic cancer. Nat Rev Dis Primers 2016; 2: 16022.
[4] Ferlay J, Shin HR, Bray F, Forman D, Mathers C and Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127: 2893-2917.
[5] Feigl P and Zelen M. Estimation of exponential survival probabilities with concomitant information. Biometrics 1965; 21: 826-838.
[6] Chakraborty A and Tsokos C. Survival analysis for pancreatic cancer patients using Cox-Proportional Hazard (CPH) model. Glob J Med Res 2021; Vol 21, No 3-F.
[7] Xu Y, Kepner J and P Tsokos C. Identify attributable variables and interactions in breast cancer. J Appl Sci 2011; 11: 1033-1038.
[8] Ilic M and Ilic I. Epidemiology of pancreatic cancer. World J Gastroenterol 2016; 22: 9694.
[9] O'brien PC. Comparing two samples: extensions of the t, rank-sum, and log-rank tests. J Am Stat Assoc 1988; 83: 52-61.
[10] Kleinbaum DG and Klein M. Kaplan-Meier survival curves and the log-rank test. Survival analysis. Springer; 2012. pp. 55-96.
[11] Choulakian V and Stephens MA. Goodness-of-fit tests for the generalized Pareto distribution. Technometrics 2001; 43: 478-484.
[12] Anderson TW and Darling DA. A test of goodness of fit. J Am Stat Assoc 1954; 49: 765-769.
[13] De Haan L, Ferreira A and Ferreira A. Extreme value theory: an introduction. Springer, 2006.
[14] Arnold BC and Press SJ. Bayesian estimation and prediction for Pareto data. J Am Stat Assoc 1989; 84: 1079-1084.
[15] Hosking JR and Wallis JR. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 1987; 29: 339-349.
[16] Grimshaw SD. Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 1993; 35: 185-191.
[17] del Castillo J and Daoudi J. Estimation of the generalized Pareto distribution. Stat Probab Lett 2009; 79: 684-688.
[18] Castillo E and Hadi AS. A method for estimating parameters and quantiles of distributions of continuous random variables. Comput Stat Data Anal 1995; 20: 421-439.
[19] Zhang J. Likelihood moment estimation for the generalized Pareto distribution. Aust N Z J Stat 2007; 49: 69-77.
[20] Castillo E and Hadi AS. Fitting the generalized Pareto distribution to data. J Am Stat Assoc 1997; 92: 1609-1620.
[21] Pham MH, Tsokos C and Choi BJ. Maximum likelihood estimation for the generalized pareto distribution and goodness-Of-Fit test with censored data. J Mod Appl Stat Methods 2019; 17: 11.

[22] Singh VP and Guo H. Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). Hydrol Sci J 1995; 40: 165-181.

[23] Greenwood JA, Landwehr JM, Matalas NC and Wallis JR. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resour Res 1979; 15: 1049-1054.

[24] Kaplan EL and Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958; 53: 457-481.

[25] Bland JM and Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ 1998; 317: 1572-1580.

[26] Strauss DJ, Shavelle RM and Ashwal S. Life expectancy and median survival time in the permanent vegetative state. Pediatr Neurol 1999; 21: 626-631.

[27] Chakraborty A and Tsokos CP. Parametric and non-parametric survival analysis of patients with acute myeloid leukemia (AML). Open J Appl Sci 2021; 11: 126.

[28] Johnson NL. Bivariate distributions based on simple translation systems. Biometrika 1949; 36: 297-304.

[29] Siekierski K. Comparison and evaluation of three methods of estimation of the Johnson SB distribution. Biom J 1992; 34: 879-895.

[30] Soukissian T. Use of multi-parameter distributions for offshore wind speed modeling: the johnson SB distribution. Appl Energy 2013; 111: 982-1000.