Scientific Research Publishing

Parametric and Non-Parametric Analysis of the Survival Times of Patients with Multiple Myeloma Cancer

Lohuwa Mamudu*, Chris P. Tsokos

Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA
Email: *lohuwa@mail.usf.edu, ctsokos@usf.edu

Abstract

Multiple myeloma (MM) is a type of cancer that remains incurable. In the last decade, most research into MM has focused on investigating the improvement in the therapeutic strategy. Our study assesses the survival probability of 48 patients diagnosed with MM based on parametric and non-parametric techniques. We performed parametric survival analysis and found a well-defined probability distribution of the survival time to follow three-parameter lognormal. We then estimated the survival probability and compared it with the commonly used non-parametric Kaplan-Meier survival analysis of the survival times. The comparison of the survival probability estimates of the two methods revealed a better survival probability estimate by the parametric method than the Kaplan-Meier. The parametric survival analysis is more robust and efficient because it is based on a well-defined parametric probabilistic distribution, hence preferred over the non-parametric Kaplan-Meier. This study offers therapeutic significance for further enhancement in the treatment strategy of multiple myeloma cancer.

Keywords

Multiple Myeloma Cancer, Three-Parameter Lognormal, Kaplan-Meier, Maximum Likelihood Estimation, Health Science

1. Introduction

Multiple Myeloma (MM), also known as Kahler disease, myelomatosis, and plasma cell myeloma is a type of cancer that starts from malignant plasma cell (Specifically the white blood cell) [1]. As part of the human immune system is antibodies produced by the plasma cell which fight against germs and other sub-

stances harmful to the human body. When the plasma cell becomes abnormal, called the myeloma cell, it causes myeloma [2]. When the myeloma cells increase, it accumulates in the bone marrow and overcrowds the active blood cells, and with time may destroy the solid part of the bone. Hence, the collection of several myeloma cells in the bones causes multiple myeloma cancer [3] [4]. The development of the myeloma cells is shown in **Figure 1** [2] [5].

Abnormal antibodies are produced by the abnormal plasma cells causing kidney problems and highly thick blood [6]. MM has no specific causes. However, some research has found obesity, radiation exposure, family history, and certain chemicals as associated with the cause of MM [7] [8] [9] [10]. There have been some treatment recommendations for multiple myeloma focused on decreasing the clonal plasma cell population and consequently decrease the symptoms of disease [11]. A preferred treatment like high-dose chemotherapy, commonly with bortezomib-based regimens, and lenalidomide-dexamethasone followed by autologous hematopoietic stem-cell transplantation (ASCT), the transplantation of a person's stem cells have been recommended for MM patients under 65 years [12]. In 2017, a meta-analysis performed has shown that post-ASCT maintenance therapy with lenalidomide has improved the progression-free survival and overall survival in persons at standard risk [13]. Whereas in 2012, it was found from a clinical trial that intermediate and high-risk disease patients benefit from a bortezomib-based maintenance regimen [14].

Statistically, approximately 30,000 new patients are diagnosed with MM in the United States (U.S.) every year, making it the second most common hematologic malignancy in the U.S. [15]. In 2019, a report by the Surveillance, Epidemiology and End Results (SEER) Cancer Institute reported that of all new cancer cases in the U.S for MM constitutes 1.8% and ranked among the top 14 list of cancer
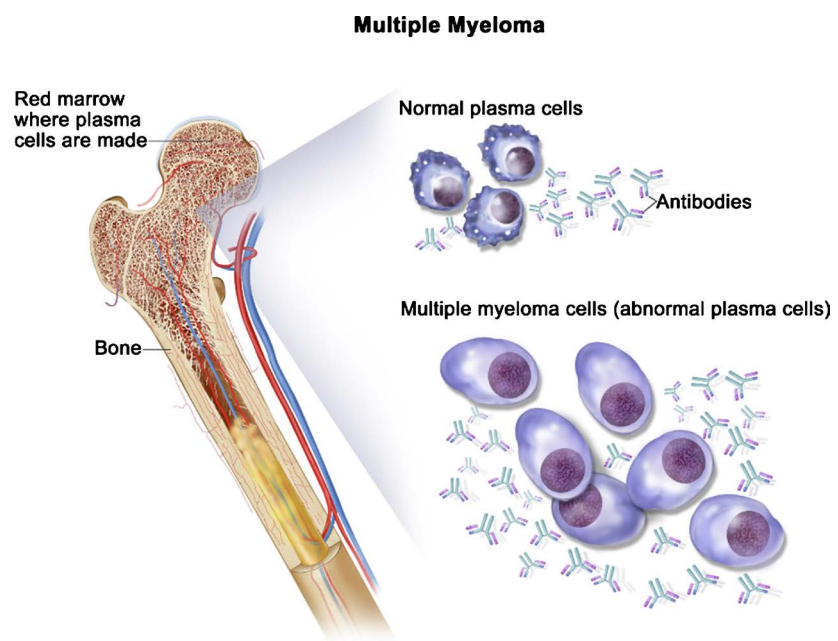

Figure 1. Development of the myeloma cell.

diseases [3]. A further projection by SEER indicates 32,110 estimated new cases of MM and an estimated 12,960 MM patients are expected to die. Those figures are scary and intriguing and cannot be overlooked. There is a sufficient increase compared with the 24,050 estimated new MM cases reported in 2014 [16]. The identified risk factors of MM are reported to be common among the black race, families with MM history, and being a male [3] [17]. SEER cancer institute reported from 2012-2016 that 63.1% of all races and sexes of MM cases are aged 65 or greater.

Though multiple myeloma cancer disease remains incurable, most researches into MM focused on how to improve the survival times of patients diagnosed with MM. The Kaplan-Meier (KM) method has been popularly used for analyzing cancer survivorship data in recent times due to the simplicity of its usage. It is often used to compare the survival difference of observations/groups based on the log-rank test of the null hypothesis that there is no difference. KM is mostly used for longitudinal studies like a cohort study [18]; an example is the present study (*i.e.* the survival time of patients diagnosed with multiple myeloma). Brain *et al.* [19] used Kaplan-Meier to test whether there was a significant difference in the overall survival duration between the categories of risk factors based on the generalized Wilcoxon test and the log-rank test. They found a significant difference in the survival duration between MM patients with LI% < 1% (*i.e.* low percentage labeling index) and LI% ≥ 1% (*i.e.* high percentage labeling index). Also, there was a significant difference in the survival duration for MM patients with the number of DNA synthesizing (S) values $< 1.0 \times 10^{11}$ and S values $\geq 1.0 \times 10^{11}$. Shaji K. Kumar *et al.* [20], used the Kaplan Meier to test for the significant difference of the overall survival from the time of post-transplantation relapse between MM group treated subsequently with one or more of the newer drugs (thalidomide, bortezomib, or lenalidomide) and those not exposed to the newer drug, and they found a significant difference between the two groups.

In the present study, we developed a parametric and non-parametric survival analysis of the survival time of patients diagnosed with multiple myeloma. We believe that being able to find the unique/correct probability distribution or probabilistic behavior that characterizes the survival time is a great step towards getting a good and accurate prediction of the survival duration. It is well known that parametric analysis is more powerful in decision analysis than its non-parametric counterpart. Feigl and Zelen ([1965] p. 835) and other authors have pointed out that the assuming exponential distribution works well for studying the survival of cancer-related cases [21]. However, almost every data given on any cancer survival problem may have an associated well-defined probability distribution. Hence, assuming an exponential distribution for a given cancer survival case without any further investigation is a serious mistake that will lead to making incorrect decisions. We compare the more powerful parametric analysis of the survival time to the commonly known non-parametric Kaplan-Meier analysis of cancer survivorship.

2. Method

2.1. Data Description

The data used in this research is from West Virginia University Medical Center provided by Harley [22] [23]. Originally, the data constituted survival times of 72 multiple myeloma (MM) patients diagnosed and treated with alkylating agents [22]. 65 out of 72 patients provided complete data for 16 contributing variables (risk factor) whiles the remaining 7 were eliminated due to missing data in at least one of the 16 risk factors. Thus, the survival or death times of the MM patients is driven by the 16 risk factors. Given that a patient is diagnosed with myeloma, the 16 risk factors were recorded, and the time up to which the patient survived the disease was noted (called the survival time from diagnosis to the nearest month). Of the 65 patients, 48 and 17 were dead and alive, respectively. In the present research, we utilized the complete data of the 48 patients with the survival times for our analysis.

Before we proceeded to perform the parametric analysis of the survival times of patients with multiple myeloma, we wanted to know whether there is a difference in the survival times for gender, *i.e.* male and female. Given that we have a small data of only 48 patients, we used the log-rank test [24] to compare the difference in survival times of male and female. From **Figure 2**, the log-rank test resulted in a large p-value = 0.45, indicating a failure to reject the null hypothesis (*i.e.* $H_0 : \mu_M = \mu_F$) that there is no difference in the survival times of males and females. Given that we have a small sample size of only 48 patients and the fact that there is no difference in the survival times of males and females, we
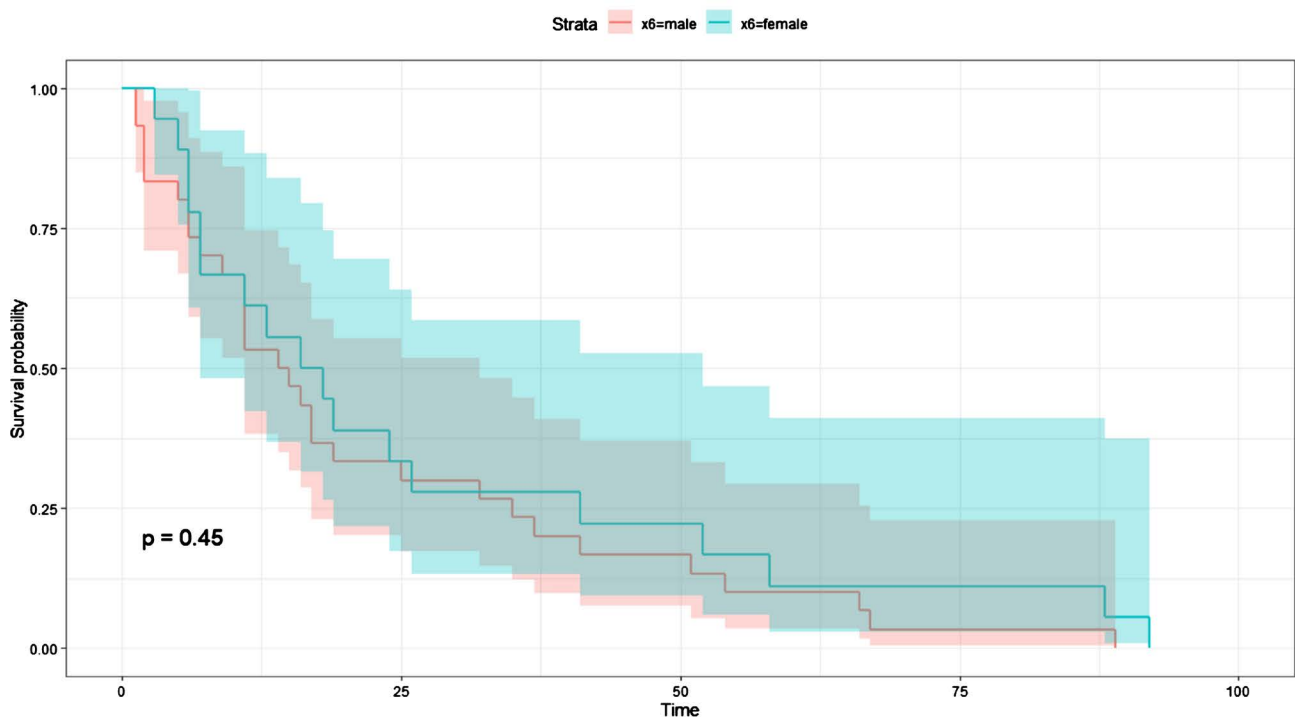

Figure 2. Log-rank test for difference in survival time of gender.

proceeded with parametric analysis without considering stratification based on gender.

2.2. Parametric Analysis of the Survival Times of Multiple Myeloma

Multiple Myeloma (MM) has been, and continues to be, the subject of many research studies. The main goals of these studies are to investigate the means of improving the therapeutic/treatment and survival of MM patients. Nearly 1 to 5 per 100,000 individuals are affected by MM each year worldwide with a higher incidence in the West [25]. The continuous research into the survival rate of MM has seen an improvement from 34.5% to 49.6% in the last 13 years, a statistic reported by the Multiple Myeloma Research Foundation (MMRF) [26]. The difficulties in the investigation of the survival of MM have been hampered by the limitation of data. Most often data collected has a small sample size and duration. This makes it tedious to undertake a parametric analysis. As a result, most analysis of survival time of MM has been based on nonparametric methods. Brian G.M. Durie *et al.* [19] used the Kaplan-Meier method to calculate the actual survival curves of MM stages and test their differences using the generalized Wilcoxon test, and then used the log-rank test to test the difference in the survival duration. Shaji K. Kumar *et al.* [20] also used Kaplan-Meier to test for the overall survival improvement and tested for statistical significance using the 2-tailed log-rank test. Nonparametric tests are preferable if there is no unique probability distribution for the given data. However, their approach is statistically less robust or not as powerful compared to parametric analysis if a pdf can be found. On the contrary, if parametric analyses are used prematurely (*i.e.* assumed a pdf), the result will be misleading.

In the last few decades, scientific transformations have made it possible to employ parametric analysis to data which initially lack a unique probability distribution, particularly for skewed data. For instance, ANOVA is a parametric analysis widely recommended for comparing statistical models and means of different data sets. ANOVA assumes normality, homoscedasticity, and random independent samples. Several suggestions on transformations have been proposed by various authors (Sokal and Rohlf 1995, Zar 1996, Hayek and Buzas 1997, Krebs 1999) to possibly achieve the required assumptions for parametric analysis. To use ANOVA, Rogers, Gilnack, and Fitz (1983) [27], and others used the arcsine-square root transformation. Brown *et al.* (2004) utilized the arcsine-square root transformation to use the paired t-test. Log transformation has often been applied to skewed data to achieve the normal distribution. Giampaolo Merlini *et al.* [28] used log transformation to reduce asymmetry variables to a low level and avoid obvious outliers. Performing parametric analysis is an imperative approach to achieve good statistical analysis and inference by knowing pdf of the data.

2.3. Descriptive Statistics of the Survival Times of Multiple Myeloma

We plotted the histogram to investigate the distribution of the survival time of multiple myeloma of our data as shown in Figure 3. We can see that the distribution of the survival time of MM is right-skewed. Table 2 displays the descriptive statistics of the survival time of MM. If a skewed value is negative (less than zero) implies the data depict left or negative skewness, if a skewed value is positive implies right or positive skewness [29]. So, the positive skewed value of 1.41 in Table 2 is further evidence to support the right-skewed behavior as shown in Figure 3. Statistics like Kurtosis allow for the assessment of highly extreme values in the data. A positive kurtosis value demonstrates a leptokurtic behavior of the distribution, and a negative value displays a platykurtic behavior of the distribution. A kurtosis value of zero implies the distribution behavior is mesokurtic [30]. Thus, the kurtosis value of 0.78 in Table 1 attests to the leptokurtic in the survival time data of MM.
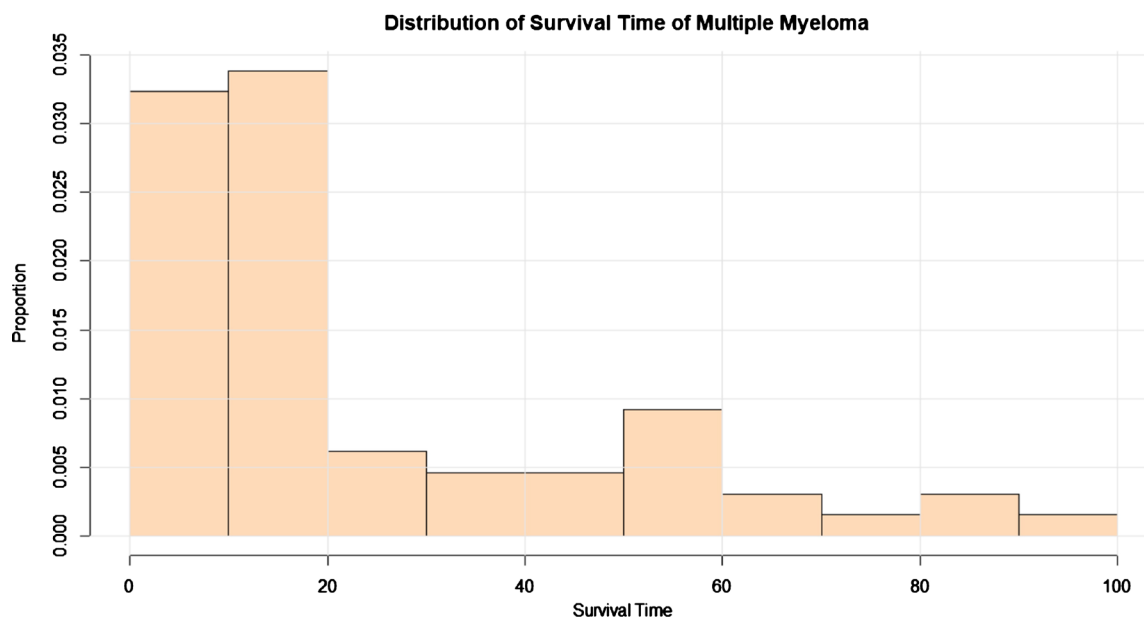

Figure 3. Histogram showing the distribution of survival time (to the nearest month) of multiple myeloma.

Table 1. Descriptive statistics of survival time (to the nearest month) of multiple myeloma.

Mean	Median	Std Err	Std Dev	Kurtosis	Skewness
24.43	15.50	3.56	24.65	0.78	1.33

Table 2. Goodness-of-fit test of the 3p-lognormal distribution of the survival time.

Type of Test	p-value
Kolmogorov-Smirnov	0.90171
Anderson-Darling	0.37878
Chi-Squared	0.69163

2.4. Three-Parameter (3P) Lognormal Probability Estimation of the Survival Times of Patients with Multiple Myeloma

After the assessment of Figure 3, and the descriptive statistics in Table 1, we find that the probability distribution of the survival time of MM follows the 3-parameter lognormal distribution. Table 2 shows three different goodness-of-fit tests of the 3p lognormal probability distribution of the survival times of the MM patients given by the Kolmogorov-Smirnov, Anderson-Darling and Chi-square test. Each of the tests resulted in a large p-value, indicating that we fail to reject the null hypothesis, H_0: the probability distribution follows the 3p-lognormal. In this section, we defined the probability density function (pdf) of the 3p-lognormal distribution and estimate its parameters. Given the survival time of MM, t as a random variable, then the pdf of the 3p-lognormal probability distribution is given by,

$$f\left(t;\gamma,\mu,\sigma^2\right) = \begin{cases} 0, & \text{if } t \le \gamma \\ \left(2\pi\sigma^2\right)^{-\frac{1}{2}}\left(t_i-\gamma\right)^{-1}\exp\left(-\frac{1}{2}\left(\frac{\ln\left(t_i-\gamma\right)-\mu}{\sigma}\right)^2\right), & \text{if } t > \gamma \end{cases} \quad (1)$$

where $\sigma > 0$ denotes continuous shape parameter, $-\infty < \mu < \infty$ represents the continuous scale parameter and γ is the continuous location parameter. If $\gamma \equiv 0$ yields the 2p-lognormal distribution. To estimate the three parameters σ, μ and γ, we would apply the maximum likelihood estimation (MLE) procedure. The MLE estimates the parameters of the probability distribution by maximizing the likelihood function. Generally, MLE is the most used parameter estimation method for statistical inference due to the robustness compared to other traditional methods like the method of moment estimation; the logic is intuitive and flexible [31]. It has some unique and important statistical properties like consistency, invariant, efficiency, sufficiency and asymptotic normality.

The MLE for three-parameter lognormal PDF may not be asymptotically efficient, it is still regarded as a better parameter estimation technique than the method of moments or quantiles because the variance is much smaller [32]. Therefore, the MLE is considered the best parameter estimation method for the 3p-lognormal probability distribution. Calitz (1973) suggested that the MLE procedure by Cohen (1951) should be adopted because it works better than other procedures. To compute the MLE estimators γ, μ, and σ^2, we first find the likelihood function. The likelihood function of the random sample of n independent observations of survival time $t = (t_1, \cdots, t_n)$ for the 3p-lognormal pdf $f(t)$ can be expressed as follows:

$$\begin{aligned} L\left(\gamma,\mu,\sigma^2 \mid t_i\right) &= \prod_{i=1}^{n} f\left(t_i \mid \gamma,\mu,\sigma^2\right) \\ &= \prod_{i=1}^{n}\left[\left(2\pi\sigma^2\right)^{-\frac{1}{2}}\left(t_i-\gamma\right)^{-1}\exp\left(-\frac{1}{2}\left(\frac{\ln\left(t_i-\gamma\right)-\mu}{\sigma}\right)^2\right)\right] \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}}\prod_{i=1}^{n}\left[\left(t_i-\gamma\right)^{-1}\exp\left(-\frac{1}{2}\left(\frac{\ln\left(t_i-\gamma\right)-\mu}{\sigma}\right)^2\right)\right], \quad \forall t_i > \gamma \end{aligned} \quad (2)$$

To find the estimates $\hat{\gamma}, \hat{\mu}$ and $\hat{\sigma}$ of γ, μ and σ, Cohen (1951) obtained local maximum likelihood estimators (LMLE) and equate the partial derivative of the logarithmic likelihood function to zero. Thus, we have

$$\ell\left(\gamma, \mu, \sigma^2; t_i\right) = L\left(\gamma, \mu, \sigma^2 \mid t_i\right)$$
$$= -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \sum_{i=1}^{n}\ln(t_i - \gamma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\ln(t_i - \gamma) - \mu\right)^2 \tag{3}$$

We find $\hat{\mu}$ by equating the partial derivative of ℓ with respect to μ to zero and solving for μ, we have

$$\frac{\partial\ell}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(\ln(t_i - \gamma) - \mu\right) = 0$$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln(t_i - \hat{\gamma}). \tag{4}$$

We find $\hat{\sigma}$ by equating the partial derivative of ℓ with respect to σ to zero and solving for σ.

$$\frac{\partial\ell}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\left(\ln(t_i - \gamma) - \mu\right)^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\ln(t_i - \hat{\gamma}) - \hat{\mu}\right)^2$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\ln(t_i - \hat{\gamma}) - \hat{\mu}\right)^2} \tag{5}$$

Similarly, we equate the partial derivative of with respect to γ and set it equal to zero, that is,

$$\frac{\partial\ell}{\partial\gamma} = \sum_{i=1}^{n}(t_i - \gamma)^{-1} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(t_i - \gamma)^{-1}\left(\ln(t_i - \gamma) - \mu\right) = 0. \tag{6}$$

Hill (1963) demonstrated that arbitrarily large likelihoods can be obtained by allowing $\hat{\gamma}$ to converge on $t_{(1)} = \min\left(t_{(1)}, \cdots, t_{(n)}\right)$, and $t_{(1)}, \cdots, t_{(n)}$ are ordered samples. Thus, the global maximum likelihood results to the inadmissible estimates $\hat{\gamma} = t_{(1)}$, $\hat{\mu} = -\infty$ and $\hat{\sigma} = +\infty$, regardless of the sample. However, Cohen (1951) [33] found the localized estimate $\hat{\gamma}$ to be sufficient in the identification of the 3p-lognormal. The validation of the estimates from such procedure has since been investigated by Calitz (1973), Cohen and Whitten (1980) [34], and Chen (2006) [35] and others. Cohen's (1951) procedure estimated the LMLE for γ by substituting $\hat{\mu}$ and $\hat{\sigma}$ from Equations (4) and (5) into Equation (6) to obtain $\hat{\gamma}$. Thus, we have

$$\left[\sum_{i=1}^{n}(t_i - \gamma)^{-1}\right]\left[\sum_{i=1}^{n}\ln(t_i - \hat{\gamma}) - \sum_{i=1}^{n}\left(\ln(t_i - \hat{\gamma})\right)^2 + \frac{1}{n}\left(\sum_{i=1}^{n}\ln(t_i - \hat{\gamma})\right)^2\right]$$
$$-\left[\sum_{i=1}^{n}(t_i - \gamma)^{-1}\ln(t_i - \hat{\gamma})\right] = 0 \tag{7}$$

To solve for $\hat{\gamma}$ in Equation (7), we can solve iteratively the local maximum likelihood estimate (LMLE) of γ to obtain $\hat{\gamma}$. Here, we take into consideration

admissible roots for which $\hat{\gamma}$ is below $t_{(1)}$. Cohen and Whitten (1980) found only one of such roots; however, in the event of multiple admissible roots, choose the root that results in the closest agreement between the sample mean $\bar{t}$ and the expected value of the 3p-lognormal probability distribution, $E(t) = \hat{\mu}_t = \hat{\gamma} + \exp(\hat{\mu} + \hat{\sigma}^2/2)$. Base on the procedure for the parameter estimation of the three-parameter lognormal probability distribution discussed above, the 3p-lognormal probability distribution of the survival time t of multiple myeloma ($\hat{\gamma}, \hat{\mu}, \hat{\sigma}$), the approximate estimates are given in **Table 3** below.

Table 3. Parameter estimates for the three-parameter lognormal probability distribution of the survival time of multiple myeloma.

Location ($\hat{\gamma}$)	Scale ($\hat{\mu}$)	Shape ($\hat{\sigma}$)
−0.17824	2.7015	1.0429

We substituted the estimates into Equation (1) to obtain the probability density function (pdf) of the 3p-lognormal distribution of the survival time of multiple myeloma given by,

$$f(t) = \begin{cases} 0, & \text{if } t \leq -0.17824 \\ 0.38253(t_i + 0.17824)^{-1} \exp\left(-\frac{1}{2}\left(\frac{\ln(t_i - 0.17824) - 2.7015}{1.0429}\right)^2\right), & \text{if } t > -0.17824 \end{cases} \quad (8)$$

The above findings that the survival times of multiple myeloma patients data follows three-parameter lognormal distribution can ensure efficient and accurate analysis of the survival times of MM patients. Given the pdf in Equation (1), we find the cumulative density function (cdf) by taking the integral of the pdf with respect to the random variable t, given by:

$$F_T(t; \gamma, \mu, \sigma^2) = \int_0^t f_T(s \mid \gamma, \mu, \sigma^2) ds = P(0 \leq t \leq t) E(t \mid 0 \leq t \leq t)$$

$$= \int_0^t (2\pi\sigma^2)^{-\frac{1}{2}} (s_i - \gamma)^{-1} \exp\left(-\frac{1}{2}\left(\frac{\ln(s_i - \gamma) - \mu}{\sigma}\right)^2\right) ds \quad (9)$$

$$= \int_0^t \frac{1}{\sigma(s_i - \gamma)\sqrt{2\pi}} \exp \frac{1}{\sigma(s_i - \gamma)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(s_i - \gamma) - \mu}{\sigma}\right)^2\right) ds.$$

Given the tediousness in integrating the 3p-lognormal pdf, we adopted the method of integration by substitution; substituting into Equation (9) the standard normal cdf given by

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t \exp\left(-\frac{1}{2}z^2\right) dz.$$

In Equation (9), we substitute $z = (\ln(t_i - \gamma) - \mu)/\sigma$ and the remaining part $1/\sigma(t_i - \gamma) \approx 1$. Hence, the 3p-lognormal cdf of the survival times can be expressed as

$$F_T(t; \gamma, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \int_0^t \exp\left(-\frac{1}{2}z^2\right) dz = \Phi\left(\frac{\ln(t_i - \gamma) - \mu}{\sigma}\right). \quad (10)$$
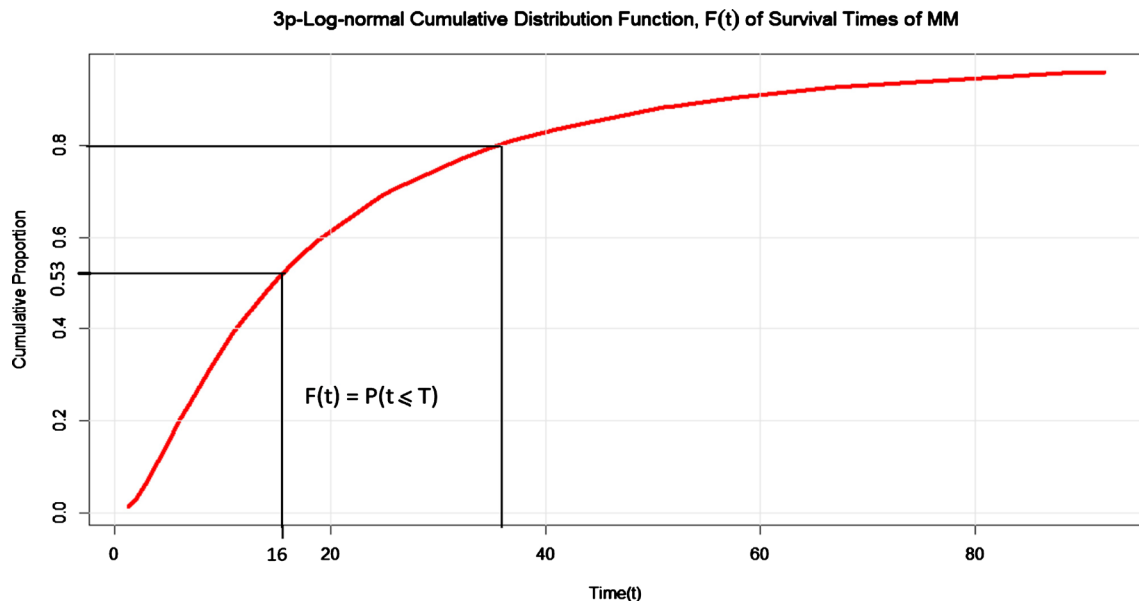
3p-Log-normal Cumulative Distribution Function, F(t) of Survival Times of MM

Figure 4. Cumulative distribution function plot for the survival time of MM.

where $\Phi(.)$ represents the standard normal cdf at a given survival time t_i. Substituting the estimates of **Table 3** into equation (10), we have

$$F_T\left(t;\gamma,\mu,\sigma^2\right) = \Phi\left(\frac{\ln\left(t_i + 0.17824\right) - 2.7015}{1.0429}\right). \tag{11}$$

$\Phi(.)$ denotes the standardized normal cumulative probability distribution. The cdf can be useful in determining the probability of a given random observation (survival time t) would be less than or equal to some value T; thus, $P(t \leq T)$. **Figure 4** shows the cdf plot of the survival times of multiple myeloma patients data. For example, we can estimate the probability that a patient with MM survives up to time $t = 16$ months is approximately 0.53.

Now, given the cdf of the survival times of MM patients in Equation (12), we obtained the reliability of the survival time t of MM patients, given by

$$\hat{S}\left(t_i;\gamma,\mu,\sigma^2\right) = 1 - F_T\left(t;\gamma,\mu,\sigma^2\right) = 1 - \Phi\left(\frac{\ln\left(t_i + 0.17824\right) - 2.7015}{1.0429}\right) \tag{12}$$

The survival function can be used to estimate the probability that a patient diagnosed with multiple myeloma would survive beyond time T; thus, $P(t > T)$. In **Figure 5**, we display the estimate of the survival function $\hat{S}(t)$ of the survival times of MM patients for the 3p-lognormal. As expected, we can see that the survival function of the survival times is decreasing and approximately zero beyond time $t = 80$. For instance, the probability that a patient survives beyond 20 months is approximately 0.40.

2.5. Kaplan-Meier Estimation of Survival Probability of the Survival Times of Patients with Multiple Myeloma

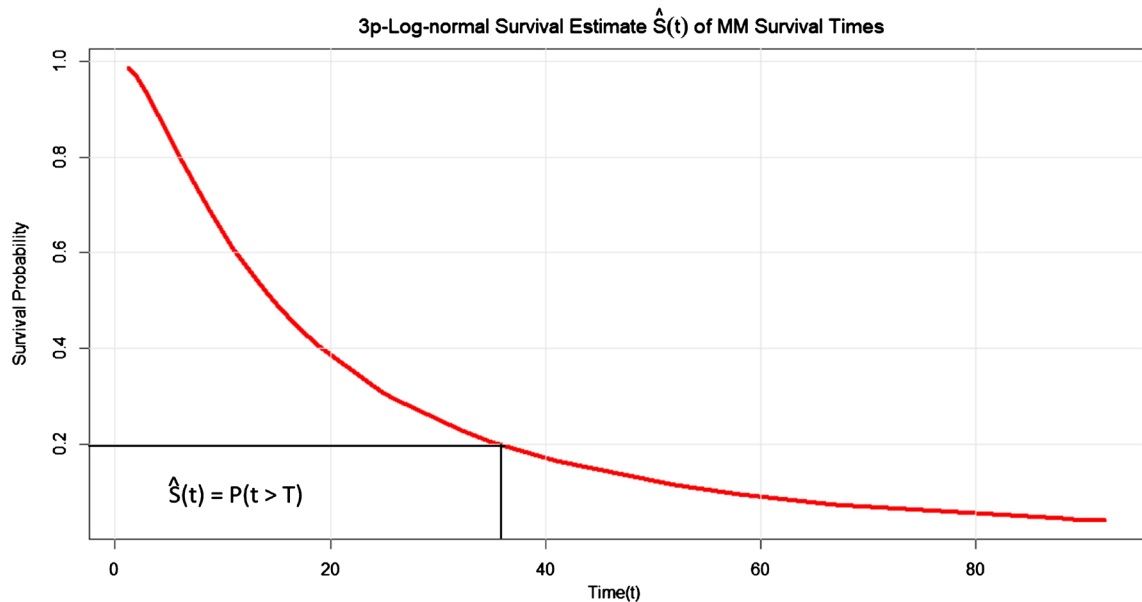Kaplan-Meier estimate (KM) was introduced by Edward L. Kaplan and Paul

Figure 5. Survival estimate for the survival time of MM.

Meier (1958) [36], which is a non-parametric analytical tool. KM defined as the probability of survivorship at a given length of time called the survival time. Graphical representations of KM estimates are used to determine the events, censoring, and survival probability. Another name for KM estimator is the product-limit estimator which estimates the proportion of survival beyond a particular time t, given that some of the observed units may not die or fail. The KM survival estimate $\hat{S}(t)$ of MM patients can be said to represent the reliability estimate $\hat{R}(t)$ of MM patients obtained by taking the product of the conditional probability of surviving at time t_{i+1}, given that a patient survived until time t_i. We estimate the survival function $\hat{S}(t)/\hat{R}(t)$ or the probability that observation last longer than time t as

$$\hat{R}(t) = \hat{S}(t) = \prod_{i:t_i \leq t}\left(1 - \frac{\tau_i}{n_i}\right),$$

where t_i is the time when at least one event happened, τ_i denotes the number of events (e.g. deaths or alive) at time t_i, n_i represents the individuals/observations at risk (not yet had an event or censored) up to time t_i, and τ_i/n_i denotes the probability of survival. In survival analysis, a nearly universal feature data is the censoring data, the commonest of which is the *right-censoring* where an individual expires or removed before the end of the study or clinical trial. The other cases of censoring which rarely happen are the *left-censoring* (the initial time at risk is unknown or individuals removed at the start of the study), and *interval-censoring* (a case of both right and left censored). A key advantage of the KM curve is that it can take into account censored data, particularly right-censoring, and is easy to estimate. The analysis of Kaplan-Meier survival curve takes into consideration the following three assumptions: 1) it assumes that censored observations have the same prospects of survival as those who continue in the
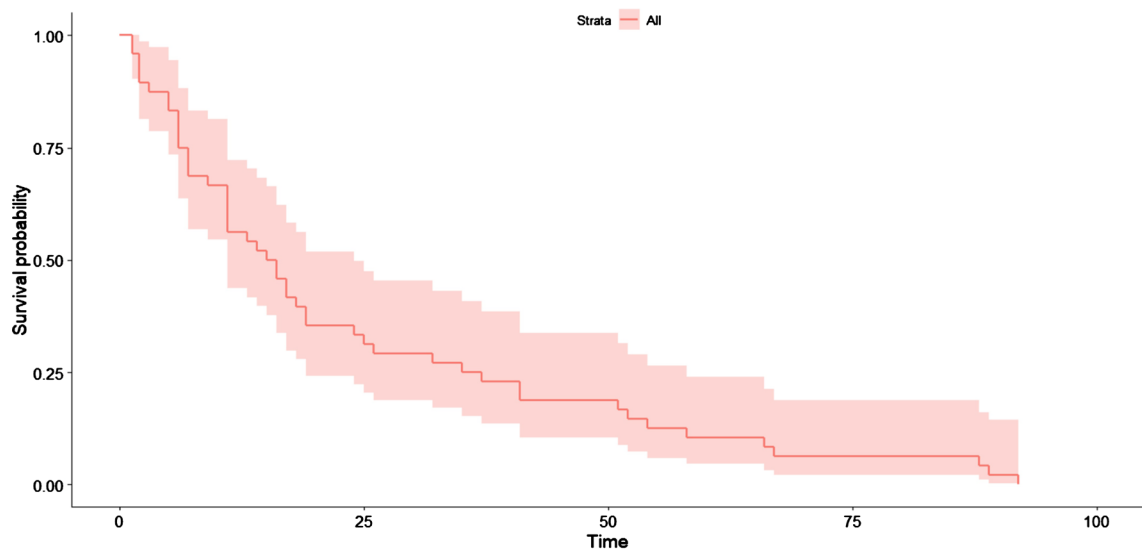
Figure 6. Kaplan Meier global estimates of the survival time with CI.

study, 2) it assumes survival probabilities to be the same for individuals recruited early and late in the study, and 3) it assumes that the event happens at the specified time. We utilized KM to assess the overall survival of MM. In **Figure 6**, we show the KM curve for the global estimates of the probability of survival times of patients diagnosed with multiple myeloma with a 95% confidence interval. We can see that all the MM patients demised by time 92 (in months). Though the KM estimator is the most commonly used technique of survival analysis, and is useful for examining recovery rates, the probability of death and effectiveness of treatment. However, as we mentioned before the KM method is not as powerful as the parametric survival analysis for decision making. We cannot obtain the hazard function to estimate the rate at which patients die with MM using KM.

2.6. Comparison of Three-Parameter Lognormal with the Kaplan-Meier Estimation of the Survival Function

In the parametric analysis, we found the survival times of patients with MM follow the three-parameter lognormal probability distribution. In section 2.5, we performed a non-parametric analysis using the Kaplan-Meier to estimate the survival probability of the MM patients. Now, we compare the survival estimate of the 3p-lognormal with the Kaplan-Meier survival estimate of the survival times of the MM patients. The relevance of the survival function of the two methods is to estimate the survival probability of a patient diagnosed with MM beyond a given time, after a given treatment. The survival times along with the survival probabilities of the two survival functions is given by **Table 4**. We see that the probability estimates by the 3p-lognormal survival function are mostly higher than that of Kaplan-Meier. Though there are times in which the KM estimates higher probabilities, the 3p-lognormal survival function from the parametric probability distribution estimates the survival proportion more accurately than the Kaplan-

Table 4. Kaplan-Meier ($\hat{S}_{KM}(t)$) vs parametric (3P-lognormal, $\hat{S}_P(t)$) survival function estimate of the survival times.

t_i	$\hat{S}_{KM}(t)$	$\hat{S}_P(t)$	t_i	$\hat{S}_{KM}(t)$	$\hat{S}_P(t)$
1.25	0.958	0.988	25.00	0.313	0.308
2.00	0.896	0.967	26.00	0.292	0.295
3.00	0.875	0.931	32.00	0.271	0.230
5.00	0.833	0.845	35.00	0.250	0.205
6.00	0.750	0.801	37.00	0.229	0.190
7.00	0.688	0.758	41.00	0.188	0.164
9.00	0.667	0.679	51.00	0.167	0.118
11.00	0.563	0.609	52.00	0.146	0.115
13.00	0.542	0.547	54.00	0.125	0.108
14.00	0.521	0.519	58.00	0.104	0.098
15.00	0.500	0.493	66.00	0.083	0.076
16.00	0.458	0.469	67.00	0.063	0.074
17.00	0.417	0.446	88.00	0.042	0.044
18.00	0.396	0.424	89.00	0.021	0.043
19.00	0.354	0.404	92.00	0.000	0.040
24.00	0.333	0.321			

Meier. Thus, because parametric methods are more powerful, robust and efficient than non-parametric methods.

3. Discussion

Given the danger posed by multiple myeloma (MM) cancer in recent years, it is imperative to investigate the prognosis and how to enhance the therapeutic/ treatment strategy of MM. While MM cancer currently remains incurable, there has been some improvement in the treatment process. The treatment progress has been largely due to the introduction of therapeutic agents such as thalidomide, lenalidomide, and bortezomib, as well as the introduction of high-dose chemotherapy and stem-cell rescue (ASCT). The quest to improve the survival time (or increase the death time) of patients with MM has resulted in adopting several different research approaches and methodologies.

In the present study, 1) we identified a well-defined probability distribution that characterizes the survival times of the 48 patients diagnosed with MM and used it to estimate the survival function. 2) we estimated the proportion of survival time utilizing the commonly used Kaplan-Meier (KM) technique of analysis of cancer survivorship. 3) we compare and contrast the relevance and efficiency of the two different methods of analysis of survival probability estimation of patients diagnosed with MM beyond a given survival time.

Firstly, we investigated utilizing the log-rank test to test the difference be-

tween the survival times of the males and females diagnosed with MM. We found that the survival times of both males and females diagnosed with MM are not different, so we performed the analysis using the combined data of the males and females.

We then found a well-defined probability distribution that characterizes the survival time of the 48 patients diagnosed with MM follows the 3p-lognormal probability distribution. We believe that being able to find the best probability distribution that characterizes the probabilistic behavior for a given cancer patient survival time can lead to estimating the survival probability with much more accuracy and efficiency. The fact that we found a unique probability distribution for our study of the survival times of patients diagnosed with MM refute the suggestion of assuming exponential distribution (as suggested by Feigl and Zelen ([1965] p. 835) and other authors) or using the non-parametric Kaplan-Meier for most cancer survivorship studies.

We found both the 3p-lognormal survival most often estimates higher survival probability than the KM survival function, given by Table 4. We know that KM is the most commonly used technique for analyzing cancer survivorship data. Statistically, the parametric technique is said to be more robust and efficient than the non-parametric counterpart. Therefore, our finding of the parametric 3p-lognormal probability distribution is better and of a higher degree of accuracy in estimating the survival probability of the patients diagnosed with MM than the Kaplan-Meier. The KM technique is more often used to compare the difference between the survival probability of the survival times of two or more entities or groups usually based on the log-rank test. However, by obtaining the best parametric probability distribution that characterizes the survival times, we can find the survival function and estimate the survival rate and compare the results of two or more entities with a high degree of accuracy. The only situation where non-parametric becomes highly applicable or recommended is when there is no parametric form (*i.e.* pdf) for the given data. One key demerit of KM is that it cannot be used to estimate the rate at which patients die or survive with MM (*i.e.* the hazard function). The hazard function can be easily calculated after finding the parametric probability distribution by dividing the probability density function, pdf by the survival function.

4. Conclusions

We estimated the survival probability of patients diagnosed with multiple myeloma (MM) using two different methods; the parametric three-parameter lognormal and the non-parametric Kaplan-Meier. We found the parametric method to often give a higher estimate of the survival probability than the non-parametric method. Even though there are times when the non-parametric survival probability estimates are higher, all-important arguments favor the parametric method. The difficulty of the parametric survival analysis is the fundamental intrinsic assumption that the survival times of the population under study follow a

specific probability distribution. But if such a limitation can be overcome, then we can obtain a more robust or powerful result from the parametric analysis. We can also estimate the hazard function to assess the rate at which patients die with MM after finding the right parametric distribution.

Base on the two different methods utilized for estimating the probability of survival of patients diagnosed with MM, we offer the following essential recommendations. 1) If the only information about patients is the survival (death) time, then estimating the survival probability using the parametric technique will yield more accurate, robust, and efficient results than the commonly used Kaplan-Meier. 2) However, if no unique or well-defined parametric probability distribution is found, then we still recommend the use of Kaplan-Meier technique for the estimation of the survival probability. Though the use of non-parametric Kaplan-Meier survival analysis may, in some cases result in a similar or higher estimate of the survival rate (such as in our case), the parametric analysis remains more powerful, robust and efficient, and hence must be considered first in the analysis of any given cancer survivorship data. The study provides an improved method for estimating the survival probability and analysis of cancer survivorship data, and further enhance the therapeutic/treatment process of the incurable multiple myeloma cancer.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Raab, M.S., Podar, K., Breitkreutz, I., Richardson, P.G. and Anderson, K.C. (2009) Multiple Myeloma. *The Lancet*, **374**, 324-339.
https://doi.org/10.1016/S0140-6736(09)60221-X

[2] Ferri, F.F. (2013) Ferri's Clinical Advisor 2014. Elsevier Health Sciences, Amsterdam, 726.

[3] SEER Cancer Facts: Myeloma. National Institute, Bethesda, MD.
https://seer.cancer.gov/statfacts/html/mulmy.html

[4] About Multiple Myeloma. American Cancer Society.
https://www.cancer.org/cancer/multiple-myeloma/about/what-is-multiple-myeloma.html

[5] Plasma Cell Neoplasms (Including Multiple Myeloma) Treatment. National Cancer Institute.
https://www.cancer.gov/types/myeloma/patient/myeloma-treatment-pdq#section/all

[6] Van de Donk, N.W., Mutis, T., Poddighe, P.J., Lokhorst, H.M. and Zweegman, S. (2016) Diagnosis, Risk Stratification and Management of Monoclonal Gammopathy of Undetermined Significance and Smoldering Multiple Myeloma. *International Journal of Laboratory Hematology*, **38**, 110-122. https://doi.org/10.1111/ijlh.12504

[7] World Health Organization (2014) World Cancer Report 2014. Chapter 5.13.

[8] Roberts, D.L., Dive, C. and Renehan, A.G. (2010) Biological Mechanisms Linking Obesity and Cancer Risk: New Perspectives. *Annual Review of Medicine*, **61**, 301-316. https://doi.org/10.1146/annurev.med.080708.082713

[9] Dutta, A.K., Hewett, D.R., Fink, J.L., Grady, J.P. and Zannettino, A.C. (2017) Cutting Edge Genomics Reveal New Insights into Tumour Development, Disease Progression and Therapeutic Impacts in Multiple Myeloma. *British Journal of Haematology*, **178**, 196-208. https://doi.org/10.1111/bjh.14649

[10] Landgren, O., Kyle, R.A., Pfeiffer, R.M., Katzmann, J.A., Caporaso, N.E., Hayes, R.B., Dispenzieri, A., Kumar, S., Clark, R.J., Baris, D., Hoover, R. and Rajkumar, S.V. (2009) Monoclonal Gammopathy of Undetermined Significance (MGUS) Consistently Precedes Multiple Myeloma: A Prospective Study. *Blood*, **113**, 5412-5417. https://doi.org/10.1182/blood-2008-12-194241

[11] Korde, N., Kristinsson, S.Y. and Landgren, O. (2011) Monoclonal Gammopathy of Undetermined Significance (MGUS) and Smoldering Multiple Myeloma (SMM): Novel Biological Insights and Development of Early Treatment Strategies. *Blood*, **117**, 5573-5581. https://doi.org/10.1182/blood-2011-01-270140

[12] Kyle, R.A. and Rajkumar, S.V. (2008) Multiple Myeloma. *Blood*, **111**, 2962-2972. https://doi.org/10.1182/blood-2007-10-078022

[13] McCarthy, P.L., Holstein, S.A., Petrucci, M.T., Richardson, P.G., Hulin, C., Tosi, P., Bringhen, S., Musto, P., Anderson, K.C., Caillot, D., Gay, F., Moreau, P., Marit, G., Jung, S.H., Yu, Z., Winograd, B., Knight, R.D., Palumbo, A. and Attal, M. (2017) Lenalidomide Maintenance after Autologous Stem Cell Transplant in Newly Diagnosed Multiple Myeloma: A Meta-Analysis. *Journal of Clinical Oncology*, **35**, 3279-3289. https://doi.org/10.1200/JCO.2017.72.6679

[14] Sonneveld, P. (2012) Bortezomib Induction and Maintenance Treatment in Patients with Newly Diagnosed Multiple Myeloma. *Journal of Clinical Oncology*, **30**, 2946-2955. https://doi.org/10.1200/JCO.2011.39.6820

[15] Siegel, R.L., Miller, K.D. and Jemal, A. (2017) Cancer Statistics, 2017. *CA: A Cancer Journal for Clinicians*, **67**, 730. https://doi.org/10.3322/caac.21387

[16] American Cancer Society (2014) Cancer Facts Figures. http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc04215.pdf

[17] Becker, N. (2011) Epidemiology of Multiple Myeloma. *Recent Results in Cancer Research*, **183**, 25-35. https://doi.org/10.1007/978-3-540-85772-3_2

[18] Meyer, B.D. (1990) Unemployment Insurance and Unemployment Spells. *Econometrica*, **58**, 757-782. https://doi.org/10.2307/2938349

[19] Brian, G.M., *et al.* (1980) Pretreatment Tumor Mass, Cell Kinetics, and Prognosis in Multiple Myeloma. *Blood*, **55**, 364-372. https://doi.org/10.1182/blood.V55.3.364.364

[20] Kumar, S.K., *et al.* (2008) Improved Survival in Multiple Myeloma and the Impact of Novel Therapies. *Blood*, **111**, 2516-2520. https://doi.org/10.1182/blood-2007-10-116129

[21] Feigl, P. and Zelen, M. (1965) Estimation of Exponential Survival Possibilities with Concomitant Information. *Biometrics*, **21**, 826-838.

https://doi.org/10.2307/2528247

[22] Krall, J.M., Uthoff, V.A. and Harley, J.B. (1975) A Set-up Procedure for Selecting Variables Associated with Survival. *Biometrics*, **31**, 49-57.
https://doi.org/10.2307/2529709

[23] Harley, J. B. (1971) Ten Years of Experience in Multiple Myeloma at the West Virginia University Hospital. Morgantown, WV.

[24] Richard, P. and Julian, P. (1972) Asymptotically Efficient Rank Invariant Test Procedures. *Journal of the Royal Statistical Society*, *Series A*, **135**, 185-207.
https://doi.org/10.2307/2344317

[25] Parkin, D.M., Bray, F., Ferlay, J. and Pisani, P. (2005) Global Cancer Statistics, 2002. *CA*: *A Cancer Journal for Clinicians*, **55**, 74-108.
https://doi.org/10.3322/canjclin.55.2.74

[26] Multiple Myeloma Research Foundation (MMRF).
https://themmrf.org/multiple-myeloma/prognosis/

[27] Rogers, C.S., Gilnack, M. and Fitz III, H.C. (1983) Monitoring of Coral Reefs with Linear Transects: A Study of Storm Damage. *Journal of Experimental Marine Biology and Ecology*, **66**, 285-300. https://doi.org/10.1016/0022-0981(83)90165-X

[28] Merlini, G., Waldenstrom, J.G. and Jayakar, S.D. (1980) A New Improved Clinical Staging System for Multiple Myeloma Based on Analysis of 123 Treated Patients. *Blood*, **55**, 1011-1019. https://doi.org/10.1182/blood.V55.6.1011.1011

[29] Doane, D.P. and Seward, L.E. (2011) Measuring Skewness: A Forgotten Statistic? *Journal of Statistics Education*, **19**, 1-18.
https://doi.org/10.1080/10691898.2011.11889611

[30] Sharma, R. and Bhandari, R. (2015) Skewness, Kurtosis and Newton's Inequality. *Rocky Mountain Journal of Mathematics*, **45**, 1639-1643.
https://doi.org/10.1216/RMJ-2015-45-5-1639

[31] Chambers, R.L., Steel, D.G., Wang, S.J. and Welsh, A. (2012) Maximum Likelihood Estimation for Sample Surveys. CRC Press, Boca Raton, FL.
https://doi.org/10.1201/b12038

[32] Calitz, F. (1973) Maximum Likelihood Estimation of the Parameters of the Three Parameter Lognormal Distribution: A Reconsideration. *Australian Journal of Statistics*, **9**, 221-226. https://doi.org/10.1111/j.1467-842X.1973.tb00199.x

[33] Cohen, A.C.J. and Whitten, B.J. (1980) Estimation in the Three-Parameter Lognormal Distribution. *Journal of the American Statistical Association*, **75**, 399-404.
https://doi.org/10.1080/01621459.1980.10477484

[34] Cohen, A.C.J. (1951) Estimating Parameters of Logarithmic-Normal Distributions by Maximum Likelihood. *Journal of the American Statistical Association*, **46**, 206-212.
https://doi.org/10.1080/01621459.1951.10500781

[35] Chen, C. (2006) Tests of Fit for the Three-Parameter Lognormal Distribution. *Computational Statistics Data Analysis*, **50**, 1418-1440.
https://doi.org/10.1016/j.csda.2005.01.012

[36] Kaplan EL. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**, 457-481.
https://doi.org/10.1080/01621459.1958.10501452