# Perspectives on Hazard Rate Functions: Concepts; Properties; Theories; Methods; Computations; and Application to Real-Life Data

## Anthony Joe Turkson

Mathematics, Statistics and Actuarial Science Department, Takoradi Technical University, Sekondi-Takoradi, Ghana
Email: anthony.turkson@ttu.edu.gh

## Abstract

A critical look at living organisms, devices, socio-economic units, and social units reveals that at any point in time of their survival, there will be a well-defined continuum of states: good $\rightleftharpoons$ bad; healthy $\rightleftharpoons$ unhealthy; and functional $\rightleftharpoons$ dysfunctional. Recent studies have shown that, the hazard function plays a crucial role in depicting the aging process. It is in trying to investigate the underlying processes of the probability distributions of survival functions, and an attempt to understand intuitively the concept of hazard rate functions that informed the conduct of this study. Simulation design and real-life tertiary data were employed in this study. The plot for the cumulative hazard function for the Weibull model gave us an intercept on the Y-axis as $-3.314$ and the intercept on the time-axis as 1.397. The curve was approximately linear, this meant that the data used for the study fitted the model. The plot of the cumulative hazard function against time for the exponential model passed through the origin, implicitly, the data fitted the model. Further results revealed that as the hazard rates decreased from $0.061 \Rightarrow 0.051$, survival probabilities increased from $0.941 \Rightarrow 0.950$ respectively; and as the hazard rates increased from $1.098 \Rightarrow 1.609$, survival probabilities decreased from $0.333 \Rightarrow 0.200$ respectively. We noted again that, the risk of death was distributed among all four BMI groupings, the effect of the BMI was not readily seen. Gender and age appeared not to contribute significantly towards death due to heart attack. We also saw that the hazard rate for the first few days for all the four categories of BMIs was about constant, on the 3rd and 4th days there was a significant increase in the hazard rates especially for the female obese category. For the male category, we noted that there were stepwise increases in the hazards of three of the BMI categories; underweight, obese and healthy

weights. This study has intuitively demonstrated, theorized, modelled, discussed and explored the relevance of the hazard model in assessing risk.

## 1. Introduction

A critical look at organisms like human beings and animals; devices like phones and computer sets; socio-economic units like organizations and labor unions; and social units like families and churches, reveals that at any point in time of their survival, there will be a well-defined state. A patient, may be alive or dead after some medical treatment, a workforce may be out of work due to injuries, and machinery may either be down or functioning. We notice that risk is inevitable to all these entities. One of the topics which are of great importance in biostatistics is the hazard rate function which assumes its importance in the calculation of risk rates. Over the years, researchers have been estimating the hazard rate function using the Kaplan-Meier and Nelson-Aalen. Hazard rate function is an important concept that can be used to postulate life distributions in the presence of several competing risk factors, it is perhaps the most popular of the techniques used in modeling and analyzing survival data. The most common use of the function is to model an entity's chance of death as a function of their age. It can also be used to model any other time-dependent event of interest. The function models the occurrence of only one event, namely the first event, whereas the intensity function models the occurrence of a sequence of events over time. This is due to its intuitive interpretation as the amount of risk associated with a unit at age $t$. Another reason for its popularity is that it is a special case of the intensity function for a non-homogeneous Poisson process. Intuitively, if we have data with discrete times in weeks, months, or years, we could get an intuitive idea of the hazard rate. The hazard rate is the unobserved rate at which events occur. For instance, if the hazard rate was constant over time and it was equal to 2, this would mean that one should expect 2 events to occur in a unit time interval. Again, if one entity had a hazard rate of 1.5 at time $t$ and a second entity had a hazard rate of 3.0 at time $t$, then we could say that the second entity's risk of an event would be twice as much as the first one at time. The most common use of the function is to model an entity's chance of death as a function of their age. It can also be used to model any other time-dependent event of interest. It is important to note that though the hazard rate is an un-observed variable, it controls both the occurrence and the timing of the events. It is the fun-

damental dependent variable in survival analysis. The derivative of the survivorship function which is the hazard function is the instantaneous risk of death. The function measures the conditional probability of a failure, given that the entity has worked past a time point $t$, apparently, the function measures the instantaneous risk of having an event at time $t$ given that the entity has survived up to $t$. In perspective, the hazard rate function is more informative about the underlying mechanism of failure than the other techniques used in analyzing lifetime distributions. The hazard function assumes other aliases in other fields: Force of mortality or force of decrement in demography and actuarial sciences; intensity function in stochastic processes; in vital statistics and in the life sciences, it is known as age-specific failure rate; inverse of the Mill's ratio in economics; in point process and extreme value theory it is known as the rate function or intensity function; in the engineering sciences, it is known as the failure rate and in reliability analysis, it is known as conditional failure rate. The failure density (pdf) measures the overall speed of failures whiles the hazard rate measures the dynamic speed of failures. The hazard rate for the occurrence of events may be increasing, decreasing, constant, bathtub-shaped or hump-shaped. Events where there are wear and tear or which are connected to aging normally produce increasing hazards with time, events like death of a child at age five and beyond normally produces decreasing hazards of deaths with time. Bathtub hazards normally occur in populations followed from birth to death. If the hazard for an event increases steadily and starts declining with time then the hazard curve is called the hump-shaped, this type of hazard is associated with hazards of death after surgery. For discrete hazard rates $h(t)$: $0 \le h(t) \le 1$. In other words, the hazard rate or function cannot be negative.

Cox-Oakes (1984) as contained in Wu, [1] provided reasons why consideration of the hazard rate function may be a good idea than other methods used in summarizing survival analysis:

- it is instructive to consider the risk attached to an entity which is alive at age $t$;
- useful in the comparison of groups of individuals;
- convenient when there is censoring or several types of failures;
- simple to compare with an exponential distribution; and
- It is the special form for the single failure system of the complete intensity.

Another key feature of the hazard function has to do with how the shape of the hazard function could influence other functions of interest such as the survival function. Figure 1, illustrates a hazard function with a 'bathtub shape' (observed failure rate). This graph is depicting the hazard function for the survival of a patient. At time $t = 0$, the patient was having a surgery (heart surgery), the risk of dying was high, therefore, the patient would have a very high hazard. After a successful surgery, the hazard function decreases. At another point in time, the patient's condition might remain stable for some time, the next time you observe, you might see the patient experiencing deterioration and the chances of dying increases again, thus, the hazard function starts increasing. Besides, the
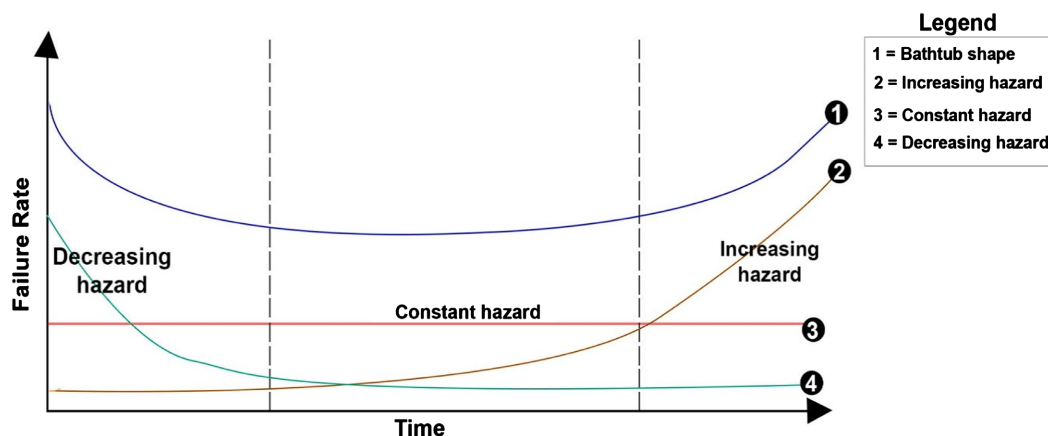
**Figure 1.** Graph showing "Bathtub shape"; increasing hazard rate; decreasing hazard rate; and constant hazard rate.

bathtub shape, Figure 1 also presents the shapes of the three other curves; increasing hazard rate (wear out failure); decreasing hazard rate (early infant mortality failure); and constant hazard rate. The hazard function may not seem exciting to model compared with some other functions, but it is of interest to note that functions such as the cumulative hazard function and survival function are derived from the hazard rate function [Equation (7) and Equation (8) respectively]. Once we model the hazard rate function, we could easily obtain these other functions. When we graph the Nelson-Aalen cumulative hazard function on the vertical axis and the Cox Snell residuals on the horizontal axis, it gives us the opportunity to compare the hazard function to the diagonal line. If the hazard function follows the 45-degree line, then, we could say that the function has approximately an exponential distribution with a hazard rate of one (1), again, the model fits the data well as in UCLA [2].

In survival analysis, some researchers, Clark *et al.* [3] and Park *et al.* [4] who study the timing and occurrence of event, often analyze the probability distributions of the time preceding the occurrence of the event. They focus mostly on the end result of the process, rather than the processes that generated the end results, but in real life, apart from chance events, most events do not just happen, there may be some underlying developments preceding the events. When researchers consider the underlying processes leading to the end result, it might improve the understanding of the concept generating the end result, some studies have revealed that the hazard function plays a crucial role in characterizing the aging process. It is in trying to find the underlying concepts which generate the end result, and an attempt to understand intuitively the concept of hazard rate functions that this study was conducted. The study was carried out as follows: Looked at some concepts underlying the hazard rate function; provided theoretical and mathematical definitions; established some theorems and properties and provided proofs; identified real-life situations that generated the various shapes of the hazard rate function; formulated the hazard rate function and hazard rate models; undertook some computations with intuitive interpretation;

simulated some studies using the R software; and finally, applied the principle governing the hazard rate function to real-life data.

## 2. Conceptual Framework

Eiser *et al.* [5], underscored the fact that understanding how people interpret risks and choose actions based on their interpretations was vital to any strategy for disaster reduction. Kurniasari *et al.* [6] have maintained that in studying survival analysis, one of the key points worthy of investigation was hazard rate. They averred that the hazard function was an alternative characterization of the distribution of $T$ ($T$ being a continuous random variable with probability density function (pdf) $f(t)$ and cumulative distribution function $F(t) = Pr(T < t)$, giving the probability that the event has occurred by time $t$). Zhang and Peng [7] established that crossing hazard functions had extensive applications in modeling survival data and maintained that existing studies in the literature mainly focused on comparing crossed hazard functions and estimating the time at which the hazard functions crossed, and that there was little theoretical work on conditions under which hazard functions from a model will have a crossing. Blackstone [8] has noted that the hazard function was the instantaneous rate of occurrence of a time-related event, such as death and indicated that there were methods, we could use to determine the hazard function from clinical outcome data, to identifying risk factors for higher hazard, and to generating patient-specific predictions.

Read and Vogel [9] have maintained that the field of hazard function analysis (HFA) involved a probabilistic assessment of the time to failure $T$ of an event of interest. They intimated that the hazard function $h(t)$, was central to HFA, and averred that for a stationary process, the probability distribution function (pdf) of the return period always followed an exponential distribution, the same could not be said for nonstationary processes. Similar views were espoused by Upadhyay [10] who noted that hazard rate function (HRF) was an important concept for researchers and practitioners working in areas such as engineering statistics, and biomedical sciences. He stated that hazard rate function had the tendency to provide an alternative characterization for the distribution of a random variable, especially when dealing with the lifetime data. Boland *et al.* [11] has adjudicated that hazard rate ordering was an ordering for random variables which compared lifetimes with their hazard rate functions. They maintained that the hazard rate ordering was particularly useful in reliability theory and survival analysis, owing to the importance of the hazard rate function in those areas. The research conducted by Greenwich [12] culminated in the use of a unimodal hazard rate function in modeling failure rate that had a relatively high rate of failure in the middle of expected life time.

### 2.1. Definition of Hazard Function

The hazard rate measures the propensity of an item to fail or die depending on

the age it has reached. It is part of a wider branch of statistics called survival analysis. Hazard rate is part of a larger equation called the hazard function. The hazard function can be defined as the instantaneous risk that the event of interest happens, within a small time period. The hazard function is a conditional failure rate, in that, it is conditioned on the premises that a person has actually survived until time *t* as in Hinchliffe [13]. In other words, the function at year ten (10) only applies to those who were actually alive in year 10; it does not take into consideration those who died in previous years. The hazard function is used to model the distribution of data in survival analysis. It is used to model a subjects' chances of death as a function of age, again, it models periods with the highest or lowest chances of an event. It can also be used to model any other time-dependent variable. The Kaplan Meier (KM) method uses rates for the hazard function with no upper limit, this is preferred in clinical trials. In actuarial method, the hazard function is stated as a proportion as stated in Glen [14]. The hazard function $h(t \mid z)$ or simply h(*t*) is defined mathematically as:

$$h(t \mid z) = \lim_{\Delta t \to 0+} \frac{1}{\Delta t} P\{t \leq T < t + \Delta t \mid T \geq t, Z = z\}. \tag{1}$$

where,

$Z$ = *m*-dimensional vector of covariates;

$h(t|z) = h(t)$ = Hazard function;

$P$ = Conditional probability of failure;

$t$ = survived up to time *t;*

$T$ = random variable representing the survival time, which is nothing but the time-to-event;

$\Delta t$ = given small time period;

$(t + \Delta)$ = Time interval.

## 2.2. Formulation of the Hazard Function

The numerator of the expression in Equation (1) is the conditional probability that the event will occur in the interval $(t, t + \Delta t)$ given that it has not occurred earlier; the denominator $\Delta t$ is the width of the interval. Dividing the numerator by the denominator gives us a rate of event occurrence per unit of time. The limit as the width approaches zero, gives us the instantaneous rate of occurrence. The conditional probability in the numerator may be written as the ratio of the joint probability that *T* is in the interval $(t, t + \Delta t)$ and $T \geq t$ to the probability of the condition $T \geq t$. The former may be written as $f(t)\Delta t$ for small $\Delta t$ while the latter is $S(t)$. We must note that the hazard function is not a probability value but a rate value because we are finding the ratio of the probability to a time interval. The units of $h(t)$ is probability/probability/time which is 1/time. It is counts per time (which gives us rate). Estimating the hazard function $h(t)$ is not straight forward. We must first estimate the cumulative hazard function $H(t)$ which is used as an intermediary to estimating $h(t)$. We can use the Nelson-Aalen estimator to first estimate $H(t)$ and then proceed from

there to calculate the hazard function. To find an expression for $h(t)$, we should realize that $h(t)$ must be a conditional probability; it is conditional on not having the event up to time $t$ (or conditional on surviving to time $t$).

From probability theory, the conditional probability is given by:

$$Pr(A \backslash B) = \frac{Pr(A \text{ and } B)}{Pr(B)}. \tag{2}$$

where $A$ = having the event at time $t$;

$B$ = not having the event by time $t$.

Here $P(A \text{ and } B) = \frac{\Delta F}{\Delta t} = \frac{dF}{dt}$. That is, the delta probability of the event per unit time of $\Delta t$ and we cannot compute the probability at exactly time $t$ is zero.

$P(B) = S(t)$, as defined earlier.

$$h(t) = \frac{\dfrac{dF(t)}{dt}}{S(t)} = \frac{df}{S(t)} \tag{3}$$

where,

$f(t)$ = the probability density function of survival time,

$S(t)$ = the survivor function (the probability of surviving beyond a certain point in time).

If $T$ is a continuous random variable then,

$$h(t \mid z) = h(t) = \frac{f(t)}{S(t)} = -d \ln [S(t)]. \tag{4}$$

To prove that $\ln S(t) = -h(t)$ (5)

We proceed using this rule of calculus $\dfrac{d}{dt}\{\ln(u)\} = \dfrac{1}{u}\dfrac{du}{dt}$ (6)

Equation (4), Equation (5) and Equation (6) gives

$$\frac{d \ln S(t)}{dt} = \frac{1 dS(t)}{S(t) dt} = \frac{dS(t)/dt}{S(t)} = \frac{d[1 - F(t)]/dt}{S(t)} = \frac{-f(t)}{S(t)} = -h(t).$$

The cumulative hazard function $H(t)$ is also defined as,

$$H(t) = \int_0^t h(u) du = -\ln[S(t)]. \tag{7}$$

Thus, for a continuous random variable,

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u) du\right]. \tag{8}$$

The results from Equation (3) to Equation (8) show that the hazard and survival functions provide alternative but equivalent characterizations of the distribution of $T$. Given the survival function we can either differentiate to obtain the density function, and then calculate the hazard rate using Equation (4). Given the hazard function, we can always integrate to obtain the cumulative hazard and then proceed to obtain the survival function using Equation (5).

**Deriving the Functions *f(t)*,*s(t)* and *h(t)* for the Exponential Function**

Let $F(t)$ denote the cumulative distribution function (cdf). $F(t)$ is the cumulative probability of an event up to time *t*.

Let $s(t) = 1 - F(t)$, $s(t)$ is the survival probability.

If we take the first derivative of a cumulative distribution function, we get the probability density function (pdf) $f(t)$. That is,

$$F'(t) = \frac{\mathrm{d}F}{\mathrm{d}t} = f(t). \tag{9}$$

$$f(t) = \frac{\mathrm{d}F}{\mathrm{d}t} = \lambda \mathrm{e}^{-\lambda t}.$$

$$s(t) = \mathrm{e}^{-\lambda t}.$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda \mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t}} = \lambda.$$

When we want to predict the chances of failure at age *t* for a newly born or produced unit having $F(t)$ as its cdf we have to use $f(t)$, *i.e.*, $f(t)$, is an unconditional predictor for risk to fail at *t*. When we know that a unit has survived up to *t*, we have to use $h(t)$ which is a conditional predictor. Comparing numerically $f(t)$ to $h(t)$ we notice that:

- $f(0) = h(0)$; and
- $f(t) \geq h(0) \quad \forall t > 0$; $S(t) \leq 1 \quad \forall t > 0$.

There is a fundamental difference between the hazard rate function $h(t)$ and the conditional failure density $f(y \mid T > t)$.

1) $h(t)$ is a function of *t*, the age reached, whereas $f(y \mid T > t)$ is a function of the future lifetime *Y* following a given age *t*.

2) Both, $h(t)$ and $f(y \mid T > t)$ are non-negative, but $h(t)$ is not a density function as it is not normalized, instead we have $\int_0^\infty h(t)\mathrm{d}t = \infty$.

## 2.3. Properties of the Hazard Rate Function

**Theorem 1**: Any function $h(t)$ is a hazard rate if and only if it satisfies the following properties:

1) For all *t* positive, $h(t) \geq 0$.

2) It has no upper bound, $h(t)$ can be greater than 1 and can go up to $\infty$, that is, $\int_0^\infty h(t)\mathrm{d}t = \infty$.

3) $h(t)$ is increasing and then decreasing or vice versa with time.

Proof

1) $f(t) \geq 0$; $s(t) > 0$, thus $h(t) = f(t)/s(t) \geq 0$.

2) $\int_0^\infty h(t)\mathrm{d}t = -\int_0^\infty \mathrm{d}[\ln s(t)] = -\ln s(t)\big|_0^\infty = \ln s(0) - \ln s(\infty) = \ln 1 - \ln 0 = \infty$.

3) $h(t) \geq 0$ and $\lim_{t\to 0} f(t) = \lim_{t\to 0} h(t)$; $f(t) \geq h(t) \quad \forall t > 0$. thus, there is at least an interval such that *h* is increasing or decreasing, or vice versa. We distinguish between:

- monotone hazard rates: which increases, when the unit is wearing out with age, or decreases, when the unit is improving with age; and

- non-monotone hazard rates: which is either bathtub-shaped as with the case of age-specific death rate or as an inverted bathtub-shape as mentioned in Rinne [15].

## 2.4. Expressing the Hazard rate Function in Terms of pdf, cdf

$$h(t) = \frac{f(t)}{\int_t^\infty f(u)\,\mathrm{d}u} = \frac{F'(t)}{1-F(t)} = -\frac{S'(t)}{S(t)} = -\frac{\mathrm{d}\ln S(t)}{\mathrm{d}t} \tag{10}$$

Integrating Equation (10), we shall get

$$\int_0^t h(u)\,\mathrm{d}u = -\int_0^t \frac{\mathrm{d}\ln s(t)}{\mathrm{d}t} = -\ln s(t)\big|_0^t = -\big[\ln s(t) - \ln s(0)\big] = -\ln s(t) \tag{11}$$

Exponentiating Equation (11), we shall obtain

$$s(t) = \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} \tag{12}$$

so

$$F(t) = 1 - \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} \tag{13}$$

Finally, differentiating Equation (12) yields $f(t)$ in terms of $h(t)$:

$$f(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t} = \frac{\mathrm{d}\left(1 - \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\}\right)}{\mathrm{d}t} = h(t)\exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} \tag{14}$$

## 2.5. Defining Distributions by Their Hazard Rate Functions

We formulate the hazard rate models from Equations (12), (13) and (14).

### 2.5.1. The Constant Hazard Model: Exponential Distribution

$$h(t) = \lambda \quad \forall t \geq 0, \quad \lambda > 0.$$

$$\begin{aligned} f(t) &= h(t)\exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} \\ &= \lambda \exp\left(-\int_0^t \lambda\,\mathrm{d}u\right) = \lambda \mathrm{e}^{-\lambda t} \end{aligned} \tag{15}$$

$$\begin{aligned} F(t) &= 1 - \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} \\ &= 1 - \exp\left\{-\int_0^t \lambda\,\mathrm{d}u\right\} = 1 - \lambda \mathrm{e}^{-\lambda t} \end{aligned} \tag{16}$$

$$S(t) = \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} = \exp\left\{-\int_0^t \lambda\,\mathrm{d}u\right\} = \mathrm{e}^{-\lambda t}. \tag{17}$$

### Implications of the Constant Hazard Model

Let us suppose that $\beta > 0$ is a constant and $\lambda = \dfrac{1}{\beta}$. A constant hazard model is usually proposed where the risk of failure does not change with time. The constant hazard model is given by $h(t) = \lambda$ for $t > 0$, The survival function associated to this is given by;

$$S(t) = \exp\left(\int_0^t -h(u)\,du\right) = \exp\left(-\int_0^t \frac{1}{\beta}\,du\right) = \exp\left(-\frac{1}{\beta}\int_0^t 1\,du\right) = \exp\left(-\frac{1}{\beta}\right).$$

The probability density function associated with it is given by

$$f(t) = -S'(t) = \frac{1}{\beta}\exp\left(-\frac{t}{\beta}\right),$$

the mean $E(T) = \beta$ and the variance $Var(T) = \beta^2$.

**Definition 1:** The survival of lifetime variable $T$ follows an exponential probability model with mean $\beta > 0$, we can write $T \approx \exp(\beta)$ when the hazard is constant with $h(t) = \frac{1}{\beta}$.

The Constant hazard model is one of the frequently used models for modeling lifetimes of components; the exponential model often fits survival models well; possible reasons may include the fact that the time between events in a Poisson process are exponentially distributed.

**Theorem 1:** If $T \approx \exp(\beta)$, then for any $T > 0$ and $s > 0$, it follows that

$$P(T > t + s \mid T > s) = P(T > s)$$

**Proof:** The probability $P(T > t + s \mid T > s)$, may be interpreted as a conditional probability, $P(A \mid B) = P(A \cap B)P(B)^{-1}$, with the identification of the events $A$ and $B$ as $A = \{T > t + s\}$ and $B = \{T > s\}$. Since $t > 0$ in order that a survival time be longer than both $t$ and $t + s$, it must exceed $t + s$. Therefore, the event $A \cap B$ may be written as $\{T > t + s\}$. Again since $P(T > r) = \exp\left(-\frac{r}{\beta}\right)$, for any positive value of $r$:

$$\begin{aligned}
P(T > t + s \mid T > s) &= \frac{P\{(T > t + s) \cap (T > s)\}}{P(T > s)} \\
&= \frac{P\{(T > t + s) \cap (T > s)\}}{P(T > s)} \\
&= \frac{\exp\left(-\dfrac{t + s}{\beta}\right)}{\exp\left(-\dfrac{s}{\beta}\right)} = \exp\left(-\frac{t}{\beta}\right) = S(t)
\end{aligned}$$

Theorem 1 says in effect that for a component with an exponentially distributed survival time, the probability that a 4-month-old component lasts 5 more weeks in operation is the same as the probability that a 7-month-old component lasts 5 more weeks in operation. This means that the components survival time do not pass through a period of old age, where there is an increased risk of failure. In this case an exponential survival time is different from a human survival time where the survival distributions shrink with age Blossfeld *et al.* and Lawless [17] [18].

**Theorem 2:** If $T \approx \exp(\beta)$, then $\frac{T}{\beta} \approx \exp(1)$.

This theorem suggests that if we multiply or divide exponential lifetimes or survival times by a constant, then the mean time to failure correspondingly multiplies or divides by the same constant.

### 2.5.2. The Linear Hazard Rate Model

$$h(t) = a + bt \quad \forall t \geq 0, a \geq 0, b > 0$$

$$
\begin{aligned}
f(t) &= h(t)\exp\left\{-\int_0^t h(u)\,du\right\} \\
&= (a+bt)\exp\left(-\int_0^t (a+bt)\,dt\right) \\
&= (a+bt)\exp\left\{-at - \frac{bt^2}{2}\right\}
\end{aligned}
\tag{18}
$$

$$F(t) = 1 - \exp\left\{-\int_0^t h(u)\,du\right\} = 1 - \exp\left\{-\int_0^t \lambda\,du\right\} = 1 - \exp\left\{-at - \frac{bt^2}{2}\right\} \tag{19}$$

$$S(t) = \exp\left\{-\int_0^t h(u)\,du\right\} = \exp\left\{-\int_0^t \lambda\,du\right\} = \exp\left\{-at - \frac{bt^2}{2}\right\} \tag{20}$$

For $a = 0$, this expression is referred to as Rayleigh distribution.

### 2.5.3. The Power Hazard Rate Model

$$h(t) = ct^{c-1} \quad \forall t \geq 0, c > 0$$

$$f(t) = h(t)\exp\left\{-\int_0^t h(u)\,du\right\} = \lambda\exp\left(-\int_0^t \lambda\,du\right) = ct^{c-1}\exp\left\{-t^c\right\} \tag{21}$$

$$F(t) = 1 - \exp\left\{-\int_0^t h(u)\,du\right\} = 1 - \exp\left\{-\int_0^t \lambda\,du\right\} = 1 - \exp\left\{-t^c\right\} \tag{22}$$

$$S(t) = \exp\left\{-\int_0^t h(u)\,du\right\} = \exp\left\{-\int_0^t \lambda\,du\right\} = \exp\left\{-t^c\right\} \tag{23}$$

This is the reduced Weibull distribution.

Let us suppose that $c > 0$ and $\beta > 0$ are constants. A power hazard model is usually proposed where imminent risk of failure is rapidly increasing with time.

The power hazard model can be rewritten as: $h(t) = \frac{c}{\beta}t^{c-1}$, for $t > 0$.

The corresponding survival function is given by;

$$
\begin{aligned}
S(t) &= \exp\left(\int_0^t -h(u)\,du\right) = \exp\left(-\int_0^t \frac{c}{\beta^c}u^{c-1}\,du\right) \\
&= \exp\left(-\frac{1}{\beta^c}\int_0^t cu^{c-1}\,du\right) = \exp\left(-\frac{t}{\beta}\right)^c
\end{aligned}
$$

Therefore, the survival function for a variable with power hazard function has a particularly simple form where the power is translated to the exponent;

$$S(t) = \exp\left(-\frac{t}{\beta}\right)^c.$$

Probability density function $f(t) = -S'(t)$, is given by;

$$f(t) = \frac{c}{\beta^c} t^{c-1} e^{-\left(\frac{t}{\beta}\right)^c}, \quad t > 0.$$

**Definition 2:** The lifetime or survival time random variable $T$ follows a Weibul distribution model with parameter $c > 0$ and $\beta > 0$, and we write $T \approx \text{Weibull}(c, \beta)$, when $T$ has a power hazard of the form; $h(t) = \frac{c}{\beta} t^{c-1}$, for $t > 0$. The parameter $\beta$ is a scale parameter.

**Theorem 3:** If $T \approx \text{Weibull}(c, \beta)$, then $\frac{T}{\beta} \approx \text{Weibull}(c, 1)$.

The Weibull model has a very simple hazard function and a simple closed form survival function, these along with its two-parameter flexibility; makes it a very useful model in many engineering contexts.

**Definition 3:** The gamma function $\Gamma(\alpha)$ is defined for all $\alpha > 0$, by the integral $\Gamma(\alpha) = \int_0^\alpha t^{\alpha-1} e^{-t} dt$.

Some properties of the gamma function are listed below for later discussion and use

a) $\Gamma(1) = 1$.

b) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

c) $\Gamma(x+1) = x\Gamma(x)$, for any real positive number $x$.

d) $\Gamma(n+1) = n!$, $n = 1, 2 \cdots$, a positive integer.

We will use the properties of gamma listed above to find the mean and variance of the Weibull.

**Theorem 4:** Using the distribution $X \approx \text{Weibull}(\alpha, 1)$, the moments about 0 are given by

$$E(X^r) = \int_0^\infty x^r \alpha x^{\alpha-1} e^{-x^\alpha} dx = \int_0^\infty u^{\frac{r}{\alpha}} e^{-u} du = \Gamma\left(1 + \frac{r}{\alpha}\right). \tag{24}$$

We achieve the above expression by substituting $u = x^\alpha$.

**Theorem 5:** If $Y = \beta X \approx \text{Wiebull}(\alpha, \beta)$, then the $r$th moment about 0 is given by;

$$E(Y^r) = E(\beta^r X^r) = \beta^r E(X^r) = \beta^r \Gamma\left(1 + \frac{r}{\alpha}\right), \tag{25}$$

The first and second moments about 0 are used to write the mean $E(Y)$ and the variance $E(Y^2) - E(Y)^2$ of $Y$.

**Theorem 6:** If $Y \approx \text{Weibull}(\alpha, \beta)$, then $\beta\Gamma\left(1 + \frac{r}{\alpha}\right)$, and variance of $Y$ is given by:

$$Var(Y) = \beta^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma\left(1 + \frac{1}{\alpha}\right)^2\right]. \tag{26}$$

We note from the Theorem 6, the following, when $\alpha = 1$, the Weibull model reduces to an exponential model with constant hazards, that is Weibull$(1, \beta) \equiv \exp(\beta)$. When $\alpha > 1$, the Weibull model is increasing, when $\alpha < 1$, the Weibull hazard function is decreasing. These results make the Weibul model very flexible in a variety of situations; that is, increasing hazards, decreasing hazards and constant hazards as in Geskus [19].

### 2.5.4. The Exponential Hazard Rate Model

$$h(t) = \mathrm{e}^t, \quad t \geq 0.$$

$$f(t) = h(t) \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} = \lambda \exp\left(-\int_0^t \lambda\,\mathrm{d}u\right) = \mathrm{e}^t \exp\left\{-\mathrm{e}^t + 1\right\}. \quad (27)$$

$$F(t) = 1 - \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} = 1 - \exp\left\{-\int_0^t \lambda\,\mathrm{d}u\right\} = 1 - \exp\left\{-\mathrm{e}^t\right\}. \quad (28)$$

$$S(t) = \exp\left\{-\int_0^t h(u)\,\mathrm{d}u\right\} = \exp\left\{-\int_0^t \lambda\,\mathrm{d}u\right\} = \exp\left\{-\mathrm{e}^t\right\}. \quad (29)$$

This is recognized as a Gompertz distribution.

## 3. Materials and Methods

A mixture of various methods and materials were employed in this study: A framework of concepts were systematically reviewed and put together; computational discourses of data meant to illustrate the exponential model and the Weibull models were done using the R Software, along with the use of the software, a manual computation of the hazard rate was also done. A tertiary data was obtained from the Worcestor Heart attack study as contained in Hosmer and Lemeshow [20].

### 3.1. Estimates of the Hazard Using Nelson-Aalen Estimate

An alternative to the Kaplan-Meier curve is the Nelson-Aalen estimator, which is based on using a counting process approach to estimate the cumulative hazard function $H(t)$. The estimate of $H(t)$ can then be used to estimate $S(t)$. Estimates of $S(t)$ derived using this method will always be greater than the K-M estimate, but the difference will be small between the two methods in large samples.

$$\tilde{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i},$$

$$\hat{V}ar\left[\tilde{H}(t)\right] = \sum_{t_i \leq t} \frac{d_i}{n_i^2}$$

The Nelson-Aalen estimate is the first order Taylor approximation of $\log(1-x)$ about $x = 0$, where $x = d_i/n_i$.

### 3.2. Intuitive Explanations

With reference to Table 1, for an entity who died (obtained the event of interest) at the 8th month they should have to be alive at the 7th month. Therefore, the

**Table 1.** Estimation of Survival probabilities $\hat{S}(t)$ and the Nelson-Aalen Hazard Rates $\tilde{H}(t)$ using a Hypothetical data showing the time in months that it took an entity to fail (obtain the event of interest).

| Time (months) $t_i$ | Frequency $f_i$ | Number of Failures $d_i$ | Number Censored $C_i$ | Risk Set $n_i$ | Survival Probabilities $\hat{S}(t)$ | Nelson-Aalen Hazard Rates $\tilde{H}(t)$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 50 | 1.00 | 0.00 |
| 1 | 8 | 6 | 2 | 50 | 0.88 | 0.12 |
| 2 | 6 | 3 | 3 | 42 | 0.82 | 0.19 |
| 3 | 5 | 3 | 2 | 36 | 0.75 | 0.27 |
| 4 | 4 | 3 | 1 | 31 | 0.68 | 0.37 |
| 5 | 4 | 3 | 1 | 27 | 0.60 | 0.48 |
| 6 | 3 | 2 | 1 | 23 | 0.55 | 0.57 |
| 7 | 3 | 2 | 1 | 20 | 0.50 | 0.67 |
| 8 | 2 | 2 | 0 | 17 | 0.44 | 0.79 |
| 9 | 2 | 2 | 0 | 15 | 0.38 | 0.92 |
| 10 | 2 | 2 | 0 | 13 | 0.32 | 1.07 |
| 11 | 2 | 1 | 1 | 11 | 0.29 | 1.17 |
| 12 | 1 | 1 | 0 | 9 | 0.26 | 1.28 |
| 13 | 1 | 1 | 0 | 8 | 0.23 | 1.40 |
| 14 | 2 | 2 | 0 | 7 | 0.16 | 1.69 |
| 15 | 2 | 2 | 0 | 5 | 0.10 | 2.09 |
| 19 | 1 | 0* | 1 | 3 | − | − |
| 22 | 1 | 1 | 0 | 2 | 0.05 | 2.57 |
| 24 | 1 | 0* | 1 | 1 | − | − |

hazard at the 8th month (0.79) was the failure rate per that month, conditioned on the fact that the entity survived to the 7th month. As the months progressed, the hazard for an entity who obtained the event of interest also increased. There is a sharp contrast between the survival probability and the Hazard rate; whereas the survival probability focused on the probability that an entity was not failing (not obtaining the event of interest), the hazard rate focuses on failing, that is, it focuses on the fact that the event of interest occurs. Thus, in some sense the hazard rate function could be considered as giving the opposite side of the information giving by the survival probabilities. For instance, the probability that an entity will survive for 12 months or more was 0.26, whereas the hazard rate or the risk that the entity will fail (die) at the 12th month given that they survived to the 11th month was 1.28.

**NB: Table 1 was arrived at using the formula below:** $\tilde{H} = \sum_{t_i \leq t} \dfrac{d_i}{n_i}$

Month 0;  $\bar{H}(t) = 0 = 0.00$
Month 1;  $\bar{H}(t) = 6/50 = 0.12$
Month 2;  $\bar{H}(t) = 0.12 + 3/42 = 0.19$
Month 3;  $\bar{H}(t) = 0.19 + 3/36 = 0.27$
Month 4;  $\bar{H}(t) = 0.27 + 3/31 = 0.37$
Month 5;  $\bar{H}(t) = 0.37 + 3/27 = 0.48$
Month 6;  $\bar{H}(t) = 0.48 + 2/23 = 0.57$
Month 7;  $\bar{H}(t) = 0.57 + 2/20 = 0.67$
Month 8;  $\bar{H}(t) = 0.67 + 2/17 = 0.79$
Month 9;  $\bar{H}(t) = 0.79 + 2/15 = 0.92$
Month 10;  $\bar{H}(t) = 0.92 + 2/13 = 1.07$
Month 11;  $\bar{H}(t) = 0.92 + 2/13 = 1.07$
Month 12;  $\bar{H}(t) = 1.17 + 1/9 = 1.28$
Month 13;  $\bar{H}(t) = 1.28 + 1/8 = 1.40$
Month 14;  $\bar{H}(t) = 1.40 + 2/7 = 1.69$
Month 15;  $\bar{H}(t) = 1.69 + 2/5 = 2.09$
Month 22;  $\bar{H}(t) = 2.09 + 1/2 = 2.59$

We note that months 19 and 24 are left out because in these two months no person failed.

### 3.3. Intuitive Interpretation of *f(t)*, *F(t)*, *S(t)*, *h(t)* and *H(t)*

$f(t)$: This is the unconditional predictor for risk to occur at time $t$. Beginning from the point of commencement. This gives us the probability distribution of time an entity is at risk of dying.

$F(t)$  This is the cumulative distribution. This function gives us the percentage of the entities that will be dead at time *T*.

$S(t) = I - F(t)$: This is the survival function; it gives us the percentage of the species that will be alive at time *T*.

$h(t)$: This is the hazard function. The function measures the conditional probability of a failure, given that the entity has worked past a time point $t$.

$H(t)$: This is the cumulative hazard function. It is the sum of all the hazard values for failed units with ranks up to and including that failed unit. it measures the total amount of risk that has been accumulated up to time *t* as explained in NIST/SEMATECH and Allison [21] [22].

### 3.4. Calculation of Cumulative Hazards

The steps needed to arrive at the cumulative hazards is shown clearly in Table 2. Using data from Table 2, a plot of the cumulative hazard for the exponential distribution $H(t) = \alpha t$ was done by plotting figures in column seven on the ordinate axis against figures in column two on the abscissa axis, a linear plot was obtained with the gradient equal to the value of $\alpha$. If the data fits the model, the plot would be a straight line passing through the origin with slope equal to. The plot for the cumulative hazard function for the exponential model is shown

Table 2. Display of the calculation of the cumulative hazard rates of thirteen light bulbs tested for 380 hours, within the period, seven failures were observed at times 40, 57, 120, 135, 150, 178, 290 and 380 hours. Five bulbs could not survive the testing process; they were therefore censored at various times as follows: 55, 60, and 70, 240 and 350 hours.

| Items | Survival time $T$ | Censoring Status (3) | Increasing order of Ranking | Decreasing order of Ranking (5) | Hazard (3)/(5) | Cum. Hazard |
|---|---|---|---|---|---|---|
| 1 | 40 | 1 | 1 | 13 | 1/13 = 0.08 | 0.08 |
| 2 | 55 | 0 | 2 | 12 | 0 | 0.00 |
| 3 | 57 | 1 | 3 | 11 | 1/11 = 0.09 | 0.17 |
| 4 | 60 | 0 | 4 | 10 | 0 | 0.00 |
| 5 | 70 | 0 | 5 | 9 | 0 | 0.00 |
| 6 | 120 | 1 | 6 | 8 | 1/8 = 0.13 | 0.30 |
| 7 | 135 | 1 | 7 | 7 | 1/7 = 0.14 | 0.44 |
| 8 | 150 | 1 | 8 | 6 | 1/6 = 0.17 | 0.61 |
| 9 | 178 | 1 | 9 | 5 | 1/5 = 0.20 | 0.81 |
| 10 | 240 | 0 | 10 | 4 | 0 | 0.00 |
| 11 | 290 | 1 | 11 | 3 | 1/3 = 0.33 | 1.14 |
| 12 | 350 | 0 | 12 | 2 | 0 | 0.00 |
| 13 | 380 | 1 | 13 | 1 | 1/1 = 1.00 | 2.14 |

in Figure 2. We can see from the plot that the plot passed through the origin. From Table 2, a plot of the cumulative hazard for the Weibull distribution $H(t) = \left(\dfrac{t}{\alpha}\right)^{\gamma}$ was also done by plotting figures in column seven on the ordinate axis against figures in column two on the abscissa axis. This plot was done on a log-log scale. If the data fits the Weibull model, then a linear plot will be obtained with the gradient equal to the value of $\gamma$. The plot for the cumulative hazard function for the Weibull model is shown in Figure 3. The intercept on the Y-axis was −3.314 while that on the time-axis (x-axis) was 1.397.

## 3.5. Simulation Design

Simulation studies are essential for appreciating and appraising statistical models. In simulating survival times, we often assume an exponential or Weibull distribution for the baseline hazard function. Based on this, a simulation-based comparison of the exponential model and the Weibull model was done. The graphs of the simulation results are plotted as in Figure 2 and Figure 3.

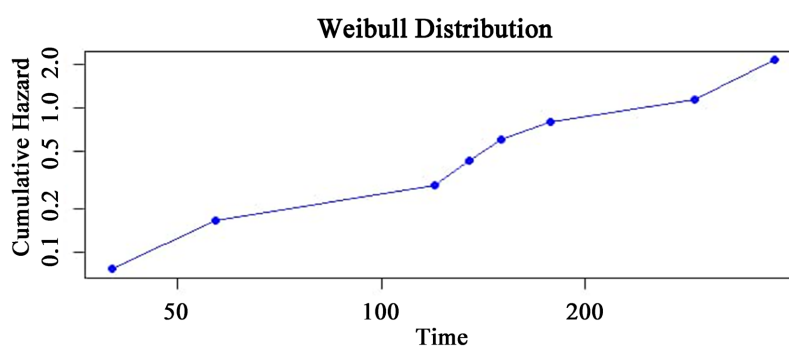## 3.6. Applications of the Hazard Rate Function to Real Data

The importance of the hazard rate function is demonstrated. In this demonstration, hundred (100) subjects of Worcestor heart attack study from [18] was used.
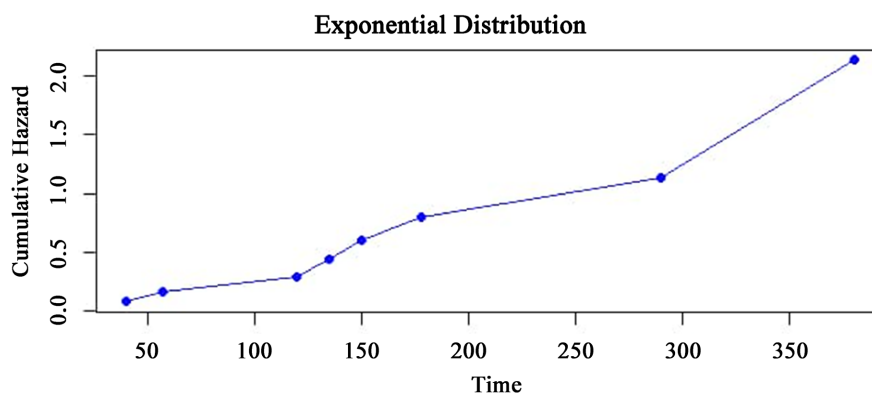
**Figure 2.** Plot of the cumulative hazard rates against Survival time for the Exponential model.



**Figure 3.** Plot of the cumulative hazard rates against Survival time for the Weibull model.

The data examined several factors including age at hospitalization, gender, and Body Mass Index that may influence survival times after heart attack. The follow-up time for all subjects begun after hospital admission and ends with death or loss to follow-up. The data was keyed into IBM SPSS version 21 and the survival analysis function, specifically, the Kaplan-Meier Operator was used to obtain both the hazard function rates and the plots, the results of the first forty (40) subjects were shown in this paper in Table 5.

## 4. Results and Discussion

Table 3 gives us the shape and application of the various types of hazards applied in survival analysis. Table 4 gives us the percentage contribution of each of the variables used to do the analysis. With the age variable, a greater percentage of them were in the 81 - 90 years bracket, with the gender, majority of the subjects were males, with respect to the BMI, a greater percentage of the subjects were within the healthy weight bracket, finally, about the same percentage obtained the event of interest. A cursory look at Table 5 reveals the following: As the hazard rate decreases (0.061), (0.051) survival probabilities increase (0.941), (0.950) respectively; as the hazard rate increases (1.098), (1.609), survival probabilities decrease (0.333), (0.200) respectively. We note again that, the risk of death was distributed among all four BMI groupings, this goes to tell us that the

**Table 3.** Various Shapes of Hazard Functions and their Applications Aalen and Gjessing [14] [16].

| S/N | Shapes | Applications | Aging Factor |
|---|---|---|---|
| 1. | **Constant (Random) Failure Rate** | • Applied to entities who have the propensity to fail randomly or by chance<br>• unpredictable failures<br>• Third-party damage<br>• Material defects | • Random |
| 2. | **Increasing failure Rate** | • In most situations this is the hazard we experience<br>• Machinery, organism,<br>• Consumer products<br>• Corrosion<br>• Fatigue | • Time dependent |
| 3. | **Decreasing failure Rate** | • Not very common<br>• New typist employed in a company | • Time dependent |
| 4. | **Bathtub-shape failure rate** | • Patients who undergo surgery<br>• infant mortality | • Time dependent |

**Table 4.** Demographic Display of Variables and their Percentages in the Hazard Rate Function.

| Variable | Frequency | Percentage |
|---|---|---|
| **Age** | | |
| 31 - 40 | 4 | 4.0 |
| 41 - 50 | 11 | 11.0 |
| 51 - 60 | 13 | 13.0 |
| 61 - 70 | 23 | 23.0 |
| 71 - 80 | 24 | 24.0 |
| 81 - 90 | 25 | 25.0 |
| Total | 100 | 100.0 |
| **Gender** | | |
| Males | 65 | 65.0 |
| Females | 35 | 35.0 |
| Total | 100 | 100.0 |
| **Body Mass Index** | | |
| Underweight | 8 | 8.0 |
| Healthy Weight | 33 | 33.0 |
| Overweight | 31 | 31.0 |
| Obese | 28 | 28.0 |
| Total | 100 | 100.0 |
| **Censoring Status** | | |
| Censored | 49 | 49.0 |
| Uncensored | 51 | 51.0 |
| Total | 100 | 100.0 |

**Table 5.** Display of the first forty (40) subjects from the Worcestor Heart attack study obtained from [18] showcasing the survival function and corresponding Hazard rate functions.

| S/N | Survival Time(days) | Status 1 = Death 0 = Censored | Age(years) | Gender F = Female M = Male | Body Mass Index | Survival Function $S(t)$ | Hazard Function $h(t)$ |
|---|---|---|---|---|---|---|---|
| 1 | 4 | 1 | 61 - 70 | M | Obese | 0.850 | 0.163 |
| 2 | 4 | 1 | 71 - 80 | M | Overweight | 0.816 | 0.204 |
| 3 | 11 | 1 | 61 - 70 | F | Underweight | 0.000 | |
| 4 | 3 | 1 | 71 - 80 | M | Underweight | 0.667 | 0.405 |
| 5 | 8 | 1 | 81 - 90 | F | Underweight | 0.400 | 0.916 |
| 6 | 3 | 1 | 81 - 90 | F | Underweight | 0.800 | 0.223 |
| 7 | 9 | 1 | 81 - 90 | F | Healthy weight | 0.381 | 0.965 |
| 8 | 4 | 1 | 81 - 90 | F | Obese | 0.500 | 0.693 |
| 9 | 6 | 1 | 81 - 90 | M | Healthy weight | 0.542 | 0.613 |
| 10 | 7 | 1 | 71 - 80 | M | Obese | 0.468 | 0.760 |
| 11 | 6 | 1 | 81 - 90 | M | Underweight | 0.333 | 10.098 |
| 12 | 7 | 1 | 51 - 60 | F | Underweight | 0.600 | 0.511 |
| 13 | 6 | 1 | 81 - 90 | F | Healthy weight | 0.635 | 0.454 |
| 14 | 5 | 1 | 51 - 60 | M | Healthy weight | 0.650 | 0.431 |
| 15 | 10 | 1 | 71 - 80 | F | Underweight | 0.200 | 1.609 |
| 16 | 5 | 1 | 71 - 80 | M | Overweight | 0.742 | 0.299 |
| 17 | 7 | 1 | 71 - 80 | M | Overweight | 0.494 | 0.705 |
| 18 | 5 | 1 | 81 - 90 | M | Healthy weight | 0.650 | 0.431 |
| 19 | 6 | 1 | 61 - 70 | F | Overweight | 0.589 | 0.529 |
| 20 | 5 | 1 | 81 - 90 | F | Overweight | 0.707 | 0.347 |
| 21 | 5 | 1 | 81 - 90 | F | Healthy weight | 0.741 | 0.300 |
| 22 | 4 | 1 | 81 - 90 | M | Healthy weight | 0.950 | 0.051 |
| 23 | 8 | 1 | 81 - 90 | F | Healthy weight | 0.508 | 0.677 |
| 24 | 11 | 1 | 81 - 90 | M | Healthy weight | 0.385 | 0.954 |
| 25 | 5 | 1 | 51 - 60 | M | Healthy weight | 0.650 | 0.431 |
| 26 | 4 | 1 | 51 - 60 | M | Overweight | 0.816 | 0.204 |
| 27 | 4 | 1 | 81 - 90 | F | Healthy weight | 0.833 | 0.182 |
| 28 | 3 | 1 | 81 - 90 | M | Overweight | 0.941 | 0.061 |
| 29 | 3 | 1 | 61 - 70 | F | Obese | 0.667 | 0.405 |
| 30 | 5 | 1 | 51 - 60 | F | Overweight | 0.707 | 0.347 |
| 31 | 5 | 1 | 41 - 50 | M | Obese | 0.789 | 0.237 |

Continued

| 32 | 7 | 1 | 71 - 80 | M | Obese | 0.468 | 0.760 |
| 33 | 4 | 1 | 41 - 50 | F | Overweight | 0.909 | 0.095 |
| 34 | 4 | 1 | 71 - 80 | M | Obese | 0.850 | 0.162 |
| 35 | 5 | 1 | 51 - 60 | M | Healthy weight | 0.650 | 0.431 |
| 36 | 5 | 1 | 41 - 50 | M | Healthy weight | 0.650 | 0.431 |
| 37 | 6 | 1 | 81 - 90 | M | Healthy weight | 0.542 | 0.613 |
| 38 | 7 | 1 | 71 - 80 | M | Healthy weight | 0.481 | 0.731 |
| 39 | 3 | 1 | 31 - 40 | F | Obese | 0.667 | 0.405 |
| 40 | 4 | 1 | 71 - 80 | M | Obese | 0.850 | 0.162 |
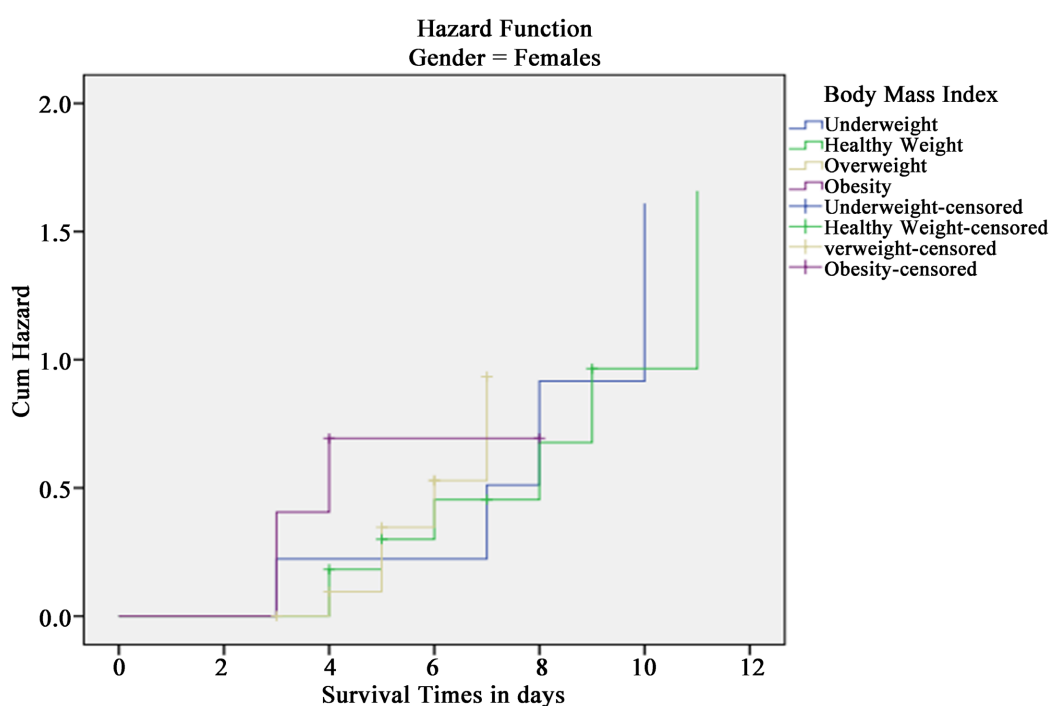


**Figure 4.** Display of the Kaplan-Meier Hazard rate curve for females who participated in the Worcestor Heart attack study.

risk of death was dependent mostly on the identified disease, that is, heart attack, the effect of the BMI was not readily seen. Gender and age appear not to contribute significantly towards death due to heart attack. In **Figure 4**, we see that the hazard rate for the first few days for all the four categories of the BMI were about constant, on the 3rd and 4th days there was a significant increase in the hazard rates especially for the female obese category. For the male category (**Figure 5**), we note that there were stepwise increases in the Hazards of three of the BMI categories; underweight, obese and healthy weights. With the overweight category, there was a stepwise increase in the first few days after which the hazard rate became constant.
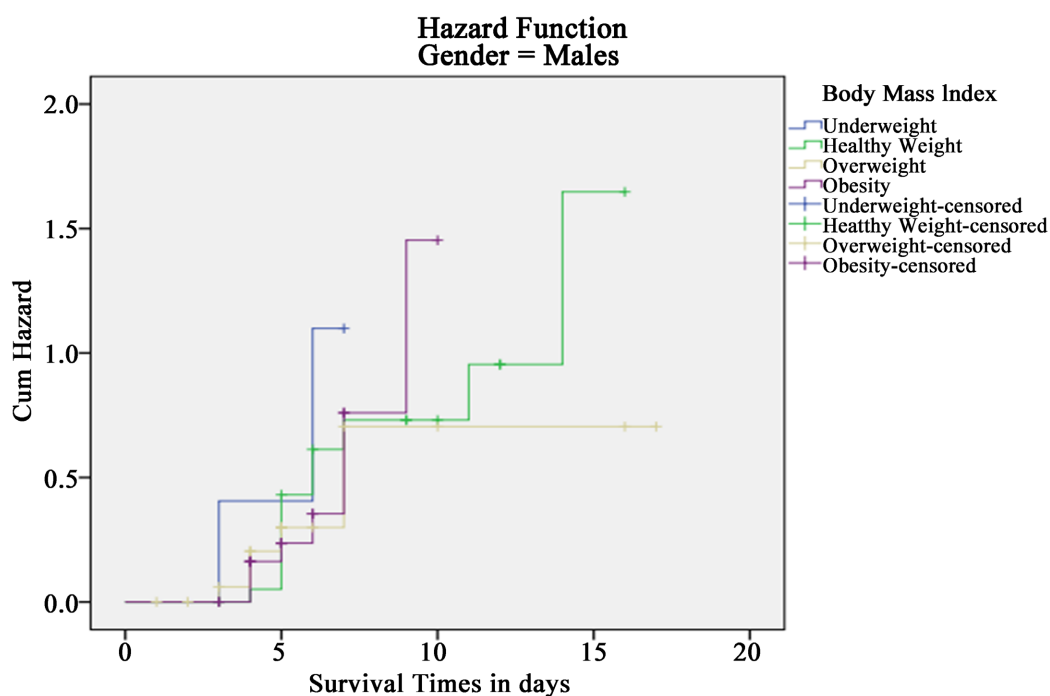
**Figure 5.** Display of the Kaplan-Meier Hazard rate curve for males who participated in the Worcestor Heart attack study.

## 5. Conclusion

In this study, it was conjectured that, in survival analysis, researchers studying the timing and occurrence of events, often analyze the probability distributions of the time preceding the occurrence of the event. They focus mostly on the end result of the process, rather than the processes that generated the end results. Research has it that, the hazard function plays a key role in characterizing the process of aging. It is in trying to investigate the underlying processes that lead to the end result, and an attempt to understand intuitively the concept of hazard rate functions which this study was conducted. The study was undertaken by conceptualizing the hazard function, providing a working definition of hazard rates and for the matter hazard functions, exploring the properties of the function with proofs, formulating the hazard function; appreciating the hazard models; computing and intuitively interpreting some principles governing the function, conducting simulation studies using the R software; and finally, application of the function to real-life data. It is concluded that this study has intuitively demonstrated, theorized, discussed and explored the relevance of the hazard model in assessing risk.

## Conflicts of Interest

The author declares no conflicts of interest.

## References

[1]  Wu, L.L. (1989) Issues in Smoothing Empirical Hazard Rates. *Sociological Metho-*

*dology*, **19**, 127-159. https://doi.org/10.2307/270950

[2] UCLA Statistical Consulting Group (2021) Survival Analysis with STATA. https://stats.idre.ucla.edu/stata/seminars/stata-survival

[3] Clark, T.G., Bradburn, M.J., Love, S.B. and Altman, D.G. (2003) Survival Analysis Part I: Basic Concepts and First Analyses. *British Journal of Cancer*, **89**, 232-238. https://doi.org/10.1038/sj.bjc.6601118

[4] Park, S.Y., Park, J.E., Kim, H. and Park, S.H. (2021) Review of Statistical Methods for Evaluating the Performance of Survival or Other Time-to-Event Prediction Models (from Conventional to Deep Learning Approaches). *Korean Journal of Radiology*, **22**, 1697-1707. https://doi.org/10.3348/kjr.2021.0223

[5] Eiser, J.R., Bostrom, A., Burton, I., Johnston, D.M., John McClure, J., Paton, D., Pligt, J.V. and White, M.P. (2012) Risk Interpretation and Action: A Conceptual Framework for Responses to Natural Hazards. *International Journal of Disaster Risk Reduction*, **1**, 5-16. https://doi.org/10.1016/j.ijdrr.2012.05.002

[6] Kurniasari, D., Widyarini, R., Warsono and Antonio, Y. (2019) Characteristics of Hazard Rate Functions of Log Normal Distributions. *Journal of Physics: Conference Series*, **1338**, Article ID: 012036. https://doi.org/10.1088/1742-6596/1338/1/012036

[7] Zhang, J. and Peng, Y. (2009) Crossing Hazard Functions in Common Survival Models. *Statistics & Probability Letters*, **79**, 2124-2130. https://doi.org/10.1016/j.spl.2009.07.002

[8] Blackstone, E.H. (1996) Outcome Analysis Using Hazard Function Methodology. *The Annals of Thoracic Surgery*, **61**, S2-S7. https://doi.org/10.1016/0003-4975(95)01075-0

[9] Read, L.K. and Vogel, R.M. (2016) Hazard Function Analysis for Flood Planning under Nonstationarity. *Water Resources Research*, **52**, 4116-4131. https://doi.org/10.1002/2015WR018370

[10] Upadhyay, S.K. (2010) Hazard Rate Function. In: *Wiley Encyclopedia of Operations Research and Management Science*, John Wiley & Sons, Inc., Hoboken. https://doi.org/10.1002/9780470400531.eorms0371

[11] Boland, P., El-Neweihi, E. and Ptoschan, F. (1994) Applications of the Hazard Rate Ordering in Reliability and Order Statistics. *Journal of Applied Probability*, **31**, 180-192. https://doi.org/10.2307/3215245

[12] Greenwich, M.A. (1992) Unimodal Hazard Rate Function and Its Failure Distribution. *Statistical Papers*, **33**, 187-202. https://doi.org/10.1007/BF02925324

[13] Hinchliffe, S.R. and Lambert, P.C. (2013) Flexible Parametric Modelling of Cause-Specific Hazards to estimate Cumulative Incidence Functions. *BMC Medical Research Methodology*, **13**, 13. https://doi.org/10.1186/1471-2288-13-13

[14] Glen, S. (n.d.) Hazard Function: Simple Definition from StatisticsHowTo.com: Elementary Statistics for the Rest of Us. https://www.statisticshowto.com/hazard-function

[15] Rinne, H. (2014) The Hazard Rate—Theory and Inference.

[16] Aalen, O.O. and Gjessing, H.K. (2001) Understanding the Shape of the Hazard Rate: A Process Point of View. *Statistical Science*, **16**, 1-14. https://doi.org/10.1214/ss/998929473 http://www.jstor.org/stable/2676773

[17] Blossfeld, H.-P., Hamerle, A. and Mayer, K.U. (1989) Event History Analys*is* Statistical Theory and Application in the Social Sciences. Erlbaum, Hillsdale.

[18] Lawless, J.F. (1982) Statistical Models and Methods for Lifetime Data. John Wiley

and Sons, New York.

[19] Geskus, R.B. (2010) Causes Specific Cumulative Incidence Estimation and the Fine and Gray Model under both Left Truncation and Right Censoring. *Biometrics*, **67**, 39-49. https://doi.org/10.1111/j.1541-0420.2010.01420.x

[20] Hosmer, D.W. and Lemeshow, S. (1999) Applied Survival Analysis, Regression Modeling of Time to Event Data. John Wiley and Sons, Inc., New York.

[21] NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/,10/11/21

[22] Allison, P.D. (1995) Survival Analysis Using the SAS System: A Practical Guide. SAS Institute, Inc., Cary.