Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 29 (2026), No. 3, pp. 231–254
DOI : 10.47974/JSMS-1356

Chen proportional hazards model with application to lifetime data

Kudugu Atiah *
Suleman Nasiru †
Department of Statistics and Actuarial Science
School of Mathematical Sciences
C. K. Tedam University of Technology and Applied Sciences
Navrongo
Ghana

Salifu Katara §
Department of Statistics
School of Physical Sciences
University for Development Studies
Nyankpala Campus
Tamale
Ghana

Abstract

In this study, we have introduced the Chen distribution as an underlying hazard rate function for a parametric proportional hazards model called the Chen proportional hazards model. The maximum likelihood estimation approach was considered for determining the model parameter values. The developed model's use was demonstrated with lifetime data. Results of the model selection criteria and the graphs of Cox-Snell residuals from the study disclosed that the newly introduced model demonstrates superiority over the Weibull and Gompertz proportional hazards models and might be useful when examining different kinds of survival statistics.

Subject Classification (2020): *62N02.*

Keywords: *Proportional hazards model, Chen distributions, Estimation, Hazard rate function, Survival function.*

* *E-mail:* **katiah.stu@cktutas.edu.gh; (Corresponding Author)**
kudugua@gmail.com
† *E-mail:* **sulemanstat@gmail.com;**
snasiru@cktutas.edu.gh
§ *E-mail:* **skatara@uds.edu.gh;**
katadeous@yahoo.com

1. Introduction

Several disciplines, including medical research, biology, environmental studies, economics, epidemiological studies, social sciences, and engineering, depend on statistical evaluation of survival information (Abdisalam *et al*., [1]). The basis of parametric survival models is probability distributions. If the distributional assumptions are attained, parametric survival models produce estimates that are more accurate and tend to yield more precise estimates, with standard errors smaller than those from semi-parametric as well as non-parametric models (Collett, [10]).

Cox proportional hazards model is chosen automatically when modeling survival data in medical research (Royston, [25]). Unfortunately, for a single data set, this model does not offer a comprehensive probability specification. This calls for the development of adaptable models that can describe data of varied shapes while taking into account these distributional assumptions. Nevertheless, due to the Cox model's undefined baseline hazard function and insufficient observations from the lifespan data analysis, estimating the model's parameters becomes computationally difficult (Gonzalez *et al*., [13]).

Parametric survival models are essential for evaluating survival statistics. The following are some advantages of utilizing parametric survival models: They can apply random effects and frailty structures (Legrand, [22]) and can also be used to estimate expected lives, not only hazard ratios like the accelerated failure time framework (Lawless, [21]), (Muse, [24]), and (Abdisalam *et al*., [2]). They can handle all types of censored data and produce better estimates when the baseline hazard function can be determined mathematically (Abdisalam *et al*., [2]).

Wang *et al*. [29] and Abonongo *et al*. [3] indicated that the two predominantly adopted approaches used in formulating parametric regression structures for survival outcomes are the accelerated failure time (AFT) model, where factors impact both the hazard rate considered alongside the time dimension, as well as the PH analytical framework, whereby covariates influence the failure rate Adam *et al.* [4].

In order to model survival data using a parametric method, one of the first tasks is to select a baseline distribution that can represent the important characteristics of the observed data. In the modeling of survival data, specific probability distributions are frequently utilized. The proportional hazards model permits only a limited set of fixed models, none of which are sufficiently flexible to effectively represent various

survival data. The majority of distributions adhering to the PH assumption cannot simulate survival statistics that are non-monotone in nature (Khan and Khosa, [19] and Abdisalam *et al*., [2]). However, Tarvirdizade and Nematollahi [28] addressed this in their recent work titled, The Power-Exponential Failure Rate Distribution, which is capable of modeling datasets with "bathtub-shaped" hazard rates.

The Chen distribution can simulate lifespan data with rising and bathtub hazard rate shapes and has numerous uses (Chen, [8]). It is closed under the PH assumption. Over the past few decades, significant efforts have been made to extend classical distributions such that they can serve as the regression models based on parametric hazards and their baseline distributions.

The Cox PH model's shortcomings are resolved by the parametric proportional hazards method, which provides more accurate estimates. Chen's distribution may also incorporate lifespan data with rising and bathtub hazard rate forms. Due to this, the new model stands out and is more appealing to scholars from many academic fields.

The structure of the paper appears as outlined: The introduction is found in Section 1. The Chen distribution is covered in Section 2, while the formulation, along with the underlying assumptions regarding the PH model, is covered in Section 3. The distributions closed under the PH assumption are outlined in Section 4. Section 5 provides the new model's formulation, while Sections 6 and 7 examine the method of estimation as well as "model selection" criteria, respectively. Section 8 demonstrates how to use this new model on an actual data set, and Section 9 provides some concluding remarks.

2. Chen Distribution

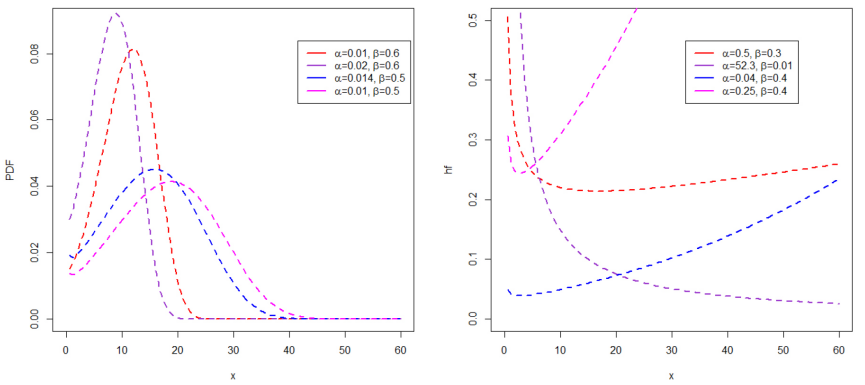
The Chen distribution's cumulative distribution together with its probability density functions (PDF) can therefore be obtained in the following form:

$$F(x) = 1 - e^{m(1-e^{x^z})} \tag{1}$$

and

$$f(x) = mzx^{z-1}e^{x^z}e^{m(1-e^{x^z})}, \tag{2}$$

respectively, where $x > 0, m > 0, z > 0$ (Chen, [8]). The parameters' m and z are both the distribution's shape parameters.



(a) **(b)**

Figure 1

Probability density (a) and hazard rate (b) functionis of the Chen distribution

The hazard rate and survival functions are defined as

$$h(x) = mzx^{z-1}\mathrm{e}^{x^z}, \tag{3}$$

and

$$S(x) = e^{m(1-e^{x^z})}. \tag{4}$$

Figure 1 shows the density together with its hazard rate graphs for various parameter values. The graph indicates that the PDF is right-tailed. The hazard function that describes the Chen distribution exhibits distinct shapes considering different parameter arrangements, such as increasing, decreasing, and bathtub shapes.

3. Proportional Hazards Model Formulation and Assumptions

Explanatory variable(s) influence survival times in numerous real-world situations. A regression model links the explanatory variable(s) with the response variable. Explanatory variable(s) must be included in survival modeling since they are crucial. Adam *et al.* [4] indicated that many approaches are available for use to establish regression models based on hazard functions. The formulation of a proportional hazards (PH) framework is one of the most widely utilized techniques (Abdisalam *et al.*, [1]). In health research, the Cox proportional hazards model is widely preferred, as it does not rely on specific assumptions about the distribution of survival times (Abdisalam et al., [1]). According to Collett [10], Cox PH

and the PH models are presented in a similar format. It is the Cox model in its parametric version.

3.1 *Formulation of the Proportional Hazards Approach*

The parametric PH structure is generated by employing a specified underlying hazard function (Abdisalam *et al*., [1]). In mathematical terms, it can be represented as follows:

$$\xi(t;\boldsymbol{\nu}) = \xi_0(t)\exp(\boldsymbol{V}'\boldsymbol{\beta}), \tag{5}$$

"in which $\xi_0(t;\boldsymbol{\nu})$ stands for the baseline hazard function, $\boldsymbol{V}$ appears in row-vector form containing the covariates, with $\boldsymbol{\beta}$ appearing in column-vector form consisting of regression parameters" (Abdisalam *et al*., [1]).

3.2 *Proportional Hazards Assumptions*

According to Abdisalam *et al*. [1], the PH framework makes the assumption that the impact associated with the covariates will cause the hazard function to either increase or decrease proportionately, regardless of time. The proportional hazard–based model makes the following assumptions:

$$\begin{aligned}\xi(t;\boldsymbol{\nu}) &= \xi_0(t)\exp(\boldsymbol{V}'\boldsymbol{\beta}) \\ &= \xi_0(t).e^{v'\beta},\end{aligned} \tag{6}$$

where this function $\xi_0(t);$ serves as the baseline hazard. Simplifying, equation (6) can be written as:

$$\xi(t\,|\,\boldsymbol{\nu}) = \xi_0(t)\exp(\beta_1 v_1 + \beta_2 v_2 + \ldots + \beta_z v_z). \tag{7}$$

Comparing any two subjects or observations with covariates $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, for instance, the hazard ratio (HR) of these two subjects or observations is defined as

$$\begin{aligned}HR(\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\xi_0,\boldsymbol{\beta}) &= \frac{\xi(t\,|\,\boldsymbol{\nu}_1,\boldsymbol{\beta})}{\xi(t\,|\,\boldsymbol{\nu}_2,\boldsymbol{\beta})} = \frac{\xi_0(t\,|\,\boldsymbol{\nu})\exp(\boldsymbol{V}_1^{'}\boldsymbol{\beta})}{\xi_0(t\,|\,\boldsymbol{\nu})\exp(\boldsymbol{V}_2^{'}\boldsymbol{\beta})} \\ &= \exp(\boldsymbol{\beta}(\boldsymbol{V}_1^{'} - \boldsymbol{V}_2^{'})).\end{aligned} \tag{8}$$

According to the aforementioned expression, the hazard rate for an observation is directly related to the hazard rate for every other observation.

3.3 *Lifetime-Related Functions for the Proportional Hazards Framework*

The following are frequently used life span distribution functions associated with the PH model.

1. Hazard Rates in the Proportional Hazards Model

The failure rate function is a fundamental lifetime distribution used within the PH modeling structure. It refers to the rate at which survivors suffer death over time t (David, [11]). Consequently, the failure-rate function for the proportional hazards structure is given by the expression

$$\xi(t;\boldsymbol{\nu}) = \xi_0(t)\exp(\boldsymbol{V}'\boldsymbol{\beta}). \tag{9}$$

2. Cumulative Hazards in the PH Model

According to Abdisalam *et al*. [1], the cumulative hazard function (CHF) within the PH structure can be defined as follows

$$H(t\,|\,\boldsymbol{\nu}) = \int_0^t \xi(s)ds = \exp(\boldsymbol{V}'\boldsymbol{\beta})\int_0^t \xi_0(s)ds = H_0(t)\exp(\boldsymbol{V}'\boldsymbol{\beta}). \tag{10}$$

3. Survival Function Specification under the Proportional Hazards Model

One approach to formulate the function governing survival over time can be by utilizing the link between the survival function as well as its hazard function. Chin-Diew [9] stated that the failure rate is defined by the expression;

$$h(t) = \frac{f(t)}{S(t)}. \tag{11}$$

Therefore, the CDF can also be expressed as;

$$\begin{aligned} H(t\,|\,\boldsymbol{\nu}) = \int_0^t h(u)du &= \int_0^t \frac{f(u)}{S(u)}du, \\ &= -\int_0^t \frac{dS(u)}{S(u)}, \\ &= -\ln(S(t)). \end{aligned} \tag{12}$$

Hence, $f(t) = h(t).S(t) = h(t).\exp(-H(i))$. Expressing $S(t)$ as a subject of $f(t)$ and $h(t)$ and simplifying yields:

$$S(t) = \exp(-H(t)) = \exp\left(-(h_0(t)) \int_0^t \exp(\text{V}'\text{¥}\ ) dt \right),$$

$$= \exp\left(-\exp(\text{V}'\text{¥}\ ). \int_0^t \text{h}_0(t)\, dt \right). \tag{13}$$

4. Cumulative Distribution Function within the PH Framework

In most cases, the CDF for the model takes the form:

$$F(t) = 1 - S(t),$$

$$= 1 - \exp(-H(t)). \tag{14}$$

5. Probability Density in the Proportional Hazards Model

Its "PDF" is defined as

$$f(t) = h(t).S(t). \tag{5}$$

4.0 Distributions Which are Closed under the Framework of the PH Model

In the Proportional Hazards model framework, closed distributions are utilized to examine survival statistics. These distributions have been analyzed and applied in various settings.

4.1 *Weibull PH Model*

The Weibull distribution serves as an effective method in statistical modeling to analyze fatigue and the lifespan of engineering materials and devices [14]. A generation of the exponential distribution is the Weibull distribution (Justin, [17] and Abdisalam *et al.,* [1]). It is flexible and may adopt traits of other continuous distribution types. In contrast to the exponential, it possesses an extra parameter. Based on the shape parameter's value, an extra parameter tells the characteristics of the structure of the hazard rate (Alkhairy *et al*., [7]). Considering *T* to have a Weibull distribution, its PDF, CDF, survival, hazard rate, and CHF, respectively, are expressed as follows;

$$
\begin{aligned}
f(t) &= d\theta(\theta t)^{d-1}\exp(-(\theta t)^{d}),\\
F(t) &= 1-\exp(-(\theta t)^{d}),\\
S(t) &= \exp(-(\theta t)^{d}),\\
h(t) &= d\theta(\theta t)^{d-1},\\
H(t) &= -\ln S(t) = -\ln\exp(-(\theta t)^{d}),\\
&= (\theta t)^{d},
\end{aligned}
\tag{16}
$$

in which "the shape parameter is denoted by $d>0$ and the rate parameter by $\theta>0$" (Abdisalam *et al.*, [1]). It is essential to recognize that non-monotone (also known as unimodal or bathtub) hazard rates cannot be accommodated by the Weibull distribution.

For a Proportional Hazards framework with a Weibull baseline, the rate of failure is expressed as

$$h_0(t) = d\theta(\theta t)^{d-1}. \tag{17}$$

Based on the PH structure's formulation, the hazard rate function of one person whose covariate set $\boldsymbol{V}$, combined with the link term $\exp(\boldsymbol{V}'\boldsymbol{\beta})$, forms

$$h(t) = h_0(t)\exp(\boldsymbol{V}'\boldsymbol{\beta}) = d\theta(d\theta)^{d-1}\exp(\boldsymbol{V}'\boldsymbol{\beta}). \tag{18}$$

Simplifying equation (18) yields

$$
\begin{aligned}
h_{WPH}(t) &= d\theta(\theta t)^{d-1}\exp(\boldsymbol{V}'\beta)\\
&= d\theta(\theta t)^{d-1}\exp(\beta_1 v_1+\beta_2 v_2+\ldots+\beta_z v_z).
\end{aligned}
\tag{19}
$$

Rewriting equation (19) as

$$h_{WPH}(t) = d\theta^{*}(\theta t)^{d-1}. \tag{20}$$

Considering that the distribution follows a Weibull form, the rate parameter $\theta^{*}=\theta\exp(\boldsymbol{V}'\beta)$ and shape parameter d in equation (20) shows that the Weibull distribution fulfils the PH assumption.

Abdisalam *et al.* [1] stated that the Weibull-based PH structure has the following lifetime distributions: the Weibull-based PH's survival term is

$$S_{WPH}(t) = \exp(-(\theta t)^{d}.\exp(\boldsymbol{V}'\beta)). \tag{21}$$

According to Muse [24], the Weibull-based PH structure has a probability density expressed as:

$$f_{WPH}(t) = d\theta(\theta t)^{d-1} \exp(\boldsymbol{V}'\boldsymbol{\beta}) \exp(-(\theta t)^{d} \exp(\boldsymbol{V}'\boldsymbol{\beta})). \tag{22}$$

"The CDF and CHF for the Weibull PH structure may be stated as"

$$F_{WPH}(t) = 1 - \exp(-(\theta t)^{d} \exp(\boldsymbol{V}'\boldsymbol{\beta})) \tag{23}$$

and

$$H_{WPH}(t) = (\theta t)^{d} \exp(\boldsymbol{V}'\boldsymbol{\beta}) \tag{24}$$

respectively.

4.2 *Gompertz PH Model*

Benjamin Gompertz introduced the Gompertz distribution in relation to human mortality and actuarial sciences (Gompertz, [12]). This distribution, which is a modification of the exponential distribution, is used in a variety of disciplines, including reliability analysis, survival, actuarial science and demography, among others, to simulate human life span (Abdisalam *et al.,* [2]). It is a flexible distribution that can be skewed to the left or the right. Accordingly, the Gompertz distribution's PDF, CDF, survival function, failure rate function, and CHF are given by:

Suppose $X \sim \text{Gompertz}(\vartheta, q)$, then,

$$\begin{aligned} f(t) &= q\vartheta e^{t\vartheta} e^{-q(e^{t\vartheta}-1)}, t \in (0,\infty), \\ F(t) &= 1 - e^{-q(e^{t\vartheta}-1)}, \\ S(t) &= e^{-q(e^{t\vartheta}-1)}, \\ h(t) &= q\vartheta e^{t\vartheta}, \\ H(t) &= -\ln S(t) = -\ln(e^{-q(e^{t\vartheta}-1)}) = q(e^{t\vartheta}-1), \end{aligned} \tag{25}$$

where $q > 0$ denotes its shape parameter, while its rate parameter is stated as $\vartheta > 0$. Abdisalam *et al.* [2] expressed the Gompertz baseline failure rate for the PH mode as;

$$h_0(t) = q\vartheta e^{t\vartheta}. \tag{16}$$

Thus, the PH framework is formulated as indicated by Muse [24], the hazard function for a subject with covariates $\boldsymbol{V}$ together with its link term $\exp(\boldsymbol{V}'\boldsymbol{\beta})$ is represented as

$$\begin{aligned} h(t) &= h_0(t)\exp(\boldsymbol{V}'\boldsymbol{\beta}) \\ &= q\vartheta e^{t\vartheta} \exp(\boldsymbol{V}'\boldsymbol{\beta}). \end{aligned} \tag{27}$$

Using the exponential link expression $\exp(\boldsymbol{V}'\boldsymbol{\beta})$ to simplify, we get

$$h_{GPH}(t) = q\vartheta e^{t\vartheta} \exp(\beta_1 v_1 + \beta_2 v_2 + ... + \beta_z v_z). \tag{28}$$

The preceding equation clearly satisfies the PH condition. The Gompertz proportional hazards model, however, is scarcely applied to practice situations. The Gompertz PH model also has the following other lifespan distributions. "The Gompertz PH model's survival function" is defined as

$$S_{GPH}(t) = \exp(-q(e^{t\vartheta} - 1).\exp(\boldsymbol{V}' \hat{a}). \tag{29}$$

(Abdisalam *et al.,* [2]) expressed the PDF corresponding to the Gompertz-based PH framework as

$$f_{GPH}(t) = q\vartheta e^{t\vartheta} \exp(\boldsymbol{V}'\beta)\exp(-q(e^{t\vartheta} - 1)\exp(\boldsymbol{V}'\beta)). \tag{30}$$

In the Gompertz PH framework, the cumulative distribution function as well as the cumulative hazard function are represented as

$$F_{GPH}(t) = 1 - \exp(-q(e^{t\vartheta} - 1)\exp(\boldsymbol{V}'\beta)) \tag{32}$$

and

$$H_{GPH}(t) = q(e^{t\vartheta} - 1)\exp(V'\beta) \tag{32}$$

respectively.

4.3 *Exponential PH Model*

This distribution has a single unidentified parameter, $\boldsymbol{s}$. When it comes to lifetime distribution models, it is the most understandable distribution. However, it cannot adequately fit the frequently occurring failure rate forms of survival observations. Abdisalam *et al.* [2] indicated that "the PDF, CHF, CDF, survival function, with corresponding hazard rate of the exponential distribution" are, respectively, expressed as follows.

Suppose a random variable $X \sim \text{exponential}\,(s)$, then,

$$\begin{aligned} f(t) &= se^{-st}, \\ H(t) &= -\ln S(t) = -\ln(e^{-st}) = st, \\ F(t) &= 1 - e^{-st}, \\ S(t) &= e^{-st}, \\ h(t) &= s, \end{aligned} \tag{33}$$

where $s > 0$ is a scale parameter with survival or failure time t. The exponential baseline failure rate for the PH model is

$$h_0(t) = s. \tag{34}$$

Abdisalam *et al*. [1] noted that "the hazard rate for a person with covariate set $\boldsymbol{V}$ and its associated link term $\exp(\boldsymbol{V}'\beta)$" is

$$h(t) = h_0(t)\exp(\boldsymbol{V}'\beta) = s\exp(\boldsymbol{V}'\beta). \tag{35}$$

Through the use of the exponential transformation $\exp(\boldsymbol{V}'\beta)$ we can simplify and obtain

$$h_{EPH}(t) = s\exp(\boldsymbol{V}'\beta) = s\exp(\beta_1 v_1 + \beta_2 v_2 + ... + \beta_z v_z). \tag{36}$$

Suppose $s\exp(\boldsymbol{V}'\beta) = s^*$ then equation (36) can be written as

$$h_{EPH}(t) = s^*. \tag{37}$$

This indicates that the exponential distribution has satisfied the PH assumption.

The exponential PH model also has the following lifetime distributions: For a PH model with an exponential baseline, the survival function takes the form of

$$S_{EPH}(t) = \exp(-(st)\exp(\boldsymbol{V}'\beta)). \tag{38}$$

The exponential-based PH model has a PDF represented by

$$f_{EPH}(t) = s\exp(\boldsymbol{V}'\beta)\exp(-(st)\exp(\boldsymbol{V}'\beta)). \tag{39}$$

Its CDF and CHF are expressed as

$$F_{EPH}(t) = 1 - \exp(-(st).\exp(\boldsymbol{V}'\beta)) \tag{40}$$

and

$$H_{(EPH)} = (st)\exp(\boldsymbol{V}'\beta) \tag{41}$$

respectively.

5. Chen Proportional Hazards Model

Given the link function $\exp(\boldsymbol{V}'\beta)$, the Chen PH (CPH) regression modeling takes the form of

$$h(t;v) = h_0(t).\exp(\boldsymbol{V}'\beta), \tag{42}$$

where the regression coefficients connected to a vector of covariance $\boldsymbol{V}$, is represented by $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^{\mathrm{T}}$ and $h_0(t)$ represents the underlying hazard rate defined in equation (3), (John *et al*., [16]). This expression satisfies the condition that $\exp(\boldsymbol{V}'\beta) > 0$ for all values of $\boldsymbol{V}$. The CPH model can, therefore, be defined by substituting equation (3) into equation (42) as;

$$h(t;v) = mzt^{z-1}e^{t^z}.\exp(\boldsymbol{V}'\beta). \tag{43}$$

Rewriting equation (43) in the form of

$$\begin{aligned} h(t;v) &= (zm)t^{z-1}e^{t^z}.\exp(\boldsymbol{V}'\beta) \\ &= m\exp(\boldsymbol{V}'\beta).(zt^{z-1}e^{t^z}) \\ &= m^* zt^{z-1}e^{t^z}. \end{aligned} \tag{44}$$

From the equation (44), the failure rate for the CPH model looks similar to that of the equation (3) where $m^* = m\exp(\boldsymbol{V}'\beta)$ under the proportional hazards framework assumption, the Chen distribution exhibits closure.

A general representation of the Chen proportional hazards framework is given below

$$h(t;v) = mzt^{z-1}e^{t^z}.\exp(\boldsymbol{V}'\beta), \tag{45}$$

where m, and z are parameters of the baseline failure rate function, $\boldsymbol{V} = (v_1, v_2, ..., v_k)'$ define the covariate vector, $\boldsymbol{\beta}$ denotes the vector of associated regression parameters, and t is survival time or censoring time.

Let $h(t;v) = mzt^{z-1}\ e^{t^z}\ \exp(\boldsymbol{V}'\beta)$, $S(t) = \exp\left(-\int_0^t h(s)ds\right)$, then the survival function $S(t)$ can be described by;

$$\begin{aligned} S(t) &= \exp\left\{-\int_0^t mzs^{z-1}e^{s^z}\ \exp(\boldsymbol{V}'\beta)ds\right\} \\ &= \exp\left\{\exp(\boldsymbol{V}'\beta)\int_0^t -mzs^{z-1}e^{s^z}ds\right\}. \end{aligned} \tag{46}$$

By integrating the equation (46), the survival expression within the CPH framework can be derived in the form,

$$S(t) = \exp(m(1-e^{t^z}).\exp(\boldsymbol{V}'\beta)). \tag{47}$$

The CPH model yields the following expression for its "PDF"

$$f(t) = h(t).S(t).$$

Therefore,

$$f(t) = mzt^{z-1}e^{t^z}\exp(\boldsymbol{V}'\beta).\exp(m(1-e^{t^z})\exp(\boldsymbol{V}'\beta)). \tag{48}$$

The corresponding CDF takes the form

$$F(t) = 1 - S(t).$$

Therefore, the CDF and cumulative hazard for the CPH framework take the form;

$$F(t) = 1 - \exp(m(1-e^{t^z})\exp(\boldsymbol{V}'\beta)) \tag{49}$$

and

$$H(t) = -m(1-e^{t^z})\exp(\boldsymbol{V}'\beta) \tag{50}$$

respectively.

6. Maximum Likelihood Estimation

The likelihood for the CPH framework may be stated in the form

$$L(\kappa \mid t_i) = \prod_{i=1}^{n}(mzt_i^{z-1}e^{t_i^z}\exp(\boldsymbol{V}_i'\beta)\exp(m(1-e^{t_i^z})\exp(\boldsymbol{V}_i'\beta)))^{\delta_i} \times \prod_{i=1}^{n}(\exp(m(1-e^{t_i^z})\exp(\boldsymbol{V}_i'\beta)))^{(1-\delta_i)}, \tag{51}$$

δ_i represents the censoring indicator with sample size n.

Hence, the CPH model's log-likelihood term corresponding to the i^{th} subject may be represented as;

$$\ell_i(\kappa \mid t_i) = \sum_{i=1}^{n}\delta_i \ln(mzt_i^{z-1}e^{t_i^z}\exp(V_i'\beta).\exp(m(1-e^{t_i^z})\exp(\boldsymbol{V}_i'\beta))) + \sum_{i=1}^{n}(1-\delta_i)\ln(\exp(m(1-e^{t_i^z})\exp(\boldsymbol{V}_i'\beta))), \tag{52}$$

where κ represents all the parameters of the model including the regression parameter, $\boldsymbol{\beta}$, as well as the parameters for the distribution, $\boldsymbol{\psi} = (m, z)'$ that feature the baseline hazard component $h_0(t)$, resulting in the parameter set $\kappa = (\psi', \beta')'$.

Therefore, the first partial derivative for the equation (52) will be represented by ℓ_m, ℓ_z and ℓ_β for easy notation. Now equating $\ell_m = 0$, $\ell_z = 0$, $\ell_\beta = 0$, we obtain

$$u\left(\frac{1}{m}+1-\exp(t_i^z)\right)+w(1-\exp(t_i^z))=0, \tag{53}$$

$$u\left(\frac{1}{z}+\ln(t_i)+t_i^z\ln(t_i)-mt_i^z\exp(t_i^z)\ln(t_i)\right)-wmt_i^z\exp(t_i^z)\ln(t_i)=0, \tag{54}$$

$$2uV_{ij}+wV_{ij}=0, \tag{55}$$

where $u=\sum_{i=1}^{n}\delta_i$ and $w=\sum_{i=1}^{n}(1-\delta_i)\ j=1,2,...,p.$

Khosa [20] noted that the maximum likelihood estimates $\hat{m}$, $\hat{z}$ and $\hat{\beta}$ for m, z and β respectively, are obtained through solving these three nonlinear equations simultaneously.

7. Model Selection Criteria

The criteria for model selection are used as a means to identify the best-fitting model among those available for the dataset analysis. The most common statistical methods for model comparison and selection are "the Akaike information criteria (AIC) formulated by Akaike" [5] and cited by Khosa [20], the corrected Akaike information criteria (AICc) derived by Sugiura [27], and the Bayesian information criteria (BIC) by Schwarz [26]. The test statistics of these are expressed as

$$AIC=-2\ell+2(\nu+\kappa), \tag{56}$$

$$AICc=AIC+\frac{2\kappa(\kappa+1)}{n-\kappa+1}, \tag{57}$$

and

$$BIC=-2\ln(L)+\kappa\ln(n), \tag{58}$$

where ℓ is log-likelihood, ν denotes the quantity of covariates in a given model, with n representing the total sample, $\ln(L)$ defines the natural logarithm associated with the likelihood expression, while κ denotes the total number of model parameters involved in the probability distribution (Schwarz [26]).

8. Chen Proportional Hazards Model Application

"Right-censored survival statistics are presented in this section to illustrate the flexibility as well as usefulness for the developed CPH model in modeling survival datasets" (Abdisalam *et al.* [1]).

8.1 *Lung Cancer Survival Data*

To highlight its application, a dataset representing the survival outcome of lung cancer individuals, as cited by Kalbfleisch and Prentice [18], was used. It consists of a sample size $n = 137$ patients who were studied by the Veterans Administration Lung Cancer Study. Nine out of these observations (representing 6.5%) were censored, where the response was survival duration (t) and the predictors included: the individual's age (x_1), diagnosis time (x_2), and treatment (x_3). This data set was fitted to the CPH, GOMPH, and WPH models. Table 1 summarizes the estimates of the parameter, related standard errors, as well as associated P-values with respect to the fitted models. The parameters for these frameworks are statistically significant at 0.05. The regression parameter (β_1) of the WPH model is also significant at 0.05 and negatively affects the observed survival durations among individuals suffering from lung cancer, as indicated within Table 1.

Table 1

Parameter estimates for lung cancer dataset

Model	Parameter	Estimate	Std Error	*P*-Value
CPH	m	0.0583	0.0194	0.0013
	z	0.2322	0.0084	$< 2.0 \times 10^{-16}$
	β_1	0.0022	0.0064	0.7290
	β_2	0.0040	0.0101	0.6899
	β_3	–0.1634	0.1809	0.3665
GOMPH	α	3.6887	0.0017	$< 2.2 \times 10^{-16}$
	γ	0.0015	0.0003	< 0.0001
	β_1	0.0089	0.0052	0.0888
	β_2	0.0032	0.0117	0.7810
	β_3	–0.3446	0.1911	0.07134

Contd...

WPH	ϑ	0.6809	0.0471	$< 2.0 \times 10^{-16}$
	ω	0.2059	0.0683	0.0026
	β_1	–0.0308	0.0063	< 0.0001
	β_2	0.0019	0.0108	0.8541
	β_3	–0.3494	0.1861	0.0604

The model selection criteria and the log-likelihood estimate for the lung cancer patient's dataset are outlined within Table 2. The recommended model was the one that yielded the most favourable AIC, AICc, and BIC results. The computed AIC, AICc, as well as the BIC statistics indicated in Table 2 show that this dataset is better captured by the CPH model compared to the GOMPH and WPH models, as it has the lowest values of these measures.

Table 2

Goodness-of-fit measures associated with lung cancer dataset

Model	-2ℓ	AIC	AICc	BIC
CPH	**1519.9190**	**1529.9190**	**1530.3701**	**1540.2120***
GOMPH	1524.1350	1534.1350	1534.5861	1548.7350
WPH	1530.1340	1540.1340	1540.5851	1554.7330

*: Means the model that fits the data well

The residual plots based on Cox–Snell measures for both the CPH and alternative models are used to assess how well they fit the lung cancer observations. Figure 2 shows the residual plots for these models developed for analyzing lung cancer observations. Evaluation of the residual plots demonstrates that the CPH model yields the most satisfactory and outperforms competing models in fitting the lung cancer data. While the fits of the WPH as well as GOMPH residuals diverge from the diagonal except at the bottom and upper ends, the residuals of the CPH model fit lie closely along the diagonals. This outcome points to the CPH model being the most appropriate choice for modeling this dataset.

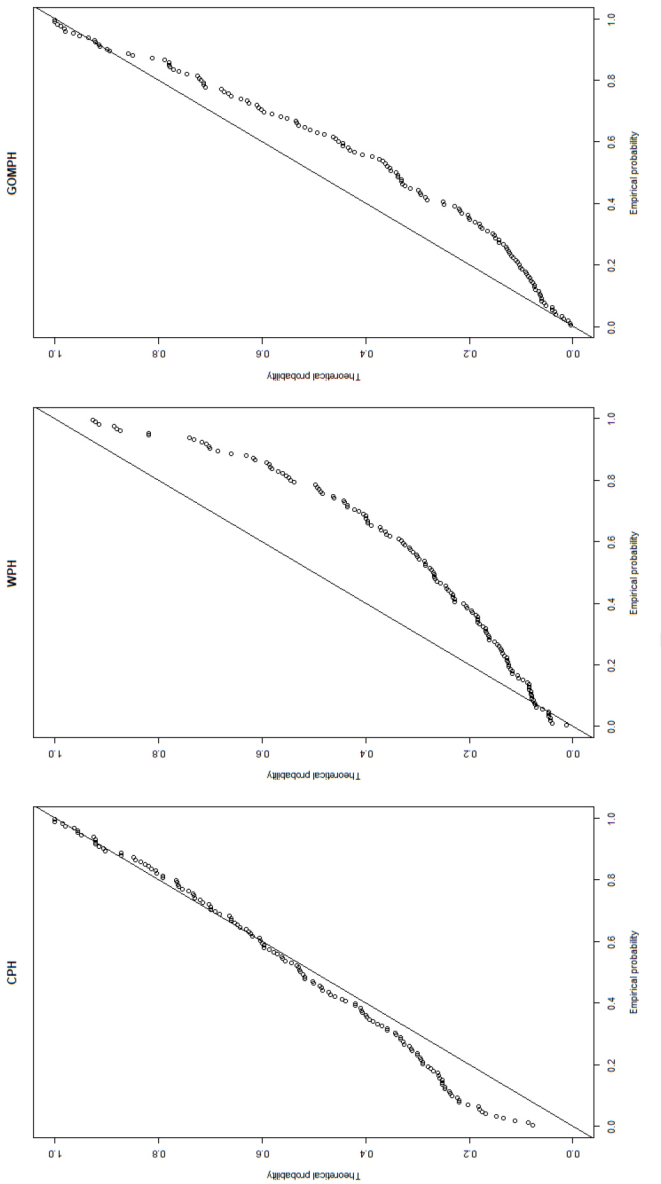


Figure 2
Cox-Snell residuals Plots for the lung cancer data

With regard to the purpose of verifying whether its proportional hazards assumption appeared true, the study took into account the numerical analysis of Schoenfeld residuals. The treatment (Trt), age, and diagnostic time (diagtime) covariates were checked for the PH assumption of each covariate as well as the overall model for the CPH model. As shown within Table 3, the *P*-values exceed 0.05 significance level, indicating that the PH assumption is met.

Table 3

Test statistics of the PH assumption test

Covariate	df	Chi-sq	*P*-Value
Trt	1	3.5263	0.0600
Age	1	1.6871	0.01900
Diagtime	1	0.0091	0.9200
GLOBAL	3	5.9971	0.1100

8.2 *Application with the Diabetes Dataset*

The second application considered a dataset made up of 394 Individuals deemed to have "high-risk" diabetic retinopathy as per the guidelines of the Diabetic Retinopathy Study (DRS) [15]. They indicated that every patient had at random one eye receiving laser treatment as well as, the remaining had no treatment. "For every eye, the event of interest was the time from the start of treatment to the time when the visual acuity dropped below 5/200 two visits in a row" [15] and Maru and Sangbum, [23]. "Thus, there is a built-in lag time of approximately 6 months". Therefore, the actual time to blindness in months less than the shortest time-to-event (6.5 months) represents the survival durations in this dataset [15]. According to Huster, Brookmeyer, and Self [15], censorship was brought about by death, dropout, or study completion (Alba and Pablo, [6]). In this data, the response variable y (time to event or last follow-up) and "covariates are: age at diagnosis (x_1), treatment (x_2), and risk group (x_3)", (Maru and Sangbum, [23]).

Table 4 displays the *P*-values, standard errors, and parameter estimates for the various models. The findings in Table 4 demonstrate that each of the CPH model regression parameters is significant at 0.05. Nevertheless, the majority of the GOMPH and WPH parameters are also significant.

Table 4

Estimates, standard errors and P-Value for the diabetic dataset

Model	Parameter	Estimate	Std Error	*P*-Value
CPH	ϕ	0.0060	0.0045	0.1069
	θ	0.3930	0.0126	$< 2.2 \times 10^{-16}$
	β_1	0.0039	0.0055	0.3087
	β_2	–0.6071	0.1690	< 0.0001
	β_3	0.1067	0.0571	0.0099
GOMPH	α	99.7650	0.0000	$< 2.2 \times 10^{-16}$
	γ	0.0027	0.0011	0.0128
	β_1	–0.0096	0.0058	0.0991
	β_2	–0.7383	0.1679	< 0.0001
	β_3	–0.2959	0.0453	< 0.0001
WPH	ϑ	0.4481	0.0287	$< 2.2 \times 10^{-16}$
	ω	90.4060	0.0000	$< 2.2 \times 10^{-16}$
	β_1	–0.0124	0.0053	0.0195
	β_2	–0.9271	0.1631	< 0.0001
	β_3	–0.4053	0.0275	$< 2.2 \times 10^{-16}$

Table 5 presents the -2ℓ, AIC, AICc, along with the BIC values. The CPH model exhibited the minimum -2ℓ, AIC, AICc, as well as BIC values when the models were compared, as found in Table 5. Therefore, the CPH model is the most suitable for analyzing the diabetes data.

Table 5

Goodness-of-fit metrics for the diabetic dataset

Model	-2ℓ	AIC	AICc	BIC
CPH	**1687.6020**	**1697.6020**	**1697.7558**	**1717.4830***
GOMPH	1751.4180	1761.4180	1761.5718	1781.3000
WPH	1829.4410	1839.4410	1839.5948	1859.3230

*; Means the model that fits the data well.

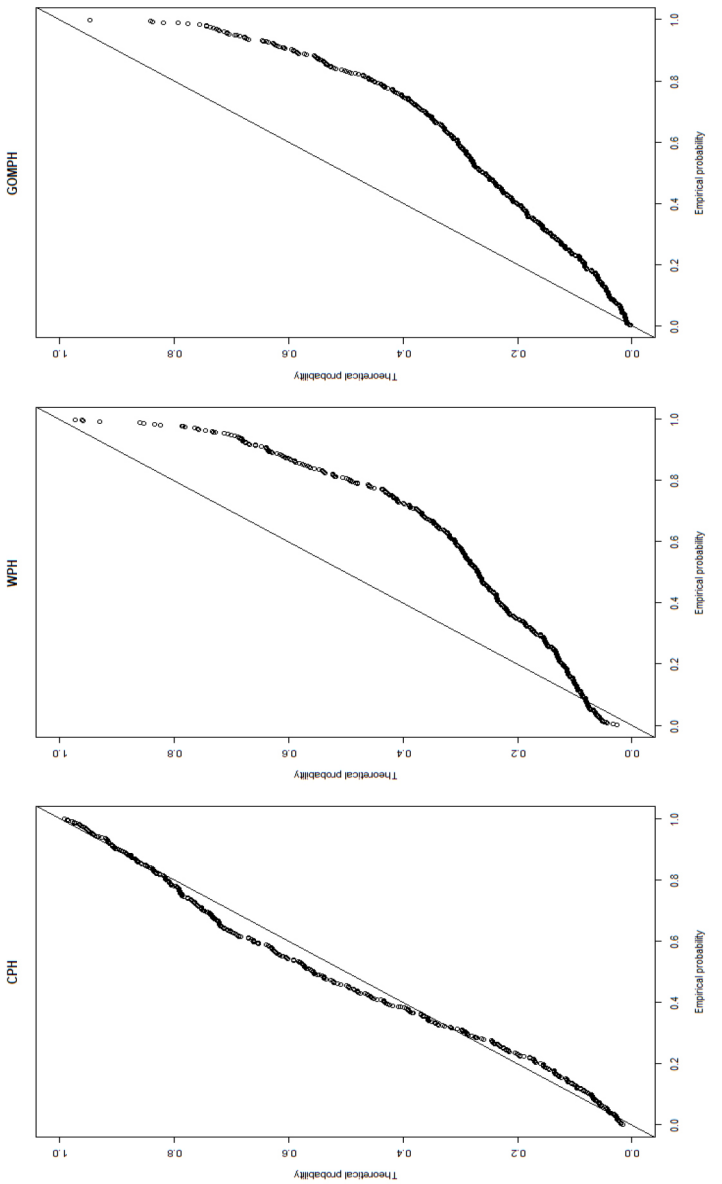


Figure 3
Cox-Snell residuals Plots for the diabetes data

The CPH, together with other competing models' Cox-Snell residuals plots, were utilized to evaluate how well these models model the diabetic data set. In Figure 3, the results of these plots are displayed. Figure 3 shows that the CPH model residuals follow the standard exponential distribution, where the residuals are closely packed along the diagonal. However, the residuals of the GOMPH and WPH models totally veer off from the diagonals. This demonstrates that the CPH model is superior to the GOMPH and WPH models. As a result, the CPH model, which is the best of the three, adequately fits the diabetes data.

The assumption that the hazards are proportional over time within the CPH framework for the diabetes data was tested using Schoenfeld residuals numerical analysis techniques (Abdisalam *et al.* [1]). The degree of freedom (df), Chi-square (Chi-sq), and *P*-value of the PH assumption test are displayed in Table 6.

Both the specific covariates and the overall fitted model were taken into account by the PH assumption test. The PH test results shown in Table 6 showed that the model as a whole and each covariate met the PH assumption of the diabetic data set. This is clear from the fact that both the *P*-values for the general model and each of the individual covariates are all higher than the 0.05 level of significance.

Table 6

Test statistics of the PH assumption test

Covariate	df	Chi-sq	*P*-Value
x_1	1	0.3640	0.5500
x_2	1	0.6020	0.4400
x_3	1	1.6340	0.2000
GLOBAL	3	2.5790	0.4600

9.0 Conclusion

To provide a better fit for modeling lifespan data, researchers are constantly interested in developing new models to enhance the ones that are already in existence. As a result, the CPH model was established as part of this research. "The maximum likelihood estimation" method was adopted to generate estimators for the parameter values defining the new model structure (Adam *et al.* [4]).

Two collections of lifetime observations were used to highlight the new model's adaptability and applicability. The log-likelihood, AIC, AICc, and BIC were obtained. The CPH model's result was compared with that of the WPH and GOMPH model structures. Considering the results provided by the model selection criteria together with the log-likelihood and the Cox-Snell residuals plots, the parametric fit provided by the CPH model surpassed that of the remaining models across the datasets.

Declaration of Interest

The authors confirm that the publication of this paper is not affected by any conflict of interest.

References

[1] H. M. Abdisalam, N. Oscar, M. Samuel, M. A. Huda, and H. E. Abdal-Aziz, "A flexible parametric proportional hazard model: Simulation and applications to right-censored healthcare data," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–28 (2022).

[2] H. M. Abdisalam, N. Oscar, M. Samuel, M. A. Huda, and H. E. Abdal-Aziz, "A flexible Bayesian parametric proportional hazard model: Simulation and applications to right-censored healthcare data," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–28 (2022).

[3] J. Abonongo, A. I. L. Abonong, A. Aljadani, M. M. Mansour, and H. M. Yousof, "Accelerated failure model with empirical analysis and application to colon cancer data: Testing and validation," *Alexandria Engineering Journal*, vol. 113, pp. 391–408 (2025).

[4] B. S. M. Adam, S. A. Abdulaziz, N. Oscar, M. Joseph, C. Christophe, and Z. A. Ahmed, "The extended exponential–Weibull accelerated failure time model with application to Sudan COVID-19 data," *Mathematics*, vol. 11, no. 2 (2023).

[5] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723 (1974).

[6] M. F. Alba and M. Pablo, "Testing ageing notions through percentiles of the residual life," *Journal of Statistical Computation and Simulation*, vol. 90, no. 5, pp. 869–888 (2020), doi: 10.1080/00949655.2019.1710150.

[7] I. Alkhairy, M. Nagy, A. H. Muse, and E. Hussam, "The Arctan-X family of distributions: Properties and application to actuarial and sciences," *Complexity*, vol. 2021, pp. 1–14 (2021).

[8] Z. Chen, "A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function," *Statistics and Probability Letters*, vol. 49, pp. 155–161 (2000).

[9] L. Chin-Diew, *Generalized Weibull Distributions*. Cham, Switzerland: Springer (2014).

[10] D. Collett, *Modeling Survival Data in Medical Research*, 3rd ed. Boca Raton, FL, USA: CRC Press (2015).

[11] D. D. H. David, *Modeling Survival Data Using Frailty Models*. New York, NY, USA: Springer (2019).

[12] B. Gompertz, "On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies," *Philosophical Transactions of the Royal Society of London*, vol. 115, pp. 513–583 (1825).

[13] D. S. Gonzalez-Gonzalez, M. R. Pina-Monarrez, and L. Torres-Trevino, "Estimation of parameters in Cox's proportional hazard model: Comparison between evolutionary and the Newton–Raphson approach," in *Proc. Computational Intelligence*, pp. 513–523 (2008), doi: 10.1007/978-3-540-88636-5_49.

[14] E. S. Hatem, M. G. Babu, and I. Elbatal, "A generalization of Weibull distribution: Theory and applications," *Advances and Applications in Statistics*, vol. 83, pp. 1–25 (2022).

[15] W. J. Huster, R. Brookmeyer, and S. G. Self, "Modeling paired survival data with covariates," *Biometrics*, vol. 45, no. 1, pp. 145–156 (1989).

[16] P. K. John, C. V. Hans, G. I. Joseph, and H. S. Thomas, *Handbook of Survival Analysis*. Boca Raton, FL, USA: Chapman & Hall/CRC (2014).

[17] A. B. Justin, "Survival analysis: A modified Kaplan–Meier estimator," M.S. thesis, Dept. Math., Missouri State Univ., USA (2017).

[18] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, 2nd ed. Hoboken, NJ, USA: Wiley (2011).

[19] S. A. Khan and S. K. Khosa, "Generalized log-logistic proportional hazard model with applications in survival analysis," *Journal of Statistical Distributions and Applications*, vol. 3, no. 16, pp. 1–18 (2016).

[20] S. K. Khosa, "Parametric proportional hazard models with application in survival analysis," Ph.D. dissertation, Dept. Math. and Stat., Univ. of Saskatchewan, Saskatoon, Canada (2019).

[21] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, 2nd ed. Hoboken, NJ, USA: Wiley (2011).

[22] C. Legrand, *Advanced Survival Models*. Boca Raton, FL, USA: CRC Press (2021).

[23] K. Maru and C. Sangbum, "A GEE approach for the semiparametric accelerated lifetime model with multivariate censored data," *Communications for Statistical Applications and Methods*, vol. 30, no. 4, pp. 389–402 (2023).

[24] A. H. Muse, "Bayesian and frequentist approach for flexible parametric hazard-based regression models with generalized log-logistic baseline distribution: An application to right-censored oncology data sets," Ph.D. dissertation, Dept. Statistics, Pan African Univ. Inst. for Basic Science, Technology and Innovation, Kenya (2023).

[25] P. Royston, "Flexible parametric alternatives to the Cox model, and more," *The Stata Journal*, vol. 1, no. 1, pp. 1–29 (2001).

[26] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464 (1978).

[27] N. Sugiura, "Further analysis of the data by Akaike information criterion and the finite corrections," *Communications in Statistics—Theory and Methods*, vol. 7, no. 1, pp. 13–26 (1978).

[28] B. Tarvirdizade and N. Nematollahi, "Inference for the power-exponential hazard rate distribution under progressive type-II censored data," *Journal of Statistics and Management Systems*, vol. 24, no. 6, pp. 1169–1212 (2021).

[29] X. Wang, Y. Yue, and J. J. Faraway, *Bayesian Regression Modeling with INLA*. Boca Raton, FL, USA: Chapman & Hall/CRC (2018).

Received October, 2023