Earthline Journal of Mathematical Sciences
E-ISSN: 2581-8147
Volume 14, Number 4, 2024, Pages 747-771
https://doi.org/10.34198/ejms.14424.747771

A Parametric Cox Proportional Hazard Model with Application

Precious O. Ibeakuzie[1,*] and Sidney I. Onyeagu[2]

[1] Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria
e-mail: op.ibeakuzie@unizik.edu.ng

[2] Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Nigeria

Abstract

Survival analysis has become integral to clinical studies, especially in emerging diseases and terminal ailments. This study focused on improving the popular Cox PH model. The new method developed is a parametric type, incorporating the hazard rate of the exponential distribution. It was noted that though the functional form of the Cox PH model was altered, the assumptions were upheld. Additionally, the new model parameters were estimated using the same maximum partial likelihood as the Cox model. Data on the survival times of 137 patients who underwent bone marrow transplants were deployed, and the proposed parametric Cox PH model proved superior to the Cox PH model.

1 Introduction

Many articles have explored the clinical characteristics of surviving patients from diseases such as leukaemia. Modelling the longitudinal (data that is collected

Received: April 26, 2024; Accepted: May 31, 2024; Published: June 3, 2024
2020 Mathematics Subject Classification: 62N02.
Keywords and phrases: exponential hazard function, Cox PH model, parametric model, estimation.
*Corresponding author

through a series of repeated observations of the same subjects over some extended time frame) and event-time outcomes separately, for example, using linear mixed models [1] or Cox regression models [2] can therefore be inefficient, and can lead to biased effect size estimates if the two outcome processes are correlated [3]. Research into joint modelling of longitudinal and time-to-event data has received considerable attention during the past two decades [3–7]. The motivation behind this an active field of research has stemmed from three broad scientific objectives:

1. Improving inference for a repeated measurement outcome subject to an informative dropout mechanism that is not of direct interest [8].

2. Improving inference for a time-to-event outcome, whilst taking account of an intermittently and error-prone measured endogenous time-dependent variable [7].

3. Studying the relationship between the two correlated processes [5].

The Cox proportional hazards (PH) model, often called the Cox regression model, is a statistical technique for analyzing survival data. It's named after the statistician Sir David Cox, who developed it in 1972. The Cox PH model allows us to assess the relationship between the survival time of subjects and predictor variables (covariates) while assuming that the hazard (risk of event occurrence) for any individual is proportional to the hazard for any other individual.

This model is commonly used in medical research, epidemiology, and other fields where survival analysis is important. It's especially valuable when studying the effects of various factors on the time until an event occurs, such as death, relapse of a disease, or failure of a mechanical system.

The Cox PH model estimates the hazard function as a product of a baseline hazard function and an exponential function of the covariates. This allows for assessing how the hazard changes as predictor variables change while maintaining the assumption of proportional hazards.

The Cox PH model is a powerful tool for analyzing survival data. It provides insights into the factors that influence the time until an event of interest occurs while accommodating the assumption of proportional hazards.

Unlike parametric methods, Coxâs method does not require some particular probability distribution to represent survival times. Thatâs why we call it a semi-parametric model, which makes Coxâs method more robust. Another advantage of using Coxâs method is that itâs relatively easy to incorporate time-dependent covariates. We use the maximum likelihood method to estimate the regression parameter in the parametric model. In contrast, we use the method of maximum partial likelihood to estimate the parameters in Coxâs model. Whatâs remarkable about partial likelihood is that you can estimate the coefficients without specifying the baseline hazard function $h_0(t)$.

Cox Proportional Hazard also known as Cox PH model for survival analysis has come a long way in the literature with some studies both in application and software implementation being remarkable such as [9, 10], [2, 11], [12, 13], [14], [15, 16]. The survival time of a particular event is called the time-to-event, [17]. The time of death and time to develop a disease are examples of survival data. Statistical methods for survival analysis have been applied to many vital fields of research. Generally, survival analysis uses data to predict survival probability and identify risk and/or prognostic factors related to subjectsâ survival and disease progression. An essential aspect of survival data is not usually fully observed in all subjects under study, leading to different censored data types. Subjects in a study are usually assumed to be selected randomly (interred the study randomly) in the sense of simple random sample (SRS) [18].

The Cox PH model is the most commonly used survival data analysis technique that simultaneously allows one to include and to assess the effect of multiple covariates [19]. These model covariates can include the variables of specific research interest (treatment groups) and potential confounders for which the researcher wants to control (demographic and other clinical factors). Multiple strategies for covariate selection have been described, and the aim of

the studyâmost often to determine the effect of a covariate while controlling for confounding versus prediction of survival using a set of predictor variablesâshould be considered in choosing a strategy [14, 20, 21].

Cox PH regression does not directly model survival probabilities or times but the hazard function. Herein, it is assumed that all patients have a common baseline hazard function that only depends on time. Each subjectâs hazard function is a multiple of this common baseline hazard, and the individual multiplicator is a constant determined by a time-independent function of a patientâs covariate values [2]. This implies that the ratio of the hazard rates between different patients (the HR) is assumed to be constant over timeâin other words, the effect of a covariate is assumed to be the same at all time points. This is the PH assumption of the Cox PH model, [22], which is discussed in more detail below.

Under this assumption, exponentiated regression coefficients for each covariate can be interpreted as the HR for a 1-unit change in the respective covariate value. This is akin to interpreting exponentiated regression coefficients as odds ratios in logistic regression [14].

While the Cox PH model estimates regression coefficients without making assumptions about the shape of the hazard function, it is possible to work backwards and use the parameter estimates to estimate the adjusted hazard or survival function. This allows the plotting of adjusted curves for different groups, which are very similar to Kaplan-Meier curves, but instead show or predict the probability of survival in each group while keeping the other covariates fixed at their mean values [23]. Survival proportions can also be predicted for each arbitrary combination of covariate values [19]. The Cox PH model is very popular among clinical researchers for numerous reasons. It does not need the researcher to specify the function of the baseline hazard. Provided proportional hazard assumptions are met, the results are robust. With results from the Cox PH model, the coefficients obtained can be used to model and predict the expected survival of patients with specific values of covariates included in the model. To

understand this, we will return to the example dataset of 228 stage III lung cancer patients who underwent surgery. We would like to understand the association of patient sex and age at surgery with all-cause mortality. For this purpose, we will fit a Cox PH, including these two covariates in the model.

The Cox PH model is a widely used regression model in survival analysis to investigate the relationship between subjects' survival time and predictor variables.

The Cox PH model assumes that the hazard function for any individual at any time is the product of an underlying baseline hazard function and an exponential function of the predictor variables. Mathematically, it can be represented as:

$$h(t|X) = h_0(t) \times \exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right), \tag{1}$$

where $h(t|X)$ is the hazard function at time t given the predictor variables X. $h_0(t)$ is the baseline hazard function representing the hazard when all predictor variables are zero. $\beta_1, \beta_2, \cdots, \beta_p$ are the regression coefficients associated with the predictor variables. In the Cox PH model, the baseline hazard function $h_0(t)$ is unspecified and is absorbed into the estimation process, allowing for the comparison of hazard rates between different groups while remaining agnostic about the shape of the baseline hazard function.

So, while a distribution's hazard function is not directly used in the Cox PH model, the model is based on the proportional hazards assumption, which involves the hazard function. This gap in functional form motivates this research.

This study aims to develop a new Cox PH model by incorporating the hazard function of the exponential distribution, altering the classical proportional hazard function structure but not violating the assumptions of the Cox PH model. It is, therefore, motivated by the need

1. To utilize the hazard function of the exponential distribution in designing a parametric Cox PH structure.

2. To estimate the parametric Cox PH structure parameters using the classical maximum likelihood estimation.
3. To deploy the parametric Cox PH structure in modelling time to event data.
4. To Compare the parametric Cox PH structure with the classical Cox PH regression model using the survival data.

2 Material and Methods

We begin this chapter by considering the assumptions and conditions for implementing the Cox PH model.

1. Independence of survival times between distinct individuals in the sample,
2. A multiplicative relationship between the predictors and the hazard (as opposed to a linear one as was the case with multiple linear regression analysis, discussed in more detail below).
3. A constant hazard ratio over time.
4. There is no intercept in the PH model because an intercept can ve absorbed into the baseline $h_0(t)$.
5. The PH assumption states that $h(t \mid x_i) = h_0(t) \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$, where $h_0(t) \geq 0$ is an unspecified function known as the baseline hazard function. It is a hazard function for an individual with features $x_{i1} = \cdots = x_{ip} = 0$.
6. The name PH arises from the fact that the hazard function for an individual with feature vector x_i is some unknown function $h_0(t)$ times the factor $\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$. The quantity $\exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$ is called the relative risk for the feature vector $x_i = (x_{i1}, \cdots, x_{ip})$, relative to that for the feature vector $x_i = (0, \cdots, 0)$.

7. There are no tied failure times. In the case of ties, the exact form of the partial likelihood is more complicated, and a number of computational approximations must be used.

Parametric models assume a specific distribution of the survival times. Advantages of a parametric model include a higher efficiency (i.e., greater power), [19], which can be particularly useful with smaller sample sizes. Furthermore, various parametric techniques can model survival times when the PH assumption is unmet.

However, it can be quite challenging to identify the most appropriate data distribution, and parametric models have the drawback of providing misleading inferences if the distributional assumptions are unmet. In contrast, the semi-parametric Cox model is a safe and proven method without specifying a specific data distribution,36, which is why this model is most common in analyzing survival data. For a more detailed discussion on parametric models, we refer to previously published literature on the topic [19, 24].

The exponential distribution is one classical distribution popular in the literature for modelling lifetime data sets. The hazard function of the exponential distribution is expressed as

$$h_0(t) = \lambda; \tag{2}$$

where $\lambda > 0$ is the rate or scale parameter independent of the component's time to failure in the life testing experiment. In the construction of an improved Cox PH model, we assume the baseline hazard function $h_0(t)$ in eq. 1 is the hazard function of the exponential distribution represented in eq. 2 and assume it is so when at least one predictor X is different from zero. That is, $\beta_i \neq 0$ for at least one i.

Theorem 1 (**Parametric Cox PH model**). *Let X_i be predictor variables (covariates) for survival data with coefficients β_i, $i = 1, 2, \cdots, p$. Define the baseline hazard function $h_0(t)$ of the Cox PH model as the hazard function of the exponential distribution that is $h_0(t) = \lambda$; an improved Cox PH model can be*

constructed as

$$h_I(t|X) = \lambda \times \exp\left(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p\right) \tag{3}$$

where $h_0(t) = \lambda$ is the baseline hazard function provided not all the predictors are zeros.

Proof. The proof of theorem 1 easily follows from substituting eq. 2 into eq. 1 □

Corollary 1.1 (**Parameter Estimation using Partial Least Squares**). *Cox's derivation of an estimator of β can be loosely described as follows. Let $t_1, t_2, \cdots, t_k$ represent the unique ordered failure times in the sample of n subjects; assume for now that there are no tied failure times (tied censoring times are allowed) so that $k = n$. Consider the individuals at risk of failing an instant before failure time t_i. This set of individuals is called the risk set at time t_i, and we use R_i to denote this risk set. R_i is the set of subjects j such that the subject had not failed or been censored by time t_i; the risk set R_i includes subjects with failure/censoring time $Y_j \geq t_i$. The conditional probability that individual i is the one that failed at t_i, given that the subjects in the set R_i are at risk of failing and given further that exactly one failure occurs at t_i, is*

$$Prob\{subject\ i\ fails\ at\quad t_i \mid R_i\ and\ one\ failure\ at\quad t_i\} = \frac{Prob\,(subject\ i\ fails\ at\quad t_i \mid R_i)}{Prob\,(one\ failure\ at\quad t_i \mid R_i)} \tag{4}$$

Though the functional form of the baseline hazard function is known, we cannot simply plug $h_I(t|X)$ into the likelihood and then estimate λ, and $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^T$ by maximum likelihood.

We use the same "sequential in time" logic to estimate the parameters and derive the Kaplan-Meier survival curve and the log-rank test. Then, the total

hazard at failure times y_i for the at-risk observations is

$$\sum_{i':y_{i'}\geq y_i} \lambda \exp\left(\sum_{j=1}^{p} x_{i'j}\beta_j\right) \tag{5}$$

Therefore, the probability that the ith observation is the one to fail at time y_i (as opposed to one of the other observations in the risk set is)

$$\frac{\lambda \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)}{\sum_{i':y_{i'}\geq y_i} \lambda \exp\left(\sum_{j=1}^{p} x_{i'j}\beta_j\right)} \tag{6}$$

The partial likelihood is simply the product of these probabilities over all of the uncensored observations

$$PL(\beta) = \prod_{i:\delta_i=1} \frac{\lambda \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)}{\sum_{i':y_{i'}\geq y_i} \lambda \exp\left(\sum_{j=1}^{p} x_{i'j}\beta_j\right)} \tag{7}$$

Critically, the partial likelihood is valid regardless of the time value of $h_0(t)$, making the model flexible and robust. Eq. 7 does not support analytical manipulation; therefore, any numerical iterations will produce estimates of the parameters. Eq. 6 can be used to obtain the relative risk function at each failure time. Suppose the failures occurred at time $t = 7, 10$ and 12; the relative risk (RR) can be derived by substituting the time component in eq. 6.

$$RR_7(\beta) = \frac{\lambda \exp\left(\sum_{j=1}^{p} x_{7j}\beta_j\right)}{\sum_{i':y_{i'}\geq y_i} \lambda \exp\left(\sum_{j=1}^{p} x_{i'j}\beta_j\right)}, \tag{8}$$

$$RR_{10}(\beta) = \frac{\lambda \exp\left(\sum_{j=1}^{p} x_{10j}\beta_j\right)}{\sum_{i':y_{i'}\geq y_i} \lambda \exp\left(\sum_{j=1}^{p} x_{i'j}\beta_j\right)}, \tag{9}$$

and

$$RR_{12}(\beta) = \frac{\lambda \exp\left(\sum_{j=1}^{p} x_{12j}\beta_j\right)}{\sum_{i':y_{i'} \geq y_i} \lambda \exp\left(\sum_{j=1}^{p} x_{i'j}\beta_j\right)} \tag{10}$$

In this sense, the partial likelihood will be expressed as $PL(\beta) = RR_7(\beta) \times RR_{10}(\beta) \times RR_{12}(\beta)$. *To estimate the parameters here, we simply maximize the partial likelihood with respect to the parameters. As is the case for logistic regression, no closed-form solution is available, so iterative algorithms are required.*

In addition to estimating the parameters, we can obtain other model outputs like those in least squares regression and logistic regression. For example, we can obtain the p-values corresponding to particular null hypotheses (e.g., $H_0 : beta_j = 0$*) and estimate the standard errors and confidence intervals associated with the coefficients.*

Suppose that we have just a single predictor ($p = 1$) *with* $x_i \in \{0, 1\}$. *We can consider taking two possible approaches to test whether there is a difference between the survival times of the observation in the two groups.*

1. *Fit an improved Cox PH model and let the null hypothesis* $H_0 : \beta = 0$ *(since* $p = 1, \beta$ *is a scalar).*
2. *Perform a log-rank test to compare the two groups.*

Now, when taking the first approach, there are a number of possible ways to test H_0. *One way is known as a score test. In the case of a single binary covariate, the score test for* $H_0 : \beta = 0$ *in the Cox PH model is exactly equal to the log-rank test.*

The area-under-the-curve (AUC) is an appealing method for assessing a fitted Cox model on a test set. In this method, for each observation, we calculate the estimated risk score, $\hat{y}_i = \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}$, *for* $i = 1, \cdots, n$ *using the estimated Cox model coefficients. The Harrel's concordance index or (C-index) computes*

the proportion of observation pairs for which $\hat{\zeta}_i' > \hat{\zeta}_i$ *and* $y_i > y_{i'}$*, so that*

$$C = \frac{\sum_{i,i':y_i>y_{i'}} I\left(\hat{\zeta}_i' > \hat{\zeta}_i\right)\delta_{i'}}{\sum_{i,i':y_i>y_{i'}} \delta_{i'}} \tag{11}$$

This is the proportion of pairs for which the model correctly predicts the relative survival time among all pairs for which this can be determined.

Analogous to comparing groups of continuous data using a t-test or analysis of variance, the survival curves for 2 or more different groups (e.g., treatments or prognostic factors) can also be compared with hypothesis testing. Most commonly, the log-rank test is applied, which tests the null hypothesis that there is no difference in the probability of an event at any time point [25].

When reporting a log-rank P value comparing Kaplan-Meier curves, the entire distribution is being tested, not a particular time, such as 5-year survival. The log-rank test is based on the same assumptions as the Kaplan-Meier survival curve and makes no explicit assumptions about the distribution of the survival curves.

However, when the survival curves of different groups crossâindicating that 1 group has a more favourable survival in a certain time interval and less favourable survival in another time intervalâthe power to detect such differences is very low [25]. Moreover, the log-rank test cannot adjust for other covariates that might affect survival time. While it can determine whether observed differences are significant, it cannot estimate the difference between groups [19]. Other techniques, described below, can be used to address these issues.

To perform this analysis, a survival dataset with survival times and covariates will be required. Typically, survival datasets include information on the time to event (survival time), the event indicator (whether the event occurred or not), and covariates (predictor variables).

3 Application

This section deploys the proposed parametric Cox PH model to data on 137 Bone Marrow Transplant Patients. The data has been studied by [26] and [27].

g —Disease group

1. ALL
2. AML low-risk
3. AML high-risk

T_1 —Time (in days) to death or on study time

T_2 —Disease-Free survival time (time to relapse, death or end of study)

δ_1 —Death indicator

1 —Dead [0] —Alive

δ_2 —Relapse indicator

1 —Relapsed [0] —Disease-Free

δ_3 —Disease-Free survival indicator

1 —Dead or relapsed [0] —Alive disease-free

T_A —Time (in days) to acute graft-versus-host disease

δ_A —Acute graft-versus-host disease indicator

1 —Developed acute graft-versus-host disease

0 —Never developed acute graft-versus-host disease

T_C —Time (in days) to chronic graft-versus-host disease

δ_C —Chronic graft-versus-host disease indicator

1 —Developed Chronic graft-versus-host disease

0 —Never developed Chronic graft-versus-host disease

T_P —Time (in days) to return of platelets to normal levels

δ_P —Platelet recovery indicator

1 —Platelets returned to normal levels

0 —Platelets never returned to normal levels

Z_1 —Patient age in years

Z_2 —Donor age in years

Z_3 —Patient sex

1 —Male [0] —Female

Z_4 —Donor Sex

1 —Male [0] —Female

Z_5 —Patient CMV status

1 —CMV positive [0] —CMV negative

Z_6 —Donor CMV status

1 —CMV positive [0] —CMV negative

Z_7 —Waiting time to transplant in days

Z_8 —FAB

1 —FAB Grade 4 Or 5 and AML [0] —Otherwise

Z_9 —Hospital

1 —The Ohio State University [2] —Alfred

3 —St. Vincent [2] —Hahnemann

Z_{10} —MTX used as a graft-versus-host-prophylactic

1 —Yes [0] —No

The results in Tables 1 and 2 reveal that the parametric Cox PH model has a higher concordance ratio of 0.9810 while the classical Cox PH model has o.7810 concordance ratio. In both model scenarios, the variable Disease-Free survival indicator did not produce any value. The mean square error of the parametric Cox PH is lower than that of the classical Cox PH model. More covariates are significant in the parametric Cox PH model than in the classical Cox PH model. This tells that the proposed parametric Cox PH model improves the classical Cox PH model. The confidence interval for both models is seemingly the same.

Kaplan-Meier curves in Figure 1 visually represent survival data over time, showing the proportion of subjects surviving. Disease group 2 has a higher survival probability at most time points. The disease group 2 has a steady-state death rate within the interval of 2300 to 4000 days (median survival time is 2300), while groups 1 and 2 showed constant death rates from 1300 days (median survival time is 1300) till the termination of the study. The censored data points were many in group 2 before the steady-state death occurred and fewer after, while there were many after the constant death rate in group 3. This is indicated by the small vertical ticks and shows that the patient left the study or was lost to follow-up before experiencing the event. The small table below the plot shows the number of subjects still at risk at various times. This shows that about 137 patients contributed to the survival estimates at the commencement of the study, 54 on the 1000th day, 14 on the 2000th day, and 1 on both the 3000 and 4000th day. The shaded regions around the survival curve represent confidence intervals at the 95% level. They indicate the uncertainty around the survival estimates, which is higher for group 2 and least for group 3. In other words, the confidence intervals for group 3 are narrower, suggesting more precise survival estimates. For comparing the three disease groups, the plot includes a p-value from the

g	T_1	T_2	δ_1	δ_2	δ_3	T_A	δ_A	T_C	δ_C	T_P	δ_P	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
1	2081	2081	0	0	0	67	1	121	1	13	1	26	33	1	0	1	1	98	0	1	0
1	1602	1602	0	0	0	1602	0	139	1	18	1	21	37	1	1	0	0	1720	0	1	0
1	1496	1496	0	0	0	1496	0	207	1	12	1	26	35	1	1	1	0	127	0	1	0
1	1462	1462	0	0	0	70	1	95	1	13	1	17	21	0	1	0	0	168	0	1	0
1	1433	1433	0	0	0	1433	0	236	1	12	1	32	36	1	1	1	1	93	0	1	0
1	1377	1377	0	0	0	1377	0	123	1	12	1	32	31	1	1	1	1	2187	0	1	0
1	1330	1330	0	0	0	1330	0	96	1	17	1	20	17	1	0	1	1	1006	0	1	0
1	996	996	0	0	0	72	1	121	1	12	1	22	24	1	0	0	0	1319	0	1	0
1	226	226	0	0	0	226	0	226	0	10	1	18	21	0	1	0	0	208	0	1	0
1	1199	1199	0	0	0	1199	0	91	1	29	1	24	40	1	1	0	1	174	0	3	1
1	1111	1111	0	0	0	1111	0	111	0	22	1	19	28	1	1	0	1	236	0	3	1
1	530	530	0	0	0	38	1	84	1	34	1	17	28	1	1	0	0	159	0	3	1
1	1182	1182	0	0	0	1182	0	112	1	22	1	24	23	0	0	0	1	203	0	2	1
1	1167	1167	0	0	0	39	1	487	1	1167	0	27	22	0	1	1	1	191	0	2	1
1	418	418	1	0	1	418	0	220	1	21	1	18	14	1	1	0	0	110	0	1	0
1	417	383	1	1	1	417	0	417	0	16	1	15	20	1	1	0	0	824	0	1	0
1	276	276	1	0	1	276	0	81	1	21	1	18	5	0	0	0	0	146	0	1	0
1	156	104	1	1	1	28	1	156	0	20	1	20	33	1	1	0	1	85	0	1	0
1	781	609	1	1	1	781	0	781	0	26	1	27	27	1	0	1	1	187	0	1	0
1	172	172	1	0	1	22	1	172	0	37	1	40	37	0	0	0	1	129	0	1	0
1	487	487	1	0	1	487	0	76	1	22	1	22	20	1	1	0	0	128	0	1	0
1	716	662	1	1	1	716	0	716	0	17	1	28	32	1	1	0	0	84	0	1	0
1	194	194	1	0	1	194	0	94	1	25	1	26	32	0	1	0	0	329	0	1	0
1	371	230	1	1	1	371	0	184	1	9	1	39	31	0	1	0	1	147	0	1	0
1	526	526	1	0	1	526	0	121	1	11	1	15	20	1	1	0	0	943	0	1	0
1	122	122	1	0	1	88	1	122	0	13	1	20	26	1	0	0	1	2616	0	1	0
1	1279	129	1	1	1	1279	0	1279	0	22	1	17	20	0	0	0	0	937	0	3	1
1	110	74	1	1	1	110	0	110	0	49	1	28	25	1	0	1	0	303	0	3	1
1	243	122	1	1	1	243	0	243	0	23	1	37	38	0	1	1	1	170	0	3	1
1	86	86	1	0	1	86	0	86	1	86	0	17	26	1	0	1	0	239	0	3	1
1	466	466	1	0	1	466	0	119	1	100	1	15	13	1	1	0	0	508	0	3	1
1	262	192	1	1	1	10	1	84	0	59	1	29	32	1	1	1	0	74	0	3	1
1	162	109	1	1	1	162	0	162	0	40	1	36	43	1	1	1	0	393	0	2	1
1	262	55	1	1	1	262	0	262	0	24	1	23	16	0	1	1	1	331	0	2	1
1	1	1	1	0	1	1	0	1	0	1	0	42	48	1	1	0	0	196	0	2	1
1	107	107	1	0	1	107	0	107	1	107	0	30	19	1	1	1	1	178	0	2	1
1	269	110	1	1	1	269	0	120	0	27	1	29	20	0	1	1	1	361	0	2	1
1	350	332	0	1	1	350	0	350	0	33	1	22	20	1	0	0	0	834	0	2	1

g	T_1	T_2	δ_1	δ_2	δ_3	T_A	δ_A	T_C	δ_C	T_P	δ_P	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
2	2569	2569	0	0	0	2569	0	2569	0	21	1	19	13	1	1	1	0	270	1	1	0
2	2506	2506	0	0	0	2506	0	2506	0	17	1	31	34	1	1	0	0	60	0	1	0
2	2409	2409	0	0	0	2409	0	2409	0	16	1	35	31	1	1	1	0	120	0	1	0
2	2218	2218	0	0	0	2218	0	2218	0	11	1	16	16	1	1	1	0	60	1	1	0
2	1857	1857	0	0	0	1857	0	260	0	15	1	29	35	0	0	1	0	90	0	1	0
2	1829	1829	0	0	0	1829	0	1829	0	19	1	19	18	1	1	1	0	210	0	1	0
2	1562	1562	0	0	0	1562	0	1562	0	18	1	26	30	1	1	1	1	90	0	1	0
2	1470	1470	0	0	0	1470	0	180	1	14	1	27	34	1	1	0	1	240	0	1	0
2	1363	1363	0	0	0	1363	0	200	1	12	1	13	24	1	1	1	0	90	0	1	0
2	1030	1030	0	0	0	1030	0	210	1	14	1	25	29	0	0	0	0	210	0	1	0
2	860	860	0	0	0	860	0	860	0	15	1	25	31	0	1	0	1	180	0	1	0
2	1256	1256	0	0	0	1256	0	120	1	66	1	30	16	0	1	1	0	180	0	2	1
2	2246	2246	0	0	0	52	1	380	1	15	1	45	39	0	0	0	0	105	0	4	0
2	1870	1870	0	0	0	1870	0	230	1	16	1	33	30	0	0	1	1	225	0	4	0
2	1799	1799	0	0	0	1799	0	140	1	12	1	32	23	1	0	0	0	120	0	4	0
2	1709	1709	0	0	0	20	1	348	1	19	1	23	28	0	1	1	0	90	1	4	0
2	1674	1674	0	0	0	1674	0	1674	0	24	1	37	34	1	1	0	0	60	1	4	0
2	1568	1568	0	0	0	1568	0	1568	0	14	1	15	19	1	0	0	0	90	0	4	0
2	1527	1527	0	0	0	1527	0	1527	0	13	1	22	12	0	1	0	1	450	1	4	0
2	1324	1324	0	0	0	25	1	1324	0	15	1	46	31	1	1	1	1	75	0	4	0
2	957	957	0	0	0	957	0	957	0	69	1	18	17	1	1	0	0	90	0	4	0
2	932	932	0	0	0	29	1	932	0	7	1	27	30	0	0	0	0	60	1	4	0
2	847	847	0	0	0	847	0	847	0	16	1	28	29	1	1	0	0	75	0	4	0
2	848	848	0	0	0	848	0	155	1	16	1	23	26	1	1	0	0	180	0	4	0
2	1850	1850	0	0	0	1850	0	9	1	37	36	0	0	0	0	1	0	180	0	3	1
2	1843	1843	0	0	0	1843	0	19	1	34	32	0	0	0	1	1	1	270	0	3	1
2	1535	1535	0	0	0	1535	0	1535	0	21	1	35	32	0	1	0	0	180	1	3	1
2	1447	1447	0	0	0	1447	0	220	1	24	1	33	28	0	1	1	1	150	0	3	1
2	1384	1384	0	0	0	1384	0	200	1	19	1	21	18	0	0	0	0	120	0	3	1
2	414	414	1	0	1	414	0	414	0	27	1	21	15	1	1	0	1	120	1	1	0
2	2204	2204	1	0	1	2204	0	2204	0	12	1	25	19	0	0	0	1	60	0	1	0
2	1063	1063	1	0	1	1063	0	240	1	16	1	50	38	1	0	1	0	270	1	1	0
2	481	481	1	0	1	30	1	481	1	24	1	35	36	1	0	1	1	90	1	1	0
2	105	105	1	0	1	21	1	120	0	15	1	37	34	1	0	1	1	120	0	1	0
2	641	641	1	0	1	641	0	641	0	11	1	26	24	1	1	0	0	90	0	1	0
2	390	390	1	0	1	390	0	390	0	11	1	50	48	1	1	0	0	120	0	1	0
2	288	288	1	0	1	18	1	100	1	288	0	45	43	1	1	1	1	90	0	1	0
2	522	522	1	1	1	25	1	140	1	20	1	28	30	1	1	0	1	90	1	1	0

g	T_1	T_2	δ_1	δ_2	δ_3	T_A	δ_A	T_C	δ_C	T_P	δ_P	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
2	79	79	1	0	1	16	1	79	0	79	0	43	43	0	0	0	0	90	0	1	0
2	1156	748	1	1	1	1156	0	180	1	18	1	14	19	1	0	0	0	60	0	1	0
2	583	486	1	1	1	583	0	583	0	11	1	17	14	0	1	0	0	120	0	1	0
2	48	48	1	0	1	48	0	48	0	14	1	32	33	0	1	1	0	150	1	1	0
2	4311	272	1	1	1	4311	0	4311	0	12	1	30	23	0	1	1	0	120	1	1	0
2	1074	1074	1	0	1	1074	0	120	1	19	1	30	32	1	1	1	0	150	1	1	0
2	393	381	1	1	1	393	0	100	1	16	1	33	28	0	0	0	0	120	1	1	0
2	10	10	1	0	1	10	0	10	0	10	0	34	54	1	0	1	1	240	0	2	1
2	53	53	1	0	1	53	0	53	0	53	0	33	41	0	1	1	1	180	0	2	1
2	80	80	1	0	1	10	1	80	0	80	0	30	35	0	0	0	1	150	0	2	1
2	35	35	1	0	1	35	0	35	0	35	1	23	24	0	1	1	1	150	0	2	1
2	1499	248	1	1	1	1499	0	1499	0	9	1	35	18	1	1	0	1	30	1	4	0
2	704	704	1	0	1	36	1	155	1	18	1	29	21	0	1	1	0	105	1	4	0
2	653	211	1	1	1	653	0	653	0	23	1	23	16	1	0	0	0	90	1	4	0
2	222	219	1	1	1	222	0	132	1	52	1	28	30	1	1	1	1	120	1	3	1
2	1356	606	1	1	1	1356	0	1356	0	14	1	33	22	1	1	1	0	210	1	3	1
3	2640	2640	0	0	0	2640	0	2640	0	22	1	18	23	1	1	0	0	750	0	1	0
3	2430	2430	0	0	0	2430	0	2430	0	14	1	29	26	1	1	0	1	24	0	1	0
3	2252	2252	0	0	0	2252	0	150	1	17	1	35	31	1	0	0	0	120	0	1	0
3	2140	2140	0	0	0	2140	0	220	1	18	1	27	17	1	1	1	1	210	0	1	0
3	2133	2133	0	0	0	2133	0	250	1	17	1	36	39	0	1	0	0	240	0	1	0
3	1238	1238	0	0	0	1238	0	250	1	18	1	24	28	1	0	1	1	240	0	1	0
3	1631	1631	0	0	0	1631	0	150	1	40	1	27	21	1	0	1	0	690	1	2	1
3	2024	2024	0	0	0	2024	0	180	1	16	1	35	41	0	1	0	0	105	1	4	0
3	1345	1345	0	0	0	32	1	360	1	14	1	50	36	1	1	1	1	120	0	4	0
3	1136	1136	0	0	0	1136	0	140	1	15	1	47	27	1	0	1	0	900	0	3	1
3	845	845	0	0	0	845	0	845	0	20	1	40	39	0	0	1	1	210	1	3	1
3	491	422	1	1	1	491	0	180	1	491	0	22	21	0	0	0	0	210	1	1	0
3	162	162	1	0	1	162	0	162	0	13	1	22	23	1	0	0	1	300	0	1	0
3	1298	84	1	1	1	1298	0	1298	0	1298	0	8	2	0	0	1	0	105	1	1	0
3	121	100	1	1	1	28	1	121	0	65	1	39	48	1	1	1	1	210	1	1	0
3	2	2	1	0	1	2	0	2	0	2	0	20	19	1	1	0	0	75	1	1	0
3	62	47	1	1	1	62	0	62	0	11	1	27	25	1	1	0	0	90	1	1	0
3	265	242	1	1	1	265	0	210	1	14	1	32	32	1	0	0	0	180	1	1	0
3	547	456	1	1	1	547	0	130	1	24	1	31	28	1	0	1	1	630	1	1	0
3	341	268	1	1	1	21	1	100	1	17	1	20	23	0	1	1	1	180	1	1	0
3	318	318	1	0	1	318	0	140	1	12	1	35	40	0	1	1	1	300	0	1	0
3	195	32	1	1	1	195	0	195	0	16	1	36	39	1	1	0	0	90	1	1	0

g	T_1	T_2	δ_1	δ_2	δ_3	T_A	δ_A	T_C	δ_C	T_P	δ_P	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}
3	469	467	1	1	1	469	0	90	1	20	1	35	33	0	0	1	0	120	0	1	0
3	93	47	1	1	1	93	0	93	0	28	1	7	2	1	1	0	0	135	1	1	0
3	515	390	1	1	1	515	0	515	0	31	1	23	25	1	1	1	0	210	1	1	0
3	183	183	1	0	1	183	0	130	1	21	1	11	7	0	1	0	0	120	1	1	0
3	105	105	1	0	1	105	0	105	0	105	0	14	18	1	0	0	0	150	1	1	0
3	128	115	1	1	1	128	0	128	0	12	1	37	35	0	0	1	1	270	0	1	0
3	164	164	1	0	1	164	0	164	0	164	0	19	32	0	0	0	1	285	1	1	0
3	129	93	1	1	1	129	0	129	0	51	1	37	34	0	1	1	0	240	1	1	0
3	122	120	1	1	1	122	0	122	0	12	1	25	29	0	1	1	1	510	1	1	0
3	80	80	1	0	1	21	1	80	0	0	1	35	28	1	0	0	0	780	1	1	0
3	677	677	1	0	1	677	0	150	1	8	1	15	14	1	1	1	0	150	1	1	0
3	73	64	1	1	1	73	0	73	0	38	1	45	42	0	1	1	0	180	1	2	1
3	168	168	1	0	1	168	0	200	1	48	1	32	43	0	1	1	1	150	1	2	1
3	74	74	1	0	1	29	1	74	0	24	1	41	29	0	1	1	1	750	0	2	1
3	16	16	1	0	1	16	0	16	0	16	0	27	36	0	0	1	0	180	0	4	0
3	248	157	1	1	1	248	0	100	1	52	1	33	39	0	0	1	1	180	1	4	0
3	732	625	1	1	1	732	0	732	0	18	1	39	43	0	1	1	1	150	1	4	0
3	105	48	1	1	1	105	0	105	0	30	1	17	14	0	1	0	0	210	1	4	0
3	392	273	1	1	1	392	0	122	1	24	1	43	50	1	1	1	0	240	0	3	1
3	63	63	1	0	1	38	1	63	0	16	1	44	37	1	1	0	0	360	1	3	1
3	97	76	1	1	1	97	0	97	0	97	0	48	56	1	1	1	1	330	0	3	1
3	153	113	1	1	1	153	0	153	0	59	1	31	25	0	1	1	1	240	0	3	1
3	363	363	1	0	1	363	0	363	0	19	1	52	48	1	1	1	0	180	0	3	1

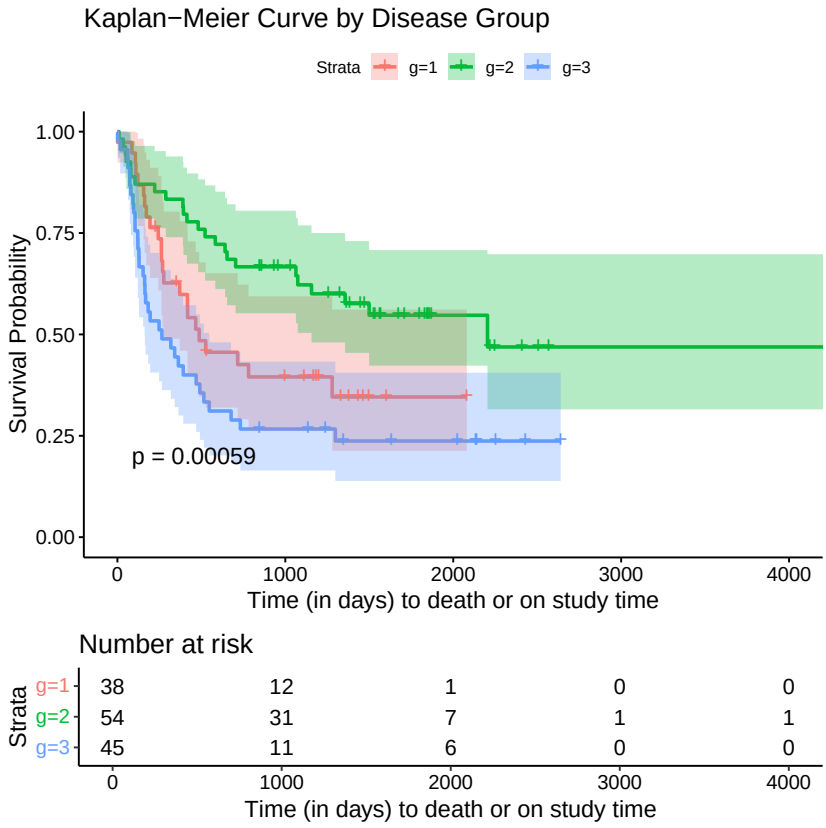

Figure 1: Kaplan-Meier Curve for the 137 Bone Marrow Transplant Patient Data

Table 1: Parametric COX PH model fitted on 137 Bone Marrow Transplant Patients Data

covariates	$\hat{\beta}_j$	e^{coefs}	e^{-coefs}	L-95	U-95	LR	Wald	p-value	con.	MSE
g	0.44132	1.50680	0.71083	0.877205	2.28949			0.16186		
T_2	-0.01482	0.98528	1.01494	0.98094	0.98965			0.00000		
δ_2	-2.07528	0.02552	7.96680	0.05391	0.29227			0.00000		
δ_3	NA	NA	NA	NA	NA			NA		
T_A	-0.02054	0.98952	1.01059	0.98584	0.99421			0.00000		
δ_A	-2.80910	0.08134	12.29380	0.02340	0.28271			0.00008		
T_C	-0.00891	0.99113	1.00895	0.98819	0.99408			0.00000		
δ_C	-1.69170	0.18421	5.42869	0.07125	0.47625			0.00048		
T_P	-0.00282	0.99718	1.00283	0.99285	1.00153			0.20345		
δ_P	-0.17990	0.83536	1.19710	0.25500	2.73649	1.22974	1.02974	0.76635	0.98101	13.231
Z_1	0.04751	1.01766	0.98264	0.95681	1.08238			0.57783		
Z_2	0.01470	1.01481	0.98541	0.96104	1.07158			0.59664		
Z_3	-0.01917	0.99087	1.00921	0.54709	1.79465			0.00586		
Z_4	-0.17895	0.83698	1.19477	0.43096	1.72553			0.00928		
Z_5	0.09902	1.10409	0.90573	0.49699	2.45276			0.00790		
Z_6	-0.46149	0.63034	1.58644	0.33376	1.19049			0.00588		
Z_7	0.00029	1.00029	0.99971	0.99924	1.00134			0.00877		
Z_8	-0.08159	0.93091	1.07422	0.40865	2.12062			0.00466		
Z_9	-0.64225	0.53140	1.88184	0.33364	0.84636			0.00776		
Z_{10}	1.09666	2.99616	0.33398	1.04067	8.61462			0.03196		

log-rank test, assessing whether the differences between the survival curves are statistically significant. With a p-value of 0.00059, the differences between the curves are significant.

Table 2: Classical COX PH model fitted on 137 Bone Marrow Transplant Patients Data

covariates	$\hat{\beta}_j$	e^{coefs}	e^{-coefs}	L-95	U-95	LR	Wald	p-value	con.	MSE
g	0.3413	1.4068	0.7108	0.8720	2.2695			0.1619		
T_2	-0.0148	0.9853	1.0149	0.9809	0.9897			0.0000		
δ_2	-2.0753	0.1255	7.9668	0.0539	0.2923			0.0000		
δ_3	NA	NA	NA	NA	NA			NA		
T_A	-0.0105	0.9895	1.0106	0.9858	0.9932			0.0000		
δ_A	-2.5091	0.0813	12.2938	0.0234	0.2827			0.0001		
T_C	-0.0089	0.9911	1.0089	0.9882	0.9941			0.0000		
δ_C	-1.6917	0.1842	5.4287	0.0712	0.4763			0.0005		
T_P	-0.0028	0.9972	1.0028	0.9928	1.0015			0.2034		
δ_P	-0.1799	0.8354	1.1971	0.2550	2.7365	3.0297	3.0297	0.7663	0.7810	17.6610
Z_1	0.0175	1.0177	0.9826	0.9568	1.0824			0.5778		
Z_2	0.0147	1.0148	0.9854	0.9610	1.0716			0.5966		
Z_3	-0.0092	0.9909	1.0092	0.5471	1.7946			0.9759		
Z_4	-0.1779	0.8370	1.1948	0.4310	1.6255			0.5993		
Z_5	0.0990	1.1041	0.9057	0.4970	2.4528			0.8079		
Z_6	-0.4615	0.6303	1.5864	0.3338	1.1905			0.1549		
Z_7	0.0003	1.0003	0.9997	0.9992	1.0013			0.5877		
Z_8	-0.0716	0.9309	1.0742	0.4087	2.1206			0.8647		
Z_9	-0.6322	0.5314	1.8818	0.3336	0.8464			0.0078		
Z_{10}	1.0967	2.9942	0.3340	1.0407	8.6146			0.0420		

4 Conclusion and Remarks for Future Studies

Survival analysis has become integral to clinical studies, especially in emerging diseases and terminal ailments. This study focused on improving the popular Cox PH model. The new method developed is a parametric type, incorporating the hazard rate of the exponential distribution. It was noted that though the functional form of the Cox PH model was altered, the assumptions were upheld. Additionally, the new model parameters were estimated using the same maximum partial likelihood as the Cox model. Data on the survival times of 137 patients who underwent bone marrow transplants were deployed, and the proposed parametric Cox PH model proved superior to the Cox PH model. It is clear that the

parametric Cox PH model outperformed the classical Cox model. However, a number of things can be further established;

1. Because the assumptions of the Cox PH model were not violated in this study, given that the exponential distribution has a constant hazard rate, it is therefore recommended that other choices of non-constant hazard rate functions be made and deployed in the classical Cox PH model to attain some variant parametric Cox PH models.

2. Further studies can explore various estimation procedures for models that violate the assumptions of the Cox PH model.

3. It has been shown in the literature that standardized variables are more mathematically tractable than non-standardized scores. Therefore, any future proposed variant of this parametric Cox PH model should use standardized covariates.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974. https://doi.org/10.2307/2529876

[2] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[3] Ibrahim, J. G., Chu, H., & Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology, 28*(16), 2796. https://doi.org/10.1200/JCO.2009.25.0654

[4] Asar, Ö., Ritchie, J., Kalra, P. A., & Diggle, P. J. (2015). Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International Journal of Epidemiology, 44*(1), 334-344. https://doi.org/10.1093/ije/dyu262

[5] Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics, 1*(4), 465-480. https://doi.org/10.1093/biostatistics/1.4.465

[6] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press. https://doi.org/10.1201/b12208

[7] Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330-339. https://doi.org/10.2307/2533118

[8] Dupuy, J.-F., & Mesbah, M. (2002). Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime Data Analysis, 8*, 99-115. https://doi.org/10.1023/A:1014871806118

[9] Allison, P. D. (2010). *Survival analysis using SAS: a practical guide*. SAS Institute.

[10] Allison, P. D. (2018). Event history and survival analysis. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 86-97). Routledge. https://doi.org/10.4324/9781315755649-7

[11] Cox, D. R. (2018). *Analysis of survival data*. Chapman and Hall/CRC.

[12] Fox, J. (2016). *Using the R commander: a point-and-click interface for R*. Chapman and Hall/CRC.

[13] Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage Publications.

[14] Hosmer Jr, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modeling of time-to-event data* (Vol. 618). John Wiley & Sons. https://doi.org/10.1002/9780470258019

[15] Therneau, T., et al. (2015). A package for survival analysis in S. *R package version*, 2(7), 2014.

[16] Therneau, T. M. (1997). Extending the Cox model. In *Proceedings of the first Seattle symposium in biostatistics: survival analysis* (pp. 51-84). Springer. https://doi.org/10.1007/978-1-4684-6316-3_5

[17] Samawi, H., Yu, L., & Yin, J. (2023). On Cox proportional hazards model performance under different sampling schemes. *PLOS ONE, 18*(4), e0278700. https://doi.org/10.1371/journal.pone.0278700

[18] Scheaffer, R. L., Mendenhall, W., Ott, L., & Gerow, K. (1990). *Elementary survey sampling* (Vol. 501). Duxbury Press California.

[19] Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. *British Journal of Cancer, 89*(3), 431-436. https://doi.org/10.1038/sj.bjc.6601119

[20] Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part IV: further concepts and methods in survival analysis. *British Journal of Cancer, 89*(5), 781-786. https://doi.org/10.1038/sj.bjc.6601117

[21] Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part III: multivariate data analysis–choosing a model and assessing its adequacy and fit. *British Journal of Cancer, 89*(4), 605-611. https://doi.org/10.1038/sj.bjc.6601120

[22] Kartsonaki, C. (2016). Survival analysis. *Diagnostic Histopathology*, 22(7), 263-270. https://doi.org/10.1016/j.mpdhp.2016.06.005

[23] Nieto, F. J., & Coresh, J. (1996). Adjusting survival curves for confounders: a review and a new method. *American Journal of Epidemiology, 143*(10), 1059-1068. https://doi.org/10.1093/oxfordjournals.aje.a008670

[24] Windpassinger, M., Plattner, O., Gemeiner, J., R oder, G., Baumann, A., Zimmerman, N. M., & Sessler, D. I. (2016). Pharyngeal oxygen insufflation during AirTraq laryngoscopy slows arterial desaturation in infants and small children. *Anesthesia & Analgesia, 122*(4), 1153-1157. https://doi.org/10.1213/ANE.0000000000001189

[25] Bland, J. M., & Altman, D. G. (2004). The logrank test. *BMJ, 328*(7447), 1073. https://doi.org/10.1136/bmj.328.7447.1073

[26] Dietz, K., Gail, M., Krickeberg, K., Samet, J., & Tsiatis, A. (2002). Statistics for biology and health. *Survival Analysis, 7*.

[27] Gail, M., Krickeberg, K., Samet, J., Tsiatis, A., & Wong, W. (2007). *Statistics for biology and health*. Springer.