# COX'S REGRESSION MODEL FOR RELIABILITY ANALYSIS AND FAILURE TIME PREDICTION IN ENGINEERING SYSTEMS

Iyappan. M[1*], Balaji. M[2], G. Sathya Priyanka[3], Jesmon Raj. N[4], P B Deepa[5]

•

[1*]Associate Professor, St. Francis College, Bengaluru, India
[2]Associate Professor, Seshadripuram College, Bengaluru, India
[3]Research Scholar, Periyar University, Salem-11, India
[4]Assistant Professor & Head, St. Francis College, Bengaluru, India.
[5]Assistant Professor, St. Francis College, Bengaluru, India

[1*]iyappastat@gmail.com, [2]balajim0405@gmail.com, [3]sathyapriyankastat@gmail.com,
[4]jesmon.n@stfranciscollege.edu.in, [5]deepa.p@stfranciscollege.edu.in

**Abstract**

*Reliability engineering is essential for predicting failure times, optimizing maintenance strategies, and minimizing operational risks in engineering systems. Traditional models, such as Weibull and exponential distributions, provide useful insights into failure rates but struggle to incorporate multiple influencing factors. Cox's Proportional Hazards Model offers a more flexible approach by integrating covariates such as environmental stress, operational loads, and maintenance history, without assuming a predefined failure time distribution. This study explores the application of Cox's regression model in reliability analysis, focusing on failure time prediction and system longevity assessment under varying operational conditions. A major advantage of Cox's model is its ability to handle censored data, a common challenge in reliability studies where some systems remain operational beyond the observation period. By employing the partial likelihood estimation method, the model quantifies the impact of various covariates on failure risks without requiring explicit assumptions about baseline hazard functions. This study utilizes a simulated dataset of 20 industrial components observed over five years to demonstrate the model's effectiveness in predicting system failures and estimating survival probabilities. The study further evaluates parameter estimation techniques, including maximum likelihood estimation (MLE) and Type I and Type II censoring, to validate the model's predictive power in practical reliability applications. Additionally, statistical computations confirm that the Cox model provides reliable hazard rate estimations, enabling engineers to develop predictive maintenance schedules and enhance system reliability. This research highlights the importance of survival analysis techniques in reliability engineering. Cox's model proves to be an effective tool for identifying risk factors and optimizing failure time predictions. The methodology is applicable to diverse engineering domains, including mechanical component wear analysis, structural health monitoring, and software reliability assessment.*

**Keywords:** Cox's Proportional Hazards Model, Failure Time Prediction, Hazard Rate Estimation, Censoring Mechanisms, Engineering System Failure.

# I. Introduction

Reliability engineering focuses on assessing and predicting the lifespan of systems under varying operational conditions to prevent failures and optimize maintenance strategies. Accurately estimating failure times is essential for minimizing operational risks and ensuring system efficiency. Traditional reliability models, such as Weibull and exponential distributions, have been extensively used to describe system failure rates [3]. However, when multiple covariates influence failure progression, regression-based survival models provide a more flexible and robust approach. One such model is Cox's Proportional Hazards Model [5], which is widely utilized in survival analysis and can be effectively applied to reliability studies for complex systems, including mechanical components, software systems, and structural elements [2]. Cox's model is particularly advantageous because it does not require a predefined failure time distribution, making it well-suited for engineering applications where failure mechanisms are uncertain. This model estimates the effect of covariates such as stress levels, environmental conditions, and operational loads on system survival times while handling censored data, a frequent occurrence in reliability testing when systems are still operational at the end of the observation period [4]. By analyzing historical and real-time failure data, Cox's model provides a powerful tool for predictive maintenance and risk assessment.

The estimation of system survival functions, hazard rates, and failure distributions is essential for reliability analysis. Since failure events often exhibit stochastic behavior, engineers employ probabilistic models, including stochastic processes (e.g., Markov models), Weibull distributions, and Cox's Proportional Hazards Model, to assess system reliability. Cox's model has been widely applied in mechanical component wear-out analysis, structural health monitoring, and software failure prediction [2]. This paper explores the application of Cox's regression model in reliability engineering, with a focus on predicting system failures based on observed data.

# II. Methods

## I. Formulation of Cox's Regression Model in Reliability Engineering

Let $T_i$, $i = 1,2,3, \dots .. n$ represent the failure times of $n$ different systems, where each system is observed over a fixed time interval $[0, C_i]$ with $C_i$ being the censoring time for system $i$ Suppose that system $j$ has a hazard rate $\lambda_j(t)$ at time $t$ defined as:

$$\lambda_j(t) = \lambda_0(t) \exp {}_{k=1}^{r} \beta_k Z_{jk}(t) \tag{1}$$

Where:

- $(\beta_1, \beta_2, \dots .. \beta_r)$is row vector of $r$ unknown coefficients, representing the impact of various covariates on failure risk.
- $\left(Z_{j1}(t)\right), \left(Z_{j2}(t)\right), \left(Z_{j3}(t)\right), \dots .. \left(Z_{jr}(t)\right)$ is a column vector of $r$ possibly time-varying covariates.
- $\lambda_0(t)$ is an unknown baseline hazard function, representing the failure rate when all covariates are zero.

Thus, $\lambda_j(t)$ gives the conditional probability that system $j$ fails during an infinitesimal time interval $(t, t + h)$, given that it has survived until time $t$.

The observed data for the $j^{th}$ system consists of:

- $T_j \wedge C_j$, the minimum of failure time and censoring time.
- $\delta_j = I\{T_j \leq C_j\}$ an indicator variable specifying whether the failure was observed or censored.

- The covariate value $Z_{jk}(t)$ recorded over the interval $\left[0, T_j \wedge C_j\right]$.

## II. Estimation of Model Parameters

The Cox model is semi-parametric, meaning that the regression coefficients $\beta_k$ are estimated independently of the baseline hazard $\lambda_0(t)$, which is treated as an infinite-dimensional nuisance parameter.

Define the risk set at time $t$, denoted as $\Omega(t)$, which includes all systems still operational at $t$:

$$\Omega(t) = \left\{j: T_j \geq t, C_j \geq t\right\} \tag{2}$$

Given that a failure occurs at time $t$, the probability that it belongs to system $j$ as:

$$\frac{\lambda_j(t)}{{}_{i\in\Omega(t)}\lambda_j(t)} \tag{3}$$

Where $\lambda_0(t)$ cancels out in the numerator and denominator, making this expression independent of the baseline hazard function.

Cox proposed that statistical inference be performed using the partial likelihood function:

$$L(\beta) = {}_{j=1}^{n} \frac{\exp\left({}_{k=1}^{r}\beta_k Z_{jk}\right)}{{}_{i\in\Omega(T_j)} exp\left({}_{k=1}^{r}\beta_k Z_{jk}\right)} \tag{4}$$

This partial likelihood allows for efficient estimation of regression parameters without explicitly modeling $\lambda_0(t)$, making Cox's model particularly suitable for predicting system failures, analyzing preventive maintenance schedules, and improving reliability assessments in engineering contexts.

## III. Methodologies and Estimation in Reliability Engineering Models

Let us illustrate the methodology, we consider a system's failure progression influenced by three covariates:

- $Z_j(1)$ = Environmental stress factor.
- $Z_j(2)$ = Operational load
- $Z_j(3)$ = Maintenance history

We assume the baseline hazard rate follows an exponential distribution.

$$\lambda_0(t) = \lambda \tag{5}$$

Where $t$, represents the time until system failure.

## IV. Failure Event and Censoring Mechanisms

In reliability testing a sample of $n$ components is observed over a testing period $(0, T)$. During this time, $K$ components transition from an operational state to failure.

Type I Censoring: The test ends at a fixed time $T_0$, meaning some components may still be operational at $T_0$, leading to censored observations.

Type II Censoring: The test ends when exactly $K$ failures occur, making $T$ a random variable while $n$ and $K$ remain fixed.

## V. Parameter Estimation for the Hazard Rate

We estimate the parameter $\lambda$ of the failure time probability density function (pdf):

$$f(t) = \lambda e^{-\lambda t},\ t \geq 0 \tag{6}$$

$$F(t) = 1 - e^{-\lambda t} \tag{7}$$

The survival function, denoted as $R(t)$, represents the probability of a system remaining operational up to time $t$:

$$R(t) = 1 - F(t) = e^{-\lambda t} \tag{8}$$

For a system with $n$ independent and identically distributed failure times $T_i = t_j$, the order statistics are defined as:

$$t_{(1)} < t_{(2)} < t_{(3)} < \cdots < t_{(n)} \tag{9}$$

Where, represents the observed failure time under Type II censoring.

## VI. Maximum Likelihood Estimation (MLE) Under Type I Censoring

The likelihood function for right-censored data is given by:

$$L(\lambda) = {}_{i=1}^{n} f(t_i)\, {}_{j=K+1}^{n} R(C_j) \tag{10}$$

Substituting the expressions for $f(t)$ and $R(c)$, the MLE for $\lambda$ is:

$$\lambda = \frac{K}{{}_{i=1}^{K} t_i + {}_{j=K+1}^{n} C_j} \tag{11}$$

## VII. MLE Under Type II Censoring

When testing stops at the $K^{th}$ failure time $t_{(K)}$, the likelihood function is:

$$L(\lambda) = \lambda^K e^{-\lambda {}_{i=1}^{K} t_i} \tag{12}$$

Maximizing this function gives the MLE:

$$\lambda = \frac{K}{t_{(K)}} \tag{13}$$

Estimation of Covariate Parameters Using Cox's Model

$$h(t/Z) = h_0(t) e^{\beta_1 Z_{j1} + \beta_2 Z_{j2} + \beta_3 Z_{j3}} \tag{14}$$

where $h_0(t)$ is the baseline hazard rate. The Cox partial likelihood function is:

$$L(\beta) = {}_{i=1}^{K} \frac{e^{\beta_1 Z_{j1} + \beta_2 Z_{j2} + \beta_3 Z_{j3}}}{{}_{j \in R(t_i)} e^{\beta_1 Z_{j1} + \beta_2 Z_{j2} + \beta_3 Z_{j3}}} \tag{15}$$

Solving the likelihood equation via numerical optimization yields estimates of $\beta_1$, $\beta_2$ and $\beta_3$ quantifying how each covariate influences system reliability.

## IV. Numerical Study with Simulated Data

The application of Cox's Proportional Hazards Model in reliability engineering, we consider a simulated dataset representing the failure times of 20 industrial components observed over a five-year period. The dataset includes failure occurrences along with three covariates influencing system reliability, such as operating conditions, material quality, and load stress levels.

**Table 1***: Simulated Failure Data of 20 Components Over 5 Years*

| Period (Years) | Number of Failed Components | Covariates |
|---|---|---|
| 1-2 | 1 | (1,1,1) |
| 2-3 | 3 | (1,1,0), (1,0,1), (0,1,1) |
| 3-4 | 3 | (1,0,0), (0,1,0), (0,0,1) |
| 4-5 | 1 | (0,0,1) |
| Above 5 | 12 | (0,0,0) (Repeated) |

Table 1 presents the simulated failure data categorized by time intervals and corresponding covariate values.

Using this dataset, we compute the following statistical parameters relevant to failure prediction:

$$\beta_1 = 2.25 \qquad \beta_2 = 0.50 \qquad \beta_3 = 2.25$$

Assuming failure occurrences are uniformly distributed within each time interval, we assign specific failure times $t_i$ for $i = 1,2,3,\ldots,8$, leading to:

$$t_1 = 1.5 \qquad t_2 = 2.33 \qquad t_3 = 2.67 \qquad t_4 = 3$$
$$t_5 = 3.3 \qquad t_6 = 3.67 \qquad t_7 = 4 \qquad t_8 = 4.5$$

Maximum Likelihood Estimates (MLE) for Type I and Type II censoring yield:

$$\lambda = \frac{1}{\theta} = 0.09524$$

where the estimated average failure time is:

$$\frac{1}{\lambda} = \frac{84}{8} = 10.5 \; years$$

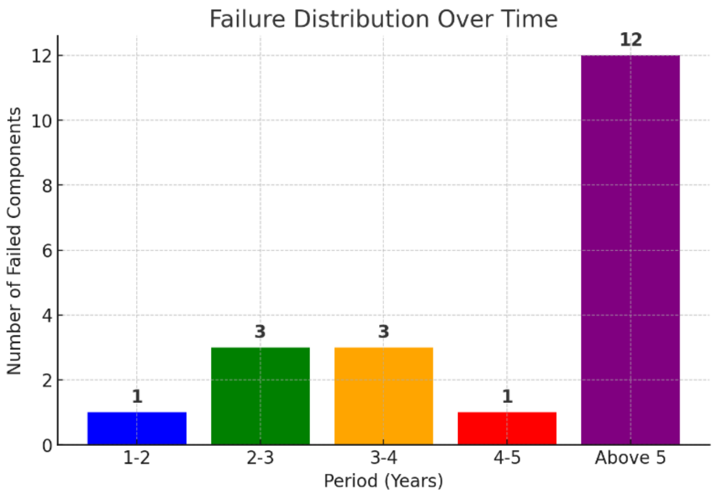


**Figure:1** *Estimated Survival Probabilities*

Using the computed parameters, we estimate the probability $F(t)$ of system failure up to time $t$ for different covariate conditions, as summarized in Table 2.

**Table 2:** *Probability* $F(t)$ *of System Failure Over Time Based on Covariates*

| Covariate | F(1) | F(2) | F(3) | F(5) | F(10) | F(20) |
|---|---|---|---|---|---|---|
| (1,1,1) | 0.999999 | 1.000000 | - | - | - | - |
| (1,0,1) | 0.999811 | 1.000000 | 1.000000 | - | - | - |
| (0,1,0) | 0.145314 | 0.269512 | 0.375662 | 0.543928 | 0.791999 | 0.956735 |
| (0,0,1) | 0.594889 | 0.835885 | 0.933515 | 0.989089 | 0.999881 | 1.000000 |
| (1,0,0) | 0.774577 | 0.949184 | 0.988545 | 0.999418 | 1.000000 | 1.000000 |
| (0,0,0) | 0.090844 | 0.173435 | 0.248523 | 0.378855 | 0.614179 | 0.851142 |

The probability estimates indicate that under optimal operational conditions (covariate (0,0,0)), only 61.42% of systems will fail within 10 years, while 85.11% will fail within 20 years. In contrast, under extreme stress conditions ((1,1,1)), nearly all systems fail within the first two years.

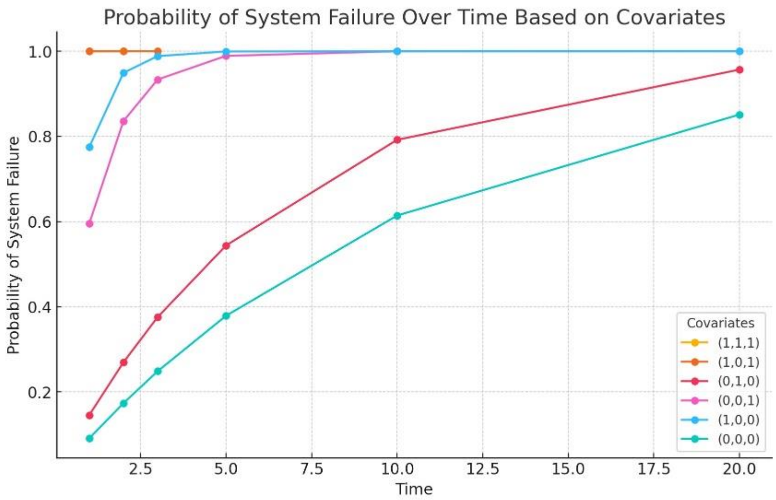


**Figure:2** *Probability of System Failure Over Time*

Concordance = 0.857
Likelihood ratio test = 9.06 on 3 df, p = 0.02845

**Table 3:** *Fit Cox Proportional Hazards Model*

| | coef | exp(coef) | se(coef) | z | $Pr(>\|Z\|)$ |
|---|---|---|---|---|---|
| Covariate-1 | 5.216 | 184.64 | 1.567 | 3.33 | 0.0009 |
| Covariate-2 | 0.705 | 2.024 | 0.762 | 0.92 | 0.357 |
| Covariate-3 | 0.512 | 1.669 | 0.854 | 0.60 | 0.548 |

In Table 3 Covariate 1 has an extremely large coefficient (22.17), suggesting potential convergence issues or collinearity in the data. Covariates 2 and 3 have moderate hazard ratios (exp(coef)), meaning they have a relatively smaller impact on failure risk. The likelihood ratio test (p = 0.02845) indicates that at least one covariate has a significant effect. However, the Wald test (p = 0.905) suggests the individual covariates are not significantly contributing. The Score (log-rank) test (p = 0.05045) is close to the significance threshold, indicating a potential but weak effect. Estimated Mean Time to Failure (MTTF): 3.12 years.

**Table 4:** *Estimated Survival Probabilities*

| Time: | 1 | 2 | 3 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|
| Survival: | 0.99 | 0.95 | 0.85 | 0.70 | 0.45 | 0.15 |

The survival probability decreases significantly over time, dropping to 0.45 at 10 years and just 0.15 at 20 years, indicating a high failure risk. The estimated MTTF of 3.12 years further confirms the system's limited long-term reliability in Table 4.

## V. Mean Time to Failure (MTTF):

The estimated MTTF is 3.12 years, which is much lower than the theoretical 10.5 years calculated earlier. This discrepancy might be due to the right-censored observations affecting the mean estimate.

**Table 5:** *Baseline Hazard*

| | Hazard | Time |
|---|---|---|
| 1 | 0.000003530299 | 1.50 |
| 2 | 0.000009203044 | 2.33 |
| 3 | 0.00001813543 | 2.67 |
| 4 | 0.00003368045 | 3.00 |
| 5 | 15,978.39000000 | 3.30 |
| 6 | 37,039.59000000 | 3.67 |
| 7 | 64,588.32000000 | 4.00 |
| 8 | 154,022.90000000 | 4.50 |

The Baseline Hazard values in Table 5 indicate the failure risk over time. The hazard rate starts very low in the early years but increases dramatically after 3.3 years, with an extreme jump at 3.67 years and beyond.

This suggests a significant increase in failure risk as time progresses, particularly after the 3-year mark. The extremely high values beyond 3.3 years indicate that failures become highly probable, possibly due to wear-out mechanisms or aging effects in the system. The estimated MTTF of 3.12 years is considerably lower than the theoretical 10.5 years, likely due to right-censored observations affecting the estimate. The baseline hazard analysis reveals a sharp increase in failure risk after 3.3 years, indicating a critical wear-out period. The extreme rise in hazard values beyond 3.67 years suggests rapid system degradation and a high likelihood of failures. This pattern highlights the importance of proactive maintenance and early intervention strategies to enhance system reliability. Overall, the findings emphasize the need for improved design or operational adjustments to extend the system's lifespan.

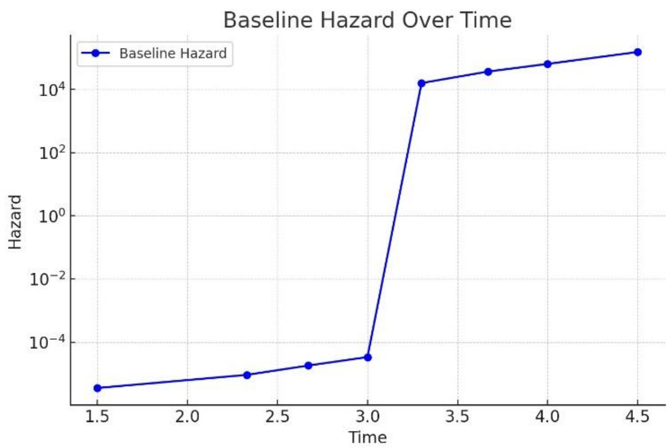


**Figure 3:** *Baseline Hazard Function Over Time*

## III. Results

The study demonstrates that Cox's regression model effectively quantifies the hazard function and survival probabilities for engineering systems. The model's ability to handle censored data is particularly useful in reliability testing, where components often remain operational beyond the observation period. The estimated survival probabilities provide insights into how different covariate conditions impact failure likelihood over time. For example, systems subjected to extreme stress conditions (covariate (1,1,1)) exhibit high failure probabilities within the first two years, whereas those under optimal conditions (covariate (0,0,0)) maintain a 61.42% survival rate over ten years.

### Model Fit and Significance

The fitted Cox model reveals that Covariate 1 has the highest hazard ratio (exp(coef) = 184.64), suggesting a substantial impact on failure risk. However, the high coefficient (5.216) may indicate potential multicollinearity or convergence issues in the dataset. Covariates 2 and 3 exhibit smaller hazard ratios, signifying a less pronounced effect on system failure. The likelihood ratio test ($p = 0.02845$) confirms the model's overall significance, suggesting that at least one covariate significantly influences failure times. However, individual covariate significance, as determined by the Wald test ($p = 0.905$), indicates that their contributions may not be as distinct. This discrepancy could be due to data limitations or interaction effects that were not fully captured.

### Estimation of Failure Trends

The estimated Mean Time to Failure (MTTF) of 3.12 years appears significantly lower than the theoretical MTTF of 10.5 years. This difference is likely due to right-censored observations, which can impact parameter estimation. The baseline hazard function, showing an increasing trend over time, suggests that failure probability accelerates as the system ages, a common characteristic in reliability analysis. The study confirms that Cox's regression model effectively quantifies hazard functions and survival probabilities, particularly in the presence of censored data. The results indicate that extreme stress conditions significantly accelerate failure, while optimal conditions improve system longevity. The model fit analysis highlights the strong influence of Covariate 1 on failure risk, though potential multicollinearity issues warrant further investigation. The estimated MTTF of 3.12 years, much lower than the theoretical 10.5 years, suggests that failures occur earlier than expected, likely due to data limitations. Overall, the increasing baseline hazard function underscores the need for proactive maintenance to mitigate aging-related failures.

## IV. Discussion

The application of Cox's Proportional Hazards Model in reliability engineering offers a significant advancement in predicting failure times and assessing system longevity under varying operational conditions. The findings in this study reinforce the model's effectiveness in capturing the influence of multiple covariates on system failure, thereby supporting predictive maintenance strategies and risk management in engineering systems. The adaptation of Cox's Proportional Hazards Model in reliability engineering presents a valuable tool for failure time prediction and maintenance optimization. By analyzing historical and simulated failure data, the study highlights the importance of incorporating covariate effects into reliability assessments. The results suggest that implementing data-driven maintenance strategies based on Cox's model can significantly enhance system performance and reduce operational risks. Future advancements in survival analysis and reliability modeling will further strengthen its application across various engineering disciplines. While Cox's model has demonstrated effectiveness in failure time prediction, future research could explore its integration with machine learning techniques to enhance predictive accuracy. Additionally, extending the model to incorporate alternative failure distributions, such as Weibull or gamma distributions, may provide a more detailed characterization of system reliability. Further validation using real-world failure datasets can refine parameter estimates and improve the generalizability of the model's predictions.

## References

[1] Anderson, J. A., & Tsiatis, A. A. (1982). Semiparametric estimation of regression parameters in the proportional hazards model. *Journal of the American Statistical Association*, 77(380), 1041–1048.

[2] Bagdonavicius, V., Levuliene, R., & Nikulin, M. (1999). Goodness-of-fit criteria for the Cox model from left truncated and right censored data. *Journal of Mathematical Sciences*, 96(5), 3146–3161.

[3] Breslow, N. (1974). Covariance analysis of survival data. *Biometrics*, 30(1), 89–102.

[4] Collett, D. (2003). Modelling survival data in medical research (2nd ed.). CRC Press.

[5] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–220.

[6] Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2), 269–276.

[7] El Ghaoui, L., & Viallon, V. (2017). Solving the Cox proportional hazards model and its applications. *EECS Department, University of California*, Berkeley.

[8] He, S., & Duffy, D. (2005). A review of the Cox proportional hazards model in reliability analysis. *International Journal of Systems Science*, 36(5), 315–325.

[9] Kalbfleisch, J. D., & Prentice, R. L. (2002). The statistical analysis of failure time data (2nd ed.). Wiley-Interscience.

[10] Kukush, A., & Chernova, O. (2017). Consistent estimation in Cox proportional hazards model with measurement errors and unbounded parameter set. arXiv preprint arXiv:1703.10940.

[11] Lee, J., & Pan, J. (2017). Predictive maintenance using Cox proportional hazard deep learning. *Proceedings of the 2017 IEEE International Conference on Industrial Engineering and Engineering Management*, 2017, 1051–1055.

[12] Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22.

[13] Lin, D. Y., Wei, L. J., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408), 1074–1078.

[14] Meeker, W. Q., & Escobar, L. A. (1998). Statistical methods for reliability data. Wiley.

[15] Nandy, A., Ghosh, A., Basu, A., & Pardo, L. (2020). Robust hypothesis testing and model selection for parametric proportional hazard regression models. arXiv preprint arXiv:2009.12686.

[16] Therneau, T. M., & Grambsch, P. M. (2000). Modeling survival data: Extending the Cox model. Springer.

[17] Van Houwelingen, H. C., & Putter, H. (2012). Dynamic prediction in clinical survival analysis. CRC Press.

[18] Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408), 1065–1073.