Estimating Confidence Intervals for Hazard Ratio with Composite Covariates in the Cox Models

Shofi Andari[1*]
[1]Department of Statistics, Faculty of Science and Data Analytics, ITS Surabaya Indonesia
*Corresponding author: sandari@its.ac.id

Received: 16 March 2025
Revised: 6 May 2025
Accepted: 1 July 2025

ABSTRACT – **Hazard ratio (HR) estimation is fundamental in survival analysis, particularly in Cox proportional hazards models, where covariates influence time-to-event outcomes. When covariates are combined into composite variables, constructing confidence intervals (CIs) for the resulting HRs becomes challenging due to potential multicollinearity, interaction effects, and violations of the proportional hazards assumption. This paper presents a systematic approach for constructing confidence intervals for HRs associated with composite covariates, comparing standard methods such as the Wald, likelihood ratio, and bootstrap-based intervals. Through simulation studies for different scenarios of Cox regression models, we evaluate the performance of these methods in terms of bias, coverage probability, and robustness under various data conditions. The findings of this study provide practical recommendations for researchers dealing with composite covariates in survival analysis, ensuring reliable inference in epidemiological and clinical studies.**

Keywords – **Composite Covariates, Confidence Intervals , Hazard Ratio, Survival Regressions**

I. INTRODUCTION

Survival analysis is a fundamental statistical framework used to model time-to-event data, with applications in fields such as medicine, epidemiology, and engineering reliability. The Cox proportional hazards model [1] remains the most widely used method for estimating hazard ratios (HRs) in the presence of covariates, allowing researchers to assess the relative risk of an event occurring across different groups. Typically, comparisons are made between two groups: one that receives the treatment and another that serves as the control group without treatment [2]. A critical aspect of survival analysis is the construction of confidence intervals (CIs) for HR estimates, which quantifies the uncertainty associated with these estimates and ensures reliable inference.

In many applications, researchers construct composite covariates, which are derived by combining multiple predictor variables into a single measure [3]. Composite covariates arise in various contexts, such as principal component scores in dimension reduction [4], index scores in clinical risk assessments[5], and interaction terms capturing joint effects of multiple exposures [6]. Despite their frequent use, the theoretical properties of confidence interval estimation for hazard ratios involving composite covariates remain underexplored. The complexity arises due to increased correlation among predictors, scaling effects, and potential violations of the proportional hazards assumption when composite measures summarize multiple risk factors.

Existing methods for constructing confidence intervals in Cox models include the Wald interval, likelihood ratio-based interval, and score-based interval [7]. The Wald method, relying on asymptotic normality, is widely used but may perform poorly in small samples or when covariate effects are weak. Likelihood ratio and score-based intervals, while often more robust, require additional computational effort and may be sensitive to model specification. The presence of composite covariates further complicates inference by altering the distributional properties of parameter estimates, potentially leading to coverage distortions in standard interval estimation techniques.

This paper aims to fill this gap by providing a rigorous theoretical framework for confidence interval estimation in Cox proportional hazards models with composite covariates. We systematically examine the asymptotic properties of different interval estimation methods, derive conditions under which they achieve nominal coverage, and explore their robustness under various data-generating mechanisms. Through theoretical derivations and simulation studies, we establish guidelines for selecting appropriate confidence interval methods when analyzing survival data involving composite covariates.

The remainder of this paper is organized as follows. Section 2 reviews the mathematical foundation of the Cox proportional hazards model and introduces the concept of composite covariates. Section 3 outlines existing confidence interval estimation methods and their theoretical properties. Section 4 presents new results on confidence interval construction for composite covariates, while Section 5 provides simulation studies to assess performance under various scenarios. Finally, Section 6 discusses implications for applied research and directions for future work.

II. LITERATURE REVIEW

A. Survival Analysis and Hazard Ratios

Survival analysis is widely used in medical research, epidemiology, and engineering to model time-to-event data, where the hazard function represents the instantaneous risk of an event occurring at a given time point [1]. The Cox proportional hazards model is one of the most used statistical methods in this field, estimating the relationship between explanatory variables (covariates) and the hazard function through the hazard ratio (HR).

In the context of survival analysis, researchers are often interested in understanding how certain covariates affect the time until an event occurs, such as death, disease progression, or equipment failure. One of the core quantities used to express this relationship is the hazard function, defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\,(t \leq T < t \mid T \geq t)}{\Delta t}, \tag{1}$$

where $h(t)$ represents the instantaneous risk of an event occurring at time t, given that the individual has survived up to that time.

To assess the effect of covariates on the hazard function, the Cox proportional hazards model is widely applied. This semi-parametric model assumes the hazard function can be expressed as hazard model (Equation 2).

$$h(t \mid X) = h_0(t)\exp\,(\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p), \tag{2}$$

with $h_0(t)$ is the baseline hazard function and $X_1, X_2, \ldots, X_p$ are covariates with corresponding regression coefficients $\beta_1, \beta_2, \ldots, \beta_p$.

B. Estimating Hazard Ratio

A hazard ratio (HR) quantifies the relative risk of an event occurring in one group compared to another [8]. It is particularly useful in clinical trials, where researchers compare treatment and control groups over time. The estimation of confidence intervals (CIs) for hazard ratios is essential for drawing reliable inferences about the effect of covariates on survival outcomes [9]. HR compares the hazard rates between two groups differing by one unit in a covariate X_j, where $j \in \{1, 2, \ldots, p\}$, all else being equal. Hence, the hazard ratio for one-unit increase in X_j is

$$HR = \exp(\beta_j). \tag{3}$$

An HR greater than one (> 1) implies an increased risk, less than one implies a decreased risk, while HR equals to one suggests no difference in risk between the groups.

C. Composite Covariates in Cox Regression Models

Composite covariates refer to variables that combine multiple predictors into a single measure. Examples include principal component scores, interaction terms, or latent variable constructs [10]. In survival analysis, composite covariates are used to reduce dimensionality by combining several covariates into one reduces the number of parameters in the Cox model, preventing overfitting [11]. Composite covariates also improve interpretability as a well-designed composite covariate provides a more meaningful summary of risk factors, as seen in clinical scores [12]. Moreover, this strategy is effective for multicollinearity mitigation as composite covariates help manage correlation among predictors, ensuring better coefficient stability [13].

When composite covariates are introduced, we define

$$C = g(X_1, X_2, \ldots, X_p), \tag{4}$$

where $g(\cdot)$ is a function that combines multiple predictors into a single covariate. The model in Equation (2) is then simplified to

$$h(t \mid C) = h_0(t)\exp(\gamma C), \tag{5}$$

with γ is the estimated coefficient for the composite covariate.

However, incorporating composite covariates into Cox regression models presents several statistical challenges, such as the distributional properties of the composite variable may differ from those of standard covariates; the scaling effect can influence the variance of $\hat{\beta}$, affecting confidence interval estimation; the correlation structure among the components of the composite covariate may violate the proportional hazards assumption [9].

D. Existing Approaches to Construct Confidence Intervals with Composite Covariates

The literature on constructing confidence intervals for composite covariates in survival models is limited. Existing studies suggest several approaches:

1) Wald confidence interval

The Wald confidence interval assumes that the estimated regression coefficient $\hat{\beta}$ follows an approximately normal distribution. Given an estimated coefficient and its standard error. The Wald CI is constructed as Equation (4), which gives the confident interval for the hazard ratio. This method is widely used due to its simplicity but has limitations in small sample sizes or when the normality assumption is violated [9] Some researchers propose adjusted standard errors to correct for dependency structures within composite covariates. However, these corrections rely on asymptotic approximations and may still lead to biased results in small samples [14].

The point estimate $\hat{\beta}_j$ of the log-hazard ratio us typically obtained via partial likelihood estimation. Under regularity conditions, $\hat{\beta}_j$ is asymptotically normal, allowing the construction of confidence intervals and hypothesis tests [1], [9]. The standard error of $\hat{\beta}_j$ plays a key role in quantifying uncertainty, especially when driving the confidence interval for HR, formulated as the following

$$CI_{Wald} = \left(\exp\left(\hat{\beta}_j - z_{\frac{\alpha}{2}}.SE(\hat{\beta}_j)\right), \exp\left(\hat{\beta}_j + z_{\frac{\alpha}{2}}.SE(\hat{\beta}_j)\right)\right). \tag{6}$$

Accurate estimation of both the point estimate and its confidence interval is essential for valid inference and decision-making [15], especially in high-stakes domains like healthcare and public policy.

2) Likelihood profiling methods

An alternative and often preferred method is the likelihood ratio (LR) confidence interval, which is constructed using the profile likelihood function. It relies on comparing the log-likelihood values of the two models, namely the *full* model, includes the composite covariates or parameter of interest and the *reduced* model, which excludes the parameter of interest. The likelihood function for a Cox proportional hazards model is derived from the partial likelihood, which focuses on the ordering of event times rather than the full likelihood. Given survival times $(t_1, t_2, \ldots, t_n)$ and covariates X_i, for n observations, the partial likelihood is

$$L(\beta) = \prod_{f\in D} \frac{\exp\,(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_f)}{\sum_{k\in R_f} \exp\,(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_k)}, \tag{7}$$

where D is the set of individuals experiencing the event and R_f is the risk set at time t_f, i.e. individuals still at risk just before t_f. Then taking the log-likelihood function, we obtain

$$\ell(\beta) = \sum_{f\in D}\left(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_f - \log\sum_{k\in R_f} \exp\,(\boldsymbol{\beta}^{\mathrm{T}}X_k)\right). \tag{8}$$

The likelihood ratio test statistic takes place by comparing the two log-likelihood of full and reduced model.

$$\Lambda = -2(\ell(\boldsymbol{\beta})_{reduced} - \ell(\boldsymbol{\beta})_{full}) \tag{9}$$

Under the null hypothesis H_0, that the composite covariate has no effect. The test statistic Λ follows a chi-square distribution with degrees of freedom (df) equal to the difference in model parameters, that is $\Lambda \sim \chi^2_{df}$ [7]. A likelihood ratio-based confidence interval for the hazard ratio (HR) is obtained by solving

$$-2(\ell(\boldsymbol{\beta}^*) - \ell(\boldsymbol{\beta})_{max}) = \chi^2_{1,1-\alpha}. \tag{10}$$

So, for a 95% CI, the critical values from the ci-square distribution with 1 degree of freedom is [16] and $\ell(\beta^*)$ would give two values of upper β_U and lower β_L bound in the interval [16]Hence, the CI for HR [16] is

$$CI_{LR} = (\exp(\boldsymbol{\beta}_L), \exp(\boldsymbol{\beta}_U)). \tag{11}$$

III. METHODOLOGY

A. Data Simulation Scenarios

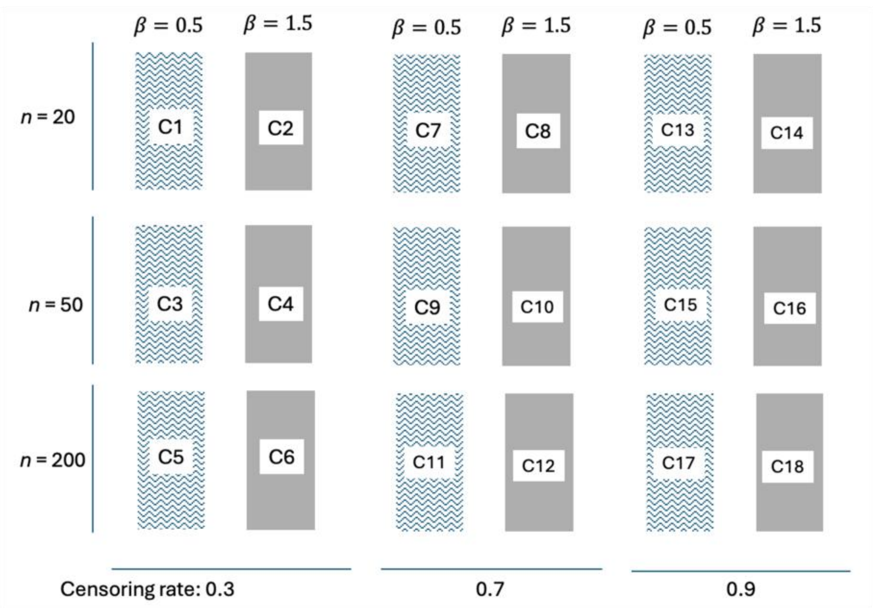

Figure 1 Simulation designs for different scenarios

To evaluate the performance of confidence interval (CI) estimation for hazard ratios (HR), we simulate survival data under different conditions. The simulation aspects include:

1) Sample size variation

Here, we analyze three different sample sizes to assess how small vs. large samples affect the precision and reliability of HR estimates. Small sample size is evaluated at $n = 20$. Medium sample is at $n = 50$ and large sample size is at $n = 200$.

2) Effect size variation

We vary the log-HR (β) to see how different strengths of the composite covariate influence CI estimation. A moderate effect at $\beta = 0.5$ represents a weak association between the composite covariate and survival outcome. While a strong effect at $\beta = 1.5$ represents a stronger relationship, expected to yield more precise HR estimates.

3) Censoring rate variation

Censoring in survival analysis affects the number of observed events, which impacts the reliability of HR estimation. There are three levels of censoring considered in this study: low censoring, where 30% are censored; moderate censoring (70%); and high censoring (90%).

Therefore, we have 18 simulated data sets. On top of that, continuous and categorical composite covariates are introduced for data simulation assuming linear combination for both covariates, with the following composition:

$$C = 0.5Z_1 + 0.3Z_2, \quad (12)$$

where $Z_1 \sim N(0,1)$ and $Z_2 \sim Binomial\ (n, 0.5)$. The weights 0.5 and 0.3 were chosen arbitrarily to create a dominant effect from Z_1, making the composite covariate more heterogeneous. This approach is to ensure correlation. Instead of treating covariates independently, the composite covariate combines information from different sources. This also is to mimic many risk scores or predictive models use linear combinations of multiple variables. By making the hazard model more complex, we can observe how Wald and LR methods behave differently.

B. Methodology and Steps of Analysis

The analysis carried out in this study consists of:

1) Data simulation based on the scenario illustrated in Figure 1.
2) Model estimation to fit Cox proportional hazards model to estimate β and obtain hazard ratio.
3) Confidence interval calculation using Wald and likelihood-ratio methods.

IV. RESULTS AND DISCUSSIONS

In this study, we evaluated the performance of Wald and Likelihood Ratio (LR) confidence intervals (CIs) for hazard ratio (HR) estimation in Cox proportional hazards models with composite covariates. The simulations were designed to assess the influence of sample size (n = 20, 50, 200), effect size (β = 0.5, 1.5), and censoring rates (low to high) on the accuracy and reliability of these interval estimation methods. The findings highlight critical distinctions between the two approaches, particularly in small sample settings and under high censoring conditions.

The results indicate that sample size plays a crucial role in determining the accuracy of CI estimation. In small samples (n = 20), both methods exhibited high variability, but Wald intervals tended to be more unstable, occasionally producing excessively narrow or wide intervals. The LR method provided more consistent results, as it does not rely on large-sample normality assumptions. At moderate sample sizes (n = 50), the difference between Wald and LR methods was less pronounced, though Wald's tendency to underestimate variance was still observed in some cases. When the sample size increased to n = 200, both methods converged, producing nearly identical results. This suggests that in large samples, the normal approximation underlying the Wald method becomes more reliable, and the likelihood surface is well-approximated, leading to similar interval coverage.

In this study, Wald intervals do not shown as be unstable in small samples due to reliance on asymptotic properties. The likelihood ratio method, which does not assume normality but rather constructs confidence bounds by profiling the likelihood function, provides a more reliable alternative in cases where sample size is limited.

The magnitude of the effect size had a significant impact on CI stability. When β = 0.5, corresponding to a weaker association between covariates and the hazard function, confidence intervals were generally wider across all methods. The reduced distinguishability between treatment groups contributed to greater variability in HR estimates, with Wald intervals occasionally failing to cover the true HR due to underestimated uncertainty in smaller samples.

Conversely, with β = 1.5, the effect of covariates on hazard became more pronounced, leading to narrower and more precise confidence intervals. The greater separation between risk groups resulted in more stable HR estimates, and both methods performed comparably when sample sizes were large. However, in smaller samples, Wald CIs still showed more variation, reinforcing the observation that LR confidence intervals may provide better control over interval coverage in cases where effect sizes are moderate to small.

These results are consistent with theoretical expectations: when the effect size is large, the information content in the likelihood function increases, making standard error estimation more stable. However, for small effect sizes, the assumptions underlying the Wald method (normality of the log-HR estimates) are more likely to be violated, leading to potential misestimation of confidence bounds.

One of the most striking differences between the two methods was observed under high censoring rates (≥70%). While both methods performed adequately under low (≤30%) censoring, Wald intervals became significantly narrower as censoring increased, often failing to adequately reflect the uncertainty in HR estimates. This is particularly concerning, as artificially narrow confidence intervals may lead to overconfidence in conclusions drawn from the model.

In contrast, the LR method provided more stable intervals across different censoring rates. This robustness arises from the fact that likelihood-based methods naturally account for data sparsity by profiling the likelihood function, rather than relying on asymptotic standard errors. In extreme censoring conditions (e.g., 90% censored), Wald intervals were sometimes unrealistically narrow, while LR intervals, though still affected, provided a more appropriate representation of variability.

Previous research has similarly noted that Wald-based standard errors tend to be underestimated in highly censored datasets, leading to misleadingly precise estimates. The likelihood ratio approach, which adjusts for the shape of the likelihood function rather than assuming symmetry, offers a more reliable alternative when censoring is substantial.

The inclusion of composite covariates added an additional layer of complexity to the analysis. Unlike simple binary or continuous covariates, composite covariates represent a weighted combination of multiple predictors, often used in risk scores or prognostic models. The simulation results showed that as complexity increased, the likelihood function became more irregular, particularly in small samples, making the Wald approximation more susceptible to bias.

For composite covariates, the LR method demonstrated greater stability across different conditions. This is likely due to the fact that profiling the likelihood function allows better adaptation to non-linear relationships and interactions between covariates. The Wald method, in contrast, assumes a fixed variance structure, which may not hold when composite covariates introduce additional heterogeneity in the data.

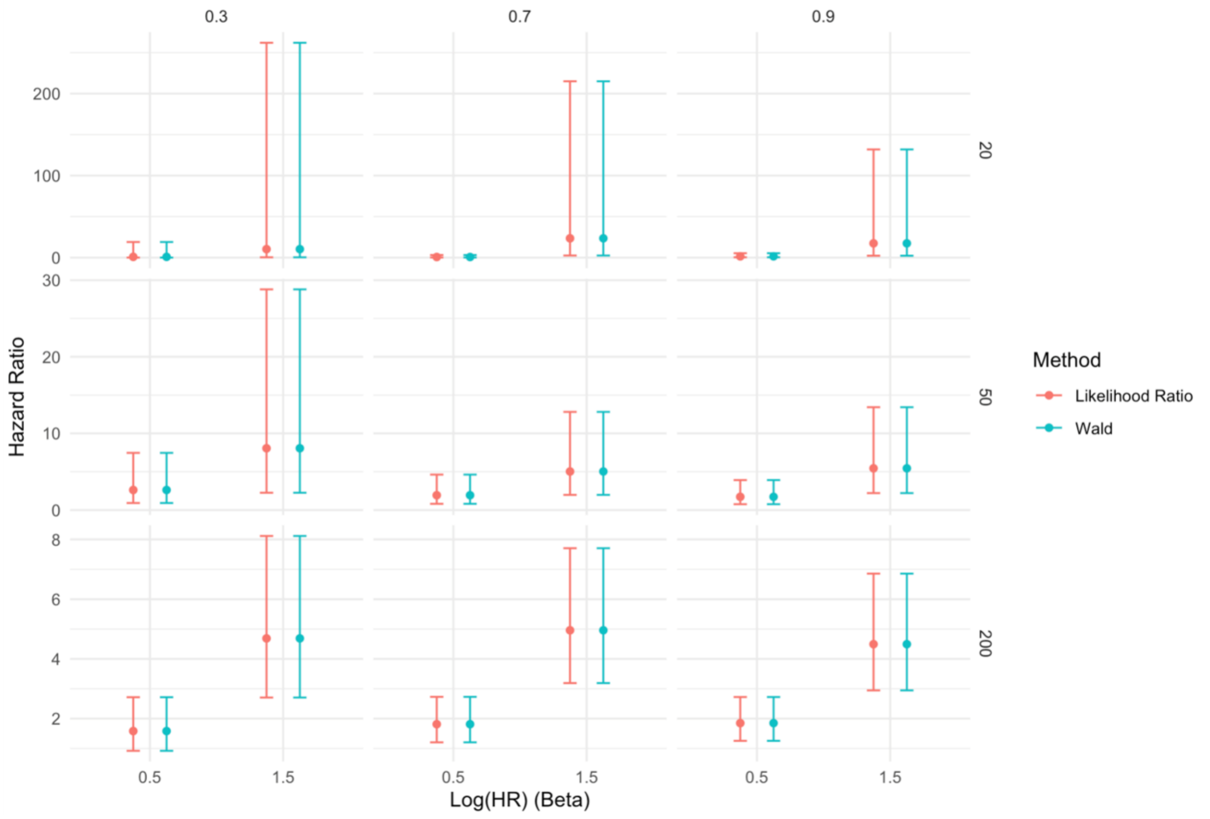

Figure 2 Confidence intervals based on Wald and likelihood-ratio (LR) methods for different modeling scenarios and sample sizes

To evaluate whether confidence intervals reliably covered the true parameter, we calculated the empirical coverage probability: the proportion of the 1000 simulation replicates in each scenario for which the constructed CI contained the true value of $\boldsymbol{\beta}$, specifically to be the number of CIs containing true $\boldsymbol{\beta}$ over the number of replicates.

Table 1 Empirical coverage of CIs

Sample size	Effect size (β)	Empirical coverage
20	0.5	0.96
	1.5	0.98
50	0.5	0.98
	1.5	0.94
200	0.5	0.92
	1.5	0.94

Table 1 presents the empirical coverage probabilities of the constructed confidence intervals (CIs) for hazard ratio estimates across different combinations of sample size and effect size. Coverage was calculated as the proportion of simulated confidence intervals containing the true effect value (β), based on 1,000 replicates per scenario. These results provide insight into the accuracy and reliability of the interval estimation methods under varying data conditions. The findings indicate that empirical coverage varies slightly across scenarios. For small samples (n = 20), empirical coverage slightly exceeded the nominal 95% level, suggesting conservative intervals that may overstate uncertainty. In contrast, for larger samples (n = 200), coverage occasionally fell below 0.95, particularly for weaker effects (β = 0.5), reflecting potential under-coverage and underestimated uncertainty. This pattern underscores the importance of considering both sample size and effect magnitude when interpreting confidence intervals, as standard asymptotic methods like Wald and likelihood-ratio intervals may yield conservative or anti-conservative results depending on data characteristics.

V. CONCLUSIONS AND SUGGESTIONS

These findings underscore the importance of adequate sample size in survival analysis, particularly when dealing with composite covariates and non-proportional hazards. While both the Wald and Likelihood Ratio (LR) methods are viable for constructing confidence intervals, their performance varies depending on the complexity of the model and available sample size. Future studies could explore alternative confidence interval estimation techniques and validate these findings using real-world data. Here, since Wald and LR methods appear to behave similarly, future studies could examine alternative CI construction techniques, such as bootstrap confidence interval using non-parametric which may flexible estimation of interval bounds, particularly in small samples. Future work might also include extending this work to non-proportional hazard models. This study assumes the proportional hazards (PH) assumption holds. However, in real-world applications, hazards may vary over time. Future research could explore time-varying coefficients in Cox models with flexible survival models that allow non-proportional hazard properties.

REFERENCES

[1] D. R. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society*, vol. 34, no. 2, pp. 187–220, 1972.

[2] D. G. Kleinbaum and M. Klein, *Statistics for Biology and Health Series Editors*, 3rd ed. Springer, 2012. [Online]. Available: http://www.springer.com/series/2848

[3] T. Emura, Y.-H. Chen, and H.-Y. Chen, "Survival Prediction Based on Compound Covariate under Cox Proportional Hazard Models," *PLoS One*, vol. 7, no. 10, 2012, doi: 10.1371/journal.pone.0047627.

[4] E. Bair, T. Hastie, D. Paul, and R. Tibshirani, "Prediction by supervised principal components," *J Am Stat Assoc*, vol. 101, no. 473, pp. 119–137, Mar. 2006, doi: 10.1198/016214505000000628.

[5] E. W. Steyerberg and Y. Vergouwe, "Towards better clinical prediction models: seven steps for development and an ABCD for validation," *Eur Heart J*, vol. 35, no. 29, pp. 1925–1931, Aug. 2014, doi: 10.1093/EURHEARTJ/EHU207.

[6] T. J. Van Der Weele and M. J. Knol, "A tutorial on interaction," *Epidemiol Methods*, vol. 3, no. 1, pp. 33–72, Dec. 2014, doi: 10.1515/em-2013-0005.

[7] D. Collett, *Modelling Survival Data in Medical Research*. Oxon: CRC Press, 2023. [Online]. Available: https://www.routledge.com/

[8] David. Hosmer, Susanne. May, and Stanley. Lemeshow, *Applied Survival Analysis : Regression Modeling of Time to Event Data, 2nd Edition*. Wiley-Interscience, 2008.

[9] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*. New Jersey: Wiley, 2002.

[10] T. M. Therneau and P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model*. New York: Springer, 2000.

[11] D. M. Witten and R. Tibshirani, "Survival analysis with high-dimensional covariates", doi: 10.1177/0962280209105024.

[12] M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie, "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation," *J Chronic Dis*, vol. 40, no. 5, pp. 373–383, 1987, doi: 10.1016/0021-9681(87)90171-8.

[13] C. F. Dormann *et al.*, "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance," *Ecography*, vol. 36, no. 1, pp. 27–46, 2013, doi: 10.1111/J.1600-0587.2012.07348.X.

[14] W. Pan and M. M. Wall, "Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations," *Stat Med*, vol. 21, no. 10, pp. 1429–1441, May 2002, doi: 10.1002/sim.1142.

[15] J. Meis, M. Pilz, B. Bokelmann, C. Herrmann, G. Rauch, and M. Kieser, "Point estimation, confidence intervals, and P-values for optimal adaptive two-stage designs with normal endpoints," *Stat Med*, vol. 43, no. 8, pp. 1577–1603, Apr. 2024, doi: 10.1002/sim.10020.

[16] M. Hollander, I. W. McKeague, and J. Yang, "Likelihood Ratio-Based Confidence Bands for Survival Functions," *J Am Stat Assoc*, vol. 92, no. 437, p. 215, Mar. 1997, doi: 10.2307/2291466.

© 2025 by the authors. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (http://creativecommons.org/licenses/by-sa/4.0/).