ORIGINAL RESEARCH ORİJİNAL ARAŞTIRMA

DOI: 10.5336/biostatic.2025-112965

Comparing Statistical Power of Cox-Based Survival Models Under Class Imbalance: Methodological Study

Sınıf Dengesizliği Altında Cox Tabanlı Sağkalım Modellerinin İstatistiksel Güç Düzeylerinin Karşılaştırılması: Metodolojik Çalışma

Merve BAŞOL GÖKSÜLÜK[a]

[a]Sakarya University Faculty of Medicine, Department of Biostatistics, Sakarya, Türkiye

ABSTRACT Objective: This study investigates the impact of class imbalance on the performance of Cox-based survival models, an important issue in clinical research where event rates (e.g., death or disease recurrence) are typically low. Unlike previous studies that apply resampling techniques to correct imbalance, we preserved the original data structure to evaluate model robustness under realistic conditions. **Material and Methods:** Six modeling approaches were compared: Cox proportional hazards model, the weighted Cox model, 3 regularized Cox models [least absolute shrinkage and selection operator (LASSO-Cox), Ridge-Cox, Elastic Net-Cox], and a Bayesian Cox model. Simulations were conducted across varying sample sizes (n= 50, 100, 250 ve 500) and imbalance ratios (r= 0.1, 0.2, 0.3, 0.4 and 0.5) to evaluate each model's statistical power and estimation accuracy. **Results**: The Bayesian Cox model consistently achieved the highest statistical power and estimation precision under conditions of severe imbalance and small sample sizes. However, its advantage diminished as sample size increased, with its power converging to that of the Cox model. Among regularized Cox models, Ridge-Cox regression demonstrated the most stable estimates, producing narrower confidence intervals than LASSO-Cox and Elastic Net-Cox. In contrast, the weighted Cox model consistently underperformed, showing lower power and unstable estimates across all scenarios. **Conclusion:** These findings emphasize the importance of selecting modeling strategies. In scenarios with few observed events, it is generally more effective to apply model-based adjustments rather than altering the original data distribution, which may distort event prevalence and compromise generalizability. The performance of Cox-based models improves as sample size increases; however, in small-sample, high-imbalance settings, the use of inherently robust models becomes more critical.

Keywords: Survival analysis; class imbalance; Cox models; simulation

ÖZET Amaç: Bu çalışma, ilgilenilen bir olayın (örneğin ölüm veya hastalığın nüks etmesi) görülme oranının genellikle düşük olduğu klinik araştırmalarda yaygın bir sorun olan grup dengesizliğinin Cox tabanlı sağkalım modellerinin performansı üzerindeki etkisini incelemektedir. Önceki çalışmaların aksine bu çalışmada, dengesizliği gidermek için yeniden örnekleme (resampling) yöntemleri uygulanmamış, bunun yerine orijinal veri yapısı korunarak modellerin gerçekçi koşullarda ne kadar dayanıklı olduğu değerlendirilmiştir. **Gereç ve Yöntemler:** Çalışmada, 6 Cox-tabanlı modelin performansı karşılaştırılmıştır: Cox orantılı hazard modeli, ağırlıklı (weighted) Cox model, Cezalandırılmış Cox regresyon modelleri [en az mutlak küçülme ve seçim operatörü (least absolute shrinkage and selection operator "LASSO-Cox"), Ridge-Cox ve Elastic Net-Cox] ve Bayesçi Cox modeli. Farklı örneklem büyüklüğü (n=50, 100, 250, 500) ve dengesizlik oranı (r=0.1, 0.2, 0.3, 0.4, 0.5) dikkate alınarak bir benzetim çalışması yapılmıştır. Elde edilen sonuçlar ile her modelin istatistiksel gücü ve tahmin doğruluğu değerlendirilmiştir. **Bulgular:** Bayesçi Cox modeli, özellikle ciddi dengesizlik ve küçük örneklem durumlarında, istatistiksel güç ve tahmin doğruluğu açısından en iyi performansı göstermiştir. Ancak örneklem büyüklüğü arttıkça, klasik Cox modeli ile benzer performanslar göstermeye başlamıştır. Cezalandırılmış Cox modeller arasında Ridge regresyonu en kararlı kestirimleri sağlamış ve LASSO ile Elastic Net-Cox ile karşılaştırıldığında daha dar güven aralıkları elde edilmiştir. Buna karşılık, ağırlıklı Cox modeli tüm senaryolarda en zayıf performansı gösteren yöntem olmuştur. **Sonuç:** Bu çalışma, modelleme stratejilerinin seçilmesinin önemini vurgulamaktadır. İlgili olayın daha az gözleme sahip olduğu durumda, orijinal veri dağılımını değiştirmek, olayın prevelansının değişmesine ve genellenebilirliğin azalmasına neden olacağı için bu durumda model tabanlı yöntemlerin kullanılması önerilir. Örneklem büyüklüğü arttıkça Cox tabanlı modellerin gücü artar, ancak, küçük gözleme ve yüksek dengesizlik oranına sahip verilerde sağlam modeller kullanılmalıdır.

Anahtar kelimeler: Sağkalım analizi; sınıf dengesizliği; Cox modeller; benzetim

TO CITE THIS ARTICLE:
Başol Göksülük M. Comparing statistical power of Cox-based survival models under class imbalance: Methodological study. Turkiye Klinikleri J Biostat. 2025;17(2):92-103.

Correspondence: Merve BAŞOL GÖKSÜLÜK
Sakarya University Faculty of Medicine, Department of Biostatistics, Sakarya, Türkiye
E-mail: mbasolgoksuluk@sakarya.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 23 Jun 2025 ***Received in revised form:*** 26 Aug 2025 ***Accepted:*** 26 Aug 2025 ***Available online:*** 08 Sep 2025

2146-8877 / Copyright © 2025 by Türkiye Klinikleri. This in an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Survival analysis encompasses a set of statistical methods used to examine the time until the occurrence of a specific event, such as death, disease progression, or recovery. In medical research, where outcomes often unfold over time and are not always fully observed within the study period, this approach is particularly valuable. Unlike standard regression models that assume complete outcome information, survival analysis is uniquely equipped to handle censored data-instances where the event has not yet occurred or remains unobserved.[1] This allows for more accurate modeling of real-world clinical scenarios and supports more reliable estimation of risks, treatment effects, and prognostic factors across varying follow-up durations.

The Cox proportional hazards (PH) regression model is a widely used statistical modeling approach in clinical research. It is particularly valued for its ability to identify prognostic biomarkers or estimate mortality risk in patients with life-threatening diseases, such as cancer.[2] Today, the Cox PH model remains the standard method for mortality risk estimation in most real-world applications.[3] Its semi-parametric nature, interpretability, and applicability to censored data make it a standard tool in survival modeling. Over time, however, with the increasing complexity of biomedical data and the growing demand for more accurate predictions, extended versions of the Cox model-such as regularized Cox models [least absolute shrinkage and selection operator (LASSO-Cox), Ridge-Cox, Elastic Net-Cox] and Bayesian formulations-have gained prominence.[4-8]

One major challenge in clinical survival analysis is class imbalance, where the number of censored cases substantially exceed the number of observed events. This often reflects high overall survival rates in many healthcare settings, resulting in relatively few patients experiencing the event of interest.[9] For instance, in acute coronary syndrome, one study reported that fewer than 2% of patients die within 90-days, resulting in a predominance of censored observations. Similarly, in a large-scale cohort of 7,606 coronavirus disease-2019 patients, only 1% experienced mortality.[9,10] In the presence of such class imbalance, the model is primarily influenced by the majority class, while the minority class (i.e., the event group) may be underrepresented. This may lead to unstable or imprecise estimates, inflated Type 1 and 2 error rates, and reduced statistical power.

There is an extensive body of research in the field of machine learning (ML) focused on addressing class imbalance. Building on these developments, researchers have begun adapting ML-based techniques to the analysis of imbalanced survival data to improve the predictive performance of trained models and to more effectively handle rare event scenarios.[3,11,12] To mitigate this problem, ML literature has proposed data-level interventions such as undersampling the majority class, oversampling the minority class, or applying the Synthetic Minority Over-sampling Technique (SMOTE).[13-15] These methods aim to balance the ratio of events to censored cases by incorporating aforementioned data processing procedures, thereby improving model performance. However, altering the natural data distribution may result in overfitting and introduces a fundamental limitation: the resulting survival probabilities may no longer represent the original population, thereby threatening the external validity of the findings.[16] Moreover, performance gains achieved in artificially balanced datasets may not generalize well to real-world, imbalanced scenarios, and may increase the risk of overfitting.

These approaches offer greater flexibility than traditional survival models by improving predictive performance and more effectively handling rare event scenarios.

Despite these concerns, most research to date has focused on manipulating the data rather than enhancing the model's ability to handle class imbalanced.[10,16-18] This raises a critical question: Is it possible to achieve accurate and reliable survival estimates under class imbalance by modifying model settings-such as weighting or regularization-without altering the original data distributions? If so, which strategies are most effective in such scenarios?

This study aims to fill this gap by systematically evaluating 6 Cox-based survival models across varying sample sizes and class imbalance ratios. These include the Cox PH model, a weighted Cox PH model using inverse probability of censoring weights (IPCW), three regularized Cox models (LASSO, Ridge, Elastic Net), and a Bayesian Cox regression model. Importantly, no resampling methods were applied; all analyses were conducted on the original, imbalanced datasets.

This study uses a simulation-based design in which both the total sample size and the degree of class imbalance are systematically varied. Sample sizes ranged from small (n=50) to highly large (n=500). Imbalance ratio was defined based on the proportion of individuals who experienced the event of interest, with the event group treated as the minority class. The imbalance ratio varied from a balanced scenario (r=0.5) to an extreme imbalanced 1 (r=0.1).

The primary goal of this study is to compare the statistical power of Cox-based models under different scenarios in estimating regression coefficients. To this end, 2 covariates-a binary variable (X1) and a continuous variable (X2)-were generated as explanatory variables in the models. Each model was assessed based on its statistical power and the width of the 95% confidence intervals for the regression coefficients, calculated as the distance between the upper and lower bounds.

MATERIAL AND METHODS

STUDY POPULATION

This study is a methodological simulation study and does not involve the inclusion of any human participants or patient data. Therefore, ethical approval was not required. Since primary objective of the study was to compare the performance of various adaptations of Cox models including weighted and regularized approaches across different sample sizes and effect sizes, no sample size or post-power calculation was performed.

STATISTICAL METHODS AND STUDY DESIGN

In survival analysis, the primary outcome is the time until the occurrence of a specified event. Let T_i denote the survival time, and C_i represent the censoring time for the i-th individual. The observed survival time is defined as $\widehat{T_i} = (T_i, C_i)$ and the event indicator is given by $\delta_i = I(T_i \leq C_i)$, where $\delta_i = 1$ indicates that the event occurred, and $\delta_i = 0$ indicates censoring.

Assuming continuous event times, the survival function at time ***t*** is defined as:

$$S(t) = P(T > t) = \int_t^{\infty} p(s)\, ds \tag{1}$$

The hazard function, which describes the instantaneous event rate at time ***t*** given survival up to that point, is expressed as:

$$h(t) = \lim_{\Delta \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}. \tag{2}$$

This study compared 6 different adaptations of the Cox model to assess their performance under conditions of class imbalance.

Cox PH REGRESSION MODEL

The Cox PH model is a semi-parametric method that estimates the effect of covariates on survival time.[1] Given the typically right-skewed distribution of survival times, the model uses a log transformation of the hazard function:

$$\log[h_i(t, X_i)] = h_0(t) + \sum_j \beta_j X_j \tag{3}$$

Here, $h_0(t)$ is the unspecified baseline hazard function, and β_j represents the regression coefficient for covariate X_j. Coefficients are estimated via partial likelihood, considering only those individuals who experience the event. Let ***F*** denote the set of failure times, then the partial likelihood is:

$$L(\beta) = \prod_{r\in F} \frac{\exp(\beta^T X_{i_r})}{\{\sum_{j\in R_r} \exp(\beta^T X_i)\}} \quad (4)$$

WEIGHTED Cox PH MODEL

To account for class imbalance, the IPCW method was used. IPCW adjusts for the potential bias by weighting observations based on the estimated probability of being uncensored up to the observed time.[19,20] These probabilities are estimated via a Cox PH model, and weights are defined as $w_i = 1 / G(\hat{T}_i)$. The IPCW-adjusted partial likelihood becomes:

$$L(\beta) = \prod_{r\in D} \frac{w_i \exp(\beta^T X_{i_r})}{\{\sum_{j\in R_r} w_i \exp(\beta^T X_i)\}} \quad (5)$$

REGULARIZED Cox REGRESSION MODELS

To enhance predictive accuracy and perform variable selection, regularization techniques were incorporated.

- **The LASSO** introduces a regularization technique that applies an L1 penalty to the regression coefficient within the framework of penalized partial likelihood. This approach shrinks some coefficients toward zero by minimizing the residual sum of squares under an L1-norm constraint, as shown in Equation (6).

$$\hat{\beta} = log\ L(\beta)\text{ , subject to } \sum\left|\beta_j\right| \leq s \quad (6)$$

Here, **s** is a positive constant term chosen by the researcher. Equation (6) is integrated into the partial likelihood function, used in the Cox PH model, as a penalty term with tuning λ,

$$\mathrm{L_{LASSO}}(\beta) = \operatorname{argmin}\left(\mathrm{L}(\beta) + \lambda \sum_{j=1}^{k}\left|\beta_j\right|\right) \quad (7)$$

- **Ridge** uses an L2 penalty, adding the squared sum of coefficients as a penalty term:[18,21]

$$\mathrm{L_{RIDGE}}(\beta) = \mathrm{L}(\beta) - \lambda \sum_{j=1}^{k} \beta_j^2 \quad (8)$$

- **Elastic Net** combines L1 and L2 penalties, enabling both variable selection and shrinkage:[18,21]

$$\mathrm{L_{ElasticNet}}(\beta) = \mathrm{L}(\beta) - \left(\lambda_1 \sum_{j=1}^{k}\left|\beta_j\right| + \lambda_2 \sum_{j=1}^{k} \beta_j^2\right) \quad (9)$$

BAYESIAN Cox MODEL

Unlike the Cox PH model, the Bayesian approach incorporates prior distributions for the model parameters. The posterior distribution is obtained by combining the likelihood with the prior using Bayes' theorem[7,8]

$$\mathrm{P}(\theta\,|D) = \frac{L(\theta|D)P(\theta)}{\int_{\theta} L(\theta|D)P(\theta)d\theta'} \quad (10)$$

Here, **D** represents the observed data, θ is the vector of unknown parameters, and ***L*(θ|D)** is the likelihood. Because posterior distributions often lack closed-form solutions, Markov Chain Monte Carlo methods-such as Gibbs sampling or Metropolis-Hastings sampling-are used to approximate them.[7]

SIMULATION SCENARIOS

To evaluate model performance under varying levels of class imbalance, a simulation study was conducted using the 6 survival models: Cox PH model, weighted Cox PH, regularized Cox models (LASSO-Cox, Ridge-Cox, Elastic Net-Cox), and the Bayesian Cox model.

A total of 1,000 datasets were generated for each simulation scenario. Survival times were simulated from an exponential distribution with a median of 24 months and a maximum follow-up of 120 months. The regression model included 2 covariates:

- **X_1**: a binary variable with equal probabilities (p=0.5),
- **X_2**: a continuous variable from a normal distribution with a mean of 40 and a standard deviation of 5.

Hazard ratios were set as $\boldsymbol{exp(\beta_1) = 2}$ for X_1 and $\boldsymbol{exp(\beta_2) = 1.1}$ for X_2. The null and alternative hypotheses were defined as H_0: $\beta_i = 0$ vs H_A: $\beta_i \neq 0$ for i = 1, 2. Statistical power was calculated as the proportion of simulations in which the model correctly rejected the null hypothesis in favor of the alternative hypothesis. Additionally, 95% confidence intervals and their sizes were also assessed, with narrower intervals indicating more precise estimates. Simulation scenarios were constructed by varying the total number of observations and the imbalance ratio between the 2 groups as follows:

- Sample sizes: n={50, 100, 250, 500}
- Imbalance ratios: r={1:1, 2:3, 3:7, 1:4, 1:9} which corresponds to the event proportions {0.5, 0.4, 0.3, 0.2 and 0.1}, respectively.

The imbalance ratios were selected to allow to evaluate the power across a spectrum ranging from balanced (r=0.5) to highly imbalanced scenarios (r=0.1). It was defined based on the proportion of individuals who experienced the event of interest, with the event group treated as the minority class. The scenarios ranged from a balanced distribution (r=0.5) to extreme imbalanced settings (r=0.1). For example, in the extreme imbalanced case, only 10% of the observations experienced the event, while 90% were censored.

All simulation were conducted in R version 4.4.2. The *simsurv* package (v1.0.0) was used for data generation.[22] Cox PH and weighted Cox PH models were implemented via the *survival* package (v3.8-3), with IPCW calculated using *pec* package (v 2023.04.12) package.[23,24] Regularized Cox models (LASSO-Cox, Ridge-Cox, Elastic Net-Cox) were estimated using *glmnet* (v 4.1-8), setting α=0 (Ridge), α=1 (LASSO), and α=0.5 (Elastic Net).[25,26] Confidence intervals were calculated using 1,000 bootstrap samples. The Bayesian Cox model was estimated via the Bayes Cox PH function in the *Bolstad2* package (v1.0-29), employing the Metropolis-Hastings algorithm.[27]

RESULTS

Figure 1 presents the power values for each model for categorical variable (X_1). The x-axis denotes sample sizes, while the y-axis displays power values. Separate panels correspond to each imbalance ratio, with the statistical models differentiated by color.

The results show that all models, except the weighted Cox PH model, demonstrated comparable performance. The weighted Cox PH model consistently failed to achieve power values above 80% across all scenarios. Among the remaining methods, the Bayesian model yielded comparatively higher power from others, particularly under high imbalance. In the most imbalanced scenario (r=0.1), only the Bayesian and Cox PH models reached power levels above 80% when the sample size was n=500.

Figure 2 shows the power values for each model with respect to the continuous variable (X_2). Again, the weighted Cox PH model demonstrated substantially lower performance compared to the other models. It achieved power above 80% only under balanced conditions (r=0.4 and r=0.5) and with a sample size of n=500. The Bayesian model showed higher statistical power, particularly in scenarios with small sample sizes. Under balanced conditions (r=0.5), the Cox PH model performed similarly to the Bayesian model.

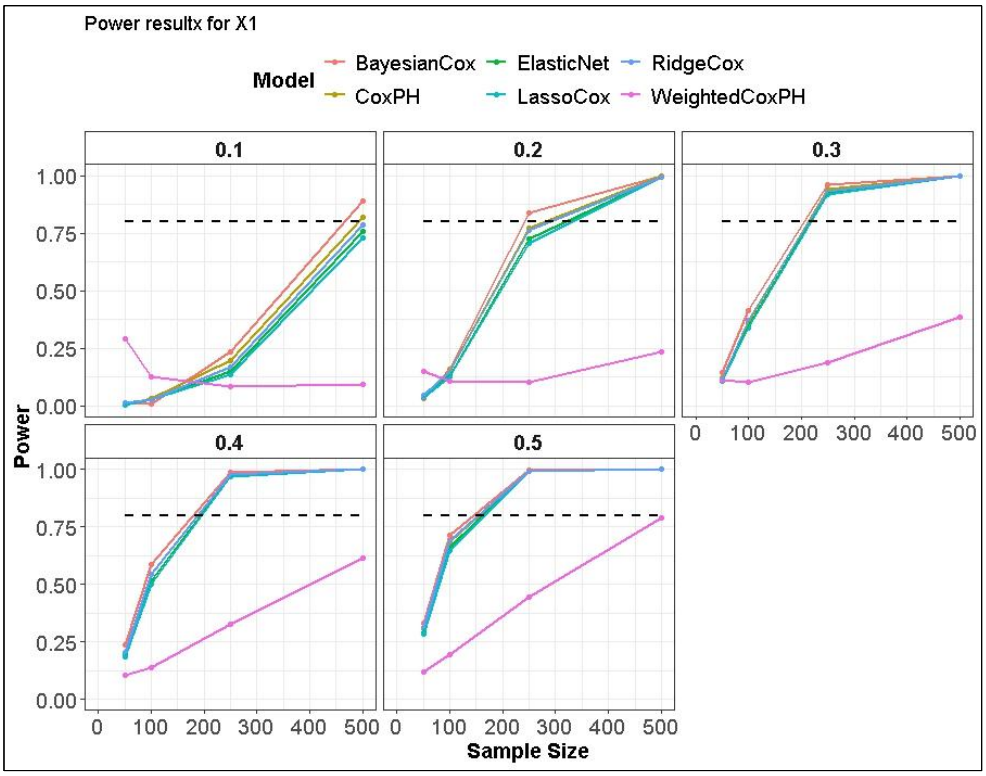


FIGURE 1: Power of 6 Cox models in each sample size and imbalance ratio for categorical independent variable X1
PH: Proportional hazards

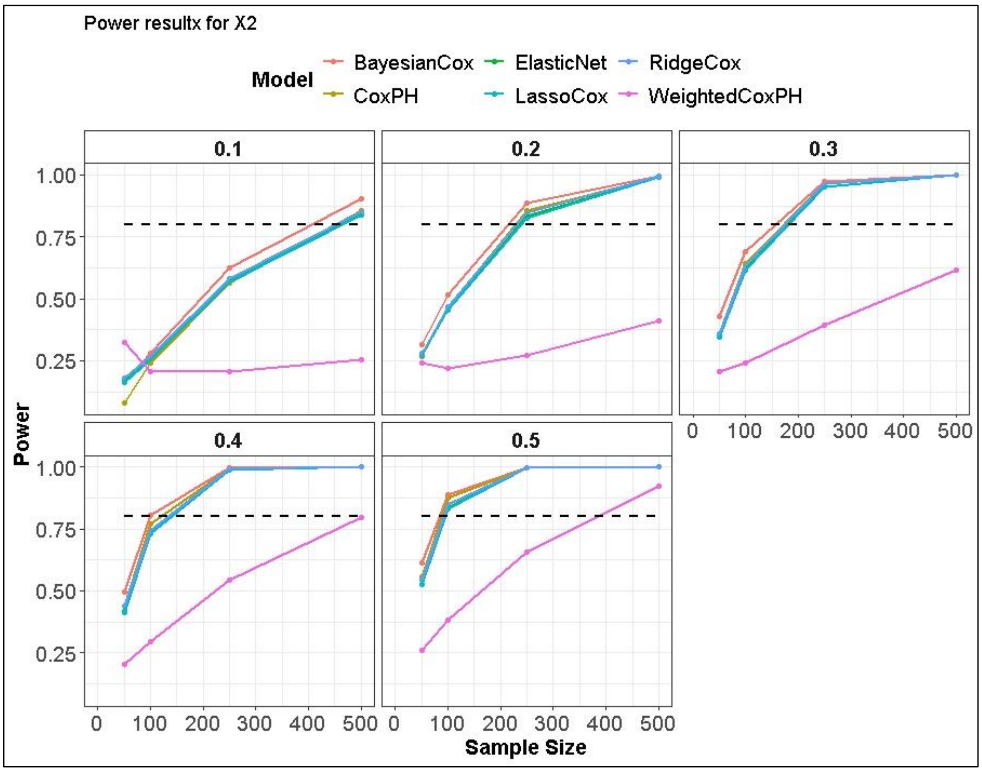
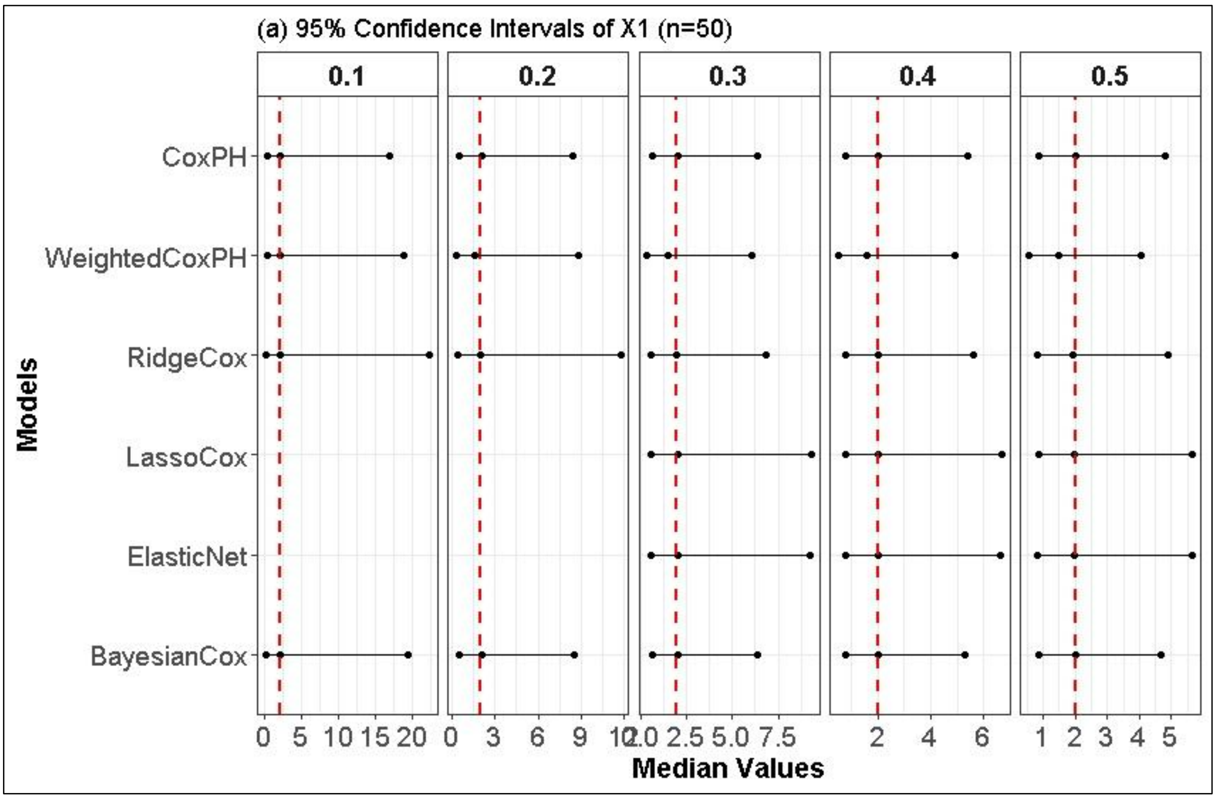


FIGURE 2: Power of six Cox models in each sample size and imbalance ratio for continuous independent variable X2
PH: Proportional hazards



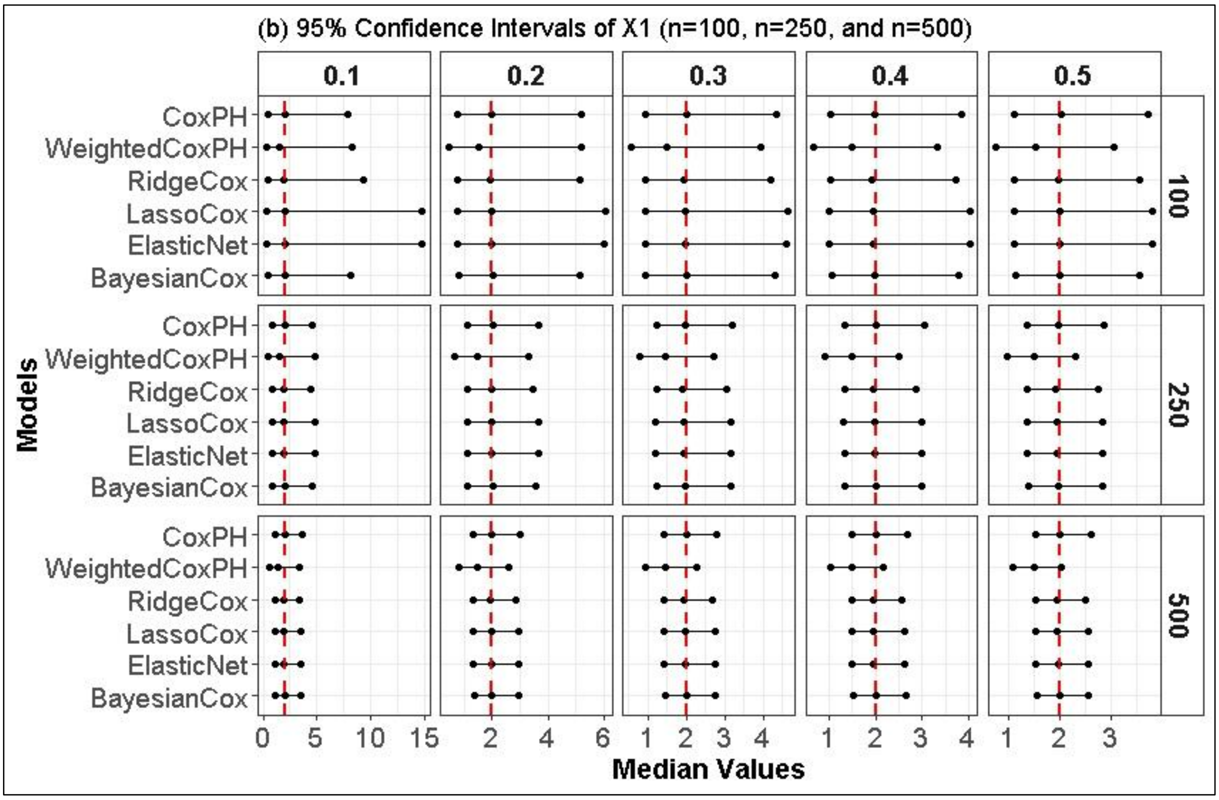


FIGURE 3: Median values of lower and upper confidence limits for X1 variable across 1,000 simulated datasets in each model. **(a)** Results sample size n = 50; **(b)** Results for sample sizes n=100, 250, and 500

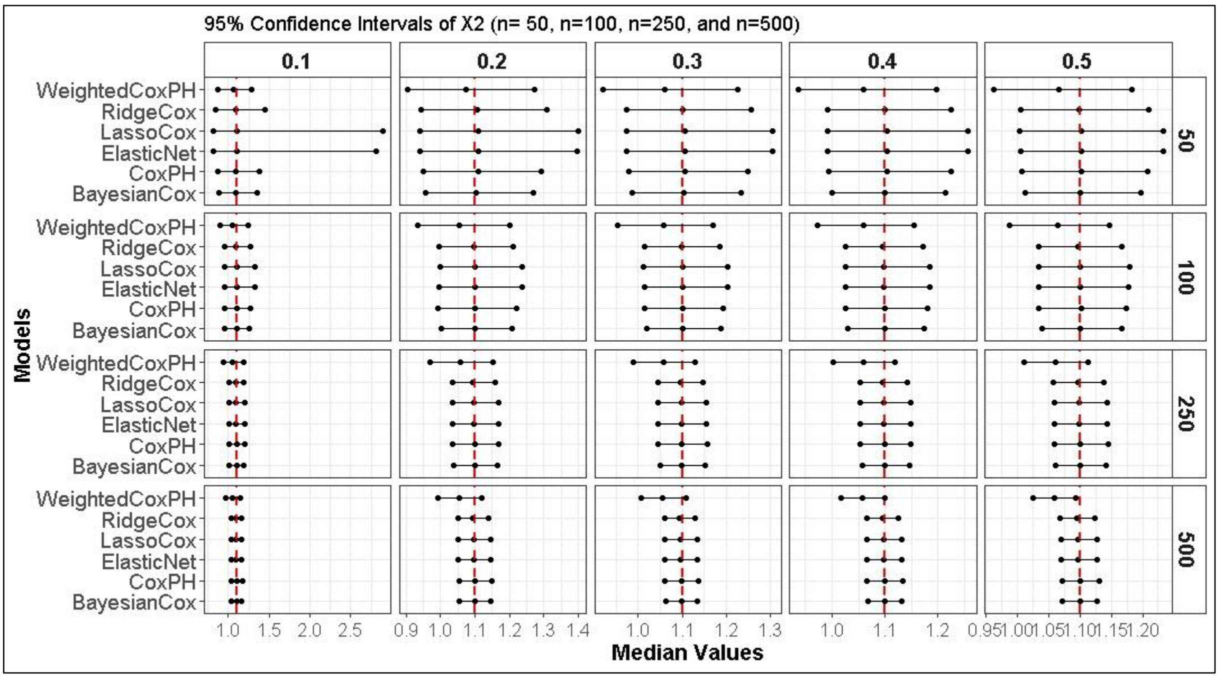


FIGURE 4: Median values of the lower and upper confidence limits for X2 variable across 1,000 simulated datasets in each model

To assess the precision of the regression coefficient estimates, Figure 3(a-b) display the median values of the lower and upper confidence interval bounds for X_1. Due to the large confidence interval observed when n=50, these results are shown separately in Figure 3(a) to preserve the scale of other results. In Figure 3(b), sample sizes are presented in rows, and imbalance ratio in columns. The red vertical line represents the true hazard ratio for X_1 scenario ($\exp(\beta_1)=2$). The lines for each model indicate the median lower and upper confidence bounds for X_1.

In Figure 3(a), the LASSO and Elastic Net models produced extremely wide upper confidence bounds in scenarios with r=0.1 and r=0.2, making them unvisualisable. These 2 models also yielded wide intervals when n=100. In contrast, the Ridge regression model provided narrower intervals under small sample sizes compared to the LASSO and Elastic Net models. The weighted Cox PH model consistently produced narrower confidence intervals across almost all scenarios. The Bayesian and Cox PH had similar confidence interval bounds in nearly all scenarios.

Figure 4 presents the median confidence interval bounds for the regression coefficients of the continuous variable X_2. As with X_1, the LASSO and Elastic Net models produced larger confidence intervals at the smallest sample size (n=50). The overall patterns observed for X_1 also held for X_2, reinforcing the consistency of the model performances across variable types.

DISCUSSION

In clinical research, where time-to-event outcomes such as death or disease progression are of critical interest, survival models-particularly the Cox PH model-remain indispensable tools. However, many real-world survival datasets suffer from substantial class imbalance, often due to low event rates. This imbalance can distort the estimation process, reduce statistical power, and limit the generalizability of model results. The primary goal of this study was to evaluate the performance of several Cox-based modeling approaches under varying degrees of class imbalance, without applying data-level modifications to alter the original distribution. The results highlight key differences in model performance, particularly in scenarios involving small sample sizes and highly imbalance.

In this study, we evaluated the performance of 6 Cox-based survival models-classical, weighted (IPCW), 3 regularized variants (LASSO, Ridge, Elastic Net), and Bayesian-under various degrees of class imbalance and sample sizes, without applying any resampling or data balancing methods. By preserving the original data structure, our simulation framework allowed a more realistic assessment of model performance in scenarios that closely mirror clinical datasets.

To address imbalance in time-to-event data, we evaluated the performance of the weighted Cox model, which assigns greater weight to uncensored observations rather than treating all participants equally. Our simulation showed that weighted Cox model consistently yielded lower statistical power and narrower confidence intervals compared to other Cox-based regression models, particularly under conditions of extreme imbalance (r=0.1) and small sample size (n=50). The reduced statistical power may be attributable to the narrower, and potentially biased, confidence intervals, which fail to capture the true parameter values.

Among the regularized Cox regression models, LASSO-Cox and Elastic Net-Cox generally produced larger confidence intervals, particularly in small samples with severe imbalance. In contrast, Ridge-Cox regression yielded narrower confidence intervals and more stable coefficient estimates compared to other regularized Cox models. In the literature, these models are often evaluated after data-level interventions such as undersampling, oversampling, or SMOTE to balanced the groups. Mulugeta et al., compared the statistical (Cox PH and regularized Cox models) and some ML methods following SMOTE-based balancing in predicting survival after kidney transplantation. While stochastic gradient boosting, one of the ML method, achieved the best overall performance, Ridge model performed better than other regularized Cox models. Although tree-based ML models showed higher predictive performance in that study, the authors suggested that Ridge regression may still be preferable in some contexts due to its ability to provide more interpretable estimates of predictor effects, particularly in clinical research.[18] Similarly, Andishgar et al., found that Ridge regression outperformed other regularized Cox models after oversampling and produced results comparable to those of the Cox PH model.[17] Our findings, based on the simulated datasets without any data-level modification (i.e., using the original imbalance structure), are consistent with these results. In contrast, Datta et al., compared the Elastic Net and Cox PH models after applying both under- and oversampling. Although, the Elastic Net model demonstrated better performance, the authors emphasized that resampling alters the prevalence of the event and may limit the generalizability of survival estimates.[16] In our study, where no resampling was performed, the Cox PH model yielded more favorable results.

In our setting, with 2 predictors having true non-zero effects, Ridge's superior power is consistent with both theory and prior evidence for penalized Cox models. The L2 penalty shrinks coefficients but does not set them to zero, thereby preserving true signals and reducing estimator variance. In Cox regression, such penalization is known to stabilize coefficient estimates and improve finite-sample behavior, particularly when incidental collinearity arises.[28,29] By contrast, the LASSO's L1 penalty performs hard selection, driving smaller effects to zero and, under correlation, often selecting only one variable from a group a behavior that can depress coverage and power when effects are modest but non-null. Elastic Net partially mitigates this via its L2 component, yet still inherits L1-induced sparsity.[5,25,30,31] These mechanisms align with how glmnet fits L1-, L2-, and Elastic Net-penalized Cox models via coordinate descent, helping explain why Ridge produced the most stable estimates and the highest power across our scenarios.

The Bayesian Cox model outperformed all other approaches in terms of statistical power, particularly under extreme imbalance (e.g., r=0.1) and small sample sizes (n=50). The advantage of Bayesian methods lies in their ability to incorporate prior distributions, which can stabilize estimates when data are sparse.[7,8] However, as the sample size increased, the performance of the Bayesian and Cox PH models converged-suggesting that the added complexity of the Bayesian approach may offer limited benefit in large sample settings. To our knowledge, there are no prior studies that have applied Bayesian Cox model specifically to class-imbalanced survival data.

A notable strength of this study is its decision to preserve the original imbalance structure throughout the analysis. Unlike previous studies that utilized oversampling, undersampling, or synthetic techniques like SMOTE, we evaluated model performance without altering data. While resampling methods can improve metrics like area under the curve or the concordance index, these improvements often come at the cost of distorting the event prevalence and reducing the interpretability of survival probabilities.[10,16-18] As noted by Datta et al., resampling may also lead to overfitting, particularly when models are applied to independent test sets that reflect real-world imbalance.[16] This reinforces the rationale for maintaining the original data distribution and instead focusing on model-based strategies that are robust to imbalance.

There were several limitations of this study: (i) Exclusive reliance on simulated datasets without validation on clinical or other real-world cohorts may restrict the generalizability of findings; (ii) The simulation framework's focus on a single binary and one continuous covariate under an exponential survival distribution limits applicability to high-dimensional settings or alternative hazard structures; (iii) These points underscore that simulation-only results may not directly translate to practical datasets and that the narrow covariate and distributional assumptions constrain the study's scope. Including them will enable readers to assess the work's validity and applicability better.

In summary, our findings suggest that:

- Although conceptually appealing, the weighted Cox model underperformed across most scenarios in terms of both statistical power and reliability.
- Among the regularized models, Ridge regression consistently yielded more stable coefficient estimates and narrower confidence intervals than LASSO and Elastic Net, particularly in small, imbalanced.
- Bayesian Cox model achieved the highest statistical power and estimation accuracy in settings with severe imbalance and small sample sizes. However, its advantage diminished as sample size increased, with performance converging toward that of the classical Cox PH model.

CONCLUSION

Class imbalance is a common yet often underappreciated issue in survival analysis, particularly in clinical research where events such as death or disease recurrences are relatively rare. While traditional data-level strategies such as oversampling or synthetic data generation (e.g., SMOTE) are frequently used to address imbalance, these approaches can distort the original data distribution and compromise the real-world applicability of model results.

This study highlights the importance of selecting modeling strategies that are inherently robust to imbalance, rather than relying on external data manipulation. For studies where modifying the dataset is not feasible-especially those aiming to produce generalizable survival probability estimates-model-based approaches such as Bayesian and Ridge regression may provide more reliable alternatives. Bayesian Cox, Cox PH and Ridge-Cox regression models consistently demonstrated higher statistical power with extreme imbalance, particularly when sample size reached at least 500. In contrast, under more balanced conditions ($r=0.4$ and $r=0.5$), all Cox-based models-except the weighted Cox model-achieved comparable power levels with sufficient sample size ($n\geq200$).

Future research should further investigate the sensitivity of Bayesian approaches to prior selection, particularly in sparse data contexts, and explore hybrid or ensemble methods that balance interpretability with robustness to imbalance. Overall, our results emphasize the need for methodological rigor and the simulation-based evaluations when modeling survival data with pronounced censoring and event rarity.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

This study is entirely author's own work and no other author contribution.

REFERENCES

1. Kleinbaum DG, Klein M. Survival analysis: a self-learning text. In: Gail M, Krickeberg K, Samet JM, Tsiatis A, Wong W, eds. 2nd ed. USA: Springer; 2005. p.1-590. [Crossref]
2. Harrell FE. Regression Modeling Strategies. 2nd ed. Switzerland: Springer; 2015. p.1-582.
3. Afrin K, Illangovan G, Srivatsa SS, Bukkapatnam ST. Balanced random survival forests for extremely unbalanced, right censored data. arXiv. 2018. [Link]
4. Zhang HH, Lu W. Adaptive-LASSO for Cox's proportional hazards model. Biometrika. 2007;94(3):691-703. [Crossref] [PubMed]
5. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385-95. [PubMed]
6. Park MY, Hastie T. L1-regularization path algorithm for generalized linear models. Journal of the Royal Statistical Society B. 2007;69(4):659-77. [Crossref] [PubMed]
7. Ibrahim JG, Chen MH, Sinha D. Bayesian Survival Analysis. 1st ed. New York: Springer; 2001. p.1-478. [Crossref] [PubMed]
8. Işık H, Karasoy D, Karabey U. A new adjusted Bayesian method in Cox regression model with covariate subject to measurement error. Hacet. J. Math. Stat. 2023;52(5):1367-78. [Link]
9. Chia CC, Rubinfeld I, Scirica BM, McMillan S, Gurm HS, Syed Z. Looking beyond historical patient outcomes to improve clinical models. Sci Transl Med. 2012;4(131):131ra49. [Crossref] [PubMed]
10. Lian J, Huang F, Huang X, Lau KY, Ng KS, Chu CCF, et al. Admission blood tests predicting survival of SARS-CoV-2 infected patients: a practical implementation of graph convolution network in imbalance dataset. BMC Infect Dis. 2024;24(1):803. [Crossref] [PubMed] [PMC]
11. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. The Annals of Applied Statistics. 2008;2(3):841-60. [Crossref] [PubMed]
12. Lyashevska O, Malone F, MacCarthy E, Fiehler J, Buhk JH, Morris L. Class imbalance in gradient boosting classification algorithms: application to experimental stroke data. Stat Methods Med Res. 2021;30(3):916-25. [Crossref] [PubMed]
13. Drummond C, Holte RC. Class Imbalance and Cost Sensitivity: Why Under-Sampling Beats Over-Sampling. 2003. [Link]
14. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. 2009;21(9):1263-84. [Crossref] [PubMed]
15. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002;16:321-57. [Link]
16. Datta G, Alexander LE, Hinterberg MA, Hagar Y. Balanced Event Prediction Through Sampled Survival Analysis. 2019;2(1):28-38. [Crossref]
17. Andishgar A, Bazmi S, Lankarani KB, Taghavi SA, Imanieh MH, Sivandzadeh G, et al. Comparison of time-to-event machine learning models in predicting biliary complication and mortality rate in liver transplant patients. Sci Rep. 2025;15(1):4768. [Crossref] [PubMed] [PMC]
18. Mulugeta G, Zewotir T, Tegegne AS, Muleta MB, Juhar LH. Developing clinical prognostic models to predict graft survival after renal transplantation: comparison of statistical and machine learning models. BMC Med Inform Decis Mak. 2025;25(1):54. [Crossref] [PubMed] [PMC]
19. Tsiatis AA. Semiparametric Theory and Missing Data. 1st ed. New York: Springer; 2006. [Crossref]
20. Willems SJW. Inverse probability censoring weights for routine outcome monitoring data [Master thesis]. The Netherlands: Leiden University Medical Center; 2014. [Link]
21. Liu L, Yang F, Fan Y, Kao C, Wang F, Yu L, et al An Improved Training algorithm based on ensemble penalized Cox regression for predicting absolute cancer risk. China CDC Wkly. 2023;5(9):206-12. [Crossref] [PubMed] [PMC]
22. Brilleman SL, Wolfe R, Moreno-Betancur M, Crowther MJ. Simulating survival data using the simsurv R package. Journal of Statistical Software. 2021;97(3):1-27. [Crossref]
23. Therneau T. A Package for Survival Analysis in R. R package version 3.8-3. 2024. [Link]
24. Gerds TA. Pec: prediction error curves for risk prediction models in survival analysis. R package version 2023.04.12. 2024. [Link]
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1-22. [Crossref] [PubMed] [PMC]
26. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. J Stat Softw. 2023;106:1. [Crossref] [PubMed] [PMC]
27. Curran J. Bolstad2: bolstad functions. R package version 1.0-29. 2022. [Link]

28. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55-67. [Link]

29. Verweij PJ, Van Houwelingen HC. Penalized likelihood in Cox regression. Stat Med. 1994;13(23-24):2427-36. [Crossref] [PubMed]

30. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2005;67(2):301-20. [Crossref]

31. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw. 2011;39(5):1-13. [Crossref] [PubMed] [PMC]