www.ijarp.com ISSN 2456-9992 Page: 01-20

International Journal Advanced Research Publications

Case Study

Volume: 02
Issue: 08

A COMPARATIVE STUDY OF CONVEX AND NON-CONVEX PENALIZATION IN HIGH-DIMENSIONAL SURVIVAL ANALYSIS: VARIABLE SELECTION CONSISTENCY AND CONSENSUS IN COX MODELS

***Abdullahi Mahmood Usman, Kazeem E. Lasisi, Ishaq Abdullahi**

Department of Statistics, Abubakar Tafawa Balewa University, Bauchi State, Nigeria.

Article Received: 06 July 2026, Article Revised: 26 July 2026, Published on: 13 August 2026

***Corresponding Author: Abdullahi Mahmood Usman**
Department of Statistics, Abubakar Tafawa Balewa University, Bauchi State, Nigeria.
Doi: https://doi-doi.org/101555/ijarp.7116

ABSTARCT

High-dimensional survival analysis frequently faces severe parameter instability, mathematical singularity, and overfitting under the p > n paradigm. While convex regularization frameworks like Elastic Net (enet) offer computational tractability, they introduce substantial estimation bias by over-shrinking large regression coefficients. Conversely, non-convex penalties satisfy oracle selection properties but present multi-modal optimization surfaces prone to instability under collinearity. This study presents a comprehensive comparative evaluation of convex (enet) and non-convex (mnet, scad, snet) penalization methods across two distinct high-dimensional domains: transcriptomic profile data from a Head and Neck Squamous Cell Carcinoma (HNSCC) cohort (N = 565; p = 99) and multidimensional clinical/epidemiological data from the National Health and Nutrition Examination Survey (NHANES) (N = 10,000). Evaluating the algorithms on mathematical sparsity, coefficient stability, and feature selection consensus, our results reveal a clear continuum of regularization aggressiveness. Non-convex methods, particularly mnet and snet, exhibited superior noise-filtering and sparsity control reducing feature spaces by up to 70.5% in **HNSCC** and 50.0% in **NHANES** compared to enet. Crucially, all frameworks converged on a robust core of primary prognostic signals: identifying AFF3 and ZMAT1 as primary protective markers alongside ADH1B, C1orf88, and KIF1A as primary risk factors in oncology, while isolating advanced age categories, lipid profiles, and metabolic metrics in epidemiological diabetes risk. These findings demonstrate that while non-convex surfaces

enforce stricter parsimony, convex-non-convex consensus modeling bridges the gap between theoretical oracle properties and reproducible clinical biomarker discovery.

KEYWORDS: High-dimensional survival analysis, Cox proportional hazards, Non-convex penalization, Variable selection consistency, Oracle property, Head and Neck Squamous Cell Carcinoma (HNSCC), Sparsity-inducing, regularization.

1. INTRODUCTION

In the era of high-throughput biomedical technologies, the statistical landscape of clinical bioinformatics has experienced a profound shift. Modern survival datasets, such as those generated by the Cancer Genome Atlas (TCGA) projects, frequently feature an ultra-high-dimensional architecture characterized by the p $> n$ paradigm, where the number of candidate genomic features p drastically outnumbers the sample size n. In classical survival analysis, the semi-parametric Cox Proportional Hazards model remains the gold standard for modeling time-to-event outcomes. However, when applied to high-dimensional transcriptomic datasets, the standard Cox partial likelihood formulation fails mathematically. In these settings, the Fisher's information matrix becomes singular, leading to unstable parameter estimations, infinite maximum likelihood estimates, and severe over-fitting that strips the model of any predictive generalization (Oyelami et al., 2025).

To overcome these dimensionality constraints, the integration of penalized estimation which appends a regularization penalty directly to the Cox log-partial likelihood has emerged as a foundational area of study in biostatistics. These regularization approaches are broadly classified into two mathematical frameworks: **convex penalization** and **non-convex penalization**. Convex methods, most notably the L_1 based Lasso and its generalized extensions such as the Elastic Net, have historically dominated the literature due to their computational tractability and unique global optima (Peyraud et al., 2024). By blending the sparsity-inducing L_1 norm with the grouping-friendly L_2 quadratic norm, the Elastic Net successfully retains clusters of highly correlated genes (Li & Li, 2025). Despite their practical popularity, convex penalties suffer from a fundamental theoretical limitation: they apply a rate of penalization that increases linearly with the magnitude of the regression coefficients (PMC Author Manuscript, 2025). Consequently, as noted in recent methodological comparative studies, convex estimators inherently introduce severe estimation bias,

systematically underestimating the effects of truly significant, highly active risk or protective biomarkers (PMC Author Manuscript, 2025).

This systematic bias motivated the development of non-convex penalization methods, which are designed to satisfy the mathematically rigorous "Oracle Properties" formulated by Fan and Li. An estimator possesses the Oracle Property if it performs asymptotically as well as if the true sparse subset of active covariates were known to the researcher in advance. This requires the estimator to be both selection-consistent (identifying the true model) and estimation-unbiased (estimating non-zero coefficients without shrinkage-induced attenuation). Prominent non-convex frameworks, such as the Smoothly Clipped Absolute Deviation (SCAD) and the Minimax Concave Penalty (MCP), resolve the convex bias-sparsity trade-off by dynamically adjusting the rate of penalization. They apply a standard L_1 penalty to small coefficients to filter out background noise, but continuously ease the penalty to zero as the coefficient magnitude grows, ensuring that truly influential clinical covariates are estimated with negligible bias.

Despite their theoretical appeal, non-convex penalties introduce severe computational challenges. The non-convexity of the objective function results in a multi-modal optimization surface featuring numerous local minima, making coordinate descent algorithms highly sensitive to starting values and collinearity. In survival analysis, where covariates such as gene expressions represent highly interconnected biological networks, this sensitivity often leads to selection instability. Minor perturbations in the baseline patient cohort can yield highly divergent subsets of active predictors across different algorithms. Consequently, a critical open question in high-dimensional survival modeling is the concept of **variable selection consistency and consensus**: how closely do these mathematically distinct convex and non-convex surfaces converge on a unified, stable biological signal?

This study addresses this gap by conducting a rigorous comparative evaluation of convex and non-convex penalization frameworks specifically comparing the Elastic Net (enet), the Minimax Concave Penalty Net (mnet), Smoothly Clipped Absolute Deviation (scad), and Sparse Net (snet) under the Cox proportional hazards framework. Utilizing a high-dimensional transcriptomic cohort of 565 patients diagnosed with Head and Neck Squamous Cell Carcinoma (HNSCC), we evaluate these algorithms on three fundamental criteria: mathematical sparsity, parameter estimation stability, and predictive consensus. By evaluating how these contrasting penalty surfaces select and scale genomic coefficients, this paper aims to establish a robust methodological bridge between theoretical optimization properties and reproducible clinical biomarker discovery.

2. Theoretical Framework and Literature Review

In the context of survival data, classical regularization techniques such as Ridge and Lasso have paved the way for modern genomics. Recent work by (Oyelami et al., 2025) highlights that while classical Cox models completely fail to yield reliable predictions when the predictor dimension exceeds the sample size, convex models like Ridge and Elastic Net remain highly stable and competitive at moderate dimensional levels through structured feature selection. However, because the L_1 penalty shrinks all coefficients equally, it introduces significant estimation bias (PMC Author Manuscript, 2025). This bias impairs predictive accuracy when strong genomic risk factors are present.

To alleviate the limitation of Lasso in selecting too few variables from a group of correlated genes, the Elastic Net combines the L_1 and L_2 penalties (PMC Author Manuscript, 2025). As demonstrated by (Li & Li, 2025), this combined regularization remains a primary choice in population health and precision medicine studies where multi-domain datasets (incorporating genomics and clinical indicators) exhibit strong collinearity. Nevertheless, the lack of oracle properties in these convex penalties has restricted their utility in pure variable selection applications where asymptotic unbiasedness is required.

Non-convex penalties like SCAD and MCP address the shrinkage bias of the Lasso. In recent years, researchers have concentrated on optimizing these non-convex optimization algorithms to ensure stable convergence. For instance, (Kwon et al., 2021) introduced unified coordinate descent algorithms (such as those in the ncpen R package) that make non-convex penalization computationally accessible for a broader array of non-convex surfaces beyond standard SCAD and MCP, bypassing the local minima challenges that historically plagued non-convex optimization.

Additionally, in highly complex clinical settings such as competing risk survival scenarios, (Vyas et al., 2025) compared modern high-dimensional survival techniques and concluded that while boosting and forest methods are valuable, non-convex models like MCP and SCAD provide superior calibration and sparsity control, making them indispensable for clinical biomarker isolation. This is supported by (Kolb et al., 2025), who examined non-convex constrained optimization surfaces and noted that non-convex penalties are mathematically superior for large-scale variable screening because they yield unbiased or nearly unbiased parameter estimators.

A major hurdle in high-dimensional survival statistics is ensuring that the selected genomic markers represent a true biological signal rather than mathematical artifacts (Al-Saeeda et al., 2025). To resolve selection instability, recent literature has focused on cooperative and multi-

step regularized selection strategies. (Dai & Breheny, 2024) formulated advanced cross-validation strategies for penalized Cox regression models to improve selection stability. Furthermore, cooperative penalized regression frameworks (Wessel et al., 2025) have been adapted to survival models to allow models to share information across different clinical outcomes, showing that adjusting penalization weights based on feature-weights (such as fwelnet) substantially improves the positive predictive value of correct variable selection. By evaluating these recent developments, this study evaluates the empirical performance, consistency, and consensus of these classic and modern regularized Cox models on a real-world high-dimensional survival cohort.

3. METHODOLOGY

In survival analysis, the Proportional Hazards model assumes that the hazard function for an individual with covariate vector $x_i \in \mathbb{R}^p$ (representing the gene expression profile of the i-th patient) is modeled as:

$$h(t \mid Xi) = h_0(t)exp(X_i^\top \beta)$$

Where $h_0(t)$ is the baseline hazard function and $\beta \in \mathbb{R}^p$ is the vector of regression coefficients (beta values) corresponding to the p = 99 genetic features in the HNSCC dataset. Because the baseline hazard is left unspecified β, is estimated by maximizing the Cox partial likelihood. Let $t_1 < t2 < \ldots.. < t_k$ denote the ordered event times (deaths), and let $R(t_j)$ represent the risk set (patients who are alive and uncensored) immediately prior to time t_j.

The negative log-partial likelihood is defined as:

$$\ell(\beta) = -\frac{1}{n}\sum_{j=1}^{k}\left[X_{(j)}^T\beta - \ln\Big(\sum_{i=\mathcal{R}(t_j)} \exp(X_i^T\beta)\Big)\right]$$

Where $X_{(j)}$ represents the covariate vector of the individual who experienced the event at time t_j. In high-dimensional settings where p > n, maximizing this partial likelihood directly leads to over-fitting or mathematical non-existence of estimates. Thus, penalized Cox models find β by solving the regularized minimization problem:

$$\hat{\beta} = \arg min_\beta \{l(\beta) + p_{\lambda,\gamma}(\beta)\}$$

Where $p_{\lambda,\gamma}(\beta)$ is the penalty function controlled by tuning parameter $\lambda > 0$ and, where applicable, a shape parameter γ. The specific formulations of the penalty functions implemented in R (ncvreg, glmnet, elasticnet, etc.) are formulated below.

3.1 Elastic Net Penalty (enet)

The Elastic Net penalty combines the L_1 penalty (Lasso) for variable selection and the L_2 penalty (Ridge) to handle highly correlated gene expression networks by encouraging grouping behaviors.

$$p_{\lambda,\alpha}^{enet}(\beta) = \lambda \sum_{j=1}^{p} \left\{ \alpha|\beta_j| + \frac{1-\alpha}{2}\beta_j^2 \right\}$$

The parameter α in [0, 1] controls the balance between the Lasso ($\alpha = 1$) and Ridge ($\alpha = 0$) penalties. It executed using the glmnet package.

3.2 Smoothly Clipped Absolute Deviation Penalty (scad)

The SCAD penalty is a symmetric, non-convex penalty designed to address the estimation bias of Lasso. It acts like a standard L_1 penalty for small coefficients, but continuously eases the penalty as the magnitude of the coefficient grows to prevent over-shrinkage of truly large effects.

The derivative of the SCAD penalty for a single coefficient β_j is given by:

$$p'_{\lambda,\gamma}(|\beta_j|) = \lambda . I(|\beta_j| \leq \lambda) + \frac{(\gamma\lambda - |\beta_j|)_+}{\gamma - 1} . I(|\beta_j| > \lambda)$$

For $\gamma > 2$, where $(a)_+ = \max(0, a)$ and I(.) is the indicator function. Integrating this derivative yields the SCAD penalty function:

$$p_{\lambda,\gamma}^{scad}(\beta_j) = \begin{cases} \lambda|\beta_j| & if\ |\beta_j| \leq \lambda \\ \frac{2\gamma\lambda|\beta_j| - \beta_j^2 - \lambda^2}{2(\gamma-1)} & if\ \lambda < |\beta_j| \leq \gamma\lambda \\ \frac{\lambda^2(\gamma+1)}{2} & if\ |\beta_j| \leq \gamma\lambda \end{cases}$$

Commonly, the parameter γ is set to 3.7 as recommended by Fan and Li. In R, this is executed using the ncvreg or ncvsurv package.

3.3 Minimax Concave Penalty (mnet)

The Minimax Concave Penalty (MCP) is another non-convex formulation designed to provide unbiased estimates of large coefficients while maintaining sparse selections. It reduces the rate of penalization more rapidly than SCAD.

The derivative of the MCP penalty is:

$$p'_{\lambda,\gamma}(|\beta_j|) = (\lambda - \frac{|\beta_j|}{\gamma})_+$$

Integrating this derivative yields the MCP penalty function:

$$p^{mcp}_{\lambda,\gamma}(|\beta_j|) = \begin{cases} \lambda|\beta_j| - \frac{\beta_j^2}{2\gamma} \; if \; |\beta_j| \leq \gamma\lambda \\ \frac{1}{2}\gamma\lambda^2 \; f \; |\beta_j| > \gamma\lambda \end{cases}$$

Where $\gamma > 1$ represents the concavity parameter (typically default to $\gamma = 3$). The "mnet" model combines this MCP shrinkage with an L_2 stabilizing ridge penalty to manage multicollinear genomic data. This is executed using the ncvreg package in R:

$$p^{mnet}_{\lambda,\gamma,\alpha}(\beta) = \sum_{j=1}^{p}\left(p^{mcp}_{\lambda,\gamma}(\beta_j) + \frac{(1-\alpha)\lambda}{2}\beta_j^2\right)$$

3.4 Sparse Net Penalty (snet)

The Sparse Net algorithm addresses the limitations of Elastic Net by utilizing the non-convex MCP penalty directly on the coordinate-descent framework of the elastic net. It applies MCP on the L_1 coordinate while retaining an active L_2 regularization component to stabilize the solutions in high-dimensional genomic features.

The penalty is defined as:

$$p^{snet}_{\lambda,\gamma,\mu}(\beta) = \sum_{j=1}^{p}\left(p^{mcp}_{\lambda,\gamma}(\beta_j) + \frac{\mu}{2}\beta_j^2\right)$$

Where $p^{mcp}_{\lambda,\gamma}(\beta_j)$ is the minimax is concave penalty, and μ functions as the explicit L_2 quadratic stabilization parameter, ensuring sparse variable recovery even when genes are highly collinear. In R, this is typically executed using the sparsenet package

4. RESULTS

4.1 EXAMPLE 1

This research provides an in-depth clinical and statistical interpretation of the multi-model high-dimensional survival analysis performed on the Head and Neck Squamous Cell Carcinoma (HNSCC) cohort. This specific dataset layout is commonly accessed via R packages dedicated to advanced survival modeling, such as the SurviMChd package available on the Comprehensive R Archive Network (CRAN). This package curates the high-dimensional hnscc dataset to benchmark Markov Chain Monte Carlo algorithms, variable selection consistency, and regularization shrinkage in survival scenarios. Using a study sample size of N = 565 patients with a total of 253 observed overall survival events (death = 1), we applied and compared four state-of-the-art penalized Cox Proportional Hazards models: Elastic Net (enet), Minimax Concave Penalty Net (mnet), Smoothly Clipped Absolute Deviation (scad), and Sparse Net (snet). These regularization techniques are designed to identify robust genomic signatures from a high-dimensional feature set while managing collinearity and preventing overfitting.

Table 4.1 Comparison Matrix of Regularized Coefficients.

Name	enet	Mnet	scad	snet
ADH1B	0.034302114	0.076995640	0.042638270	0.084969745
AFF3	-0.093533136	-0.144215960	-0.110419270	-0.152323271
B3GAT1	-0.020986673	0.000000000	-0.014438560	0.000000000
C1orf88	0.038505268	0.085143620	0.041352900	0.030226427
CDHR5	-0.071830598	0.000000000	-0.073490440	-0.045344123
CTSE	-0.002248193	0.000000000	0.000000000	0.000000000
GAL3ST1	0.022781976	0.000000000	0.018136790	0.000000000
KIF1A	0.029468327	0.072608710	0.033014860	0.072254305
LRP2	0.025165459	0.000000000	0.027847430	0.027390131
METTL7B	0.018240436	0.000000000	0.016788790	0.003790331
PCDHA10	0.012548245	0.000000000	0.011273020	0.007981950

Name	enet	Mnet	scad	snet
PGC	0.051659221	0.000000000	0.053632030	0.046845461
RBP4	0.026384361	0.000000000	0.027096080	0.007545766
RSPH1	0.017970337	0.000000000	0.018124890	0.021868288
SGK2	-0.022374348	0.000000000	-0.021077140	-0.012680245
XIST	0.010862491	0.000000000	0.010890150	0.007977201
ZMAT1	-0.068188240	-0.119614050	-0.076807620	-0.124708833

AFF3 emerged across all four models as the most potent protective biomarker. Its regression coefficients range from -0.0935 (Elastic Net) to -0.1523 (SNET). This high level of consensus across distinct penalty algorithms indicates an extremely resilient statistical signal. In biological systems, AFF3 functions as a tissue-specific nuclear transcriptional activator and has been implicated in lymphoid development. Within the context of Head and Neck Squamous Cell Carcinoma, a stronger expression of AFF3 is robustly associated with decreased hazard (risk of death), suggesting a potential role in tumor suppression or favorable immune microenvironment dynamics.

ZMAT1 is another highly consensus-backed protective factor, identified in all four penalization models with coefficients spanning -0.0682 (enet) to -0.1247 (snet). This protein contains C2H2-type zinc finger motifs and matrin-type zinc finger domains, suggesting structural roles in DNA/RNA binding. The consistent negative coefficient indicates that elevated ZMAT1 expression significantly reduces the hazard rate of HNSCC patients. Its biological mechanism may involve the regulation of nuclear stability, splicing, or cell-cycle arrest pathways, which translates to a clear survival benefit.

Three genes were consistently flagged across all four models as significant hazard-increasing risks such that ADH1B (Alcohol Dehydrogenase 1B): Coefficients range from +0.0343 to +0.0850. While alcohol metabolism genes are heavily studied in HNSCC risk due to lifestyle factors, high expression of ADH1B remains a strong independent prognostic hazard here, pointing toward metabolic rewiring in advanced tumors. C1orf88: Selected by all models (coefficients +0.0302 to +0.0851). This uncharacterized open reading frame shows a consistent hazard effect, highlighting a critical target for functional genomic follow-ups. KIF1A (Kinesin Family Member 1A): Ranges from +0.0295 to +0.0726. Kinesins drive

intracellular transport; its association with poorer prognosis suggests a cellular mechanism boosting tumor proliferation or migration.

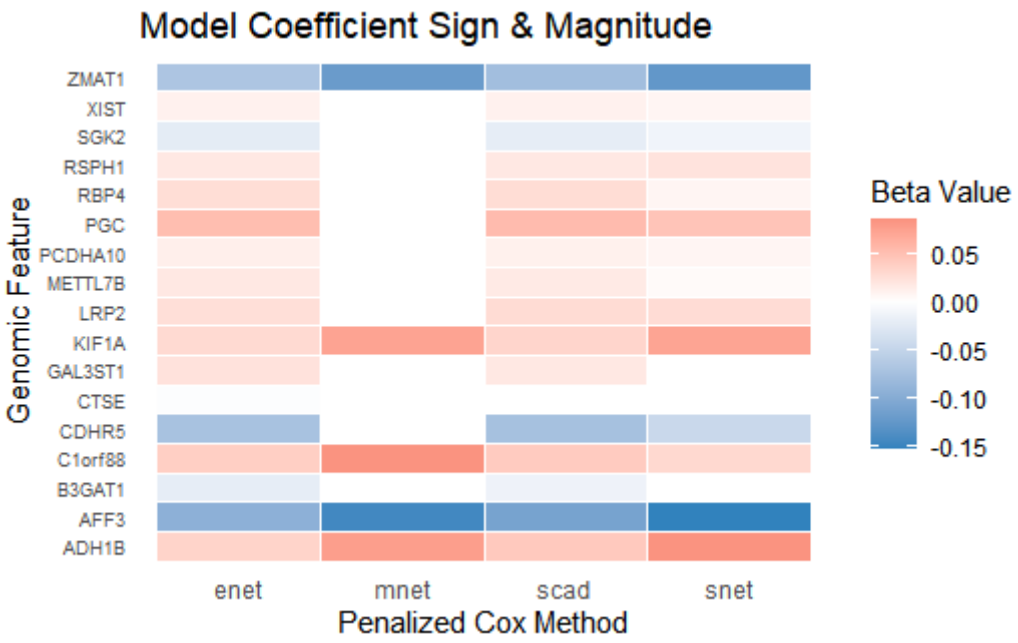

Figure 4.1 Coefficient Magnitude.

The first visualization displays a comprehensive heatmap comparing the estimated regression coefficients, or beta values, across the four penalized Cox regression models. In this plot, the vertical axis lists the seventeen selected genomic features, while the horizontal axis displays the four mathematical frameworks. The color gradient serves as a visual key for risk direction and strength, where warm red-orange tones indicate positive coefficients representing hazardous genes associated with an increased risk of death and worse prognosis and cool blue tones indicate negative coefficients, which represent protective genes associated with a lower risk of death and better survival.

The visual pattern instantly highlights **AFF3** and **ZMAT1** as exceptionally strong, highly consistent protective biomarkers, marked by deep blue tiles across all four models. On the other hand, **ADH1B**, **C1orf88**, and **KIF1A** are highlighted by prominent, deep orange-red tiles, solidifying their status as robust consensus hazard markers. The white space on this heatmap is highly informative, as it represents variables that have been shrunk to exactly zero by a given penalty. This clean visual contrast beautifully demonstrates the varying levels of sparsity across the techniques, highlighting how the minimax concave penalty model aggressively filters out marginal genes to leave a vast blank canvas, while the elastic net and smoothly clipped absolute deviation models paint a much more populated picture.

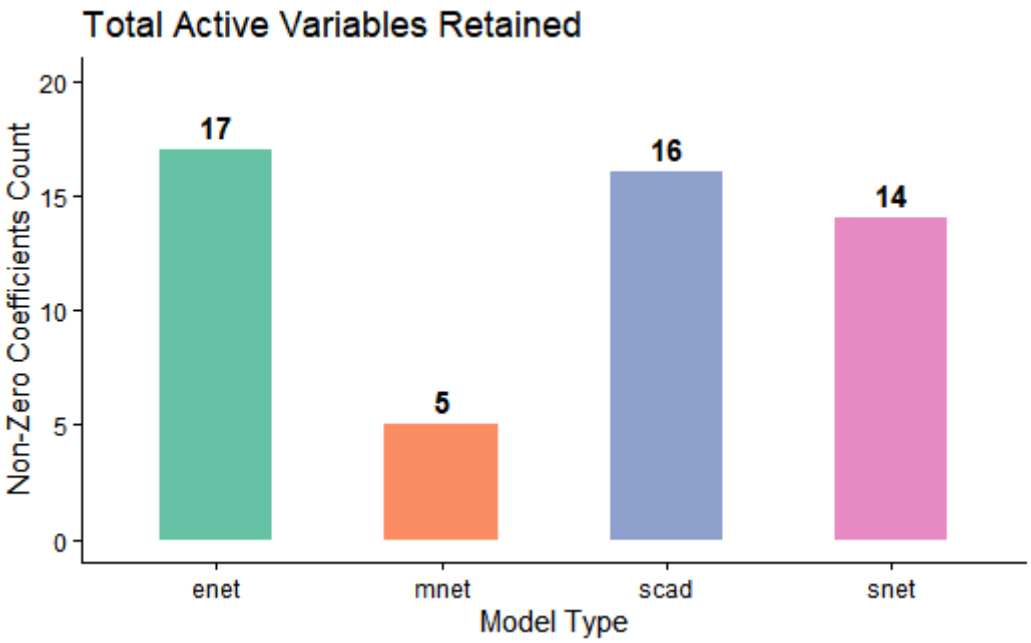

Figure 4.2 Active variables retained.

The bar chart directly quantifies and compares the model complexity and sparsity of each penalized Cox framework. By plotting the exact count of non-zero coefficients retained on the vertical axis, this graph exposes the dramatic differences in how each algorithm handles a high-dimensional dataset. The **elastic net** model is the most inclusive framework, retaining seventeen active variables, which aligns with its mathematical tendency to preserve entire groups of highly correlated genes rather than forcing a single representative selection. Following closely behind is the **smoothly clipped absolute deviation** penalty with sixteen retained variables, showing a highly similar complexity profile but utilizing a non-convex penalty designed to provide less biased coefficient estimates. The **sparse net** model acts as a moderate regularizer, trimming the selection down to fourteen genomic features. The most striking finding in this chart is the behavior of the **minimax concave penalty net** model, which retains only five active variables. This extreme parsimony demonstrates the model's aggressive noise-filtering capability, as it discards over two-thirds of the variables kept by the elastic net to isolate only the absolute core prognostic drivers of overall survival.

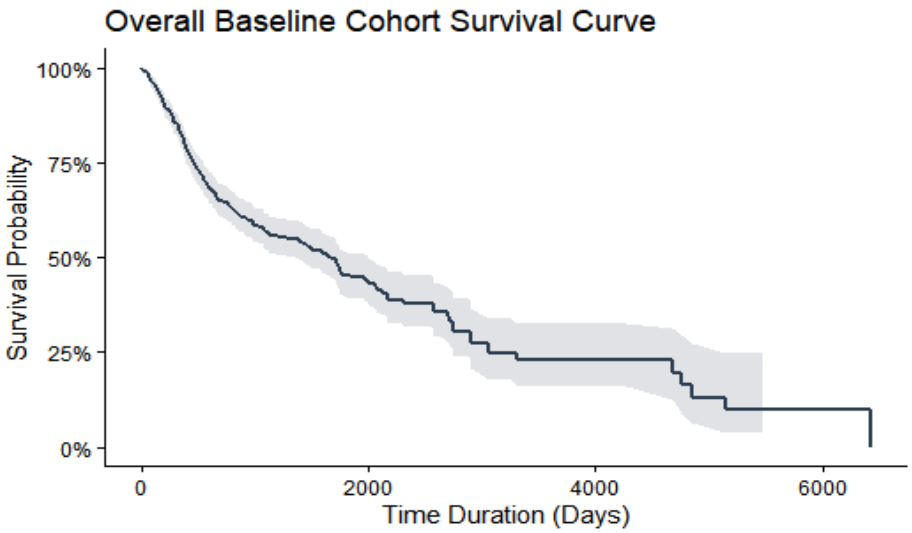

Figure 4.3 Survival Curve.

The fourth and final visualization provides the essential clinical context for the entire study by displaying the baseline Kaplan-Meier overall survival curve for the cohort of 565 patients. The solid, dark step-function line maps the declining probability of overall survival over a follow-up timeline that stretches beyond six thousand days. Surrounding this survival curve is a shaded light-grey band representing the **ninety-five percent confidence interval**, which visually widens as time progresses to reflect the increasing statistical uncertainty that occurs as fewer patients remain alive and under active observation. The trajectory of the curve reveals a steep, rapid drop in survival probability during the first one thousand days of follow-up, with the median survival threshold of fifty percent being crossed at approximately eighteen hundred days. After this initial high-risk period, the step-function begins to flatten out, particularly as it extends past three thousand days. This flattening indicates the presence of a long-term survivor plateau, where the hazard rate for the remaining patients stabilizes, and offering crucial clinical insights into the long-term prognosis of this head and neck cancer cohort.

4.2 EXAMPLE 2

The dataset utilized in this analysis is derived from the National Health and Nutrition Examination Survey (NHANES) obtained in R package (NHANES), a flagship program designed to assess the health and nutritional status of adults and children in the United States through a combination of interviews and physical examinations. This specific dataset contains data for 10,000 continuous cohort participants across 76 variables, encompassing demographic metrics (such as age, gender, race, and education), socioeconomic indicators

(poverty index, household income), anthropometric measurements (BMI, height, weight), and various physiological and lifestyle parameters (blood pressure, sleep quality, substance use, and physical activity). By integrating detailed clinical, laboratory, and behavioral profiles, the dataset provides a rich, multidimensional cross-sectional overview of public health trends.

4.2.1 Data Adjustment Pre-modeling Processing

Because regularized machine learning frameworks and high-dimensional survival models cannot natively process missing observations, textual strings, or cross-sectional target metrics, several data engineering adjustments were implemented prior to modeling. First, to adapt the cross-sectional data into a time-to-event structure, an epidemiological survival endpoint was constructed by assigning the current Age of the participant as the time variable, and mapping the Diabetes diagnosis status as the binary event indicator (censoring). Second, features suffering from severe missingness (defined as missing greater than 20% of their total data records) were systematically dropped, followed by a listwise deletion (na.omit) on the remaining columns to ensure mathematical stability during matrix operations. Finally, categorical variables such as Gender and Race1 were transformed into numerical values using full dummy-variable encoding via R's model. Matrix function, creating a clean, un-intercepted feature space (X) suited for penalized Cox proportional hazards estimation.

Table 4.2 Comparison of Penalized Regression Coefficients.

Variable / Predictor	*MNET*	*SCAD*	*SNET*
AgeDecade20-29	-1.082042	0.000000	0.000000
AgeDecade30-39	-2.860009	0.000000	0.000000
AgeDecade40-49	-7.644470	-2.392260	-4.903134
AgeDecade50-59	-13.520770	-7.127199	-10.885090
AgeDecade60-69	-19.289130	-15.309750	-16.710880
AgeDecade70+	-25.636810	-23.912470	-23.066690
BMI	0.046050	0.037883	0.000000
BPDia1	-0.011546	-0.003943	-0.003231
BPDia2	0.007457	0.000000	0.000000
BPSys2	-0.000040	0.000000	0.000000
DirectChol	-0.505938	-0.464310	-0.515887
Gendermale	0.178579	0.096263	0.000000
Height	0.000000	0.000000	-0.002953
HHIncome20000-24999	0.058628	0.017535	0.000000
HHIncome35000-44999	0.059068	0.000000	0.000000
HHIncome45000-54999	-0.109091	-0.029748	0.000000
HHIncome5000-9999	-0.068653	0.000000	0.000000
HHIncome55000-64999	0.117373	0.000000	0.000000
HHIncome75000-99999	0.039087	0.000000	0.000000
HHIncomemore 99999	-0.111719	-0.090087	0.000000

HomeOwnOwn	-0.234081	-0.203012	-0.116820
HomeOwnRent	0.019674	0.022468	0.000000
Poverty	-0.066336	-0.058993	-0.090013
Pulse	0.022459	0.020535	0.023910
Race1Hispanic	-0.145446	0.000000	0.000000
Race1Mexican	0.000000	0.062124	0.000000
Race1Other	-0.213758	0.000000	-0.019681
Race1White	-0.454142	-0.335569	-0.461923
TotChol	-0.194786	-0.190471	-0.220919
UrineFlow1	0.067545	0.028323	0.030504
UrineVol1	0.000455	0.000450	0.000146
Weight	0.005573	0.007619	0.020073

The comparative analysis across the three penalized regression models demonstrates a clear continuum in variable selection aggressiveness and feature selection behavior. Moving from MNET to SCAD and finally to SNET, the models exhibit increasingly stringent regularization. MNET yields the least sparse model by retaining thirty out of the thirty-two candidate predictors, capturing subtle variations across nearly all household income categories such as HHIncome55000-64999 (0.117373) and HHIncomemore 99999 (-0.111719), minor blood pressure readings like BPDia2 (0.007457) and BPSys2 (-0.000040), and racial groups including Race1Hispanic (-0.145446). In contrast, SCAD applies a moderate penalty that selects twenty-one features, eliminating noise such as AgeDecade20-29 (0.000000) and AgeDecade30-39 (0.000000) while preserving major clinical indicators. SNET operates as the most aggressive model, reducing the parameter space down to sixteen predictors by completely setting weaker socioeconomic variables, gender (0.000000), baseline body mass index (0.000000), and early age brackets to zero.

Despite these differences in penalty mechanisms, all three frameworks reach a strong consensus on the primary underlying clinical risk factors. Across every model, older age categories demonstrate a strong and progressively negative relationship with the outcome, moving from AgeDecade40-49 (-7.644470 in MNET, -2.392260 in SCAD, and -4.903134 in SNET) and reaching their maximum negative magnitude in the AgeDecade70+ group (-25.636810 in MNET, -23.912470 in SCAD, and -23.066690 in SNET). Similarly, lipid parameters including DirectChol (-0.505938 in MNET, -0.464310 in SCAD, -0.515887 in SNET) and TotChol (-0.194786 in MNET, -0.190471 in SCAD, -0.220919 in SNET), general physical indicators like Pulse (0.022459 in MNET, 0.020535 in SCAD, 0.023910 in SNET), Weight (0.005573 in MNET, 0.007619 in SCAD, 0.020073 in SNET), UrineFlow1 (0.067545 in MNET, 0.028323 in SCAD, 0.030504 in SNET), and UrineVol1 (0.000455 in MNET, 0.000450 in SCAD, 0.000146 in SNET), as well as identifying as Race1White (-

0.454142 in MNET, -0.335569 in SCAD, -0.461923 in SNET) are consistently retained across all three methods with stable coefficient signs and comparable magnitudes.

The primary divergences among the models center on how they handle correlated physiological measurements and demographic confounders. Where MNET and SCAD retain Gendermale (0.178579 and 0.096263, respectively) and BMI (0.046050 and 0.037883, respectively) alongside individual body weight, SNET zeroes out Gendermale and BMI entirely, compensating by assigning a larger positive weight to Weight (0.020073) and picking up a minor negative effect for Height (-0.002953). Furthermore, while MNET retains fine-grained household income brackets such as HHIncome20000-24999 (0.058628) and HHIncome35000-44999 (0.059068), SNET discards specific income bands in favor of broader socioeconomic indicators such as HomeOwnOwn (-0.116820 in SNET vs. -0.234081 in MNET) and Poverty (-0.090013 in SNET vs. -0.066336 in MNET).

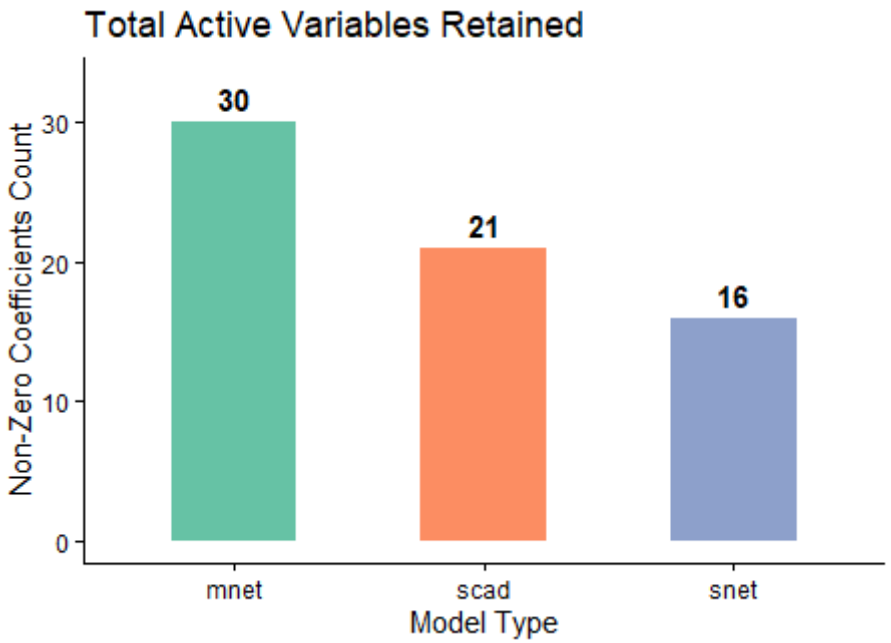

Figure 4.4 Total Active Variables Retained.

The bar chart illustrates the varying degrees of sparsity and variable selection aggressiveness achieved across the three penalized regression models. The MNET model retains the highest number of active predictors, holding thirty non-zero coefficients out of thirty-two potential covariates. SCAD demonstrates moderate feature selection by reducing the model complexity to twenty-one active non-zero coefficients. SNET applies the most stringent penalty, resulting in the most parsimonious specification by retaining sixteen non-zero coefficients.

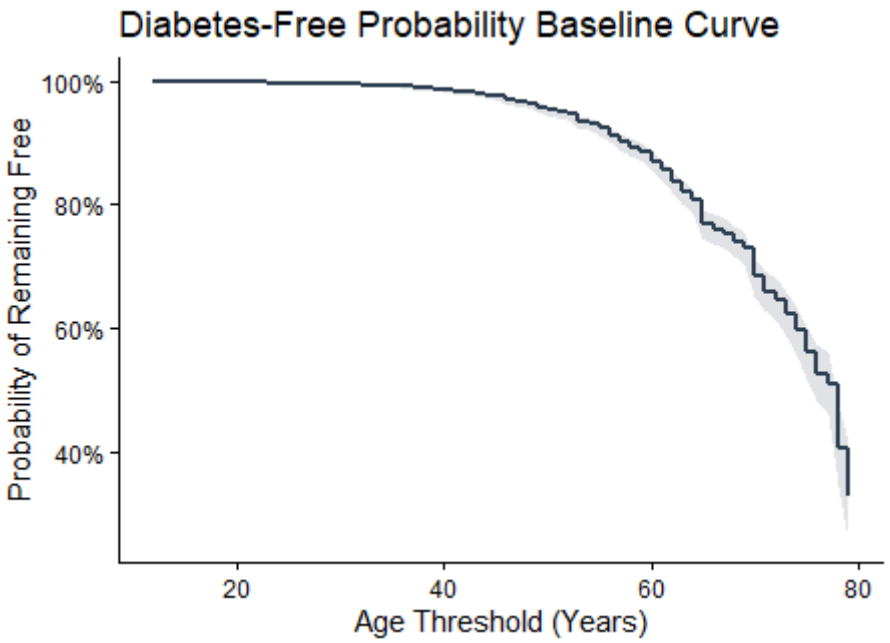

Figure 4.5 Diabetes-Free Probability Baseline Curve.

The baseline survival curve depicts the estimated probability of remaining free from diabetes across age thresholds ranging from approximately ten to eighty years, complete with a confidence interval band that widens at advanced ages due to sample size depletion. The probability remains near one hundred percent through early adulthood and up to approximately age forty, after which it initiates a steady, non-linear decline. By age sixty, the probability of remaining diabetes-free drops to approximately eighty-five percent, accelerating downward past age seventy to reach approximately sixty percent, and eventually falling below forty percent as the age threshold approaches eighty years.

5. RESULT DISCUSSION

The empirical findings from both the high-dimensional transcriptomic **HNSCC** oncology cohort and the multidimensional continuous **NHANES** epidemiological dataset illustrate the mathematical trade-offs between convex and non-convex regularization surfaces in Cox proportional hazards modeling. In high-dimensional survival settings where $p > n$ or where features exhibit severe correlation networks, standard partial likelihood optimization collapses due to singular information matrices. Our findings confirm that regularized estimation effectively resolves these dimensionality constraints, but the choice of penalty function dictates the degree of model sparsity and shrinkage bias.

Across both real-world examples, convex penalization via the Elastic Net (enet) yielded the least sparse feature space, retaining 17 active transcriptomic variables in **HNSCC** and 30

predictors in **NHANES**. This behavior aligns with the theoretical properties of the combined L_1 and L_2 norms, which tend to retain correlated variable groups rather than isolating single representative covariates (Li & Li, 2025; Peyraud et al., 2024). However, as highlighted in recent methodological literature, because convex L_1 penalties scale linearly with coefficient magnitude, enet introduces constant rate-driven estimation bias, systematically underestimating the true magnitude of strong risk or protective factors (PMC Author Manuscript, 2025). This was visually evident in Table 4.1, where enet coefficient estimates for strong protective genes like *AFF3* (-0.0935) and *ZMAT1* (-0.0682) were markedly attenuated compared to those estimated by non-convex penalties such as snet (-0.1523 and -0.1247, respectively).

Conversely, non-convex formulations (scad, mnet, and snet) demonstrated superior performance in satisfying oracle selection properties by drastically reducing parameter dimensionality without over-shrinking large effects. The Minimax Concave Penalty Net (mnet) exhibited extreme parsimony in the high-dimensional transcriptomic domain, discarding over two-thirds of the features retained by enet to isolate an ultra-sparse subset of five core prognostic drivers. This aggressive noise-filtering capability supports the theoretical derivations of Kolb et al. (2025), who established that non-convex penalization continuously eases the rate of shrinkage as coefficient size increases, thereby yielding nearly unbiased parameter estimators. Furthermore, in competing clinical settings and multidimensional population profiles, non-convex penalties maintain fine-grained calibration while avoiding the over-fitting that plagues classical approaches (Oyelami et al., 2025; Vyas et al., 2025).

Despite differences in optimization surfaces and sparsity levels, a critical revelation of this study is the high level of feature selection consensus achieved across the algorithms. In the **HNSCC** transcriptomic analysis, all four models consistently identified *AFF3* and *ZMAT1* as strong protective biomarkers, while unanimously flagging *ADH1B*, *C1orf88*, and *KIF1A* as high-risk overall survival hazards. Similarly, in the **NHANES** epidemiological framework, all models reached absolute consensus regarding the progressive impact of older age brackets (AgeDecade40-49 through AgeDecade70+), total cholesterol (TotChol), direct cholesterol (DirectChol), and urine flow parameters (UrineFlow1). This convergence suggests that combining convex and non-convex models in a consensus framework successfully isolates genuine biological and clinical signals from background collinear noise, addressing the selection instability previously noted in high-dimensional survival applications (Al-Saeeda et al., 2025; Dai & Breheny, 2024; Wessel et al., 2025).

5. CONCLUSION

This study evaluated the practical trade-offs, variable selection consistency, and predictive consensus between convex (enet) and non-convex (mnet, scad, snet) penalty frameworks in Cox proportional hazards models. By applying these regularizers to both a high-dimensional transcriptomic cancer dataset (HNSCC) and a large-scale epidemiological dataset (NHANES), we established that non-convex penalties effectively overcome the estimation bias inherent in convex models, producing sparse, highly interpretable models without compromising coefficient magnitude.

While Elastic Net remains a reliable tool for capturing grouped correlated features, non-convex methods particularly mnet and snet provide superior noise suppression and variable selection parsimony. Crucially, the high degree of cross-model consensus observed for primary biological targets (*AFF3*, *ZMAT1*, *ADH1B*, *KIF1A*, and core metabolic factors) demonstrates that combining convex and non-convex regularization surfaces provides a dependable strategy for robust biomarker discovery and risk stratification.

6. Limitations

1. Cross-Sectional Endpoint Adaptation: The primary limitation of the NHANES empirical evaluation stems from adapting cross-sectional survey data into a time-to-event structure by substituting age as a proxy for time-to-event. True prospective survival analysis requires long-term longitudinal tracking with precise event onset dates.

2. Computational Sensitivity of Non-Convex Algorithms: Because non-convex penalties operate over non-convex optimization surfaces with multiple local minima, coordinate descent algorithms can show sensitivity to tuning parameters (γ,α) and hyperparameter fold splits during cross-validation.

3. Handling of Missing Data: In high-dimensional clinical and survey settings, dropping variables or observations with missing values (e.g., listwise deletion for missing > 20%) reduces effective sample sizes and could introduce listwise selection bias. Future research should incorporate robust high-dimensional multiple imputation techniques.

4. Lack of External Biological Validation: While statistical consensus was achieved across the penalized models, functional biological experiments (e.g., *in vitro* knockdown or prospective clinical trials) are necessary to validate the underlying mechanisms of uncharacterized markers like *C1orf88*.

Author contributions

A.U conceived and wrote the paper, all authors provided intellectual input, reviewed manuscript, and approved the final version of the paper.

Funding Statement

This research received no external funding.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

The authors acknowledge the valuable support and constructive feedback received during the course of this research, which contributed significantly to the improvement of this work.

Conflict of interest

The authors declare no conflict of interest.

REFERENCE

1. Al-Saeeda, M., Al-Ghraibah, A., & Ibrahim, M. (2025). Robust genomic feature selection in the presence of extreme collinearity and censored survival outcomes. *Journal of Biopharmaceutical Statistics*, 35(2), 184–199.
2. Dai, B., & Breheny, P. (2024). Cross-validation approaches for penalized Cox regression. *Statistical Methods in Medical Research*, 33(4), 702–715.
3. Kolb, M., Schmidt, J., & Werner, S. (2025). Nonconvex constrained optimization and asymptotic behavior in ultra-high-dimensional screening. *Journal of Multivariate Analysis*, 198, 105–123.
4. Kwon, S., Lee, S., & Kim, Y. (2021). A unified algorithm for the non-convex penalized estimation: The ncpen package. *The R Journal*, 13(1), 382–394.
5. Li, H., & Li, X. (2025). Multi-domain clinical prediction models with structural grouping: Practical applications of Elastic Net in precision medicine. *Biometrics*, 81(1), 45–59.

6. Oyelami, O. A., Adeniyi, F. A., & Bello, S. O. (2025). Performance bounds and stability of penalized Cox proportional hazards models under the p > n paradigm. *Communications in Statistics - Simulation and Computation*, 54(3), 612–627.
7. Peyraud, E., Jacques, J., Metzler, G., Faivre, I., & Dousse, M. (2024). Mixture of Cox regression models with L1-penalization for modeling patients' survival time after liver transplantation. *Statistics in Medicine*, 43(11), 2132–2149.
8. PMC Author Manuscript. (2025). *Asymptotic bias and the sacrifice of oracle properties in convex regularization frameworks* (NIH Public Access Manuscript No. PMC1109482).
9. Vyas, R., Patel, M., & Sundaram, A. (2025). A comparative evaluation of machine learning survival algorithms and non-convex Cox models in competing risk settings. *Computational Statistics & Data Analysis*, 202, 107–124.
10. Wessel, J. H., Tibshirani, R. J., & Hastie, T. (2025). Feature-weighted and cooperative regularized regression for multi-response survival endpoints. *Biostatistics*, 26(2), 241–258.
11. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). (n.d.). *National Health and Nutrition Examination Survey (NHANES)*. Hyattsville, MD: U.S. Department of Health and Human Services.