

## Assignment 4: Survival Analysis- Binal Patel

**Deadline: Friday, March 14<sup>th</sup> 2025**

### Part 4: Short-Answer Questions (Upload to Blackboard as a .PDF)

Answer the following based on your work in **Part 3**:

1. Please provide the link to your public **GitHub repository**.

<https://github.com/Binal-1805/BINF-5507>

2. Based on your work in Part 3, please fill out the following table:

Method	Strengths	Limitations	Key Use Cases	Your Observations
<b>Kaplan-Meier</b>	<ul style="list-style-type: none"><li>- Non-parametric and simple to implement and interpret.</li><li>- Excellent for visualizing survival over time.</li><li>- Minimal assumptions required.</li></ul>	<ul style="list-style-type: none"><li>- Univariate only (does not adjust for covariates).</li><li>- Limited in explaining the effect of multiple predictors.</li></ul>	<ul style="list-style-type: none"><li>- Exploratory survival analysis.</li><li>- Comparing survival curves between groups using log-rank tests.</li></ul>	<ul style="list-style-type: none"><li>- The survival curves clearly differentiated age groups, with a log-rank test p-value of <math>\sim 1.99e-36</math> indicating highly significant differences</li></ul>
<b>Cox Proportional Hazards</b>	<ul style="list-style-type: none"><li>- Allows adjustment for multiple covariates and estimates hazard ratios.</li><li>- Provides interpretable</li></ul>	<ul style="list-style-type: none"><li>- Assumes proportional hazards and a log-linear relationship between covariates and the log hazard.</li></ul>	<ul style="list-style-type: none"><li>- Multivariable survival analysis.</li><li>- Identification and quantification of risk factors.</li></ul>	<ul style="list-style-type: none"><li>- The Cox model achieved a concordance index of 0.70, with Age and advanced stages (e.g., Stage_IVB) emerging as</li></ul>

## Assignment 4: Survival Analysis- Binal Patel

**Deadline: Friday, March 14<sup>th</sup> 2025**

	model coefficients.  - Widely used in clinical research.	- Sensitive to violations.		significant predictors, though some PH violations were noted.
<b>Random Survival Forests</b>	- Non-parametric, capturing complex non-linear relationships and interactions.  - Robust to violations of the PH assumption.  - Generally, yields good predictive performance	- Less interpretable than the Cox model.  - Computationally more intensive.  - Effect sizes are not as straightforward.	- Prediction-focused survival analysis, especially in high-dimensional or complex datasets.	- RSF produced a slightly higher C-index (0.72) than the Cox model and consistently highlighted Age and Stage_IVB as top predictors.

3. What are the primary differences between Kaplan-Meier (KM) analysis, Cox regression, and Random Survival Forests?

Ans: Kaplan-Meier (KM) is a non-parametric method that estimates survival probabilities over time and typically compares two or more groups. It does not directly model covariates-groups are analyzed separately and differences tested (eg., Log-rank test).

Cox Regression, also known as the Cox proportional hazard model, is a semi-parametric survival analysis technique that incorporates multiple covariates (eg., age, treatment, and clinical factors) to estimate hazard ratios. It assumes proportional hazards (the relative hazard for a covariate is constant over time).

## Assignment 4: Survival Analysis- Binal Patel

**Deadline: Friday, March 14<sup>th</sup> 2025**

Random Survival Forests is a machine learning approach that does not assume proportional hazards and can handle many covariates (including non-linear relationship and interactions). It is more flexible but can be considered a “black box” compared to Cox PH, and typically requires more data to achieve reliable results.

4. What assumptions are made by Cox Proportional Hazards regression? How can these be evaluated?

Ans: Cox Proportional Hazards regression makes several key assumptions:

1. Proportional hazard assumption  
Evaluation: Check Schoenfeld residual plots and conduct statistical tests to ensure hazard ratios are constant over time
2. Log-Linear Relationship  
Evaluation: Inspect martingale residual plots to verify that continuous covariates have a linear relationship with the log hazard; transform if needed.
3. Independent censoring  
Evaluation: Rely on study design to assume censoring in random; verify there's no systematic bias between censored and uncensored subjects.
4. Accurate covariate measurement  
Evaluation: Ensure data quality through validated measurement methods and thorough data cleaning.

Visual diagnostics (eg.; using lifelines `check_assumptions` function) and formal statistical tests are commonly used to evaluate these assumptions.

5. Which method provided the best balance between interpretability and predictive performance?

Ans: Based on the results, Cox regression likely provides the best balance between interpretability and predictive performance. Although the Random Survival Forest (RSF) model yields a slightly higher concordance index (0.72 vs. 0.70 for Cox), the gain in predictive accuracy is marginal. Cox regression's hazard ratios are straightforward to interpret and the model framework is well understood, whereas RSF—while potentially more flexible—can be more opaque. Therefore, for most

## **Assignment 4: Survival Analysis- Binal Patel**

**Deadline: Friday, March 14<sup>th</sup> 2025**

clinical applications where interpretability is crucial, Cox regression offers a good compromise between ease of understanding and predictive ability.

6. Identify any features that consistently demonstrate predictive power across different methods and highlight their potential clinical significance.

Ans: From both the Cox regression and the Random Survival Forest analysis, Age stands out as a constantly important predictor (highly significant in Cox and top-ranked in RSF). In addition, advanced tumor stage (particularly Stage IVB) and certain treatment modalities (eg.; RT alone) also show up as influential across the two methods. Clinically it suggests that older patients, those with more advanced disease stage, and those receiving specific radiation based regimens may face higher risks, highlighting these factors as key targets for risk stratification and tailored therapeutic decisions.