Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)

Modeling for Competing Risk Regression in Survival Analysis with Application in Breast Cancer Disease

[1]Sami Ali Obed, [2] Prof. Dr. Kurdistan I. Mawlood

[1]Salahaddin University -Erbil , College of Administration and Economics- Statistics and Informatics Department, Sami.obed@su.edu.krd

[2]Salahaddin University -Erbil, College of Administration and Economics- Statistics and Informatics Department , Kurdistan.mawlood@su.edu.krd

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract:

Competing risks regression is an essential component of survival analysis, particularly when there are several possible event types that prevent additional events from being seen. This paper investigates the modelling and analysis of competing risks in time-to-event data. The study was applied to a sample size of (4420) patients with Breast cancer. The data was obtained from Rizgari Hospital in the period from 1st of January 2019 to 31st of August 2024.

In survival analysis, a competing risk occurs when an event (such as death from a cause other than breast cancer) precludes the occurrence of the primary event of interest (e.g., breast cancer-specific mortality).

The aim f this study is to model the Competing Risk Regression, which treatment effect or risk on our patients which classic survival models such as the Kaplan-Meier estimator may not sufficiently address, in order to assess the likelihood of a particular occurrence in the presence of conflicting hazards. Based on the opinions of the doctors the Age and Family History factors that the risk variable on breast cancer disease, our results show that laterality, family history as a significant predictor in two models (sub distribution hazard model and cause specific hazard model) age is significant effects in sub distribution hazard and hormone presents as a significant predictor in cause specific hazard model. Furthermore, we compare sub-distribution hazard models with cause-specific hazard models by using AIC and BIC measures so as to determine which model better matches with our data, and we find that the cause-specific hazard models a more reliable match to the data. This study emphasizes how crucial it is to choose the right models for analyzing Competing risks in order to agreement precise predictor and outcomes that are easy to understand, especially in biomedical research. To support practical application, packages that implement competing risks analysis like STATA and NCSS.

Keywords: Competing Risks, Cumulative Incidence Function (CIF), Survival Analysis, Censoring, Breast Cancer.

1. Introduction to Survival Analysis

Survival analysis is an area of statistics that deals with the analysis of time-to-event data. The events of interest can be diverse, such as death, failure, relapse, default, and so on. Time-to-event data is characterized by an interest in the time until the occurrence of a particular event. The main goal of

survival analysis is to investigate the time until the occurrence of an event of interest. Survival analysis aims to characterize the occurrence of time-to-event variables and possibly identify factors associated with the occurrence(Nagpal et al., 2020a).

The presence of variables, or predictors, whose effects are altered when a competing risk event takes place, is one of the key problems with survival analysis. The marginal effect of some significant covariates may be significantly skewed if such covariate modification effects are not taken into consideration. The Kaplan-Meier method and the Cox proportional hazards model are the primary tools used in medical research analysis. However, if a study event has competing risks and only performs a single failure analysis, the Kaplan-Meier technique and Cox analysis lead to biased estimations of survival functions and marginal effects of significant covariates. Despite the advancements in multistate models that accurately address multiple hazards by taking into account all competing risks of failure, the intricacy of model assumptions and computation development make it difficult to apply these methods to clinical situations(Austin et al., 2021a; Obed et al., 2021).

1.1. Definition of Survival Analysis

Survival analysis is a branch of statistics that deals with the analysis of time-to-event data. The events of interest vary from application, including failure of mechanical systems, death of patients, default of credit card holders, and loss of customers. In the context of medical research, the time until death can be called “survival time,” the occurrence of death an “event,” and those who died “failures,” while “censored” subjects who did not experience the event (death) during the observation period(Austin et al., 2021b; Kurdistan Ibrahim Mawlood et al., 2019; Nagpal et al., 2020b).

$S(t) = 1 - F(t)$

$$S(t) = P(T > t), \qquad 0 < t < \infty \qquad \dots\dots (1)$$

$$S(t) = \int_t^{\infty} f(u)du$$

$$\hat{S}_{(t)} = \frac{1}{n}\sum_{i=1}^{n} I(t_i < t)$$

Where I is the indicator function and t_i is the survival time for the i^{th}individual, i = 1, …, n(Latouche et al., 2007).

2. Competing Risk models in Survival Data

Concerning the event of interest, subjects can be either observed (either event occurrence or censoring) or unobserved (in case of censoring prior to the end of the follow-up). Observed times can be either continuous or discrete, while unobserved times are always treated as larger than the maximum observed time. besides the event of interest, it is possible to have other events that, if they occur, will preclude the observation of the event of interest. These other events are known as competing risks. Several examples can be outlined, an interesting example arises from breast cancer studies where the event of interest is death from the disease; however, there are other possible events that may preclude observation of this event, namely death from other causes, loss to follow-up, and administrative censorship, i.e., when the study data are obtained from a health system with a time limit. Another interesting example arises from studies with patients subjected to cardiovascular catheterization and either returning to the hospital or death caused by cardiac or non-cardiac events, where the event of

interest is the death caused by a cardiac event. Competing risks in survival data can only be analyzed by appropriate statistical methods(Austin et al., 2021a).

When studying the primary event of interest, a variety of other events may occur within a population which has important implications. First, individuals may be censored due to reasons such as administrative end of follow-up, competing event occurrence, or other reasons not of interest. If subjects are only censored due to loss of follow-up, then the risk of censorship is not affected by the occurrence of the competing event; the competing risks are non-informative with respect to the censorship process. Second, individuals may experience a different event of interest, which likely has different associated risk factors or profiles, and which may prevent the future occurrence of the event of interest(Bellach et al., 2020; Ulanowski et al., 2024).

Competing-risks analysis is an extension of traditional survival analysis. For the latter, statistical techniques such as the log-rank test, the Kaplan-Meier estimate, and Cox regression are widely used in many applications.

3.1 Types of Competing Risks

1. Censoring

Censoring is the term used to describe conditions in which certain people do not observe the main event of interest throughout the research period. This may happen if a patient takes out from the study, if they are still alive at the conclusion, or if other unrelated conditions make it impossible to see the main event. To prevent bias in the analysis, censoring must be handled independently even though it is not a competitive event.

2. Competing Events

Alternative outcomes that could stop the main event of interest from happening are known as competing events. For example, death from other causes, such cardiovascular disease, is a competitive risk in breast cancer survival studies because it prevents the measurement of mortality from breast cancer. Since they must be taken into consideration in order to produce accurate survival probability, these competing events are essential to competing risk regression models.

3. Cumulative Incidence

The Cumulative Incidence Function (CIF) calculates the probability of observing a particular event over time. By calculating the probability of each type of event occurring across time, CIF provides an improved understanding of the probability of each outcome than the classic survival function, which may overstate risk by ignoring competing events(Chihara et al., 2011; Mozumder et al., 2023).

3.2 Competing Risks Regression Models:

A competing risks scenario occurs when individuals can experience one of several different events, each of which precludes the occurrence of the other events. In survival analysis, censoring and left truncation are common mechanisms that remove information on the survival times of at least some individuals. A right-censored time is one in which follow-up ended prior to the relevant event occurring. A left-truncated time is one at which observation of individuals began after the survival time of interest occurred. The most common motivations for studying time-to-event data arise from

time-to-death studies, where the event of interest is the death of an individual. The primary tool of analysis of such data in the absence of competing risks is the Cox proportional hazards model, which relates a vector of independent variables to a time-to-event outcome(Cheng & Fine, 2012).

3.2.1 Cumulative Incidence Function (CIF): is a statistical method used in survival analysis that calculates, while accounting for competing risks. The likelihood that a certain event will occur by a given time. Other possible events that could stop the event of interest from happening or change its likelihood are known as competing risks. Given that the occurrence of one event might influence the likelihood of witnessing another, CIF offers a more precise evaluation of the likelihood of each type of event(Lee et al., 2014).

$$CIF_k(t) = P(T \leq t, event\ type = k) \quad \dots (1)$$

Where:

T is the time to the occurrence of the event.
k denotes the specific event type among the competing risks.
The CIF can be expressed in terms of the cause-specific hazard function $h_k(t)$ as follows:

$$CIF_k(t) = \int_0^t S(u)\, h_k(u) du \ \dots (2)$$

Where:

- S (u) = $\exp(\int_0^t \sum_{j=1}^{k} h_j(s) ds$) is the overall survival function, representing the probability of surviving (not experiencing any event) up to time u.
- $h_k(u)$ is the cause-specific hazard function for the event type k at time u.
- K is the total number of competing risks(Austin & Fine, 2017; Shen & Yang, 2010).

3.2.2 Cause-Specific Hazard Model:

is a method used in survival analysis to simulate the likelihood of a certain kind of event when there are conflicting hazards. It functions on the premise that after one possible event happens, all subsequent ones are censored and are not witnessed after that, models the impact of factors on the risk associated with each kind of event independently. Every event type's risk (or hazard) is determined by the model separately from other event types. It treats the other competing events as filtered and only takes into account the occurrences of the individual event of interest(Austin & Fine, 2017).

The cause-specific hazard function for event type k at time t is given by:

$$h_k(t) = \lim_{\Delta t \to 0} \frac{pr(t \leq T \leq t + \Delta t, event\ type = k | T \geq t)}{\Delta t} \dots (3)$$

Where:

T is the time to the occurrence of any event.

k denotes the specific event type among the competing risks.

Δt is a small time interval.

For event type k, the cause-specific hazard model can be expressed as:

$$h_k(t|X) = h_{k0}(t) \exp(\beta_k^T X) \dots (4)$$

Where:
$h_k(t|X)$ is the cause-specific hazard for event type k at time t given covariates X.
$h_{k0}(t)$is the baseline hazard for event type k.
β_k The vector of regression coefficients for event type k.

Benefits: Easy to understand and permits independent evaluation of every conflicting risk(Austin et al., 2020).

3.2.3 Sub distribution Hazard Model (Fine and Gray's):

Sometimes referred to as the Fine-Gray model, is a survival analysis technique that models the Cumulative Incidence Function (CIF) directly when there are conflicting risks. The Fine-Gray model provides more intuitive measurements for assessing the effect of variables since it clearly predicts the chance of each event type occurring over time, in contrast to the cause-specific hazard model, which focuses on the immediate risk of each event type independently(Mohammad et al., 2017).

The Sub-Distribution Hazard Model's Main Features as the Following:

• Direct Modelling of CIF: The Fine-Gray model uses the sub distribution hazard function to explain the occurrence of a particular event type while taking the influence of competing events on that event's likelihood.

• Preservation of Competing Events: This model preserves competing events, as opposed to the cause-specific hazard model, by accounting for them and estimating the cumulative incidence more precisely.

$$\lambda_k^{sub}(t) = \lim_{\Delta t \to 0} \frac{pr(t \le T \le t + \Delta t, event\ type = k | T \ge t\ or\ competing\ event\ occurred\ before\ t)}{\Delta t} \dots (5)$$

where:

T is the time to the occurrence of any event.

k represents the specific event type of interest.

Δt is a small time interval.

It is important to note that censoring in the presence of competing risks is not informative and does not perform an implicit comparison between subjects. Unlike censoring, competing risks are informative, and to separate the risks based on the event types is taxing. The no proportionality of hazards can be modelled by the F&G model by incorporating time-dependent coefficients. Both the Cox Regression (death) and the (event) of interest can be covered by this paradigm(Austin & Fine, 2017).

$$y_{i(t,x)} = \hat{\beta}_0(t) * \exp\left(\beta_i'(x_{ij}\right) \dots (6)$$

The formulation looks very similar to Cox regression model but it applies to the sub-hazard ratio.

Where:

$y_{i(t,x)}$**:** estimate hazard sub distribution function event type i (i = 1,2, …)

$\hat{\beta}_0(t)$**:** estimate baseline hazard function event type i (i = 1,2, …)

β_1' ,……, β_1'**:** regression parameter estimation event type i (i = 1,2, …)

x_{i1} … … , x_{ij}**:** predictor variable object i event type j (j = 1,2, …)

The sub distribution function is where the curve originates. In order to verify the proportionate hazard assumption, proportionate hazard model is essentially an extension of the Cox model. The opportunity density function, represented, shows that the opportunity density function for an event of type "c" can be expressed as follows(Austin et al., 2020; Mohammad Mahmood Faqe Asst Samira Muhammad Salh Dana Taha Mohammed Salih, n.d.).

$$f_c = \lim_{\Delta t} \frac{P(t \le T < t+\Delta t, C=c)}{\Delta t} \quad \dots (7)$$

Fine-Gray sub distribution hazard function for event type c can be written as follows(Bellach et al., 2020):

$$h_c(t) = \lim_{\Delta t} \frac{P(t \le T < t+\Delta t, C=c | T>t)}{\Delta t} \quad \dots (8)$$

3.3 Assumption of computing risk:

1- Competing risks arise when there is interest in a particular event at a given time, but other occurrences may prevent that event from happening. In general, competing risks techniques can be applied when studying various event kinds with an emphasis on the timing and nature of the initial occurrence.

2- For competing risks data, the cumulative incidence function (CIF), which expresses the absolute risk of an event of interest over time, is the fundamental descriptive statistic. Since the Kaplan-Meier approach overestimates the true absolute risk, it should not be applied when competing events are present.

3- Covariates can have differing effects on the absolute risk and the rate of an event of interest, which complicates competing risks. Because they investigate the relationship between variables and the absolute risk, regression models based on the CIF (such as Fine-Gray models) are crucial for prognostic research concerns and medical decision-making.

4- All event types should be modelled in a comprehensive description of competing risks data, not just the event of primary significance.

5- Computing risk models can evaluate how an intervention affects each element of a composite endpoint separately(Zhang et al., 2008).

4. Model Assessment and Selection in Competing Risks Regression

In competing risks regression, many modeling alternatives can be considered according to the form of the regression function or the adopted link function. Thus, it is crucial to assess how well the selected model represents the data. Analogous to standard survival data, a range of goodness-of-fit tests can be derived to evaluate whether a chosen competing risks regression model constitutes a good approximation for the underlying events. These are based on martingale-type residuals. Regardless of the testing instrument, model selection is not an all-or-nothing choice. Thus, individual competing risks models can be compared among themselves. Similar testing instruments to evaluate goodness of fit can be employed for this purpose. Nevertheless, competing risks regression has its peculiarities that

need to be dealt with when investigating the overall mean function, which determines the usual alternative hypothesis settings. Asymptotic distributions for the testing statistics can be complex. Recent works have derived both semi- and non-parametric approximations for the limiting distributions(Wu et al., 2021).

4.1. Goodness-of-Fit Tests

Goodness-of-fit tests are a pivotal aspect of model assessment that gauges how well a model accurately reflects the data from which it is derived. In survival data analysis, this function translates into an evaluation of how well the models suitably reflect the underlying hazard or transition functions that govern the process generating the data(Wu et al., 2021; Zhang et al., 2008).

4.2. Model Comparison Criteria

Two popular metrics for model selection, particularly in statistical and machine learning contexts, are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). By measuring goodness-of-fit for model complexity, they both evaluate the quality of the model. When comparing models, it is better to rely on the 2 log-likelihood score. The 2 log-likelihood score checks the fitted model against the saturated (most complex) model, not one which is less flexible(Shi et al., 2013).

5. Applications of Competing Risks Regression

Competing risks is a common phenomenon in many fields, especially biomedical and epidemiological research. Survival data often involves several types or causes of events, and knowing which event occurs first is of great interest. For example, in a clinical trial of a new cancer drug, the event of interest may be disease progression in case a patient dies from other causes, such as a heart attack, age or other tumors. Competing risks regression is a very important modeling tool for survival data with competing events. In such modeling, the event or cause of primary interest is the topmost focus, and all other events are treated as the competing events(Shi et al., 2013).

5.1 Data Collection:

The data for this study were collected from the Rizgari Hospital for Cancer Disease in Erbil cancer registry department for those patients that had Breast Cancer records prior to that. The data comprised 4420 cases of all Breast cancer patients and were collected throughout a five-year and eight months, period beginning in January the 1^{st} , 2019, and concluding on August 31, 2024. Of those, 3974 patients either survived the experiment or currently alive, 446 patients passed away. In this duration of time, STATA, and NCSS that the statistical techniques used to analyses the data.

Table 1: The Data Obtained (13) Variables which are Described below:

Variable	Description	Count
Gender	(1) male	45
	(2)female	4375
Grade	(1) Grade I	216
	(2) Grade II	2198
	(3) Grade III	1322
	(4) Grade IV	31

	(9) Unknown	653
Laterality	(1) Right	1700
	(2) Left	1721
	(3) Bilateral	59
	(4) Not Applicate	14
	(9) Unknown	926
Surgery	(0) No	266
	(1) Yes	1630
	(2) Suggested	152
	(9) Unknown	2372
Chemo	(0) No	208
	(1) Yes	1199
	(2) Suggested	381
	(9) Unknown	2632
Radio	(0) No	308
	(1) Yes	2147
	(2) Suggested	455
	(9) Unknown	1510
Hormone	(0) No	351
	(1) Yes	904
	(2) Suggested	272
	(9) Unknown	2893
Isotope	(0) No	1273
	(1) Yes	27
	(2) Suggested	7
	(9) Unknown	3113
Target	(0) No	940
	(1) Yes	203
	(2) Suggested	83
	(9) Unknown	3194
Family History	(0) No	989
	(1) Yes	385
	(9) Unknown	3046
Nationality	(1) Iraqi	4250
	(2) Arabi	155
	(3) Ajnabi	9
	(9) unknown	6
Occupation	(1) **Wife House**	1884
	(2) **Jobless**	2
	(3) **Farmer**	1

	(4) **Employee**	342
	(5) **Craftsman**	4
	(6) **Child**	3
	(7) **Retired**	93
	(9) **Unknown**	2041
Status	(0) Death	446
	(1) Alive	3974

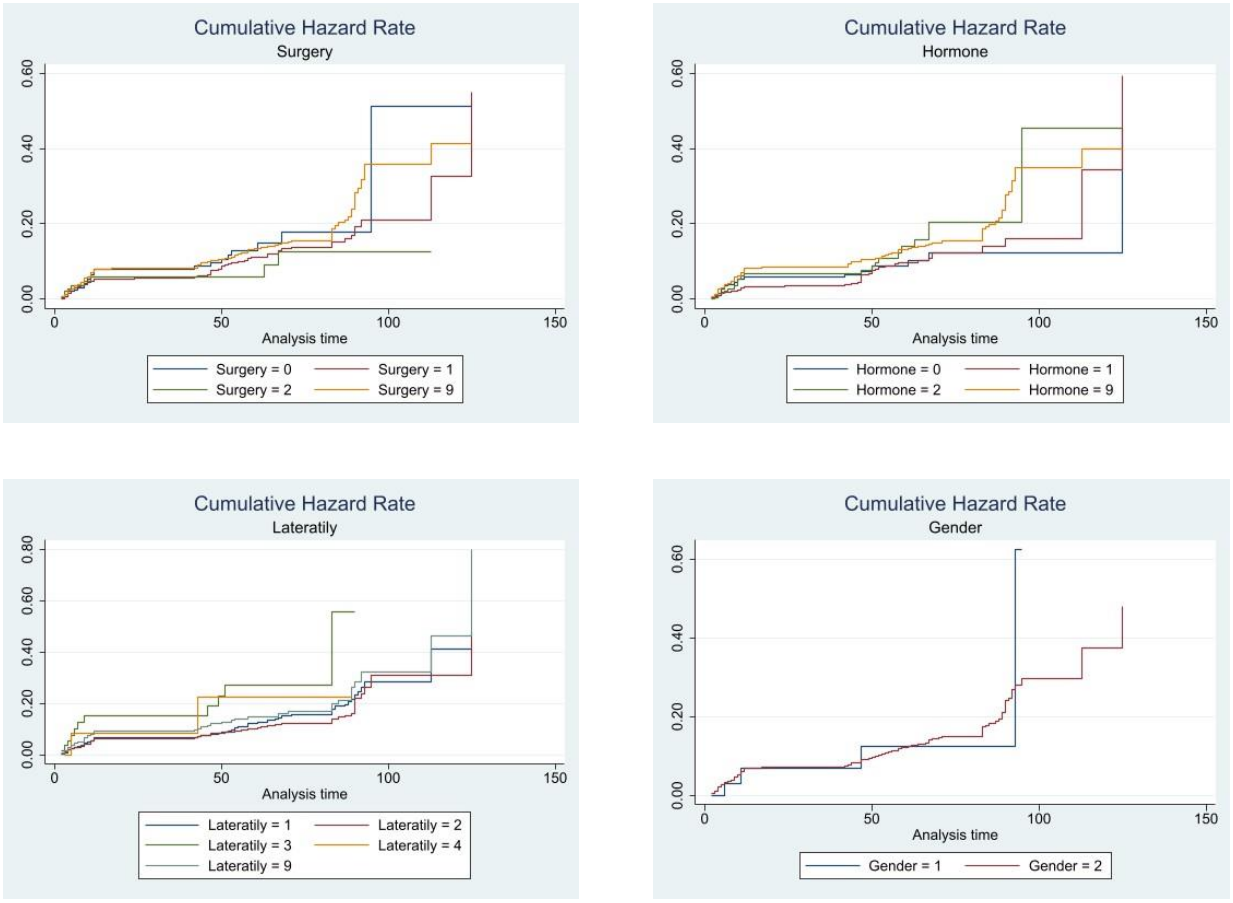

Figure 1 Cumulative Hazard Plot for Breast Cancer Data:

Figure 1 A graphical representation showing the cumulative hazard function over time, commonly used to display probabilities over time(such as the development or recurrence of breast cancer). It shows how the likelihood of an event (such as death or failure) decreases over time for different groups, with each line or symbol potentially representing a different group or treatment.

This survival plot shows survival over time for several groups (represented by different colors), with censored individuals marked by the open symbols. The shape of each curve tells us how quickly events occur in each group, while the height of the curve at any time point gives the proportion of individuals still surviving or event-free at that moment. The plot helps visualize the accumulated risk over time for the studied population to compare different patient groups (e.g., by treatment type, age, stage of cancer).

5.2 Hypothesis testing for distribution of Data

H_0: The data ~ specific distribution

$H1$: The data ~ specific distribution

Table 2: The outcomes of the survival times and distribution type analysis using chi-square and Kolmogorov-Smirnov Test

Chi-square test				Kolmogorov-Smirnov Test		
Type of distribution	**Chi square value**	**d.f**	**p-value**	**Anderson-Darling**	**d.f**	**p-value**
Weibull	**16639.8**	**97**	**0.258**	**220.114**	**98**	**0.2020**

Based on these results, the Weibull distribution appears to provide the best fit to the data, including the chi-square test, and the Anderson-Darling test, with related values such as chi-square value, degrees of freedom (d.f.), and p-value for different types of distributions (like the Weibull distribution).

1. **Chi-Square Test**: A statistical test used to examine the association between categorical variables. The chi-square value, along with degrees of freedom and p-value, helps determine if there's a statistically significant difference between expected and observed frequencies.

2. **Anderson-Darling Test**: Similar to the Kolmogorov-Smirnov test, but with higher sensitivity to deviations in the tails of the distribution. It's used to test if a sample comes from a specific distribution.

As p-value suggests no significant deviation from the expected values. The p-values (0.258) and (0.2020) for chi-square and Anderson-darling test greater than alpha=0.05 suggest that, for Weibull distribution are statistically not significant, indicating a fit of the model to the observed data or significant deviations from expected outcomes based on the tested distributions. The Weibull distribution provides the best fit to the data.

5.3 Competing Risk Regression:

Table 3: Sub distribution Hazard Model (SHR)

Factors	SHR	SE	Z	P-Value	Confidence Interval	
					lower	upper
Gender	**0.7292346**	**.1800439**	**-1.28**	**0.201**	**0.4494801**	**1.183107**
Grade	**1.020366**	**.0218827**	**0.94**	**0.347**	**0.9783654**	**1.064169**
Laterality	**1.040114**	**.0179363**	**2.28**	**0.023**	**1.005547**	**1.07587**
Surgery	**0.9844355**	**0.024757**	**-0.62**	**0.533**	**0.9370891**	**1.034174**
Chemo	**0.9810962**	**0.0280245**	**-0.67**	**0.504**	**0.9276785**	**1.03759**

Radio	**1.022119**	**0.0152605**	**1.47**	**0.143**	**0.992642**	**1.052471**
Hormone	**1.037302**	**0.0256078**	**1.48**	**0.138**	**0.9883061**	**1.088726**
Istope	**0.985886**	**0.0250894**	**-0.56**	**0.576**	**0.9379179**	**1.036307**
Target	**1.024999**	**0.0279646**	**0.91**	**0.365**	**0.9716288**	**1.0813**
Family History	**1.03894**	**0.0222414**	**1.78**	**0.004**	**0.9962496**	**1.08346**
Occupation	**0.9862976**	**0.0177633**	**-0.77**	**0.444**	**0.9520894**	**1.021735**

The table displays the findings of a competing risks regression model, which specifically used sub distribution hazard ratios (SHRs) to assess the effects of different factors on the clinical outcome's cumulative incidence function.

1. Gender: The p-value for gender is 0.201 and the SHR is 0.7292. This number indicates that there is no discernible relationship between gender and the result. There is insufficient evidence to support a gender effect, as evidenced by the inclusion of 1 in the confidence interval (0.4495 to 1.1831).
2. Grade: This factor is also not significantly linked with the cumulative incidence of the event, as indicated by the SHR of 1.0204 and p-value of 0.347. The conclusion that grade has little bearing on the result is supported by the confidence interval, which contains 1 (0.9784 to 1.0642).
3. Laterality: At the 0.05 level, laterality is statistically significant, with a SHR of 1.0401 and a p-value of 0.023. Given that 1 is not included in the confidence interval (1.0055 to 1.0759), laterality may be a potential risk factor and has a positive correlation with the result.
4. Surgery: There is no discernible impact on the result, as indicated by the SHR of 0.9844 and p-value of 0.533 for surgery. The confidence interval (0.9371 to 1.0342) includes 1, further suggesting surgery is not a significant predictor in this model.
5. Chemotherapy (Chemo): The SHR for chemotherapy is 0.9811, and it is not statistically significant, with a p-value of 0.504. There is no discernible effect of chemotherapy on the result, as indicated by the inclusion of 1 in the confidence interval (0.9277 to 1.0376).
6. Radiotherapy (Radio): With a p-value of 0.143, the SHR for radiotherapy is 1.0221. This finding indicates a minor trend towards increasing risk, but it is not statistically significant. There is not enough evidence to support an effect, as indicated by the inclusion of 1 in the confidence interval (0.9926 to 1.0525).
7. Hormone Therapy: With a p-value of 0.138, the SHR for hormone therapy is 1.0373. Although there appears to be a minor trend towards greater risk, this factor does not achieve statistical significance either, according to the confidence interval (0.9883 to 1.0887).
8. Isotope Therapy (Isotope): Isotope therapy has a statistically non-significant SHR of 0.9859 and a p-value of 0.576. One is included in the confidence interval (0.9379 to 1.0363), which suggests no discernible impact.
9. Targeted Therapy (Target): There appears to be no significant effect, as indicated by the SHR of 1.0250 and p-value of 0.365. This factor is not a significant predictor in the model, as indicated by the fact that the confidence interval (0.9716 to 1.0813) includes 1.

10. Family History: The SHR for family history is 1.0389, with a p-value of 0.004, which is less than alpha then statistical significance at the 0.05 level. Additional research is necessary due to the possible trend towards increasing risk indicated by the confidence interval (0.9962 to 1.0835).

Most factors in the model, including, gender, grade, surgery, chemotherapy, radiotherapy, hormone therapy, isotope therapy, targeted and therapy, do not show statistically significant associations with the outcome. However, laterality family history and age are identified as a significant predictor, suggesting its potential importance in the clinical context.

$$\boldsymbol{Yi_t} = \beta_0 * \exp(0.72gender + 1.02Grade + 1.04\text{Lateratily} + 0.984\text{Surgery} + 0.981\text{Chemo} + 1.02\text{Radio} + 1.037\text{Hormone} + 0.985\text{Istope} + 1.024\text{target} + 1.038\text{FamilyHistory} + 0.986\ \text{Occupation}$$

Table 4: Cause-Specific Hazard Model(CSHR)

Factors	CSHR	SE	Z	P-Value	Confidence Interval	
					lower	upper
Gender	**0.5160988**	**0.3131726**	**-1.09**	**0.276**	**0.1571149**	**1.695307**
Grade	**1.017748**	**0.0224034**	**0.80**	**0.424**	**0.9747716**	**1.062619**
Laterality	**1.037896**	**0.0177992**	**2.17**	**0.030**	**1.00359**	**1.073374**
Surgery	**0.9915479**	**0.0256609**	**-0.33**	**0.743**	**0.9425078**	**1.04314**
Chemo	**0.9829749**	**0.0281556**	**-0.60**	**0.549**	**0.9293115**	**1.039737**
Radio	**1.020308**	**0.0153669**	**1.33**	**0.182**	**0.9906298**	**1.050876**
Hormone	**1.037374**	**0.0255717**	**1.49**	**0.037**	**0.9884458**	**1.088724**
Istope	**0.9865357**	**0.0249469**	**-0.54**	**0.592**	**0.9388325**	**1.036663**
target	**1.025882**	**0.0279892**	**0.94**	**0.349**	**0.9724656**	**1.082234**
Family History	**1.039867**	**0.0221379**	**1.84**	**0.006**	**0.9973702**	**1.084174**
Occupation	**0.9862976**	**0.0177633**	**-0.77**	**0.444**	**0.9520894**	**1.021735**

The results from a statistical analysis using a Cause-Specific Hazard Model (CSHR). This type of model is often used in survival analysis to assess the effects of different variables (or covariates) on the time to a specific event, like death or disease recurrence, while accounting for other possible outcomes.

1. **Gender**:

- **CSHR = 0.516**: Suggests that gender might decrease the hazard, but it is not statistically significant (p = 0.276).
- **Interpretation**: Gender does not have a statistically significant effect on the hazard.

2. **Laterality**:

- **CSHR = 1.038**, **p = 0.030**: Indicates a significant effect, with a higher value suggesting increased risk when the event occurs on a specific side (laterality).
- **Interpretation**: Laterality has a statistically significant impact on the hazard.

3. **Hormone Therapy**:

- **CSHR = 1.037**, **p = 0.037**: Statistically significant, implying that hormone therapy is associated with a slight increase in the hazard.
- **Interpretation**: Hormone therapy has a statistically significant effect on increasing the hazard.

4. **Family History**:

- **CSHR = 1.040**, **p = 0.006**: Strongly significant, indicating that individuals with a family history have a higher hazard of the event occurring.
- **Interpretation**: Family history is a significant risk factor for the event.

Non-significant Factors

Several factors have p-values greater than 0.05, indicating no statistically significant effect on the hazard:

Grade, **Surgery**, **Chemotherapy**, **Radiotherapy**, **Isotope Treatment**, **Targeted Therapy**, and **Occupation** all have p-values greater than 0.05, suggesting these factors do not significantly influence the hazard in this study.

Statistically significant factors: Laterality, Hormone Therapy, Family History (p-values < 0.05).

Non-significant factors: Gender, Grade, Surgery, Chemotherapy, Radiotherapy, Isotope Treatment, Targeted Therapy, Occupation (p-values > 0.05).

the analysis shows that **Laterality**, **Hormone Therapy**, and **Family History** are significant predictors of the hazard, meaning they have a statistically meaningful impact on the risk of the event occurring, while the other factors do not show significant effects.

$$\boldsymbol{Yi_t} = \beta_0 * \exp\,(0.0.516gender + 1.0177Grade + 1.0378\text{Lateratily} + 0.984\text{Surgery} + 0.981\text{Chemo} + 1.02\text{Radio} + 1.037\text{Hormone} + 0.985\text{Istope} + 1.024\text{target} + 1.038\text{FamilyHistory} + 0.986\ \text{Occupation})$$

Table 5: Comparing two models Cause-Specific Hazard Model and Sub Distribution Hazard Model by AIC and BIC

Models	NO. Parameter	Log-likelihood ratio test	AIC	BIC
Cause-Specific Hazard Model	**11**	**-2816.391**	**5654.782**	**200.5521**
Sub distribution Hazard Model	**11**	**-2838.5454**	**5677.0908**	**200.5677**

Basically, the objective of table (11) aims to identify which of the two models the **Cause-Specific Hazard Model** and the **Sub-distribution Hazard Model** is better suited to our data. The AIC and BIC were calculated for each model based on two measurements, and the lowest values of both metrics were chosen. In order to compare the two models, the Bayesian information criteria and Akaike's information criterion are computed for each model.

1. **Number of Parameters**: Both models have the same number of parameters (11), which means they use the same number of variables or covariates in the analysis.

2. **Log-Likelihood Ratio Test**:

- For the **Cause-Specific Hazard Model**, the log-likelihood is **-2816.391**.
- For the **Sub-distribution Hazard Model**, the log-likelihood is **-2838.5454**.
- A higher (less negative) log-likelihood indicates a better fit to the data. Therefore, the Cause-Specific Hazard Model fits the data better than the Sub-distribution Hazard Model.

3. **Akaike Information Criterion (AIC)**:

- AIC for the **Cause-Specific Hazard Model**: **5654.782**
- AIC for the **Sub-distribution Hazard Model**: **5677.0908**
- The AIC helps compare models by penalizing the number of parameters to prevent overfitting. The model with the lower AIC is considered better. In this case, the Cause-Specific Hazard Model has a lower AIC, indicating a better balance of goodness-of-fit and complexity.

4. **Bayesian Information Criterion (BIC)**:

- BIC for the **Cause-Specific Hazard Model**: **200.5521**
- BIC for the **Sub-distribution Hazard Model**: **200.5677**
- The lower of BIC, is the better model. Again, the Cause-Specific Hazard Model has a slightly lower BIC, making it the preferred model.

Based on the comparison using AIC and BIC, the **Cause-Specific Hazard Model** is statistically better than the **Sub-distribution Hazard Model** for the given data. This indicates that the Cause-Specific Hazard Model provides a better fit to the data while appropriately balancing model complexity and goodness-of-fit.

The Cause-Specific Hazard Model with significant factors as follows:

$$\boldsymbol{Yi_t} = \beta_0 * \exp(1.04\text{Lateratily} + 1.037\text{Hormone} + +1.038\text{FamilyHistory})$$

6. Conclusion

In this study, a model on the cause-specific hazard using the Fine-Gray model for the competing risk models, and regression based on the approach as a special case of expanding the categories for transition-specific sub-distributions. For the model approach, it is noted that the necessity of model specification and verification is usually not addressed except for the Fine-Gray model. There are many possible subjects for future work to improve on the models proposed. One direction is to model the

cumulative hazard directly. A mostly intuitive advantage of the models based on the cumulative hazard is mathematical tractability.

This study explored the use of competing risks regression models, specifically the Fine-Gray model for cumulative incidence function (CIF) and the Cause-Specific Hazard Model, in analyzing time-to-event data with competing risks. Through the application of these models to a breast cancer dataset, we found that laterality, hormone therapy, and family history were significant predictors influencing the risk of event occurrence. The comparison of the Cause-Specific Hazard Model and the Sub distribution Hazard Model using AIC and BIC criteria revealed that the Cause-Specific Hazard Model provided a better fit for the data.

References:

1. Austin, P. C., & Fine, J. P. (2017). Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Statistics in Medicine*, *36*(27), 4391–4400. https://doi.org/10.1002/sim.7501
2. Austin, P. C., Latouche, A., & Fine, J. P. (2020). A review of the use of time-varying covariates in the Fine-Gray subdistribution hazard competing risk regression model. *Statistics in Medicine*, *39*(2), 103–113. https://doi.org/10.1002/sim.8399
3. Austin, P. C., Steyerberg, E. W., & Putter, H. (2021a). Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1. *Statistics in Medicine*, *40*(19), 4200–4212. https://doi.org/10.1002/sim.9023
4. Bellach, A., Kosorok, M. R., Gilbert, P. B., & Fine, J. P. (2020). General regression model for the subdistribution of a competing risk under left-truncation and right-censoring. *Biometrika*, *107*(4), 949–964. https://doi.org/10.1093/biomet/asaa034
5. Cheng, Y., & Fine, J. P. (2012). Cumulative incidence association models for bivariate competing risks data. In *J. R. Statist. Soc. B* (Vol. 74, Issue 2). www.stat.pitt.edu/yucheng.
6. Chihara, D., Oki, Y., Matsuo, K., Onoda, H., Taji, H., Yamamoto, K., & Morishima, Y. (2011). Incidence and risk factors for central nervous system relapse in patients with diffuse large B-cell lymphoma: Analyses with competing risk regression model. *Leukemia and Lymphoma*, *52*(12), 2270–2275. https://doi.org/10.3109/10428194.2011.596966
7. Kurdistan Ibrahim Mawlood, A., Sami Ali Obed, R., & History, A. (2019). QALAAI ZANIST SCIENTIFIC JOURNAL Study and Analysis of the Chest Cancer Data Using Survival Models ARTICLE INFO ABSTRACT. *A Scientific Quarterly Refereed Journal Issued by Lebanese French University-Erbil*, *4*. https://doi.org/10.25212/lfu.qzj.4.2.24
8. Latouche, A., Boisson, V., Chevret, S., & Porcher, R. (2007). Misspecified regression model for the subdistribution hazard of a competing risk. *Statistics in Medicine*, *26*(5), 965–974. https://doi.org/10.1002/sim.2600
9. Lee, J. Y., Lee, D. H., Cho, N. H., Rha, K. H., Choi, Y. D., Hong, S. J., Yang, S. C., & Cho, K. S. (2014). Impact of charlson comorbidity index varies by age in patients with prostate cancer treated by radical prostatectomy: A competing risk regression analysis. *Annals of Surgical Oncology*, *21*(2), 677–683. https://doi.org/10.1245/s10434-013-3326-6

10. Mohammad, K. A., Fatima-Tuz-Zahura, M., & Bari, W. (2017). Fine and Gray competing risk regression model to study the cause-specific under-five child mortality in Bangladesh. *BMC International Health and Human Rights*, *17*(1), 1–8. https://doi.org/10.1186/s12914-017-0112-8
11. Mohammad Mahmood Faqe Asst Samira Muhammad Salh Dana Taha Mohammed Salih, A. (n.d.). *التطبيق Estimate ث(مع (وقت احلد ة تقدير منوذج االحندار املعلمى باالعتماد على توزيع وقت البقاء على قيد احليا Parametric Regression Models Depend on (Time-To-Event) Survival Time Distributions with Application ط دانا ه حممد صاحل .آ.م.د مسرية حممد حممد صاحل .آ.م.د ف حممود حممد قى جامعة حليجة جامعة السليمانية جامعة الس ليمانية*.
12. Mozumder, S., Booth, S., Riley, R. D., Rutherford, M. J., & Lambert, P. C. (2023). *Assessment of prognostic model performance in the presence of competing risks using a cause-specific hazards approach*. https://doi.org/10.21203/rs.3.rs-3416448/v1
13. Nagpal, C., Li, X. R., & Dubrawski, A. (2020a). *Deep Survival Machines: Fully Parametric Survival Regression and Representation Learning for Censored Data with Competing Risks*. http://arxiv.org/abs/2003.01176
14. Nagpal, C., Li, X. R., & Dubrawski, A. (2020b). *Deep Survival Machines: Fully Parametric Survival Regression and Representation Learning for Censored Data with Competing Risks*. http://arxiv.org/abs/2003.01176
15. Obed, S. A., Mohammed, P. A., & Kadir, D. H. (2021). The Estimation of (Covid-19) Cases in Kurdistan Region Using Nelson Aalen Estimator. *Cihan University-Erbil Scientific Journal*, *5*(2), 24–31. https://doi.org/10.24086/cuesj.v5n2y2021.pp24-31
16. Shen, P. sheng, & Yang, Y. chen. (2010). Semiparametric estimation of the cumulative incidence function under competing risks and left-truncated sampling. *Journal of Statistical Planning and Inference*, *140*(7), 1863–1873. https://doi.org/10.1016/j.jspi.2010.01.032
17. Shi, H., Cheng, Y., & Jeong, J. H. (2013). Constrained parametric model for simultaneous inference of two cumulative incidence functions. *Biometrical Journal*, *55*(1), 82–96. https://doi.org/10.1002/bimj.201200011
18. Ulanowski, A., Ban, N., Ozasa, K., Rühm, W., Semones, E., Shavers, M., & Vaillant, L. (2024). Time-integrated radiation risk metrics and interpopulation variability of survival. *Zeitschrift Fur Medizinische Physik*, *34*(1), 64–82. https://doi.org/10.1016/j.zemedi.2023.08.002
19. Wu, W. T., Li, Y. J., Feng, A. Z., Li, L., Huang, T., Xu, A. D., & Lyu, J. (2021). Data mining in clinical big data: the frequently used databases, steps, and methodological models. In *Military Medical Research* (Vol. 8, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s40779-021-00338-z
20. Zhang, M. J., Zhang, X., & Scheike, T. H. (2008). Modeling cumulative incidence function for competing risks data. *Expert Review of Clinical Pharmacology*, *1*(3), 391–400. https://doi.org/10.1586/17512433.1.3.391