Proportional Hazard Models: Estimation, Prediction and Evaluation

Jinyi Luo[1]

[1]Statistics, George Washington University,
Email: jinyi@gwu.edu

Myelodysplastic syndromes (MDS) are group of cancers where blood cells are not maturing properly in the bone marrow. This article studies the whether IST is effective for MDS patients. Among 945 patients, 129 patients received Immunosuppressive Treatment (IST) by using antithymocyte globulin (ATG) or cyclosporine (CSA) in combination or individually. Our results show that IST is extremely helpful in treating MDS, which could extend patient's survival especially in the long run. KM and NA estimators are used When the distribution is unknown. Then proportional hazard models is built to consider multivariate case, and calculated the survival functions. Among all these models, we find that the IST treatment has a positive effect for the treatment. On the other hand, variables like AGE, NEUTRO and PLATE have tiny effect with treatment. Although the C-statistics and Net Reclassification Improvement (NRI) show that the improvement in risk prediction offered by the IST treatment is small, the competing risk model proves that IST treatment has long run positive effect on patients.

KEYWORDS
Cox Proportional Hazard Model, Estimation, Prediction, Evaluation

1 | INTRODUCTION

1.1 | Background of the Data

Myelodysplastic syndromes (MDS) are group of cancers where blood cells are not maturing properly in the bone marrow and thus do not become healthy blood cells, and patients with MDS are typically older adults with comorbidity. Shown in 2005, allogeneic stem cell transplantation is the only effective treatment that can extend survival (Wong R, Shahjahan M, Wang X, et al, 2005). In a 2008 study, Sloand et al proposed an alternative approach, which suggested that Imunosuppressive Treatment (IST) has significant positive effect on the MDS patients (Sloand, Wu, Greenberg, Young, & Barrett, 2008). Another study was able to confirm that IST is effective, and suggests the most effective form is the combination of ATG with CSA (Stahl, Deveaux, Witte, et al, 2018). 945 patients are divided into two groups, 816 patients received only supportive care reported to International Myelodysplasia Risk Analysis Workshop (IMRAW), while 129 patients received Immunosuppressive Treatment (IST) by using antithymocyte globulin (ATG) or cyclosporine (CSA) in combination or individually (Sloand et al 2008). The International Myelodysplasia Risk Analysis Workshop - Immunosuppressive Treatment (IMRAW-IST) dataset (Sloand et al 2008) contains 13 variables, which are AGE, GENDER, absolute neutrophil count (NEUTRO), platelets (PLATE), blasts cell (BLASTS), DIAGNOS, DIED, actue myeloid leukemia (AML), survival, AMLTIME, International Prognostic Scoring System (IPSS), cytogenetics (cyto.s), National Institutes of Health (NIH). There are three binary variables (DIED, AML, and NIH) with value 0 or 1 only. For variables DIED and AML, value 0 indicates the subject is censored, while 1 indicates the subject is failure. NIH =0 indicates the subject is from IMRAW, which only receive the support care, while NIH=1 indicates the subject is from NIH, and received Immunosuppressive Treatment.

1.2 | Objectives and Study Questions

In this study we evaluated the effectiveness of IST in treating MDS patients, and use various estimators as well as models to show our findings.

1.3 | Approaches and Major findings

Regardless all the other factors, we started with nonparametric models by using Kaplan-Meier (KM) and Nelson-Aalen (NA) estimators to compute the survival functions of IST patients, and compare the results with the IMRAW patients. KM estimators show that the rate of survival for IST patients after 3 years of observation are much higher than IMRAW patients. Further analyze the survival rate by incorporating AML among women and men, the result consistent with survival rate only consider time to death. NA estimators agrees with KM estimators. Given the specific distribution, parametric model like Maximum Likelihood Estimator (MLE) would be more efficient compare to the nonparametric ones, since it minimize the mean squared error. Moreover, Cox Proportional Hazard Model is used to explain all the different factors that affect the survival function. Our finding shows that IST is effective, and it decrease the hazard rate for the patients. Moreover, by adding the interaction term of NIH X GENDERM, the results show that IST effects are the same for men and women based on the IMRAW-IST dataset. By using C−Statistics, Net Reclassification Improvement (NRI), and Competing Risk Model to evaluate the performance of the cox proportional hazard model, we find that the C−Statistics are close to 0.7, which indicate that the model is functional. However, the increment value for the first 5 years is 0.0011, which is minuscule. This suggests that the improvement in risk prediction offered by IST is tiny. By calculating the NRI, the highest positive values for IST patients is of 5 years ($NRI_5^{(2)}$), which suggests that it has positive

effect in risk prediction. Last, cumulative incidence function (CIF) is used to describe event occurrence instead of survival function, and the results shows that IST treatment has positive effect on Event 2 (time to Non-AML death). On the other hand, the variables like AGE, NEUTRO, PLATE are very close to 1, which means they have little effect on Non-AML disease.

2 | METHODS

2.1 | Variables and Assumptions

In order to construct the survival models, the following variables and assumptions are denoted:
T: event time, e.g., time to death, time to disease, etc.
C: censoring time
X = T ∧ C = min{T, C}: observed time
Without censoring, i.e., C = ∞, X = T
Δ = I (T < C): event indicator
Z: covariate vectore, e.g., baseline characteristics
n: sample size, number of subjects in the study
i : subject, i = 1, . . . , n
Key assumption: $T_i \perp C_i$ for all i , given Z_i
Survival Function S(x) is a function that gives the probability an individual live beyond time x. Assume for now $C = \infty$, so that T=X. Let F(x) be the cumulative distribution function (CDF), then

$$S(x) = P(X > x) = \int_x^{\infty} f(t)dt = 1 - F(x) \tag{1}$$

Hazard Function also known as the hazard rate is:

$$\lambda(x) = h(x) = \lim_{\delta \to 0} \frac{P(x \leq X \leq x + \delta | X \geq x)}{\delta} = \lim_{\delta \to 0} \frac{P(x \leq X \leq x + \delta)}{\delta P(x \geq X)} = \frac{f(x)}{S(x)} \tag{2}$$

The cumulative hazard function is:

$$\Lambda(x) = H(x) = \int_0^x h(t)dt = -log[S(x)] \tag{3}$$

In this article, the mostly common used right censoring has been used in this article, where T_i is not known, but is known to be greater than some time.

2.2 | Kaplan-Meier Estimators

Computation of the Kaplan-Meier estimators of S(t):
Observed failure times: $t_1 < t_2 < \ldots t_{n_D}$, where n_D is the number of unique times at which deaths are observed. The

Kaplan-Meier estimator of S(t) is given by:

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} \{1 - \frac{D_j}{Y_j}\} \tag{4}$$

where Y_j=number "at risk" at $t = t_j$, D_j=number of failures at $t = t_j$, $\frac{D_j}{Y_j}$ is an estimator of $\lambda(t_j)$

2.3 | Nelson-Aalen estimators

Computation of the Nelson-Aalen estimators of $\Lambda(t)$:
Observe unique death times: $t_1 < t_2 < \ldots t_{n_D}$, at each death time, number at risk, Y_j and number of events, D_j. The Nelson-Aalen estimators of $\Lambda(t)$ is given by:

$$\hat{\Lambda}(t) = \sum_{j:t_j \leq t} \frac{D_j}{Y_j} \tag{5}$$

where Y_j=number "at risk" at $t = t_j$, D_j=number of failures at $t = t_j$, $\lambda(\hat{t}_j) = \frac{D_j}{Y_j}$

2.4 | Maximum Likelihood Estimator (MLE)

Maximum Likelihood Estimator (MLE) obtains the parameter estimates by finding the parameter values that maximize the likelihood function, and the Maximum Likelihood Estimatoion of β can be obtained by $\hat{\beta}$ which maximizes $L(\beta)$. If T_i are independent identically distributed (i.i.d.), we can write $L(\beta) = \prod_{i=1}^{n} f(T_i; \beta)$. To simplify the computation, a monotone transformation has been applied to the likelihood that converts multiplication to addition, thus the logarithm function: $\log L(\beta)$ has been calculated. Setting the score function to 0, $\hat{\beta}$ that maximizing the likelihood can be derived:

$$u(\hat{\beta}) = \frac{d log L(\beta)}{d\beta}|_{\beta=\hat{\beta}} = 0 \tag{6}$$

2.5 | Cox Proportional Hazards Model

Cox model is more flexible compare to the parametric models we discussed previously, and with this semi-parametric model, the nature of the covariate effects is specified, but not the distribution for T_i. The general expression for multiplicative proportional hazards model is:

$$\lambda_i(t) = \lambda_0(t) g(\beta, Z_i) \tag{7}$$

where g(.) is the link function, and when $g(\beta, Z_i)$=exp$\{\beta^T Z_i\}$, it is known as Cox Model.

By using Cox Model, when $\lambda_0(t)$ is unknown, the expression of the conditional survival function

$$Si(t|Zi) = P(Ti > t|Zi) = S_0(t)^{c(\beta^T Z_i)}, \text{where } c(\beta^T Z_i) = exp\{\beta^T Z_i\} \tag{8}$$

$$= exp[-\int_0^t \lambda_0(\mu)d\mu]^{exp\{\beta^T Z_i\}} \tag{9}$$

Joint Likelihood Function can be written as:

$$L(\beta) = \prod_{i=1}^{n}[\lambda_0(t)exp\{\beta^T Z_i\}]^{\Delta_i}[exp(-\int_0^t \lambda_0(\mu)d\mu)]^{exp\{\beta^T Z_i\}} \tag{10}$$

2.6 | C−Statistics has been widely used in model selection.

C−Statistics has been widely used in model selection, which gives the probability a randomly selected patient who experienced an event (DIED) had a higher risk score than a patient who had not experienced the event. Values over 0.7 indicate a good model. Let T be the event time, Z be a p*1 covariate vector, g(z) estimated risk score with Z. Given two independent copies $(T_1, g(Z_1))^T, (T_2, g(Z_2))^T$ of $(T, g(Z))^T$, a concordance measure (C−Statistics) is defined:

$$C = P[g(Z_1) > g(Z_2|T_2 > T_1)] \tag{11}$$

2.7 | Net Reclassification Improvement (NRI)

Net Reclassification Improvement (NRI) initially proposed by Pencina et al (Pencina, D' Agostino and Vasan, 2008) can be defined as follows:

$$NRI = \frac{\sum_i \text{in events v(i)}}{\text{events}} - \frac{\sum_j \text{in nonevents v(j)}}{\text{nonevents}} \tag{12}$$

where v(i)=1 represent the upward movement, and v(i)=-1 represent the downward movement, v(i)=0 means no movement.

Estimators of the standard errors of $\hat{NRI}$:

$$\hat{se} = \sqrt{\frac{\hat{P}_{up,events} + \hat{P}_{down,events}}{number of events} + \frac{\hat{P}_{up,nonevents} + \hat{P}_{down,nonevents}}{number of nonevents}} \tag{13}$$

Asymptotic test statistic for H_0: NRI=0, which means the new biomarker is useless for prediction. The goal is to find a positive NRI, which indicates that the new biomarker is useful for prediction.

$$z = \frac{\hat{NRI}}{\hat{se}} \tag{14}$$

At significance level $\alpha = 0.05$, reject H_0 if |z|>1.96.

Asymptotic 95th confidence intervals for NRI:

$$\hat{NRI} \pm 1.96 * \hat{se} \tag{15}$$

2.8 | Competing Risk Model

The KM estimated probability of the event of interest in the presence of competing risks is biased since the fundamental noninformative censoring assumption underlying the KM estimator is violated. Besides, the time to event and the censoring distributions are no longer independent: The time to event is dependent on the mechanism that would cause the patient to be censored. Therefore, cumulative incidence function (CIF) can be used to describe event occurrence instead of survival function, and CIF is defined:

$$F_j(t) = Pr[T \leq t, C = j] = \int_o^t h_j(\tau)S(\tau)d\tau, \text{for } j = 1, \ldots, J \tag{16}$$

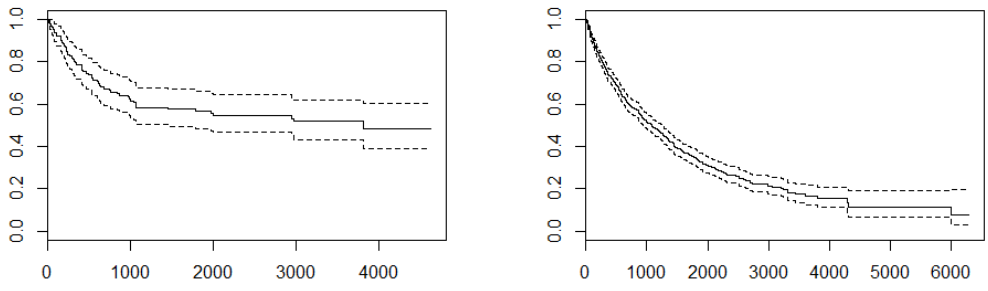
But the estimate from KM is:

$$F_j(t) = 1 - S_j(t) = \int_o^t h_j(\tau)S_j(\tau)d\tau \tag{17}$$

where $S_j(t)$ stands for the death due to jth cause. Thus, this showing that simply use KM estimator is wrong.

3 | RESULTS

KM estimators has been computed, and the survival functions of IST patients has been plotted and compared with the results of IMRAW patients. From Figure 1, after 1000 days (about 3 years), we find that the survival rate for IST patients is much higher compare to IMRAW patients.

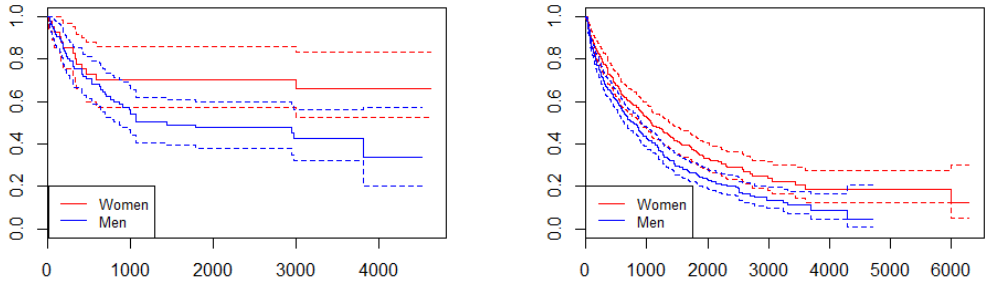


(a) IST patients

(b) IMRAW patients

FIGURE 1 The KM estimate and its 95% pointwise confidence interval of S(t) for time to death of the IST vs IMRAW patients

Besides, from Figure 2, by setting time to AML or death as the event among women and men, the survival rate of IST patients is also higher compare to IMRAW patients after about 3 years.



(a) IST patients

(b) IMRAW patients

FIGURE 2 The KM estimate and its 95% pointwise confidence interval of S(t) for time to AML or death of the IST vs IMRAW patients for women and men

From Figure 3, the results of NA estimators consistent with the results of KM estimators. Knowing the specific distribution, MLE would be more efficient compare to KM and NA estimators, since it minimize the mean squared error (MSE).

	NA estimator	Lower bound	Upper bound
$\Lambda(1)$	0.2391	0.1486	0.3296
$\Lambda(3)$	0.5352	0.3877	0.6827
$\Lambda(5)$	0.5649	0.4118	0.7181

	NA estimator	Lower bound	Upper bound
$\Lambda(1)$	0.3013	0.2598	0.3428
$\Lambda(2)$	0.6998	0.6257	0.7739
$\Lambda(5)$	1.0723	0.9613	1.1832

FIGURE 3 NA estimates and 95% pointwise confidence interval of $\Lambda(1)$, $\Lambda(3)$ and $\Lambda(5)$ for time to death of the IST vs IMRAW patients

The p-value of all the covariates of AGE, GENDER, ANC, PLATE and IST treatment are less than 0.05, which means all the covariates are significant. However, the harzard ratios of IST treatment is 0.5729, which is less than 1. This means hazard rate of having IST Treatment = 1 is 0.5729 times compared to the baseline hazard. Thus IST treatment decrease the hazard rate for the patient.

TABLE 1 Estimates of β and hazard ratios for the corresponding covariates of AGE, GENDER, ANC, PLATE, and IST treatment

	$\hat{\beta}$	hazard ratios	p-value
AGE	0.0314	1.0319	$3.40e^{-16}$
GENDER	0.3172	1.3733	0.0004
ANC	0.0292	1.0296	0.0712
PLATE	-0.0022	0.9978	$1.14e^{-09}$
IST treatment	-0.5570	**0.5729**	0.0004

Shown on Table 2, the p-value of the interaction term NIH:GENDERM is 0.2521, which is larger than 0.05, thus we fail to reject the null hypothesis, and conclude that IST effects are the same for men and women based on the IMRAW-IST dataset.

TABLE 2 Cox Model with interaction term NIH X GENDERM

	$\hat{\beta}$	hazard ratios	$Pr(>\|z\|)$
AGE	0.0311	1.0315	0.0000
NEUTRO	0.0290	1.0295	0.0737
PLATE	-0.0022	0.9978	0.0000
NIH	-0.8650	0.4211	0.0069
GENDERM	0.2862	1.3313	0.0022
NIH:GENDERM	0.4039	1.4977	**0.2521**

C−Statistics are frequently used in model selection. To test whether IST treatment is predictive for overall survival after adjusting $Z_i^{(0)} = (A_i, S_i, N_i, P_i, I_i)^T$, we would like to compute the C-statistics $C_\tau^{(1)}$ and $C_\tau^{(2)}$ for $\lambda_i(t|Z_i^{(0)})$ (Model 1) and $\lambda_i(t|Z_i)$ (Model 2). However, from Table 3, the incremental value $C_5^{(2)}$- $C_5^{(1)}$ is 0.0011, which is negligible.

TABLE 3 For $\tau = 3, 5$ years, the estimates, corresponding SE and 95% CI for C − statistics

	Estimates	**Corresponding SE**	**95% Lower Bound**	**95% Upper Bound**
$C_3^{(1)}$	0.6409	0.0138	0.6138	0.6679
$C_3^{(2)}$	0.6370	0.0143	0.6090	0.6649
$C_3^{(2)}$- $C_3^{(1)}$	-0.0039	0.0045	-0.0127	0.0049
$C_5^{(1)}$	0.6428	0.0130	0.6174	0.6682
$C_5^{(2)}$	0.6439	0.0136	0.6172	0.6706
$C_5^{(2)}$- $C_5^{(1)}$	**0.0011**	0.0041	-0.0070	0.0092

NRI has been calculated in Table 4, for each model of 3 and 5 years. $NRI_3^{(1)}$ is 0.0027, which approaches 0. This is indicative that the new biomarker is useless for prediction. For $\tau = 3$ years, the risk categories are [0; 10%) ("Low"), [10%; 30%) ("Low-Medium") and [30%; 50%) ("Medium-High"), [50%; 100%) ("High"), and the corresponding $NRI_3^{(2)}$ is -0.0246, which means it has negative effect in risk prediction. $NRI_5^{(2)}$ is 0.0752, which is the highest positive value among all these models, and this suggests that it has positive effect in risk prediction.

TABLE 4 The estimate for each models' NRI of 3 and 5 years

	Estimates
$NRI_3^{(1)}$	0.0027
$NRI_3^{(2)}$	-0.0246
$NRI_5^{(1)}$	0.0576
$NRI_5^{(2)}$	0.0752

From Table 5, the cumulative incidence functions (CIF) of competing risk model for IST patients are less than a half of IMRAW patients, and this confirms that IST has positive effect to patients, especially in the long run.

TABLE 5 For t = 3 and 5 years, the estimates, corresponding SE and 95% CI for CIF

CIF	Estimates	Corresponding SE	95% Lower Bound	95% Upper Bound
$F_1(3\|\text{IST})$	**0.1617**	0.0333	0.1029	0.2322
$F_2(3\|\text{IMRAW})$	0.3708	0.0186	0.3343	0.4073
$F_1(5\|\text{IST})$	**0.1617**	0.0333	0.1029	0.2322
$F_2(5\|\text{IMRAW})$	0.4789	0.0206	0.4381	0.5186

From Table 6, when we consider covariates Z_i for Event 2, all the p-value are much less than 0.05. Therefore, all of these variables are significant, so we reject the null hypothesis, and conclude that Z_i has effect for Event 2. On one hand, the Hazard Ratio of variable NIH is 0.6568, which is less than 1, and it means that the IST treatment provides 34.32% risk reduction compared to the control treatment. Therefore, IST treatment has positive effect on Event 2 (time to Non-AML death). On the other hand, the variables like AGE, NEUTRO, PLATE are very close to 1, which means they have little effect on Non-AML disease.

TABLE 6 The effects (i.e., coeffcients) of Z_i for Event 2

Variables	Coefficients	Hazard Ratio [exp(coef)]	Covariates Z	p-value
AGE	0.0351	1.0358	6.9049	$5.0e^{-12}$
GENDERM	0.3383	1.4026	3.1290	0.0018
NEUTRO	0.0535	1.0549	3.7859	0.0001
PLATE	-0.0015	0.9985	-3.1767	0.0015
NIH	-0.4204	0.6568	-2.0858	0.0370

4 | CONCLUSIONS

Our results show that IST would be extremely helpful in treating MDS, which could extend patient's survival especially in the long run. Standard regression methods are not applicable to the survival data is because the failure time is usually heavily right skewed. This against the normally distributed assumption of the regression model. When the distribution is unknown, we tried the nonparametric models, like KM and NA. For given distribution, MLE could be used. In order to consider multivariate case, we apply the proportional hazard models to calculate survival functions. Among all these models, we find that the IST treatment has a positive effect for the treatment. On the other hand, variables like AGE, NEUTRO and PLATE have tiny effect with treatment. Although the C-statistics and NRI show that the improvement in risk prediction offered by the IST treatment is small, the competing risk model proves that IST treatment has long run positive effect to patients.

REFERENCES

Pencina, M., D' Agostino, R., & Vasan, R. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, *27*(2), 157–172. https://doi.org/10.1002/sim.2929

Sloand, E. M., Wu, C. O., Greenberg, P., Young, N., & Barrett, J. (2008). Factors Affecting Response and Survival in Patients With Myelodysplasia Treated With Immunosuppressive Therapy. *Journal of Clinical Oncology*, *26*(15), 2505-2511. doi:10.1200/jco.2007.11.9214

Stahl, M., Deveaux, M., Witte, T. D., Neukirchen, J., Sekeres, M. A., Brunner, A. M., . . . Zeidan, A. M. (2018). The use of immunosuppressive therapy in MDS: Clinical outcomes and their predictors in a large international patient cohort. *Blood Advances*, *2*(14), 1765-1772. doi:10.1182/bloodadvances.2018019414

Wong, R., Shahjahan, M., Wang, X., Thall, P., de Lima, M., Khouri, I., … Giralt, S. (2005). Prognostic factors for outcomes of patients with refractory or relapsed acute myelogenous leukemia or myelodysplastic syndromes undergoing allogeneic progenitor cell transplantation. *Biology of Blood and Marrow Transplantation*, *11*(2), 108–114. https://doi.org/10.1016/j.bbmt.2004.10.008