Cox Regression for Two Competing Risks

Hanna Damarjian; Dr. Perry; Survival Analysis

12/4/2021

Table of Contents

The pieces of this presentation are organized as follows:

Introduction

- Individuals face death from a number of risks. These risks *compete* to become the actual cause of death, which gives rise to the situation of **competing risks**.

- More generally, this term applies when an individual may experience one of a number of different end-points, where the occurrence of any one of these hinders or eliminates the potential for others to occur.

- For example, suppose an individual with risk factors heart disease and type 2 diabetes dies. These two risks are **competing** with each other and one will be the occurrence of death.

Outline of Paper

- The paper I have chosen, *Applying Cox Regression to Competing Risks* proposes a way to analyze **two competing risks** and develops two methods to estimate the parameters. This is done under **Cox Regression modelling** and shows that it is possible to analyze competing risks using readily available standard programs.

- Assuming each method incorporates *two types* of failures and some censored data, I discover that the parameter estimations for the first method combines (1) the **Additive Theorem of Probability** for two events and (2) the likelihood function that Cox proposed. The second method combines (1) the definition of **independence** for two events along with (2) noted prior.

- There is one application of the theory presented.

Data Set-Up

Assumptions (major) for both methods include:

1. $\forall i = 1, 2, ..., m$ indivdiuals, we define failure type, δ_i, as $\delta_i = 0$ (if failure type 1 caused death) or $\delta_i = 1$ (if failure type 2 caused death).

2. Data for response failure time, t_i, is duplicated twice. If an individual has failure type 1 (for example), then failure type 2 is censored, *, and vice versa. *What happens if we do not know the failure type for both?*

3. If subject i has failure time, t_i and $\delta_i = 0$, then the covariates include $x_i, (\delta_i = 0)x_i = 0$. As a pair, the paper specifies this as (**x,0**). Analogously, if $\delta_i = 1$, then the covariate pair would be (**x,x**).

4. The hazard functions for the two types of risks are assumed to be additive. That is, the hazard of failure 1 or 2 is $(b'x_i) + (\delta_i b_0 + b'x_i + \theta'\delta_i x_i)$. **What are** b_0, b', θ'?

Method 1 Theory (Part A) presented

Assuming no ties, the contribution to the partial likelihood if observation i results in a failure (type 1 or type 2) is:

$$L(b_0, b', \theta') = \frac{e^{b_0\delta_i + b'x_i + \theta'\delta_i x_i}}{\sum\limits_{R_i} e^{b_0\delta_i + b'x_i + \theta'\delta_i x_i}} \tag{1}$$

where the summation is over all survival times (including each appropriate second entry) which have neither failed nor been censored at t_i. Note that the contribution to the denominator from the two duplicated entries (from Assumption 3) is

$$e^{b'x} + e^{b_0 + b'x + \theta'x}.$$

Interpretation of L? We also assume, for example, that the hazard function for failure type 1 is proportional to $e^{b'x}h_{00}(t_i)$.

Method 1 Theory A Implementation

Recall that if we want to compare two groups, then we use the partial likelihood function with: $h_1(t_i) = e^{\beta x_i} h_0(t_i)$. Also under the assumption that no ties occurred for failure times, Cox shows that

$$L(\beta) = \prod_{i=1}^{m} \frac{\exp(\beta x_i)}{\sum\limits_{k \in R_{t_i}} \exp(\beta x_k)} \tag{2}$$

where $R(t_i)$ is the risk set of subjects at t_i. We want to only estimate (β) and not (β, t_i). Hence the name *partial likelihood*.

The difference is coming from Assumption 4 noted prior. That is, the hazard functions for the two types of risks are assumed to be additive where the hazard of failure 1 or 2 is $(b' x_i) + (\delta_i b_0 + b' x_i + \theta' \delta_i x_i)$.

Method 1 Theory B Presented

Theorem. If $b' := b_I$ and $b' + \theta' := b_{II}$ and assuming no ties, then the full partial log-likelihood for two competing risks is:

$$\ln(L) = \sum_{i,I} b_I x_i + \sum_{i,II} (b_0 + b_{II} x_i) - \sum_i \ln(\sum_{R_i} e^{(b_I x_i)} + e^{(b_0 + b_{II} x_i)}) \qquad (3)$$

where $\sum\limits_{i,I} b_I x_i$ is the sum of all type 1 failures, $\sum\limits_{i,II} (b_0 + b_{II} x_i)$ is over the type II failures, and $\sum\limits_i \ln(\sum\limits_{R_i} e^{b_I} + e^{b_0 + b_{II} x_i})$ is over all failures (type 1 or type 2).

Note: parameter estimates by applying $\frac{\partial L}{\partial b_I}, \frac{\partial L}{\partial b_{II}}$, and $\frac{\partial L}{\partial b_0}$ and setting them all equal to 0.

Method 1 Theory B Implemented

Proof. Denote X to be a collection of random variables that takes discrete or continuous values based upon $i = 1, 2, ..., m$ individuals. Thus, $X = (X_1, X_2, ...X_m)$ where X_i represents the arbitrary covariate(s) for the ith individual.

Now, let U be the universal set of hazard ratios for two types of failures $-$ A if failure type 1 is the leading cause $(\delta_i = 0)$ and B if failure type 2 is the leading cause $(\delta_i = 1)$ $-$ such that $A, B \subset U$. So, A, B, $A \cap B$, or $A \cup B$ occur and $(A \cup B)' = \emptyset$. Then:

Method 1 Theory B Implemented (Continued)

$$
\begin{aligned}
P(A \cup B; b_0, b_I, b_{II}) =& P(A; b_0, b_I, b_{II}) + P(B; b_0, b_I, b_{II}) - P(A \cap B; b_0, b_I, b_{II}) \\
P(A \cap B; b_0, b_I, b_{II}) =& P(A; b_0, b_I, b_{II}) + P(B; b_0, b_I, b_{II}) - P(A \cup B; b_0, b_I, b_{II}) \\
L(b_0, b_I, b_{II}; A \cap B) =& L(b_0, b_I, b_{II}; A) + L(b_0, b_I, b_{II}; B) - L(b_0, b_I, b_{II}; A \cup B) \\
=& \prod_{i=1}^{m} \frac{e^{b_I x_i}}{\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i})} + \prod_{i=1}^{m} \frac{e^{b_0 + b_{II} x_i}}{\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i})} \\
& - \prod_{i=1}^{m} \frac{\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i})}{\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i})}. \\
=& \frac{\prod_{i=1}^{m} e^{b_I x_i}}{\prod_{i=1}^{m}(\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i}))} + \frac{\prod_{i=1}^{m} e^{b_0 + b_{II} x_i}}{\prod_{i=1}^{m}(\sum_{R_i}(e^{b_1 x_i} + e^{b_0 + b_{II} x_i}))} \\
& - \frac{\prod_{i=1}^{m}(\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i}))}{\prod_{i=1}^{m}(\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i}))}.
\end{aligned}
$$

Method 1 Theory B Implemented (Continued)

Now, the denominator of $L(b_0, b_I, b_{II}; A \cap B)$ is $d := \prod_{i=1}^{m}(\sum\limits_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i}))$. We multiply out both sides by d, and then let $L := d * L(b_0, b_I, b_{II}; A \cap B)$. Finally, if we take the natural logarithm *for each event* A, B, and $A \cup B$, then:

$$\begin{aligned}
L &= \prod_{i=1}^{m} e^{b_I x_i} + \prod_{i=1}^{m} e^{b_0 + b_{II} x_i} - \prod_{i=1}^{m}(\sum_{R_i}(e^{b_I x_i} + e^{b_0 + b_{II} x_i})) \\
\ln(L) &= \ln(\prod_{i,I} e^{b_I x_i}) + \ln(\prod_{i,II} e^{(b_0 + b_{II} x_i)}) - \ln(\prod_{i,I,II}(\sum_{R_i} e^{b_I} + e^{b_0 + b_{II} x_i})) \\
&= \ln(e^{\sum\limits_{i,I} b_I x_i}) + \ln(e^{\sum\limits_{i,II} (b_0 + b_{II} x_i)}) - \sum_{i,I,II} \ln(\sum_{R_i}(e^{b_I} + e^{b_0 + b_{II} x_i})) \\
&= \sum_{i,I} b_I x_i + \sum_{i,II}(b_0 + b_{II} x_i) - \sum_{i,I,II} \ln(\sum_{R_i}(e^{b_I} + e^{b_0 + b_{II} x_i})).
\end{aligned}$$

Method 2 Theory Presented - Stratification

Another method discussed in the paper states to run a Cox regression on the covariates $\mathbf{x}$ and $\delta\mathbf{x}$ while stratifying by failure type, $\delta = 0$ or 1.

Theorem. When treating the survival times of the two types of failures separately, the **partial likelihood** for two failure types with regression coefficients b', θ, and unknown baseline hazard functions, h_{00}, h_{01}, is:

$$L(b', b' + \theta) = \prod_{t_i, \delta_i = 0} \frac{e^{b' x_i}}{\sum_{R_i} e^{b' x_i}} \prod_{t_i, \delta_i = 1} \frac{e^{b' x_i + \theta x_i}}{\sum_{R_i} e^{b' x_i + \theta x_i}}. \tag{4}$$

Remark: the risk set for each case consists of those subjects with the appropriate stratum identifier, $\delta = 0$ for the first and $\delta = 1$ for the second.

Method 2 Theory Implemented - Stratification

Proof. Denote event C to be the hazard ratio for a patient led to death by failure type 1 and event D the hazard ratio for a patient led to death by failure type 2. Given sample points $(x_i, \delta_i x_i)$ for all $i = 1, 2, ..., m$ individuals, we want to find the likelihood function corresponding to $P(C_i \cap D_i)$.

From the assumptions, the likelihood function according to C is
$L(b', \theta; C) = P(C; b', \theta) = \prod_{i=1}^{m} P(C_i; b', \theta) = \prod_{i=1}^{m} \frac{e^{b' x_i + \theta(\delta_i = 0) x_i}}{\sum\limits_{R_i} e^{b' x_i + \theta(\delta_i = 0) x_i}} =$

$\prod_{i=1}^{m} \frac{e^{b' x_i}}{\sum\limits_{R_i} e^{b' x_i}}$ where b' is the unknown parameter estimate for the discrete or continuous covariate x_i, θ is the unknown parameter estimate for failure type 2, $\sum\limits_{R_i}$ represents the risk set $\forall i \in [1, m]$ that are uncensored, and $\delta_i = 0$ here for failure type 1.

Method 2 Theory Implemented - Stratification (Cont.)

Similarly, the likelihood function according to D is $L(b', \theta; D) = P(D; b', \theta) = \prod_{i=1}^{m} P(D_i; b', \theta) = \prod_{i=1}^{m} \frac{e^{b' x_i + \theta(\delta_i = 1) x_i}}{\sum\limits_{R_i} e^{b' x_i + \theta(\delta_i = 1) x_i}} = \frac{e^{b' x_i + \theta x_i}}{\sum\limits_{R_i} e^{b' x_i + \theta x_i}}$.

Since $C = (C_1, C_2, ..., C_m)$ and $D = (D_1, D_2, .., D_m)$ are independent (due to the stratification) $\rightarrow$ C_i and D_i are independent $\rightarrow$ $P(C_i)$ and $P(D_i)$ are also independent. Finally, if $L(b', \theta; C \cap D) := P(C \cap D; b', \theta)$, then:

$$\begin{aligned}
P(C \cap D; b', \theta) =& P(C; b', \theta) P(D; b', \theta) \\
=& \prod_{i=1}^{m} P(C_i; b', \theta) \prod_{i=1}^{m} P(D_i; b', \theta) \\
=& \prod_{i=1}^{m} \frac{e^{b' x_i}}{\sum\limits_{R_i} e^{b' x_i}} \prod_{i=1}^{m} \frac{e^{b' x_i + \theta x_i}}{\sum\limits_{R_i} e^{b' x_i + \theta x_i}}.
\end{aligned}$$

Critiquing the Paper's Implementation

Prostate Cancer

Survival times of $m = 506$ patients with prostate cancer were randomly allocated to different levels of treatment with the drug diethylstilbestrol (DES). There were eight specific risk factors (including drug treatment), and 23 subjects had incomplete data. The failures or causes of death were classifed as (1) cancer, (2) cardiovascular, or (3) other.

There are two important graphs of information to analyze in competing risks data: the **cause-specific survival curve** and the **proportion hazards curves** to ensure the Cox Proportional Hazards Model can be applied.

Question: What is the difference between the cause-specific survival curve(s) and the general KM survival curve we used?

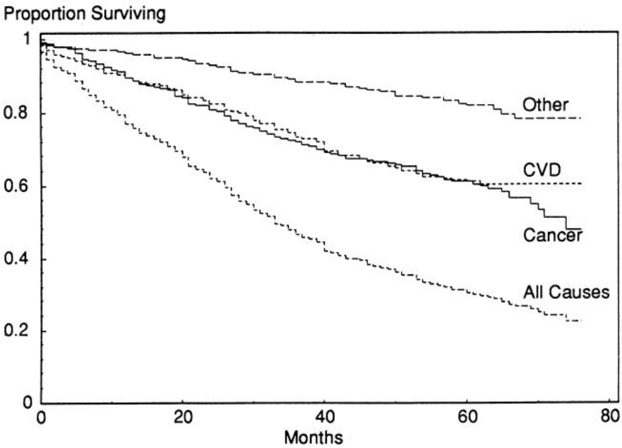

Figure 2. Cause-specific survival curves for 483 patients with prostate cancer

Figure 1: Method B is preferred

Graph 2: Proportional Hazards Curve (Major Assumption)

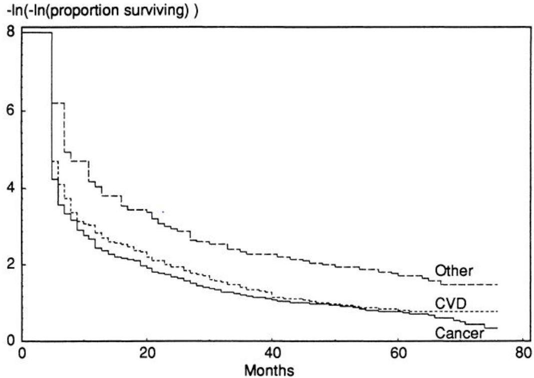

Figure 3. Re-expressed cause-specific Kaplan–Meier curves for 440 patients with prostate cancer surviving 5 months or more

Figure 2: Method A is preferred

Graph 3: Model after Graph 2 meets PH assumption

Competing Risks 531

Table 7

Complex cause-specific PH model with proportional baseline hazard functions (Method A): prostate cancer patients surviving for 5 months or more

	Cancer			CVD			Other		
Covariate	Coeff	SE	p(PH)	Coeff	SE	p(PH)	Coeff	SE	p(PH)
Cause-cancer				−.225	.325	.146	−.554	.391	.162
Treatment	−.564	.177	.367	.330	.190	.139	−.505	.302	.207
AG	−.061	.151	.358	.369	.149	.280	.821	.224	.367
WT	.209	.143	.065	.096	.164	.357	.509	.252	.429
PF	.270	.274	.306	.481	.307	.484	−.074	.608	.497
HX	−.063	.186	.285	1.104	.201	.384	.150	.307	.006
HG	.465	.185	.162	−.061	.236	.070	−.243	.411	.232
SZ	1.131	.214	.040	−.305	.438	.291	.731	.489	.312
SG	1.335	.206	.353	−.047	.203	.236	−.611	.336	.474

Partial likelihood (log times −2): 3779.04.

Figure 3: Method A Model Coefficients

Conclusion

The overall theme of the research paper was to theorize and implement two types of methods for fitting a Cox Proportional Hazards Model to cause-specific survival data. The methods involved the assumption of two different failure types (competing risks) and any number of non-failure covariates.

Thank you!