

# Survival Regression Models

David M. Rocke

May 6, 2021

# Background on the Proportional Hazards Model

The exponential distribution has constant hazard

$$f(t) = \lambda e^{-\lambda t}$$

$$S(t) = e^{-\lambda t}$$

$$h(t) = \lambda$$

Let's make two generalizations. First, let's let the hazard depend on covariates  $x_1, x_2, \dots, x_p$ . Second, let's let the base hazard depend on  $t$ . We can do this using either parametric or semi-parametric approaches.

# The Cox Model

The generalization is that the hazard function is

$$\begin{aligned}\eta &= \beta_1 x_1 + \cdots + \beta_p x_p \\ h(t|\text{covariates}) &= h_0(t)e^\eta\end{aligned}$$

This has a log link as in a generalized linear model. It is semi-parametric because the linear predictor depends on estimated parameters but the base hazard function is unspecified. There is no constant term because it is absorbed in the base hazard. Note that for two different individuals with possibly different covariates, the ratio of the hazard functions is  $\exp(\eta_1) / \exp(\eta_2) = \exp(\eta_1 - \eta_2)$  which does not depend on  $t$ .

# The Cox Model

How do we fit this model? We need to estimate the coefficients of the covariates, and we need to estimate the base hazard  $h_0(t)$ . For the covariates, supposing for simplicity that there are no tied event times, let the event times for the whole data set be  $t_1, t_2, \dots, t_D$ . Let the risk set at time  $t_i$  be  $R(t_i)$  and

$$\eta_j = \beta_1 x_{j1} + \dots + \beta_p x_{jp}$$

$$\theta_j = e^{\eta_j}$$

$$h(t|\text{covariates}) = h_0(t)e^{\eta} = \theta h_0(t)$$

# The Cox Model

Conditional on a single failure at time  $t_i$ , the probability that the event is due to subject  $j \in R(t_i)$  is

$$\begin{aligned}\Pr(j \text{ fails} | 1 \text{ failure at } t_i) &= \frac{h_0(t_i)e^{\eta_j}}{\sum_{k \in R(t_i)} h_0(t_i)e^{\eta_k}} \\ &= \frac{\theta_j}{\sum_{k \in R(t_i)} \theta_k}\end{aligned}$$

# The Cox Model

- The logic behind this has several steps. We first fix (ex post) the failure times and note that in this discrete context, the probability  $p_j$  that a subject  $j$  in the risk set fails at time  $t$  is just the hazard of that subject at that time.
- If all of the  $p_j$  are small, the chance that exactly one subject fails is

$$\sum_{k \in R(t)} p_k \left[ \prod_{m \in R(t), m \neq k} (1 - p_m) \right] \approx \sum_{k \in R(t)} p_k$$

# The Cox Model

If subject  $j(i)$  is the one who fails at time  $t_i$ , then the *partial likelihood* is

$$L(\beta | T) = \prod_i \frac{\theta_{j(i)}}{\sum_{k \in R(t_i)} \theta_k}$$

and we can numerically maximize this with respect to the coefficients. When there are tied event times adjustments need to be made, but the likelihood is still similar. Note that we don't need to know the base hazard to solve for the coefficients.

# The Cox Model

Once we have coefficient estimates  $\hat{\beta}_1, \dots, \hat{\beta}_p$ , this also defines  $\hat{\eta}_j$  and  $\hat{\theta}_j$  and then the estimated base cumulative hazard function is

$$\hat{H}(t) = \sum_{t_j < t} \frac{d_j}{\sum_{k \in R(t_j)} \theta_k}$$

which reduces to the Nelson-Aalen estimate when there are no covariates. There are numerous other estimates that have been proposed as well.



# Adjustment for Ties

At each time  $t_i$  at which more than one of the subjects has an event, let  $d_i$  be the number of events at that time,  $D_i$  the set of subjects with events at that time, and let  $s_i$  be a covariate vector for an artificial subject obtained by adding up the covariate values for the subjects with an event at time  $t_i$ . Let

$$\bar{\eta}_i = \beta_1 s_{i1} + \cdots + \beta_p s_{ip}$$

and  $\bar{\theta}_i = \exp(\bar{\eta}_i)$ .

# Adjustment for Ties

Let  $s_i$  be a covariate vector for an artificial subject obtained by adding up the covariate values for the subjects with an event at time  $t_i$ . Note that

$$\begin{aligned}\bar{\eta}_i &= \sum_{j \in D_i} \beta_1 x_{j1} + \cdots + \beta_p x_{jp} \\ &= \beta_1 s_{i1} + \cdots + \beta_p s_{ip} \\ \bar{\theta}_i &= \exp(\bar{\eta}_i) \\ &= \prod_{j \in D_i} \theta_j\end{aligned}$$

# Adjustment for Ties

Breslow's method estimates the partial likelihood as

$$\begin{aligned}L(\beta|T) &= \prod_i \frac{\bar{\theta}_i}{[\sum_{k \in R(t_i)} \theta_k]^{d_i}} \\ &= \prod_i \prod_{j \in D_i} \frac{\theta_j}{\sum_{k \in R(t_i)} \theta_k}\end{aligned}$$

This method is equivalent to treating each event as distinct and using the non-ties formula. It works best when the number of ties is small. It is the default in many statistical packages, including PROC PHREG in SAS.

# Adjustment for Ties

The other common method is Efron's, which is the default in R.

$$L(\beta|T) = \prod_i \frac{\bar{\theta}_i}{\prod_{j=1}^{d_i} [\sum_{k \in R(t_i)} \theta_k - \frac{j-1}{d_i} \sum_{k \in D_i} \theta_k]}$$

This is closer to the exact discrete partial likelihood when there are many ties. The third option in R (and an option also in SAS as “discrete”) is the “exact” method, which is the same one used for matched logistic regression.

Suppose as an example we have a time  $t$  where there are 20 individuals at risk and three failures. Let the three individuals have risk parameters  $\theta_1, \theta_2, \theta_3$  and let the sum of the risk parameters of the remaining 17 individuals be  $\theta_R$ . Then the factor in the partial likelihood at time  $t$  using Breslow's method is

$$\left( \frac{\theta_1}{\theta_R + \theta_1 + \theta_2 + \theta_3} \right) \left( \frac{\theta_2}{\theta_R + \theta_1 + \theta_2 + \theta_3} \right) \left( \frac{\theta_3}{\theta_R + \theta_1 + \theta_2 + \theta_3} \right)$$

If on the other hand, they had died in the order 1,2, 3, then the contribution to the partial likelihood would be

$$\left( \frac{\theta_1}{\theta_R + \theta_1 + \theta_2 + \theta_3} \right) \left( \frac{\theta_2}{\theta_R + \theta_2 + \theta_3} \right) \left( \frac{\theta_3}{\theta_R + \theta_3} \right)$$

as the risk set got smaller with each failure. The exact method roughly averages the results for the six possible orderings of the failures.

But we don't know the order they failed in, so instead of reducing the denominator by one risk coefficient each time, we reduce it by the same fraction. This is Efron's method.

Efron's method:

$$\left( \frac{\theta_1}{\theta_R + \theta_1 + \theta_2 + \theta_3} \right) \left( \frac{\theta_2}{\theta_R + 2(\theta_1 + \theta_2 + \theta_3)/3} \right) \left( \frac{\theta_3}{\theta_R + (\theta_1 + \theta_2 + \theta_3)/3} \right)$$

# The `coxph()` Function in R

`coxph` {survival} R Documentation  
Fit Proportional Hazards Regression Model

## Description

Fits a Cox proportional hazards regression model. Time dependent variables, time dependent strata, multiple events per subject, and other extensions are incorporated using the counting process formulation of Andersen and Gill.

## Usage

```
coxph(formula, data=, weights, subset,  
       na.action, init, control,  
       ties=c("efron","breslow","exact"),  
       singular.ok=TRUE, robust=FALSE,  
       model=FALSE, x=FALSE, y=TRUE, tt, method, ...)
```



# The `coxph()` Function in R

```
coxph(formula, data=, weights, subset,  
      na.action, init, control,  
      ties=c("efron", "breslow", "exact"),  
      singular.ok=TRUE, robust=FALSE,  
      model=FALSE, x=FALSE, y=TRUE, tt, method, ...)
```

## Arguments

- formula** a formula object, with the response on the left of a `~` operator, and the terms on the right. The response must be a survival object as returned by the `Surv` function.
- data** a `data.frame` in which to interpret the variables named in the formula, or in the subset and the weights argument.
- weights** vector of case weights. If `weights` is a vector of integers, then the estimated coefficients are equivalent to estimating the model from data with the individual cases replicated as many times as indicated by weights.
- subset** expression indicating which subset of the rows of data should be used in the fit. All observations are included by default.

# The `coxph()` Function in R

```
coxph(formula, data=, weights, subset,  
      na.action, init, control,  
      ties=c("efron","breslow","exact"),  
      singular.ok=TRUE, robust=FALSE,  
      model=FALSE, x=FALSE, y=TRUE, tt, method, ...)
```

## Arguments

`ties` a character string specifying the method for tie handling. If there are no tied death times all the methods are equivalent. Nearly all Cox regression programs use the Breslow method by default, but not this one. The Efron approximation is used as the default here, it is more accurate when dealing with tied death times, and is as efficient computationally. The `"exact partial likelihood"` is equivalent to a conditional logistic model, and is appropriate when the times are a small set of discrete values. See further below.

The proportional hazards model is usually expressed in terms of a single survival time value for each person, with possible censoring. Andersen and Gill reformulated the same problem as a counting process; as time marches onward we observe the events for a subject, rather like watching a Geiger counter. The data for a subject is presented as multiple rows or “observations,” each of which applies to an interval of observation (start, stop].

There are special terms that may be used in the model equation. A strata term identifies a stratified Cox model; separate baseline hazard functions are fit for each stratum. The cluster term is used to compute a robust variance for the model. The term  $+ \text{cluster}(\text{id})$  where each value of  $\text{id}$  is unique is equivalent to specifying the  $\text{robust}=\text{T}$  argument. If the  $\text{id}$  variable is not unique, it is assumed that it identifies clusters of correlated observations. The robust estimate arises from many different arguments and thus has had many labels. It is variously known as the Huber sandwich estimator, White's estimate (linear models/econometrics), the Horvitz-Thompson estimate (survey sampling), the working independence variance (generalized estimating equations), the infinitesimal jackknife, and the Wei, Lin, Weissfeld (WLW) estimate.

# Cox Model for bmt Survival vs. Group

```
> dfsurv <- Surv(bmt$t2,bmt$d3)
> bmt.cox <- coxph(dfsurv~factor(group),data=bmt)
> summary(bmt.cox)
```

n= 137, number of events= 83

	coef	exp(coef)	se(coef)	z	Pr(> z )
factor(group)2	-0.5742	0.5632	0.2873	-1.999	0.0457 *
factor(group)3	0.3834	1.4673	0.2674	1.434	0.1516

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
factor(group)2	0.5632	1.7757	0.3207	0.989
factor(group)3	1.4673	0.6815	0.8688	2.478

Concordance= 0.625 (se = 0.031 )

Rsquare= 0.094 (max possible= 0.996 )

Likelihood ratio test= 13.45 on 2 df, p=0.001199

Wald test = 13.03 on 2 df, p=0.00148

Score (logrank) test = 13.81 on 2 df, p=0.001004

# Cox Model for bmt Survival vs. Group

Hypothesis tests for factor levels comparing groups 2 and 3 to group 1. Group 3 has the highest hazard so the most significant comparison is not directly shown.

The coefficient 0.3834 is on the log-hazard-ratio scale, as in log-risk-ratio. The next column gives the hazard ratio 1.4673, and a hypothesis (Wald) test.

The (not shown) group 3 vs. group 2 log hazard ratio is  $0.3834 + 0.5742 = 0.9576$ . The hazard ratio is then  $\exp(0.9576)$  or 2.605. Inference on all coefficients and combinations can be constructed using `coef(bmt.cox)` and `vcov(bmt.cox)` as with logistic regression.

```
          coef exp(coef) se(coef)      z Pr(>|z|)
factor(group)2 -0.5742    0.5632  0.2873 -1.999  0.0457 *
factor(group)3  0.3834    1.4673  0.2674  1.434  0.1516
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Cox Model for bmt Survival vs. Group

The next part of the output gives 95% confidence intervals for the relative risk. For the difference between groups 2 and 3 we need to use `coef(bmt.cox)` and `vcov(bmt.cox)`.

```
                exp(coef) exp(-coef) lower .95 upper .95
factor(group)2    0.5632    1.7757    0.3207    0.989
factor(group)3    1.4673    0.6815    0.8688    2.478
```

```
> c1 <- coef(bmt.cox)
> v1 <- vcov(bmt.cox)
> cont1 <- c(-1,1)
> t(cont1) %*% c1
      [,1]
[1,] 0.9576104          log hazard ratio group 3 to group2
> t(cont1) %*% v1 %*% cont1
      [,1]
[1,] 0.07040675
> sqrt(t(cont1) %*% v1 %*% cont1)
      [,1]
[1,] 0.2653427          standard error of log hazard ratio
```

# Cox Model for bmt Survival vs. Group

Concordance is agreement of first failure between pairs of subjects and higher predicted risk between those subjects, omitting non-informative pairs.

The Rsquare value is Cox and Snell's pseudo R-squared and is not very useful.

There are three tests for whether the model with the group covariate is better than the one without

--Usual likelihood ratio chi-squared

--Wald test chi-squared, obtained by adding the squares of the z-scores

--score = log-rank test, as with comparison of survival functions.

The likelihood ratio test is probably best in smaller samples, followed by the Wald test.

Concordance= 0.625 (se = 0.031 )

Rsquare= 0.094 (max possible= 0.996 )

Likelihood ratio test= 13.45 on 2 df, p=0.001199

Wald test = 13.03 on 2 df, p=0.00148

Score (logrank) test = 13.81 on 2 df, p=0.001004



# Cox Model for bmt Survival vs. Group

The `anova` function performs the likelihood ratio test for comparing models. One can use `drop1()`, `add1()`, `step()`, or compare two explicit models.

```
> anova(bmt.cox)
Analysis of Deviance Table
Cox model: response is dfsurv
Terms added sequentially (first to last)

          loglik  Chisq Df Pr(>|Chi|)
NULL          -373.30
factor(group) -366.57 13.452  2  0.001199 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Survival Curves from the Cox Model

We defined the “base hazard” but then did not use it to estimate the coefficients. In fact, there is no meaningful *base* hazard. In this case, it is the hazard for group 1. We can use `survfit` to get survival functions, but by default it produces the hazard for an average individual with average covariates. This is like the family with 1.5 children—it does not exist. Always it is best to specify the covariate level(s) for which you want the survival curve(s). In this case, we can plot the Cox model survival curves, which are by definition proportional, along with the individual KM curves for the groups.

```
plot(survfit(dfsurv~group,data=bmt))
lines(survfit(bmt.cox,data.frame(group=1:3),conf.int=F),col="red")
legend("bottomleft",c("Kaplan-Meier","Cox Model"),col=c("black","red"),lwd=1)
title("Survival Functions for Three Groups by KM and Cox Model")
```

When we use `survfit()` with a Cox model, we should include a data frame with the same named columns as the predictors in the Cox model and one or more levels of each.

For example

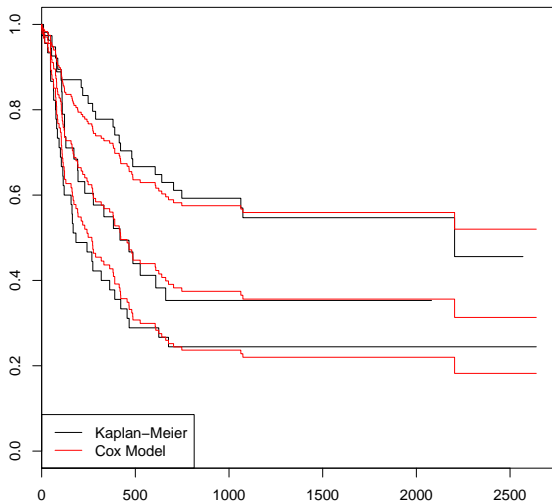
```
> data.frame(group=1:3,age=50)
```

```
  group age
1     1  50
2     2  50
3     3  50
```

```
> data.frame(group=rep(1:3,2),age=rep(c(50,60),each=3))
```

```
  group age
1     1  50
2     2  50
3     3  50
4     1  60
5     2  60
6     3  60
```

### Survival Functions for Three Groups by KM and Cox Model



# Homework

The addicts data set from the text (see page 91) is from a study by Caplehorn et al. (“Methadone Dosage and Retention of Patients in Maintenance Treatment,” *Med. J. Aust.*, 1991). These data comprise the times in days spent by heroin addicts from entry to departure from one of two methadone clinics. There are two further covariates, namely, prison record and methadone dose, believed to affect the survival times.

# Homework

The data set and R input code are on the website. The variables are as follows:

- id: Subject ID
- clinic: Clinic (1 or 2)
- status: Survival status (0 = censored, 1 = departed from clinic)
- time: Survival time in days
- prison: Prison record (0 = none, 1 = any)
- methadone: Methadone dose (mg/day)

# Homework

- 1 Plot the Kaplan-Meier survival curves for the two clinics.
- 2 Test for whether the two survival curves could have come from the same process using `survdif`.
- 3 Plot the cumulative hazards from the Nelson-Aalen estimator.

# Homework

- 4 A common comparison plot for proportional hazards is the complimentary log-log survival plot which plots  $\ln(-\ln[\hat{S}(x)]) = \ln \hat{H}(t)$  against  $\ln(t)$ . If the hazards are proportional, then so are the cumulative hazards, and after taking logs, the curves should be parallel. If the lines are straight, then the Weibull model may be appropriate. Make this plot for the two clinics using the Nelson-Aalen estimator and comment on the results. You will use the “cloglog” option in the plot command.



# Homework

- 5 Construct a Cox model using only the clinic variable. Is the “survival” different at the two clinics? What is the estimated hazard (risk) ratio, a test for significance, and a 95% confidence interval?
- 6 Pick one test for the null hypothesis that the clinics do not differ. Why would you depend on this test more than the others?
- 7 Consider adding the prison and methodone variables. Which of these covariates will improve the model?
- 8 Plot the two survival curves from your chosen Cox model and add the two KM survival curves. What do you think?