# Introduction to survival analysis and Cox proportional hazards models

Nathan Constantine-Cooke

MRC Human Genetics Unit,
MRC Institute of Genetics and Molecular Medicine,
University of Edinburgh

1 July 2020


THE UNIVERSITY *of* EDINBURGH

# Introduction to Cox proportional hazards

(These slides are primarily based on chapter 3 of Kleinbaum and Klein (2012))

# Introduction

# Introduction

- A semi-parametric class of models

# Introduction

- A semi-parametric class of models
- Established by Cox (1972) and is commonly used in medical statistics

# Introduction

- A semi-parametric class of models
- Established by Cox (1972) and is commonly used in medical statistics
- Uses covariates which do not vary with time

# Introduction

- A semi-parametric class of models
- Established by Cox (1972) and is commonly used in medical statistics
- Uses covariates which do not vary with time
- An extension for time-dependent variables exists (but is often inadvisable to use) (See Prentice (1982))

# Hazard function

The Cox proportional hazards model is typically presented via the hazard function:

$$h(t, X) = h_0(t) \cdot e^{\sum_{i=1}^{p} \beta_i x_i} \tag{1}$$

where $X = (x_1, x_2, \ldots, x_p)$ for $p$ explanatory variables

# Hazard function

The Cox proportional hazards model is typically presented via the hazard function:

$$h(t, X) = h_0(t) \cdot e^{\sum_{i=1}^{p} \beta_i x_i} \tag{1}$$

where $X = (x_1, x_2, \ldots, x_p)$ for $p$ explanatory variables

- The baseline hazard is dependent only on $t$ whilst the second term is dependent only on $X$ and a set of regression coefficients.

# Hazard function

The Cox proportional hazards model is typically presented via the hazard function:

$$h(t, X) = h_0(t) \cdot e^{\sum_{i=1}^{p} \beta_i x_i} \quad (1)$$

where $X = (x_1, x_2, \ldots, x_p)$ for $p$ explanatory variables

- The baseline hazard is dependent only on $t$ whilst the second term is dependent only on $X$ and a set of regression coefficients.
- $h_0(t)$ is unspecified but is robust: closely approximates the "correct" parametric model

# Maximum (partial) likelihood estimates 1

- Actually partial likelihood due to only considering probabilities for subjects who met the end-point (non-censored).

# Maximum (partial) likelihood estimates 1

- Actually partial likelihood due to only considering probabilities for subjects who met the end-point (non-censored).
- Can be written as product of several likelihoods (one for each event time). If we rank event times, then each contribution to the likelihood is the probability of individual j failing at the $j^{th}$ event time given there is one event from the group of individuals still at risk at this time point (the risk set).

$$L = L_1 \times L_2 \times \ldots L_k = \prod_{j=1}^{k} L_j \tag{2}$$

# Maximum likelihood estimates 2

Let $X_j = (X_{j1}, \ldots X_{jp})$ be the values of the covariates for subject $j$. Then $L_j$ can be written as:

$$L_j = \frac{h\left(Y_j|X_j\right)}{\sum_{g:Y_g \geq y_j} h(Y_j|X_g)} = \frac{h_0\left(Y_j\right)\theta_j}{\sum_{g:Y_g \geq y_j} h_0(Y_j)\theta_g} = \frac{\theta_j}{\sum_{g:Y_g \geq y_j} \theta_g}$$

where $\theta_j = \exp\left(X_j\beta\right)$.

# Maximum likelihood estimates 3

- If a subject is censored then they are still included in the risk set for $L_j$ calculations prior to the time of censoring.

# Maximum likelihood estimates 3

- If a subject is censored then they are still included in the risk set for $L_j$ calculations prior to the time of censoring.
- Note that it is possible to have ties in the data and there are numerous ways to handle this. Breslow's method does not modify the above approach whilst Efron's method is more complicated.

# Maximum likelihood estimates 3

- If a subject is censored then they are still included in the risk set for $L_j$ calculations prior to the time of censoring.
- Note that it is possible to have ties in the data and there are numerous ways to handle this. Breslow's method does not modify the above approach whilst Efron's method is more complicated.

We maximise by solving

$$\frac{\partial \ln L}{\partial \beta_i} = 0 \tag{3}$$

for $i = 1, \ldots, p$ ($p$ parameters)

# Hazard ratios 1

A hazard ratio (HR) is defined as the hazard for one individual/group divided by the hazard for a different individual/group. For example, this is a placebo group vs a treatment group.

# Hazard ratios 1

A hazard ratio (HR) is defined as the hazard for one individual/group divided by the hazard for a different individual/group. For example, this is a placebo group vs a treatment group.

$$HR = \frac{h(t, X = treatment)}{h(t, X = placebo)}$$

The group with the largest hazard is typically the numerator as hazard ratios $\geq 1$ are easier to interpret.

# Hazard ratios 2

$$HR = \frac{h_0(t) \cdot \exp\left(\sum_{i=1}^{p} \beta_i x_i^*\right)}{h_0(t) \cdot \exp\left(\sum_{i=1}^{p} \beta_i x_i\right)}$$
$$= \exp\left(\sum_{i=1}^{p} \beta_i \left(x_i^* - x_i\right)\right)$$

# Hazard ratios 2

$$HR = \frac{h_0(t) \cdot \exp\left(\sum_{i=1}^{p} \beta_i x_i^*\right)}{h_0(t) \cdot \exp\left(\sum_{i=1}^{p} \beta_i x_i\right)}$$
$$= \exp\left(\sum_{i=1}^{p} \beta_i \left(x_i^* - x_i\right)\right)$$

If $x_i$ is a (0,1) exposure variable, then $HR = e^{\beta_i}$ is the marginal effect size (in terms of HR) of an exposure - provided all other covariates are constant and there are no interaction terms.

# Hazard ratio confidence intervals

A 95% confidence interval for the hazard ratio corresponding to the $i^{th}$ covariate can be predicted using

$$\exp\left(\hat{\beta}_i \pm 1.96\sqrt{\hat{Var}\hat{\beta}_i}\right)$$

where $s_{\hat{\beta}_i} = \sqrt{\hat{Var}\hat{\beta}_i}$
This is simple to calculate when there are no interaction effects in the model, but the computational formula for the estimated standard error is more complex when there are interaction terms. Fortunately, this is already implemented in most statistical software!

# Cox survival curves

From a Cox model, survival curves can be obtained which adjust for explanatory variables. A Cox hazard function can be converted to the below survival function by using

$$S(t) = \exp\left(-\int_0^t h(t)dx\right)$$

which gives us

$$S(t, X) = [s_0(t)]^{\exp\left(\sum_{i=1}^{p} \beta_i x_i\right)}$$

# The proportional hazards assumption

HR is assumed to be constant over time. Recall that a hazard ratio does not depend on $t$.

$$HR = \frac{h(t, x^*)}{h(t, x)} = \exp\left(\sum_{i=1}^{p} \beta_i \left(x_i^* - x_i\right)\right)$$

# The proportional hazards assumption

HR is assumed to be constant over time. Recall that a hazard ratio does not depend on $t$.

$$HR = \frac{h(t, x^*)}{h(t, x)} = \exp\left(\sum_{i=1}^{p} \beta_i \left(x_i^* - x_i\right)\right)$$

We would therefore expect

$$\frac{h(t, x^*)}{h(t, x)}$$

to not depend on time. If this property is not satisfied then we should consider alternatives such as stratified Kaplan-Meier or accelerated failure time models.

# Schoenfeld residuals 1

# Schoenfeld residuals 1

- We would of course like to be able to statistically test for if the proportional hazards assumption is valid

# Schoenfeld residuals 1

- We would of course like to be able to statistically test for if the proportional hazards assumption is valid
- A popular method is via *Schoenfeld residuals*.

# Schoenfeld residuals 1

- We would of course like to be able to statistically test for if the proportional hazards assumption is valid
- A popular method is via *Schoenfeld residuals*.
- For each predictor in a model, Schoenfeld residuals exist for each individual who experiences an event.

# Schoenfeld residuals 2

Suppose subject $f$ has an event at time $Y_f$, then the Schoenfeld residual for a parameter, say $x_1$, is $x_{f1} - W_{1f}$ where $W_{1f}$ is a weighted average of $x_1$ for subjects still at risk at time $Y_f$. The weights used are each subject's hazard.

If the proportional hazards assumption is satisfied for a covariate then the Schoenfeld residuals for that covariate will not be related to survival time.

# Schoenfeld residuals 3

We implement the test as follows,

1. Obtain Schoenfeld residuals for each predictor
2. Rank event times
3. Test the correlation between Schoenfeld residuals and respective ranks

# Parametric Proportional Hazards Models 1

By making different parametric assumptions on the baseline hazard, we can formulate different kinds of proportional hazards models.

# Parametric Proportional Hazards Models 1

By making different parametric assumptions on the baseline hazard, we can formulate different kinds of proportional hazards models. For instance, if we assume exponentially distributed survival times in the baseline group, then we can use

$$h(t) = \frac{-d}{dt} \log S(t)$$

# Parametric Proportional Hazards Models 1

By making different parametric assumptions on the baseline hazard, we can formulate different kinds of proportional hazards models. For instance, if we assume exponentially distributed survival times in the baseline group, then we can use

$$h(t) = \frac{-d}{dt} \log S(t)$$

$$h_0(t) = \frac{-d}{dt} \log e^{-\lambda t} = \lambda$$

and the hazard becomes:

$$h(t, X) = \lambda \cdot e^{\sum_{i=1}^{p} \beta_i x_i}$$

# Parametric Proportional Hazards Models 2

A popular choice of distribution for survival analysis is the Weibull. If survival times follow a Weibull distribution then

$$S(t) = \exp\left(-(\lambda t)^p\right)$$

for $p, \lambda > 0$. Noting that:

$$h(t) = \frac{-d}{dt} \log S(t)$$

we have:

$$\begin{aligned} h_0(t) &= \frac{-d}{dt} - (\lambda t)^p \\ &= p\lambda(\lambda t)^{p-1} \end{aligned}$$

# Parametric Proportional Hazards Models 3

Multiplying the baseline hazard by $\theta = \exp(X\beta)$ gives:

$$\begin{aligned} h(t, X) = h_0(t)\theta &= p\lambda(\lambda t)^{p-1} \cdot \theta \\ &= p\lambda(\lambda t)^{p-1} \cdot \theta^{\left(\frac{1}{p}+1-\frac{1}{p}\right)} \\ &= p\left(\lambda\theta^{\frac{1}{p}}\right)\left(\lambda\theta^{\frac{1}{p}}t\right)^{p-1} \end{aligned}$$

# Parametric Proportional Hazards Models 3

Multiplying the baseline hazard by $\theta = \exp(X\beta)$ gives:

$$\begin{aligned} h(t, X) = h_0(t)\theta &= p\lambda(\lambda t)^{p-1} \cdot \theta \\ &= p\lambda(\lambda t)^{p-1} \cdot \theta^{\left(\frac{1}{p}+1-\frac{1}{p}\right)} \\ &= p\left(\lambda\theta^{\frac{1}{p}}\right)\left(\lambda\theta^{\frac{1}{p}}t\right)^{p-1} \end{aligned}$$

If we set $k = \lambda\theta^{\frac{1}{p}}$, we can clearly see that this is still a Weibull model which demonstrates the Weibull family is closed under proportional hazards

$$h(t, X) = pk(kt)^{p-1}$$

# References I

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Kleinbaum, D. G. and Klein, M. (2012). *Survival Analysis*. Springer New York.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–342.