Understanding the proportional hazards assumption in the Cox model

Paul Dickman

August 2020
`www.pauldickman.com`

Overview of this lecture

Target audience is students and researchers in biomedical sciences without extensive training in statistics.

- ▶ Previous lectures covered introduction to survival analysis, intro to the Cox model, and covariate by covariate interactions in the Cox model.
- ▶ This lecture builds upon those lectures and will cover
  - ▶ What is the proportional hazards assumption?
  - ▶ Assessing and testing the proportional hazards assumption
  - ▶ Relaxing the proportional hazards assumption
- ▶ Slides available at
  http:\\www.pauldickman.com\video\proportional-hazards\
- ▶ Examples use R, but Stata and SAS code available
  on the same page as the slides.

Key learning outcome – proportional hazards (PH) assumption

- ▶ The PH assumption is a familiar assumption with a special name.
- ▶ Common regression models (e.g., linear, logistic, Cox) assume estimated effects are the same for all values of other covariates. Called either
  - ▶ No interaction, or
  - ▶ No effect modification.
- ▶ The PH assumption is conceptually identical; covariate effects are the same for all values of time.
- ▶ Approaches for assessing and relaxing the PH assumption are conceptually the same as for covariate by covariate interactions. If the PH assumption doesn't hold then we add time by covariate interactions.
- ▶ Since we don't actually estimate the effect of time in the Cox model, interactions with time are technically more complicated.

The proportional hazards assumption

- The Cox model (and many other survival models) assumes that the *ratio* of the hazard functions for any two patient subgroups (i.e. two groups with different values of explanatory variables) is constant over follow-up time.
- It is possible to fit a model that allows for non-proportional hazards.
- Note that it is the hazard ratio which is assumed to be constant. The hazards can vary freely with time.

The proportional hazards assumption (2)

▶ When comparing an aggressive therapy vs a conservative therapy, for example, it is not unusual that the patients receiving the aggressive therapy do worse earlier, but then have a lower hazard (i.e., better survival) than those receiving the conservative therapy.

▶ In this situation, the ratio of the hazard functions will not be constant over time, as is assumed by the PH model.

▶ The estimated HR from the Cox model will (conceptually) be the weighted average of values that vary over time. If the hazards cross then the estimated HR may be 1, despite substantial effects (in different directions) at some time points.

Example of non-proportional hazards [1]
Limited (D1) vs. extended (D2) lymph node dissection for gastric cancer

Long-term survival with non-proportional hazards: results from
the Dutch Gastric Cancer Trial

H. Putter[1,*,†], M. Sasako[2], H. H. Hartgrink[3], C. J. H. van de Velde[3]
and J. C. van Houwelingen[1]

► Randomised study comparing the effect of an aggressive (D2) versus
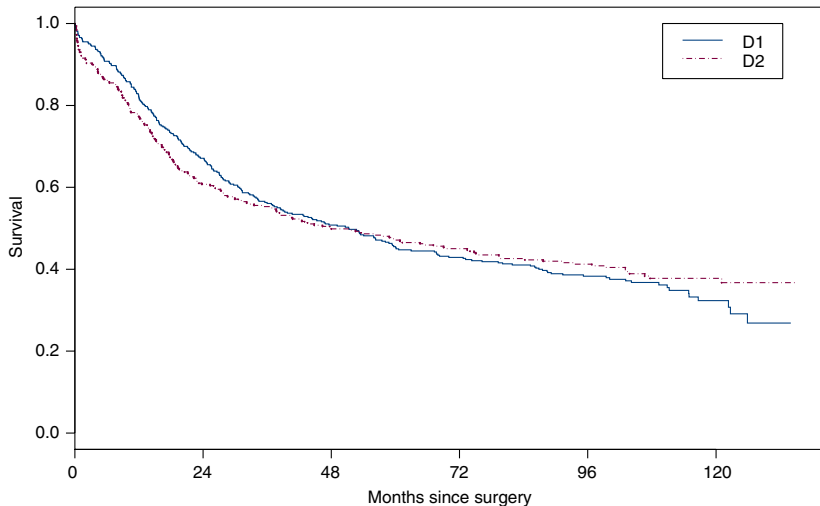conservative (D1) surgical technique on cancer-specific mortality.

Figure 1. Kaplan–Meier plots of the survival curves for D1- and D2-dissection. The survival curves cross after 53 months.

The Cox regression with only randomization as a time-fixed effect gives an estimated hazard ratio of 0.97 of D2 dissection compared to D1-dissection, with a *p*-value of 0.73. The survival
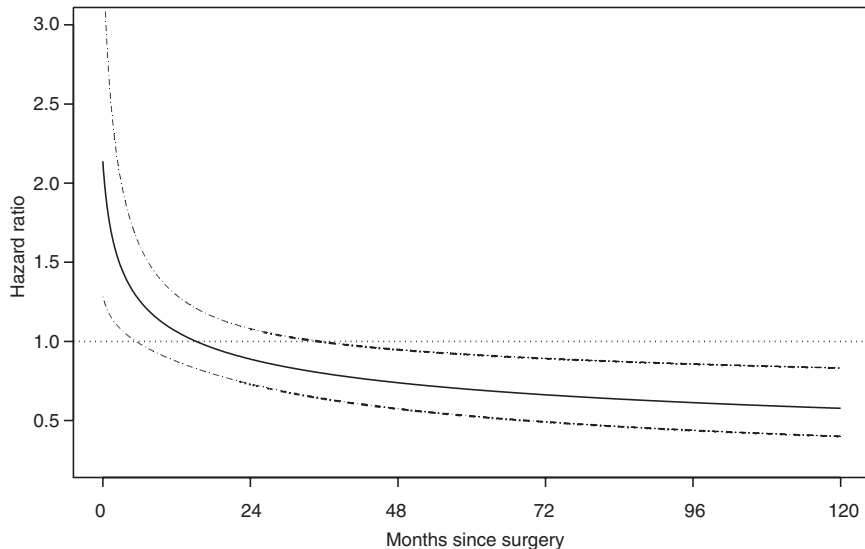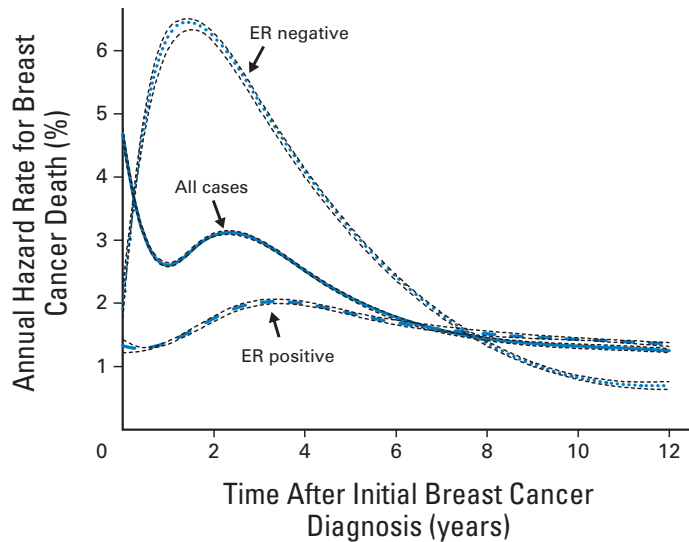
Figure 4. The estimated hazard ratio with 95 per cent confidence intervals based on Cox regression with treatment as time-dependent effect. A hazard ratio of one indicates equality of the hazard rates of D1 and D2.

# Another example: Breast cancer mortality by ER status [2]

Assessing the proportional hazards assumption

- Following is a list of some methods for assessing the appropriateness of the proportional hazards assumption (in increasing order of utility):
  1. Plotting the cumulative survivor functions and checking they do not cross. Not recommended, since the survivor functions do not have to cross for the hazards to be non-proportional.
  2. Plotting the log cumulative hazard functions over time and checking for parallelism.
  3. Plotting the log hazard functions over time and checking for parallelism.
  4. Including time-by-covariate interaction terms in the model and testing statistical significance.
  5. Plotting Schoenfeld residuals against time to identify patterns, and tests based on Schoenfeld residuals.

- The first three methods do not allow for the effect of other covariates, whereas the second two methods do.

- Including a time-by-covariate interaction in the model has the advantage that we obtain an estimate of the hazard ratio as a function of time.

K-M plots are not optimal for assessing PH

▶ These curves (from the 'intro to survival' lecture) are an example of extreme non-proportional hazards.

Which treatment is associated with the best survival?
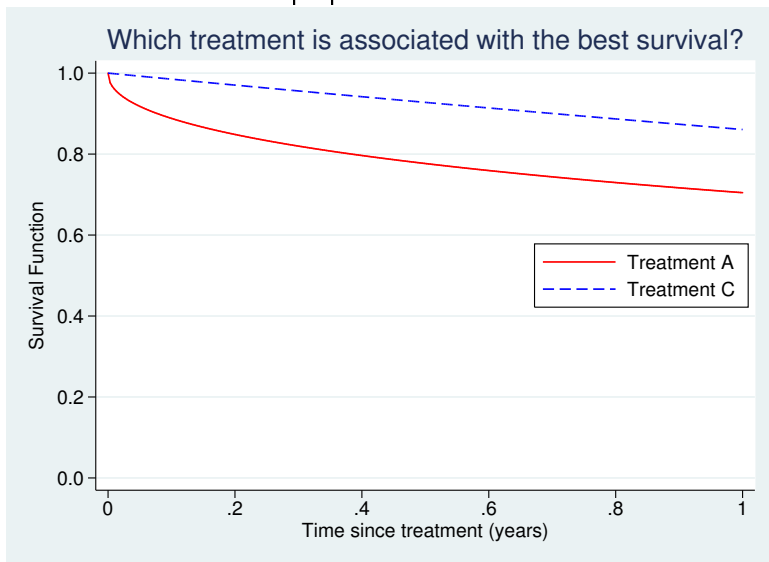
Lack of proportionality is clear when we plot the hazards
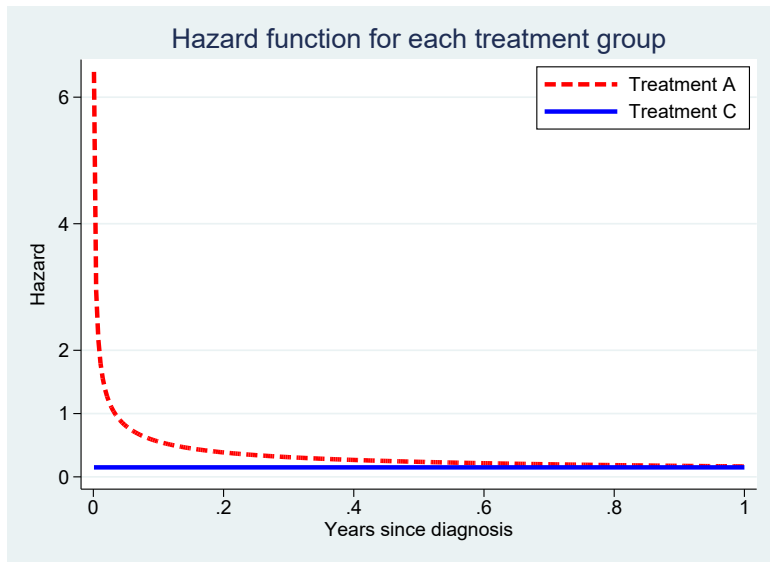
Plots of the log cumulative hazard function

▶ This method of graphically assessing the PH function is somewhat outdated. With modern computers and software it is now possible to plot the hazard (or log hazard) for simple graphical assessment and plots of Schoenfeld residuals are even better. I'll leave the slides here for completeness.

▶ The hazard function and the survivor function are related. One relationship of particular importance is

$$
\begin{aligned}
S(t) &= \exp\left(-\int_0^t \lambda(s)\,\mathrm{d}s\right) \\
&= \exp(-\Lambda(t)),
\end{aligned}
\tag{1}
$$

where $\Lambda(t)$ is called the cumulative hazard (or integrated hazard) at time t.

▶ If we use a proportional hazards model, then another way to write this equation is

$$
S(t|\boldsymbol{X}) = (S_0(t))^{\exp(\beta_1 X_1 + \cdots + \beta_k X_k)}.
$$

Plots of the log cumulative hazard function (2)

▶ Consider the situation where we have only a single binary variable, X, then

$$S(t|X = 1) = (S(t|X = 0))^r,$$

where $r = \exp(\beta)$ is the hazard ratio.

▶ Taking natural logarithms of both sides gives

$$\log S(t|X = 1) = r \log (S(t|X = 0)).$$

▶ Taking natural logarithms of the negatives of both sides gives

$$\log (- \log S(t|X = 1)) = \log r + \log (- \log (S(t|X = 0))).$$

▶ Consequently, if the proportional hazards model is appropriate, plots of $\log (- \log S(t))$ vs t for each group will be parallel, with the constant difference between them equal to $\log r$, which is the coefficient β.

▶ From equation 1, we see that $- \log S(t)$ is equivalent to the cumulative hazard function, $\Lambda(t)$.

▶ Plots of $\log[- \log S(t)]$ are often called log cumulative hazard plots.

Modelling interactions with time to test and model non-PH

- ▶ Non proportional hazards is just a special name for 'effect modification by time on a log scale'.
- ▶ Effect modification is a familiar concept; we can use interaction terms to test for effect modification and to estimate the effect of exposure in each stratum of the modifier.
- ▶ We can use one of two approaches:
  - ▶ Split by time.
  - ▶ Use the options in R for modelling 'time-varying covariates' (using the `tt()` function in `coxph()`).
- ▶ What we are actually interested in is the situation where the *effect* of a covariate varies by time, which is not the same as the value of covariate varying with time. We'll discuss the distinction in more detail on slide 25.

Modelling interactions with time to test and model non-PH (2)

▶ We do not explicitly estimate the effect of the underlying time scale in a Cox model, but we can estimate interactions with the underlying time scale.

▶ Note that it is possible, using postestimation, to obtain estimates of the underlying time-scale (baseline cumulative hazard and hazard) after fitting a Cox model (see `survival::basehaz` and `biostat3::coxphHaz`).

▶ We still allow the baseline hazard to vary freely, but relax the assumption that hazards must be proportional over time.

Modelling interactions with time, by splitting time

▶ The R function `survSplit()` divides risktime into several records, one for each timeband we specify.

▶ We will now model an interaction with time in the colon carcinoma data, to allow for different hazard ratios for calendar period before and after 2 years (24 months).

▶ We saw in a previous lecture that mortality depends on calendar period of diagnosis (HR 0.72 for recent/early period).

▶ Would we expect mortality in the recent period to be 28% lower at all points in the follow-up or is it conceivable that the effect is greater (or even restricted) to the period immediately following diagnosis?

▶ If the effect is different early in the follow-up, compared to later in the follow-up, then we have a case of non-proportional hazards.

▶ That is, the effect of calendar period is modified by time since diagnosis.

Modelling interactions with time, by splitting time (2)

▶ Based on clinical knowledge, we choose to estimate the effect separately for the first 24 months of follow-up.

▶ We start with splitting the data on time, $t < 24$ months, using `survival::survSplit`.

```
> localised <- survSplit(Surv(surv_mm, status=="Dead: cancer") ~
                         agegrp+sex+year8594,
                         cut=c(24,1000),
                         data=colon, subset=(stage=="Localised"),
                         episode="timeband")
> localised <- transform(localised, timeband = factor(timeband))
```

Modelling interactions with time, by splitting time (3)

▶ We can now fit a model containing the interaction between year of diagnosis (two categories) and time (in two categories).

```
> summary(coxph(Surv(tstart,surv_mm,event)
                ~agegrp+sex+year8594*timeband,
                data=localised))

  n= 10885, number of events= 1734

                                   coef exp(coef) se(coef)      z Pr(>|z|)
agegrp45-59                    -0.05169   0.94962  0.13845 -0.373  0.70888
agegrp60-74                     0.29122   1.33806  0.12573  2.316  0.02055 *
agegrp75+                       0.81496   2.25908  0.12605  6.465 1.01e-10 ***
sexFemale                      -0.09003   0.91390  0.04937 -1.824  0.06822 .
year8594Diagnosed 85-94        -0.42272   0.65526  0.06531 -6.473 9.63e-11 ***
timeband2                           NA        NA  0.00000     NA       NA
year8594Diagnosed 85-94:timeband2  0.32288   1.38110  0.09883  3.267  0.00109 **
```

Modelling interactions with time, by splitting time (4)

- ▶ Recall how we interpret interaction effects (in general).
  - ▶ `year8594Diagnosed 85-94`; effect of the later calendar period of diagnosis (1985–1994)
  - ▶ `timeband2`; effect of time in the second period of follow-up (after 24 months).
  - ▶ `year8594Diagnosed 85-94:timeband2`; additional (multiplicative) effect of later calendar period (1985–1994) at the second period of follow-up (after 24 months).
- ▶ `timeband2` does not have the usual interpretation because we have already adjusted for the effect of time since diagnosis (as the underlying timescale).
- ▶ We are effectively trying to adjust for the same confounder in two different ways in the same model. We should ignore this estimate and focus on the other two.

Modelling interactions with time, by splitting time (5)

- ► The estimated hazard ratio for the effect of period of diagnosis is
  - ► 0.72 when assuming proportional hazards
  - ► 0.66 for the first 24 months of follow-up
  - ► 0.91 after 24 months ($0.655 \times 1.381 = 0.90$)
- ► We see that there is evidence that the effect of period of diagnosis is more pronounced early in the follow-up.
- ► If the interaction effect was zero (HR associated with `year8594Diagnosed 85-94:timeband2` equal to one) then there would be no effect modification (proportional hazards).
- ► We can see that the interaction effect is statistically significant (p=0.001).

Modelling interactions with time, by splitting time (6)

▶ We can reparameterise the model to directly estimate the effect of period within each timeband.

```
> localised <- transform(localised,
                         later=ifelse(year8594=="Diagnosed 85-94",1,0))
> summary(coxph(Surv(tstart,surv_mm,event)~agegrp+sex+later:timeband,
                data=localised))


                   coef exp(coef) se(coef)      z Pr(>|z|)
agegrp45-59     -0.05169   0.94962  0.13845 -0.373   0.7089
agegrp60-74      0.29122   1.33806  0.12573  2.316   0.0205 *
agegrp75+        0.81496   2.25908  0.12605  6.465 1.01e-10 ***
sexFemale       -0.09003   0.91390  0.04937 -1.824   0.0682 .
later:timeband1 -0.42272   0.65526  0.06531 -6.473 9.63e-11 ***
later:timeband2 -0.09984   0.90498  0.07438 -1.342   0.1795
```

▶ The estimated hazard ratio, based on the model, for patients diagnosed 1985–94 compared to 1975–84 is 0.655 for the period up to 2 years of follow-up and 0.905 for the period after 2 years of follow-up (as we previously saw).

Modelling interactions with time, by splitting time (7)

▶ To test if this interaction is statistically significant we could perform a likelihood ratio test, comparing the model with the interaction to the model without the interaction.

```
> fit <- coxph(Surv(tstart,surv_mm,event)~agegrp+sex+
                  year8594*timeband,
               data=localised)
> anova(fit,test="Chisq")
Terms added sequentially (first to last)

                   loglik   Chisq Df Pr(>|Chi|)
NULL              -14442
agegrp            -14360 163.661  3  < 2.2e-16 ***
sex               -14358   2.784  1   0.095209 .
year8594          -14342  32.681  1  1.086e-08 ***
timeband          -14342   0.000  0   1.000000
year8594:timeband -14337  10.648  1   0.001102 **
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

Modelling interactions with time, by splitting time (8)

▶ Note that the previous z test statistic (slide 19) was 3.27. If we square this we get a test statistic that is χ^2_1.

$$3.27^2 = 10.69$$

▶ Both of these tests are testing the hypothesis that the interaction effect is zero versus it is non-zero. The reason for the small difference in the test statistic is that one is a likelihood ratio test and one is a Wald test.

Time-varying covariates

- ▶ We have been considering the situation where the *effect* of a covariate varies with time.

- ▶ It is possible that the underlying *values* of covariates can change during follow-up. For example, blood pressure, occupational exposure to carcinogens, parity, CD4 count, or cumulative exposure to cigarettes.

- ▶ Another application is in observational studies where an intervention may occur at any point in the follow-up. At the time of the intervention, the explanatory variable associated with the intervention changes value from 0 (false) to 1 (true).

- ▶ We highly recommend the time-splitting approach for modelling such data. That is, we split to obtain a separate observation at every value of the time-varying covariate.

- ▶ Care should be taken when modelling time-dependent covariates, particularly with internal variables (which relate to an individual and can only be measured while a patient is alive).

The tt option in coxph

▶ The tt() functions in coxph can also be used for estimating time-varying effects of covariates.

▶ Let's again fit the model where we allow the effect of period to differ in the first 2 years of follow-up.

```
> colon2 <- transform(colon, later=ifelse(year8594=="Diagnosed 85-94",1,0))
> summary(coxph(Surv(surv_mm,status=="Dead: cancer")~agegrp+sex+year8594+tt(later),
                data=colon2, subset=(stage=="Localised"),
                tt = function(x, t, ...) x*(t>=24)))

                         exp(coef) exp(-coef) lower .95 upper .95
agegrp45-59                 0.9496     1.0531    0.7239    1.2457
agegrp60-74                 1.3381     0.7474    1.0458    1.7120
agegrp75+                   2.2591     0.4427    1.7646    2.8922
sexFemale                   0.9139     1.0942    0.8296    1.0068
year8594Diagnosed 85-94     0.6553     1.5261    0.5765    0.7447
tt(later)                   1.3811     0.7241    1.1379    1.6763
```

The tt option in coxph (2)

▶ The cutoff at 24 months was chosen arbitrarily. For the first 6 months of follow-up the estimated hazard ratio was 0.724, for the first year it was 0.676, and for the first two years it was 0.657.

▶ Choosing the cutpoint after inspection of the data will invalidate statistical inference (i.e. reported P-values will be too low).

▶ We have examined only one possible alternative to proportional hazards (a step function with a single step at 24 months).

▶ In practice, it is possible to fit any model of the form

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t) \exp(\beta_1 X_1 + \beta_2 X_1 f(t)),$$

where $f(t)$ is a function of time.

Tests of the PH based on Schoenfeld residuals

▶ If the PH assumption holds then the Schoenfeld residuals (a diagnostic specific to the Cox model) should be independent of time.

▶ In its simplest form, when there are no ties, the unscaled Schoenfeld residual for covariate $x_u, u = 1, ..., p$, and for observation j observed to fail is

$$r_{uj} = x_{uj} - \frac{\sum_{i \in R_j} x_{ui} \exp(\mathbf{x}_i \hat{\beta}_{\mathbf{x}})}{\sum_{i \in R_j} \exp(\mathbf{x}_i \hat{\beta}_{\mathbf{x}})}$$

▶ That is, r_{uj} is the difference between the covariate value for the failed observation and the weighted average of the covariate values over all those subjects at risk of failure when subject j failed.

▶ A test of the PH assumption can be made by modelling the scaled Schoenfeld residuals as a function of time and testing the hypothesis of a zero slope.

Application to localised colon carcinoma

```
fit1 <- coxph(Surv(surv_mm/12,status=="Dead: cancer")~sex+agegrp+year8594,
              data=colon, subset=(stage=="Localised"))

> cox.zph(fit1)
          chisq df        p
sex        0.489  1  0.4845
agegrp    37.420  3 3.7e-08
year8594  11.025  1  0.0009
GLOBAL    50.705  5 9.9e-10
```

▶ The tests suggest there is evidence that the hazards are non-proportional by calendar period and age group.

▶ Rather than just fitting a straight line to the residuals and testing the hypothesis of zero slope (as is done by cox.zph) we can study a plot of the residuals along with a smoother to assist us in determining how the mean residual varies as a function of time.

Application to localised colon carcinoma (2)

▶ The smoother fitted to the residuals illustrates how the log hazard ratio varies as a function of time. We see, for example, that the effect of period is larger during the initial years of follow-up.

```
> fit2 <- coxph(Surv(surv_mm,status=="Dead: cancer")~sex+agegrp+year8594,
                data=colon, subset=(stage=="Localised"))
> plot(cox.zph(fit2,transform=log)[5])
```
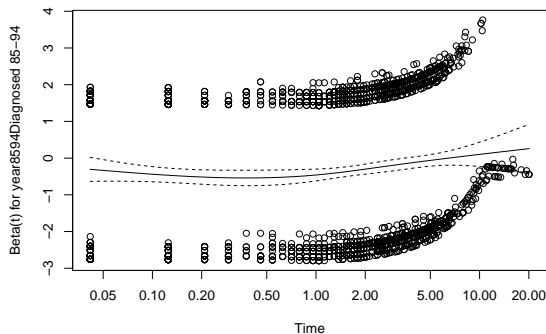
Application to localised colon carcinoma (3)

```
> plot(cox.zph(fit,transform=log)[1])
```
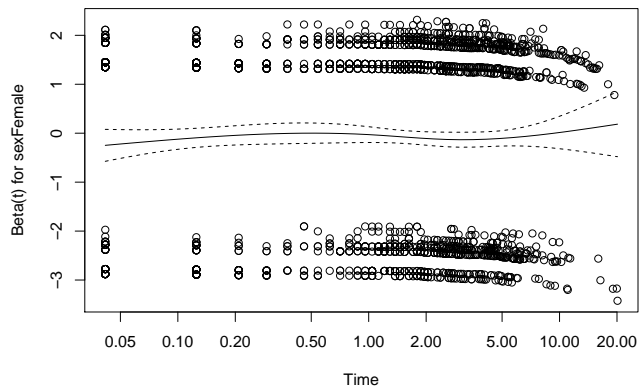
A model including stage

```
> known <- transform(colon, stage=droplevels(stage, "Unknown"))
> fit3 <- coxph(Surv(surv_mm/12,status=="Dead: cancer")~
                sex+agegrp+stage+year8594,
              data=known)
> summary(fit3)

  n= 13208, number of events= 7186
    (2356 observations deleted due to missingness)


                         coef exp(coef) se(coef)      z Pr(>|z|)
sexFemale             -0.04651   0.95456  0.02437 -1.908   0.0564 .
agegrp45-59            0.08546   1.08922  0.06382  1.339   0.1806
agegrp60-74           0.27355   1.31462  0.05868  4.662 3.13e-06 ***
agegrp75+             0.62357   1.86557  0.05937 10.504  < 2e-16 ***
stageRegional         0.83689   2.30916  0.04109 20.367  < 2e-16 ***
stageDistant          2.11896   8.32251  0.02937 72.150  < 2e-16 ***
year8594Diagnosed 85-94 -0.15366   0.85756  0.02399 -6.406 1.49e-10 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

▶ Stage is categorised into Localised, Regional and Distant tumours.
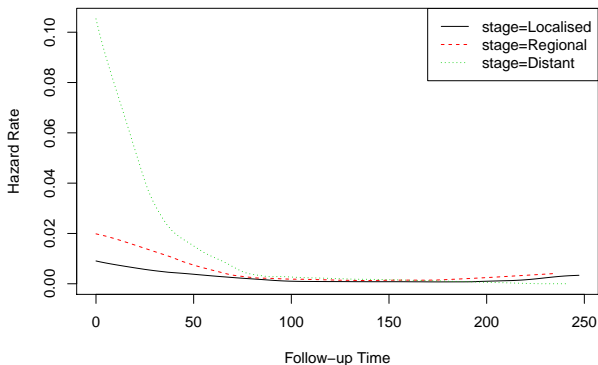
A model including stage (2)

```
> cox.zph(fit3)
           chisq df      p
sex        0.739  1   0.39
agegrp    82.798  3 <2e-16
stage    123.211  2 <2e-16
year8594   0.892  1   0.35
GLOBAL   209.888  7 <2e-16
```

▶ Evidence that the hazards are heavily non-proportional by stage.

▶ A plot of the empirical hazards (slide 34) suggests that individuals diagnosed with distant metastases have proportionally higher mortality early in the follow-up but once they have survived several years their mortality is not that much higher than the other age groups.

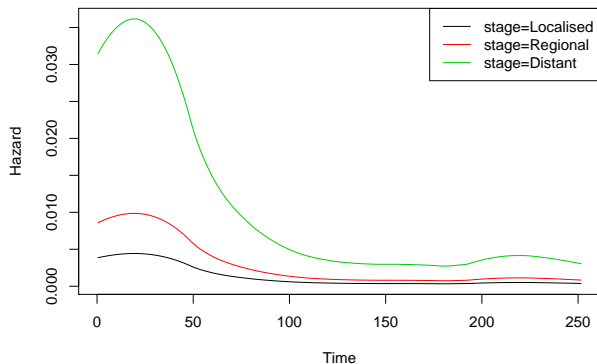▶ The plots of the fitted hazards (slide 35) show the effect of the assumption of proportional hazards.

A model including stage (3)

```
> fit4 <- muhaz2(Surv(surv_mm,status=="Dead: cancer")~stage,
                 data=known)
> plot(fit4)
```

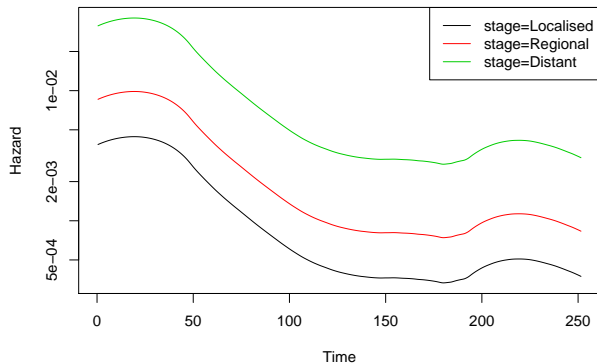# A model including stage (4)

```
> fit5 <- coxph(Surv(surv_mm,status=="Dead: cancer")~stage,
                data=known)
> plot(coxphHaz(fit5,newdata=data.frame(stage=levels(known$stage)))
```
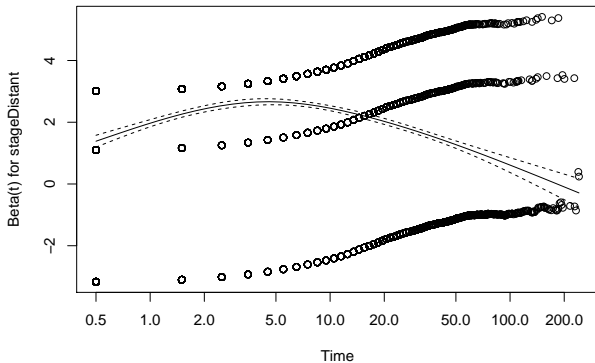
A model including stage (5)

```
> plot(coxphHaz(fit5,newdata=data.frame(stage=levels(known$stage))),
       log="y")
```

A model including stage (6)

```
> fit6 <- coxph(Surv(surv_mm,status=="Dead: cancer")~stage,
                data=known)
> plot(cox.zph(fit6)[2])
```

The stratified Cox model

- The Cox model assumes that the baseline hazard is an arbitrary function of time.
- The hazard functions for each of the other groups are assumed to be proportional to the baseline.
- It is possible to relax this assumption to allow separate baseline hazards for different groups, say for each level of age at diagnosis.
- This is known as a stratified proportional hazards model and is a useful method for modelling data where non-proportional hazards are suspected for a factor that is not of primary interest.
- Use the `strata()` term in the `coxph` formula to specify the strata variables.

The stratified Cox model (2)

```
> fit7 <- coxph(Surv(surv_mm/12,status=="Dead: cancer")~
                   sex+year8594+strata(agegrp),
               data=colon, subset=(stage=="Localised"))
> summary(fit7)

  n= 6274, number of events= 1734


                            coef exp(coef) se(coef)      z Pr(>|z|)
sexFemale               -0.08958   0.91431  0.04938 -1.814   0.0697 .
year8594Diagnosed 85-94 -0.28200   0.75427  0.04942 -5.707 1.15e-08 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
```

References

[1] Putter H, Sasako M, Hartgrink HH, van de Velde CJH, van Houwelingen JC. Long-term survival with non-proportional hazards: results from the Dutch gastric cancer trial. *Stat Med* 2005; **24**:2807–2821.

[2] Jatoi I, Anderson WF, Jeong JH, Redmond CK. Breast cancer adjuvant therapy: time to consider its time-dependent effects. *J Clin Oncol* 2011;**29**:2301–2304.