Cox regression models with multiple predictors: interactions, comparing models, model diagnostics

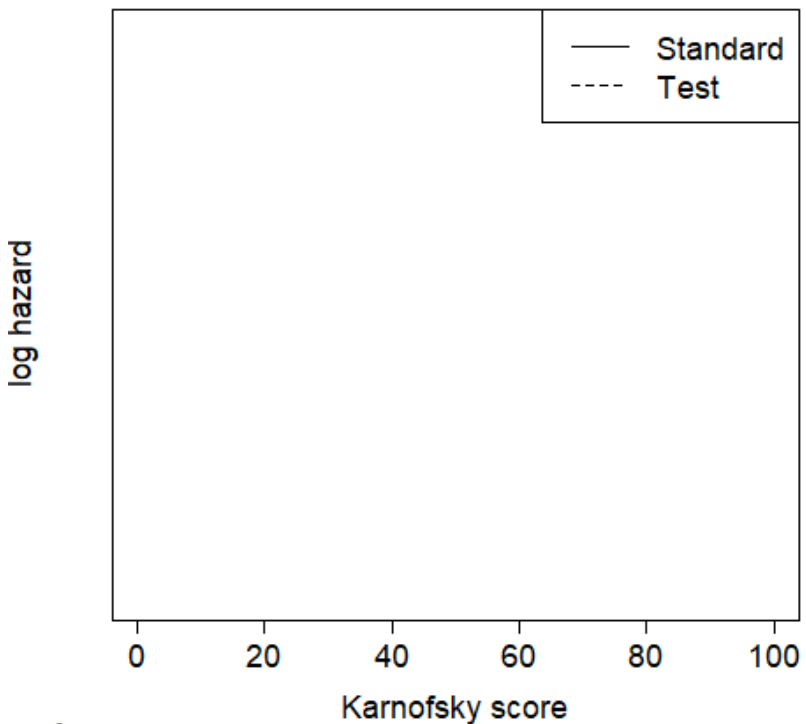
Shannon Pileggi

STAT 417

OUTLINE

Visualizing interactions

Recall the VALCG lung cancer study. Suppose we believe that the standard treatment is more effective than the test treatment for patients with lower Karnofsky scores, but less effective than the test treatment for patients with higher Karnofsky scores. Create a plausible sketch of this scenario.



Interactions

What does an interaction between X_1 and X_2 imply?

When an interaction exists:

1. Include both X_1 and X_2 in the model (main effects).
2. In addition, include *interaction terms*:

X_1	X_2	Interaction term(s)
quant.	quant	X_1X_2
quant	cat. (k)	$X_1D_1, X_1D_2, ..X_1D_{k-1}$
cat. (j)	cat. (k)	$(j-1)(k-1)$ products of dummy variables

Discussion

Suppose we are using a Cox regression model for the hazard of lung cancer with the variables $X_1 =$ Karnofsky score, $X_2 =$ treatment (where $X_2 = 0$ for standard, and $X_2 = 1$ for test treatment), and the interaction between X_1 and X_2.

How many *parameters* are estimated for this model?

1

2

3

4

5

Write out the model:

Interactions in R

In R model formula syntax:

- `:` indicates interaction
- `*` indicates main effects plus interaction

The following two model specifications are *equivalent*:

R Code

```
CR_mod1 <- coxph(Surv(time, status) ~ karno + trt + karno:trt,
                 data = veteran)

CR_mod2 <- coxph(Surv(time, status) ~ karno*trt,
                 data = veteran)
```

R Code

R output

R Output

```
Call:
coxph(formula = Surv(time, status) ~ karno * trt, data = veteran)

  n= 137, number of events= 128

               coef exp(coef)  se(coef)      z Pr(>|z|)
karno     -0.008668  0.991369  0.016704 -0.519   0.6038
trt        1.093251  2.983959  0.606419  1.803   0.0714 .
karno:trt -0.015867  0.984258  0.009954 -1.594   0.1109
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

          exp(coef) exp(-coef) lower .95 upper .95
karno        0.9914     1.0087    0.9594     1.024
trt          2.9840     0.3351    0.9091     9.794
karno:trt    0.9843     1.0160    0.9652     1.004
```

R Output

Understanding the interaction

- Estimate the hazard ratio corresponding to a 10 point increase in Karnofsky score for those on the *standard* treatment.

- Estimate the hazard ratio corresponding to a 10 point increase in Karnofsky score for those on the *test* treatment.

- What does this mean?

Assessing significance of the interaction

R Output

```
                coef exp(coef)  se(coef)      z Pr(>|z|)
karno     -0.008668  0.991369  0.016704 -0.519   0.6038
trt        1.093251  2.983959  0.606419  1.803   0.0714
karno:trt -0.015867  0.984258  0.009954 -1.594   0.1109
...
Likelihood ratio test= 45.52  on 3 df,   p=7.173e-10
Wald test            = 49.11  on 3 df,   p=1.238e-10
Score (logrank) test = 51.39  on 3 df,   p=4.04e-11
```

R Output

Is there evidence that the association between hazard and Karnofsky score depends on treatment?

Model selection

- There are many strategies to select a final set of predictors.
- One method is to compare *nested* models:
  - The **full model** contains the predictors $X_1, X_2, \ldots, X_f$.
  - The **reduced model** contains a subset of those predictors, say r out of the f predictors where $r < f$.
- For formal comparison of nested models, use the partial likelihood ratio test by comparing the evaluated log partial likelihood functions of the full and reduced models.
- This allows us to determine if the inclusion of additional predictors into a model can "improve" the model, i.e. the goodness-of-fit is significantly improved.

VALCG lung cancer potential predictors

`karno`	X_1	Karnofsky score (100=good; 0=dead); (measures cancer patients' functional impairment)
`trt`	X_2	Treatment type (0=standard, 1=test)
`celltype`	X_3	cancer cell type (squamous, smallcell, adeno, large)
`age`	X_4	patients age in years at time of treatment start
`diagtime`	X_5	months from diagnosis to treatment assignment
`prior`	X_6	prior therapy to the treatment (0=no, 1=yes)

Discussion

Consider the following model which contains:

Model 0: `celltype`, `age`, `trt`, and `celltype`×`trt`

Which of the following models would be nested in Model 0?

Model 1: `celltype`, `age`, `karno`

Model 2: `age`, `trt`

Model 3: `celltype`, `trt`, and `celltype`×`trt`

Model 4: `celltype`, `age`, `trt`, `prior`

Discussion

Model 1: `karno`, `trt`, `celltype`, `age`, `karno`×`age`, `karno`×`celltype`

Model 2: `karno`, `trt`, `celltype`, `age`

How many parameters will be estimated in

Model 1?

Model 2?

Determine if the regression model improves fit with the addition of the interaction terms.

H_0:

H_a:

Partial likelihood ratio test

$f =$

$r =$

$l_f =$

$l_r =$

$df =$

$G_l =$

Create the models

R Code

```
CR_full <-
    coxph(Surv(time, status) ~ karno + trt + celltype + age +
                               karno:age + karno:celltype,
          data = veteran)

CR_red <-
    coxph(Surv(time, status) ~ karno + trt + celltype + age,
          data = veteran)
```

R Code

Full model

R Output

	coef	exp(coef)	se(coef)	z	Pr(>
karno	-0.1135446	0.8926644	0.0366088	-3.102	0.0
trt	0.2824701	1.3264021	0.2081975	1.357	0.1
celltypelarge	-1.5886004	0.2042112	1.1776847	-1.349	0.1
celltypesmallcell	-1.4601698	0.2321968	0.8357294	-1.747	0.0
celltypesquamous	-0.9505561	0.3865260	0.9233413	-1.029	0.3
age	-0.0769218	0.9259623	0.0313400	-2.454	0.0
karno:age	0.0012390	1.0012398	0.0005325	2.327	0.0
karno:celltypelarge	0.0095336	1.0095792	0.0178371	0.534	0.5
karno:celltypesmallcell	0.0179957	1.0181586	0.0138276	1.301	0.1
karno:celltypesquamous	-0.0069418	0.9930822	0.0150711	-0.461	0.6

R Output

Reduced model

R Output

```
                    coef exp(coef)  se(coef)      z Pr(>|z|)
karno           -0.032685  0.967843  0.005409 -6.043 1.51e-09 ***
trt              0.303048  1.353980  0.205656  1.474   0.1406
celltypelarge   -0.776475  0.460025  0.297703 -2.608   0.0091 **
celltypesmallcell -0.322467  0.724360  0.268923 -1.199   0.2305
celltypesquamous -1.178807  0.307646  0.296440 -3.977 6.99e-05 ***
age             -0.008903  0.991136  0.009224 -0.965   0.3345
```

R Output

Computing the test statistic

R Output

```
> CR_full$loglik
[1] -505.4491 -469.9852

> CR_red$loglik
[1] -505.4491 -474.4578
```

R Output

Match the following values to: $l_p(0)$, l_f, l_r.

1. -505.4491
2. -469.9852
3. -474.4578

1. Calculate the partial likelihood ratio test statistic.
2. What is the degrees of freedom?

Computing the *p*-value and conclusion

R Code

```
pchisq(q = 8.945, df = 4, lower.tail = F)
```

R Code

R Output

```
0.06248902
```

R Output

What does this suggest about the inclusion of interaction terms in the model?

AIC for non-tested models

- We can use Akaike's Information Criterion (AIC) to compare non-nested models

$$AIC = -2l_p(\hat{\beta}) + 2p$$

- The *smaller* the value of AIC, the *better* the fit of the current model to the data.
- There is no formal inference procedure to compare AIC values for different models.
- Some subjective judgment must be used to determine if a model with a slightly smaller value of AIC is better.

Comparing non-nested models

R Code

```
CR_mod1 <- coxph(Surv(time, status) ~ karno + celltype,
                 data = veteran)

CR_mod2 <- coxph(Surv(time, status) ~ trt + celltype + age,
                 data = veteran)
```

R Code

R Output

```
> CR_mod1$loglik
[1] -505.4491 -475.7632

> CR_mod2$loglik
[1] -505.4491 -492.4281
```

R Output

M1: AIC =

M2: AIC =

The idea

- We can compute *predicted survival curves* for *fixed* values of explanatory variables.
- The CR model is given by:

$$h(t|X_1, \ldots, X_p) = h_0(t) \exp(\beta_1 X_1 + \cdots + \beta_p X_p)$$

- The cumulative hazard function at time t given $X_1, \ldots, X_p$:

$$H(t|X_1, \ldots, X_p)$$

- The survival function at time t given $X_1, \ldots, X_p$:

$$S(t|X_1, \ldots, X_p)$$

- Putting it together:

$$S(t) = e^{-H(t)}$$

Predictor-adjusted survival probabilities

$$
\begin{aligned}
S(t|X_1,\ldots,X_p) &= e^{H(t|X_1,\ldots,X_p)} \\
&= e^{-\int_0^t h(y|X_1,\ldots,X_p)dy} \\
&= e^{-\int_0^t h_0(y)\exp\left(\sum_{j=1}^p \beta_j X_j\right)dy} \\
&= \left[e^{-\int_0^t h_0(y)dy}\right]^{\exp\left(\sum_{j=1}^p \beta_j X_j\right)} \\
&= [S_0(t)]^{\exp\left(\sum_{j=1}^p \beta_j X_j\right)}
\end{aligned}
$$

$$
\hat{S}(t|X_1,\ldots,X_p) = \left[\hat{S}_0(t)\right]^{\exp\left(\sum_{j=1}^p \hat{\beta}_j X_j\right)}
$$

Lung cancer example

Estimate the survival probabilities for lung cancer patients with Karnofsky scores of 10 and 90.

R Code

```
KM_obj <- survfit(Surv(time, status) ~ 1,
                  conf.type = "none",
                  data = veteran)

CR_mod <- coxph(Surv(time, status) ~ karno, data = veteran)

pred_max_karno <- data.frame(karno = 90, veteran)
pred_min_karno <- data.frame(karno = 10, veteran)

adj_surv_max <- survfit(CR_mod, newdata = pred_max_karno)
adj_surv_min <- survfit(CR_mod, newdata = pred_min_karno)
```

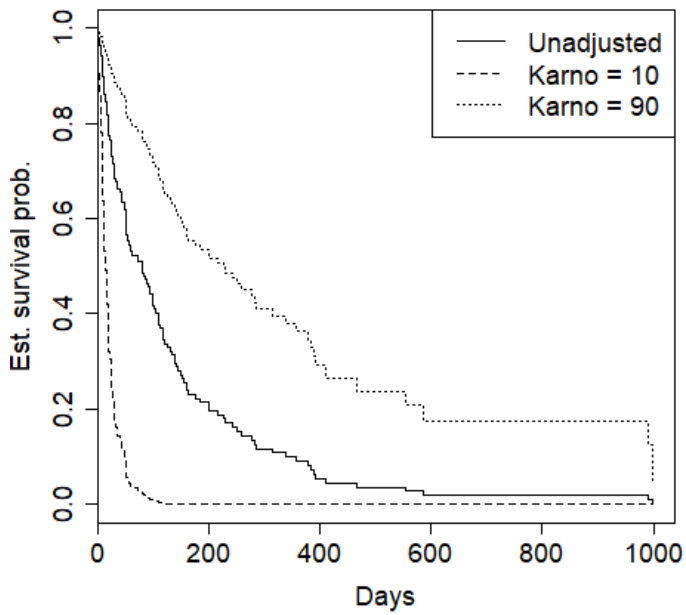
R Code

Adjusted survival curves

R Code

```
plot(KM_obj,
     xlab = "Days",
     ylab = "Est. survival prob.")
lines(adj_surv_min, lty = 2)
lines(adj_surv_max, lty = 3)
legend("topright",
       c("Unadjusted",
         "Karno = 10",
         "Karno = 90"),
       lty = 1:3)
```

R Code



Model assumptions and other checks

Assumptions to assess:

- *Proportional hazards*: the effect of the predictors on hazard does not depend on (vary by) time:

- *Effect of each predictor on log hazard is linear*:

Other checks:

- *Outliers*, i.e. identify subjects whose outcome is poorly predicted by the CR model.
- *Influential subjects*, i.e. individuals with "unusual" values for a specific predictor.

Residuals

- Recall from linear regression *residuals* are used to check key model assumptions and evaluate model adequacy.
- For linear models, the residual for the i^{th} subject is:

$$e_i = y_i - \hat{y}_i$$

- There is no exact analog to this definition when using the CR model, but there are several "developed" types of residuals that can be used for checking different aspects of model adequacy:
  - Martingale residuals
  - Deviance residuals
  - Schoenfeld residuals
  - Score residuals

Martingale residuals

Use Assessing the adequacy of the functional form of a predictor, or determining what the functional form should be.

Definition Difference between the *observed* number of events experienced by subject i and the *expected* number of events experienced by subject i (under the current model).

Applies to All subjects, for each predictor

Interpretation > 0: subject i experienced the event earlier than expected
< 0: subject i experienced the event later than expected

Result vector of length n (# of subjects)

Plot Plot values of each continuous predictor (x) versus values of the martingale residuals (y).

Check No pattern implies the functional form of the predictor is correct; distinct pattern implies a transformation may be appropriate for X.

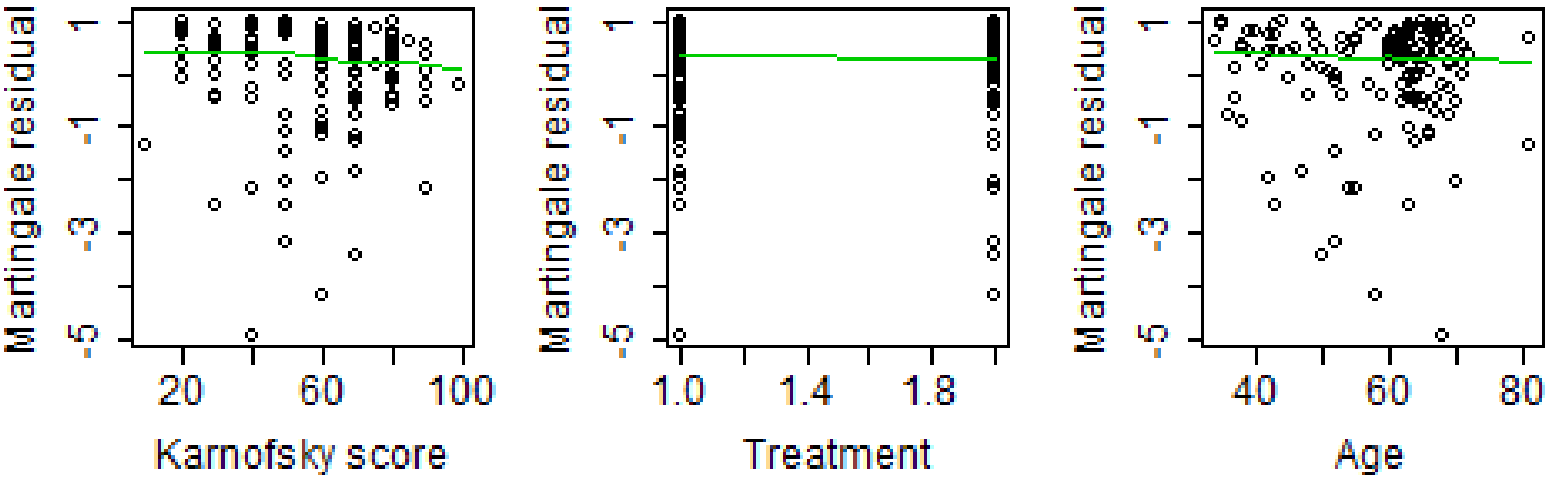
Martingale residuals in R

R Code

```
CR_mod <- coxph(Surv(time, status) ~ karno + trt + age,
                data = veteran)
veteran$mart <- residuals(CR_mod, type = "martingale")

par(mfrow = c(1,3), pty = "s")
plot(x = veteran$karno, y = veteran$mart,
     xlab = "Karnofsky score", ylab = "Martingale residual")
lines(lowess(veteran$karno, veteran$mart), col = 3)

plot(x = veteran$trt, y = veteran$mart,
     xlab = "Treatment", ylab = "Martingale residual")
lines(lowess(veteran$trt, veteran$mart), col = 3)

plot(x = veteran$age, y = veteran$mart,
     xlab = "Age", ylab = "Martingale residual")
lines(lowess(veteran$age, veteran$mart), col = 3)
```

R Code

Martingale residual plots



Deviance residuals

Use Check for outliers, i.e. identify individuals whose outcome is poorly predicted

Definition Martingale residuals transformed to have mean 0 and roughly normal distribution

Applies to All subjects, for each predictor

Interpretation > 0: subject i experienced the event earlier than expected
< 0: subject i experienced the event later than expected

Result vector of length n (# of subjects)

Plot Plot subject IDs (x) versus deviance residuals (y) to determine which individuals should be screened.

Check Subjects with $|\text{deviance residuals}| > 2.5$

Deviance residuals in R

R Code

```
CR_mod <- coxph(Surv(time, status) ~ karno + trt + age,
                data = veteran)
veteran$deviance <- residuals(CR_mod, type = "deviance")

plot(x = 1:nrow(veteran),
     y = veteran$deviance,
     xlab = "Subject index",
     ylab = "Deviance residuals")
abline(h = 0)

veteran[abs(veteran$deviance) > 2.5, ]
```

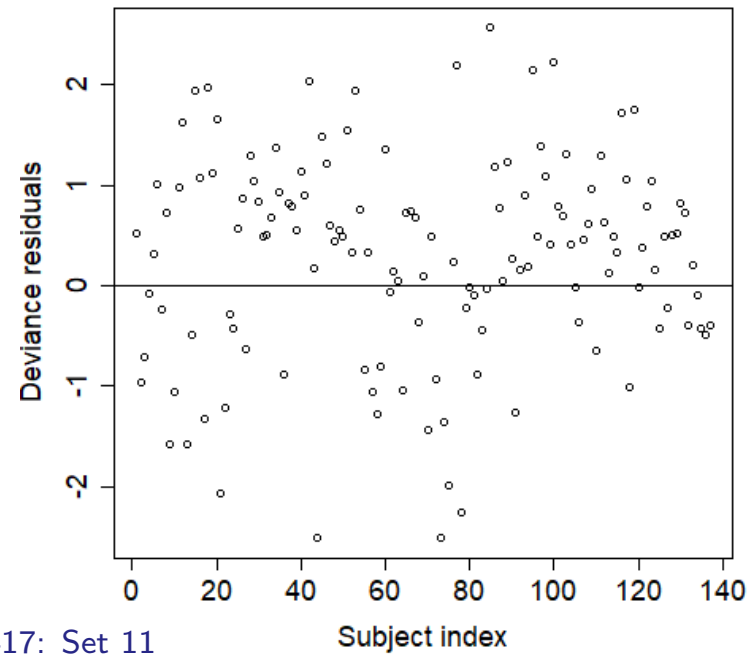
R Code

Deviance residual plots

R Output

	trt	celltype	time	status	karno	diagtime	age	prior	deviance
44	1	smallcell	392	1	40	4	68	0	-2.510903
73	2	squamous	231	0	50	8	52	10	-2.515511
85	2	squamous	1	1	50	7	35	0	2.566946

R Output



Schoenfeld residuals

Use Identify violations of the proportional hazards assumption

Definition Difference between observed and expected predictor values (average value of the predictor at the subject's event time)

Applies to Only subjects with complete event times, for each predictor

Interpretation > 0: subject i has a larger predictor value than the event time suggests
< 0: subject i has a smaller predictor value than the event time suggests

Result $(m \times p)$ matrix; $m =$ complete event times, $p =$ #predictors

Plot Plot the ordered *complete* event times (x) versus the Schoenfeld residuals (y) for each predictor

Check Random pattern (flat smoothed curve) implies the proportional hazard (PH) assumption *not* violated; otherwise, PH violated.

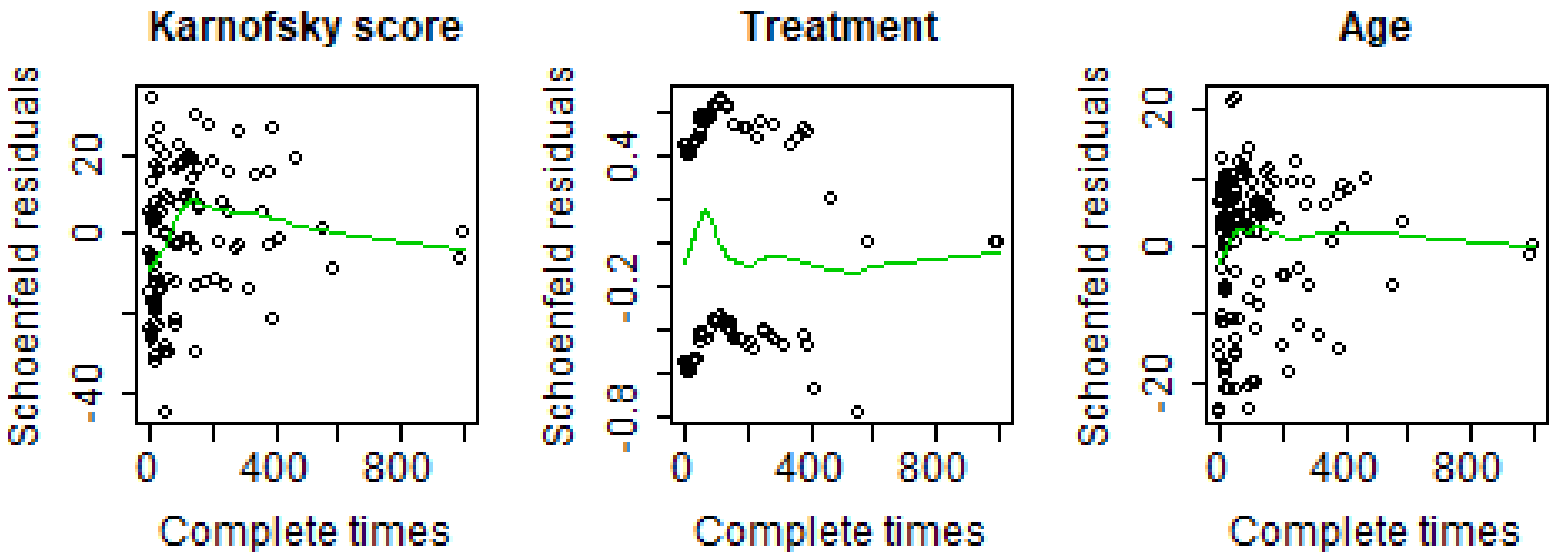
Schoenfeld residuals in R

R Code

```
schoen <- residuals(CR_mod, type = "schoenfeld")
complete_times <- sort(veteran$time[veteran$status!=0])

par(mfrow = c(1,3), pty = "s")
plot(x = complete_times, y = schoen[,1], main = "Karnofsky score",
     xlab = "Complete times", ylab = "Schoenfeld residuals")
lines(lowess(complete_times, schoen[,1]), col = 3)

plot(x = complete_times, y = schoen[,2], main = "Treatment",
     xlab = "Complete times", ylab = "Schoenfeld residuals")
lines(lowess(complete_times, schoen[,2]), col = 3)

plot(x = complete_times, y = schoen[,3], main = "Age",
     xlab = "Complete times", ylab = "Schoenfeld residuals")
lines(lowess(complete_times, schoen[,3]), col = 3)
```

R Code

Schoenfeld residual plots



Formal test for PH assumption

- It has been proposed (Grambsch and Therneau, 1994) that we can assume that β depends on a function of time t by the relationship:

  $$\beta_j(t) = \beta_j + \gamma_j g_j(t),$$

  where g is a specified function of time. A commonly used function is $g(t) = \ln(t)$.
- Under this model, we test whether γ_j is significantly different from 0, i.e:

  H_0: $\gamma_j = 0$ vs H_a: $\gamma_j \neq 0$
- Failing to reject H_0 implies:

  The PH assumption for x_j has not been violated.
- Rejecting H_0 implies:

  The PH assumption for x_j has been violated.

Formal test for PH assumption, in R

R Code

```
CR_mod <- coxph(Surv(time, status) ~ karno + trt + age,
                data = veteran)
cox.zph(CR_mod, transform = "log")
```

R Code

R Output

```
           rho chisq        p
karno   0.3154 12.13 0.000497
trt    -0.0954  1.22 0.270034
age     0.2121  6.58 0.010308
GLOBAL      NA 15.66 0.001334
```

R Output

Score residuals

Use Detecting overly influential observations

Definition A function of the approximate change in the parameter estimate that would result if the i^{th} subject were removed from the sample.

Applies to All subjects, for each predictor

Interpretation > 0: parameter estimate ($\hat{\beta}_j$) would increase without individual i
< 0: parameter estimate ($\hat{\beta}_j$) would decrease without individual i

Result $(n \times p)$ matrix; $n = \#$ subjects, $p = \#$predictors

Plot Plot subject index (x) versus the score residuals (y) for each predictor.

Check Subjects with "large" (in magnitude) score residuals are more influential on fit (no rule of thumb).

Score residuals in R

R Code

```
score <- residuals(CR_mod, type = "score")

veteran[score[,1] > 100 |
        score[,1] < -50 |
        abs(score[,2]) > 2 |
        abs(score[,3]) > 40, ]

plot(x = 1:nrow(veteran),
     y = score[,1],
     xlab = "Subject index",
     ylab = "Score residuals",
     main = "Karnofsky score")
```

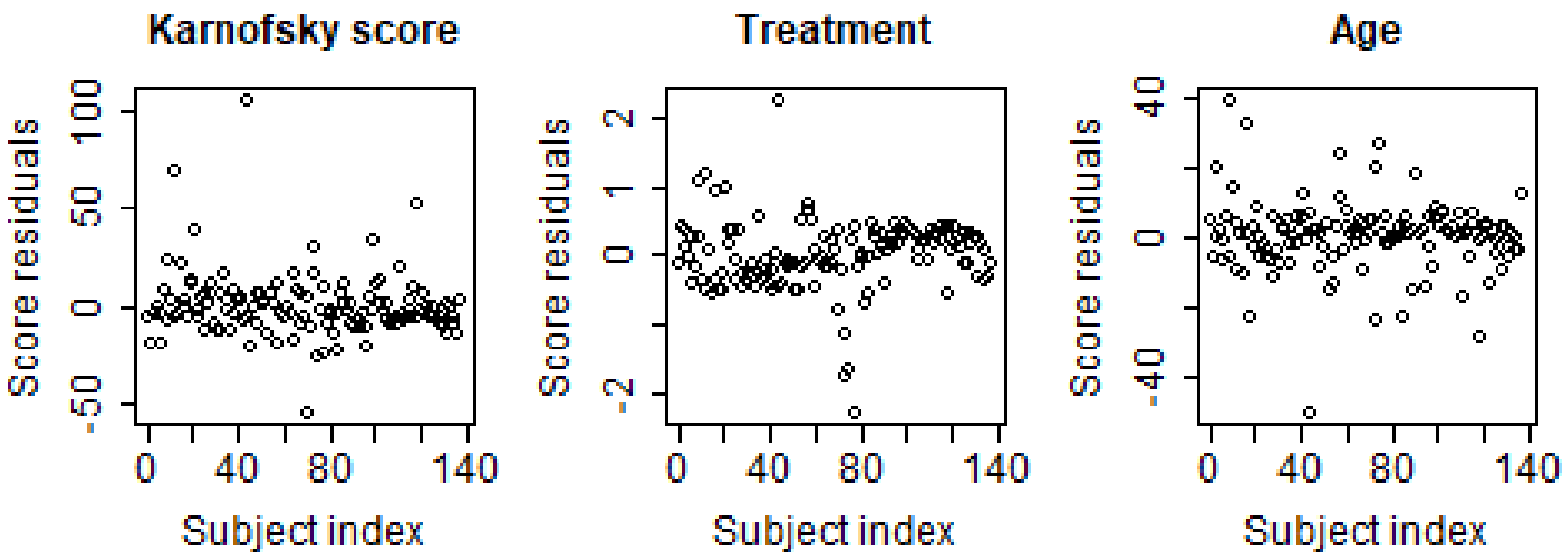
R Code

Score residual plots

R Output

	trt	celltype	time	status	karno	diagtime	age	prior
44	1	smallcell	392	1	40	4	68	0
70	2	squamous	999	1	90	12	54	10
78	2	squamous	587	1	60	3	58	0

R Output



Addressing findings

- If you have identified some outlying or influential subjects:
  1. Re-fit the model without the subjects who are influential and/or considered outliers.
  2. Carefully decide if these observations should be included or not.

- If you found a violation in the proportionality assumption:
  1. Fit a "stratified" model which posits the existence of different baseline hazard functions.
  2. Fit a model that includes an interaction term between the predictor and the time variable (or a function of time); we'll focus more on strategy # 1.

Stratified Cox model

- When one or more predictors does not satisfy the proportional hazards assumption, the standard CR model is not appropriate, since we can no longer assume an identical baseline hazard function, $h_0(t)$, for each individual.
- To accommodate different baseline hazard functions, a *stratified CR model* can be fit that allows separate baseline hazard functions (not necessarily known) for each value (stratum) of the *stratification variable*.
- Suppose the stratification variable has s levels. The stratified CR model assumes a different baseline hazard function corresponding to each value of the stratification variable, i.e. the form of the model is:

$$h_s(t|X_1, ...X_p) = h_0 g(t) \exp(\beta_1 X_1 + \cdots + \beta_1 X_1)$$

where $g = 1, \ldots, s$ and $s =$ number of levels of stratification variable.

Stratified Cox model, cont.

- The interpretation of the (estimated) coefficients and (estimated) hazard ratios are made adjusting for the other variables in the model *and* the stratification variable.
- **Violating predictor X is categorical:**

  each stratum can correspond to a different value of the predictor, and the stratification variable is identical to X.
- **Violating predictor X is quantitative:**

  create a stratification variable by dividing X into appropriate ranges, with each range corresponding to a stratum.
- **Several violating predictors $X_1, \ldots, X_k$:**

  then a stratification variable can be created based on combinations of $X_1, \ldots, X_k$.

Stratified Cox model in R

R Code

```
CR_mod_stratified <-
    coxph(Surv(time, status) ~ karno + strata(trt) + age,
          data = veteran)
summary(CR_mod_stratified)
```

R Code

Stratified Cox model in R

R Output

```
Call:
coxph(formula = Surv(time, status) ~ karno + strata(trt) + age,
    data = veteran)

  n= 137, number of events= 128

           coef exp(coef)  se(coef)      z Pr(>|z|)
karno -0.034227  0.966352  0.005417 -6.319 2.64e-10 ***
age   -0.003383  0.996622  0.009203 -0.368    0.713
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

      exp(coef) exp(-coef) lower .95 upper .95
karno    0.9664      1.035    0.9561    0.9767
age      0.9966      1.003    0.9788    1.0148
```

R Output

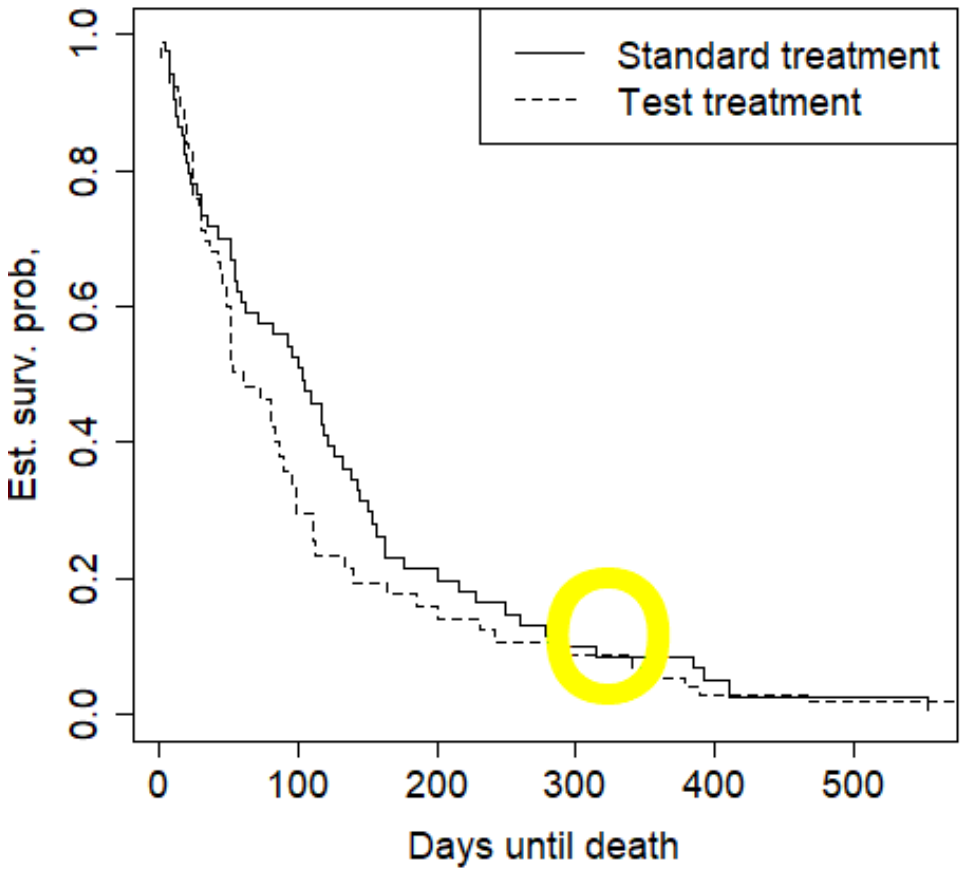
Stratified Cox model, interpretations

- Even though you do not see `trt` in the `R` output, it is not being ignored! It is taken into account in the baseline hazard function.
- The hazard of death is estimated to decrease b 3.4% for a 1 point increase in Karnofsky score, adjusting for age and treatment.
- The estimated predictor adjusted survival curves are:

test trt: $\left[\hat{S}_{01(t)}\right]^{\exp(-0.034X_1 - 0.003X_2)}$

standard trt: $\left[\hat{S}_{02(t)}\right]^{\exp(-0.034X_1 - 0.003X_2)}$

Estimated stratified predictor-adjusted survival curves



Estimated survival curves now cross!