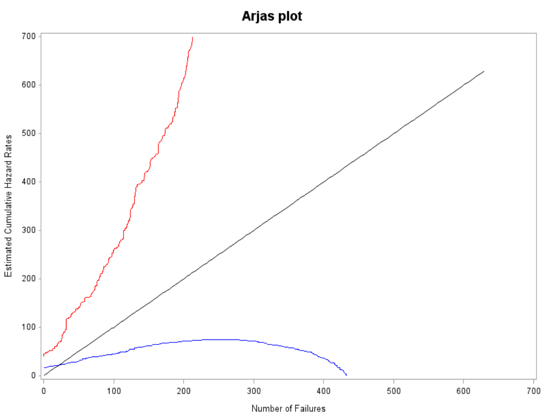
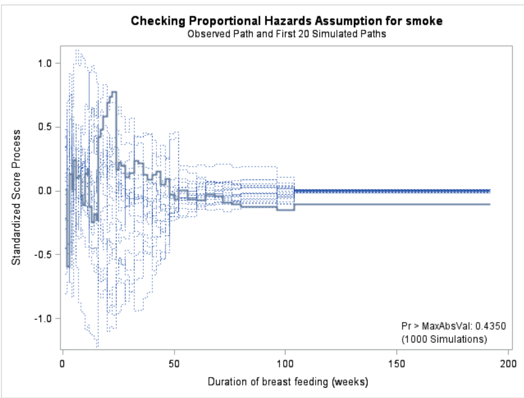
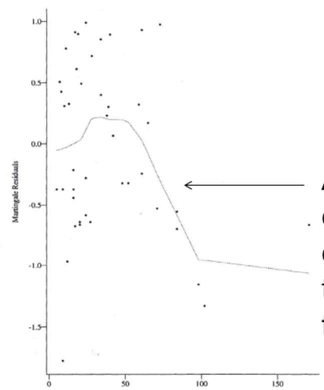
UPPSALA UNIVERSITET

# Analysis of Survival Data

## Lecture 6: Regression diagnostics



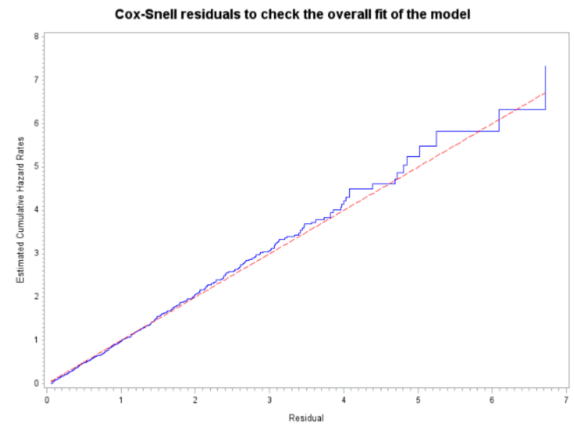
A drop, suggesting that this covariate could be categorized to better explain the relationship with the time variable

Fixed-time covariate (for which we want to test the PH assumption)

$g(t)$ is a known function of time

$$h(t \mid \mathbf{Z}) = h_0(t) \exp\left\{\beta_1 Z_1 + \beta_2 \underbrace{\left(Z_1 g(t)\right)}\right\}$$

Artificially created time-dependent covariate




Inger Persson

UPPSALA UNIVERSITET

# Program L6

- **Regression for survival data**
  - Cox's proportional hazards regression, cont'd
    - Testing the PH assumption
    - Graphical methods to investigate the PH assumption
    - What to do if the proportional hazards assumption is violated
    - Regression diagnostics

# Tests of PH assumption

A number of tests of the PH assumption available.

Most common:
Time-dependent covariate test (Cox).

UPPSALA UNIVERSITET

# Testing the assumption of proportional hazards

Time-dependent covariate methodology can be used to test the proportional hazards assumption.

One common way:

Fixed-time covariate (for which we want to test the PH assumption)

$g(t)$ is a known function of time

$$h(t \mid \mathbf{Z}) = h_0(t)\exp\{\beta_1 Z_1 + \beta_2 \underbrace{(Z_1 g(t))}_{\text{Artificially created time-dependent covariate}}\}$$

Usually: $g(t) = \ln t$

# Comparing two individuals (proportional hazards)

Comparing two individuals with distinct values of $Z_1$:

$$\frac{h(t \mid Z_1)}{h(t \mid Z_1^*)} = \exp\left(\beta_1 (Z_1 - Z_1^*) + \beta_2 g(t)(Z_1 - Z_1^*)\right)$$

If $\beta_2 = 0$ this function does not depend on $t$.

Testing $H_0$: $\beta_2 = 0$ is thus a test for the PH assumption.

Note: PH ass. not to be checked for time-dep. covariates.

UPPSALA UNIVERSITET

# Example: infection in burns (recap.)

In section 1.6 a study is described to evaluate a protocol change in disinfectant practice in a large midwestern university medical center.

Control of infection is the primary concern for the 154 patients entered into the burn unit with varying degrees of burns.

The outcome variable is the time until infection from admission to the unit. Censoring variables are discharge from the hospital without an infection or death without an infection.

84 patients were in a group which had a body-cleansing method (disinfectant: chlorhexidine) and 70 patients received the routine bathing care method (disinfectant: povidone-iodine).

UPPSALA UNIVERSITET

# Example: infection in burns

**Variables:**

*Trt* = treatment (0=routine bathing, 1=body cleansing)

*TimeStaph* = Time to staphylococcus infection (days)

*Staph* = Staphylococcus indicator (1=infection, 0=no inf.)

<u>Additional explanatory variables (found at previous lecture):</u>

*BurnType* (1=chemical, 2=scald, 3=electric, 4=flame)

*Race* (0=nonwhite, 1=white)

UPPSALA UNIVERSITET

# Example: infection in burns (recap.)

## The “best” model

| Analysis of Maximum Likelihood Estimates | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Parameter | | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | 95% Hazard Ratio Profile Likelihood Confidence Limits | |
| Trt | | 1 | -0.60100 | 0.29787 | 4.0708 | 0.0436 | 0.548 | 0.302 | 0.978 |
| BurnType | chemical | 1 | -1.00348 | 1.01626 | 0.9750 | 0.3234 | 0.367 | 0.021 | 1.706 |
| BurnType | electric | 1 | 1.16885 | 0.45143 | 6.7040 | 0.0096 | 3.218 | 1.200 | 7.274 |
| BurnType | scald | 1 | 0.57509 | 0.45294 | 1.6121 | 0.2042 | 1.777 | 0.660 | 4.018 |
| Race | | 1 | 2.28344 | 1.02559 | 4.9571 | 0.0260 | 9.810 | 2.050 | 176.405 |

# Example: infection in burns
# Testing the PH assumption

```
proc phreg data=burns;
  model TimeStaph*Staph(0)=trt lntTrt /ties=exact;
  lntTrt=log(TimeStaph)*trt;
run;
```

$Z$ (trt)

$Z\,g(t)$ (lntTrt)

$Z\,g(t) = Z\ln(t)$

# Example: infection in burns
# Testing the PH assumption

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| Trt | 1 | -0.23378 | 0.74550 | 0.0983 | 0.7538 | 0.792 |
| IntTrt | 1 | -0.15863 | 0.33385 | 0.2258 | 0.6347 | 0.853 |

OK for trt
(null hypothesis of $\beta = 0$, i.e. no time dependency, is not rejected)

UPPSALA UNIVERSITET

# Example: infection in burns
# Testing the PH assumption

```
proc phreg data=burns;
  model TimeStaph*Staph(0)=chemical scald electric
        lntChemical lntScald lntElectric/ties=exact;
  lntChemical=log(timestaph)*chemical;
  lntScald=log(timestaph)*Scald;
  lntElectric=log(timestaph)*electric;
  test lntChemical=lntScald=lntElectric=0;
run;
```

Note: Dummies must be created for categorical variables (“class” statement cannot be used for this)

# Example: infection in burns
# Testing the PH assumption

UPPSALA UNIVERSITET

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| chemical | 1 | -8.53847 | 8.16805 | 1.0928 | 0.2959 | 0.000 |
| scald | 1 | 2.90012 | 1.22275 | 5.6254 | 0.0177 | 18.176 |
| electric | 1 | 3.59420 | 1.30819 | 7.5485 | 0.0060 | 36.387 |
| IntChemical | 1 | 2.90466 | 2.81679 | 1.0634 | 0.3024 | 18.259 |
| IntScald | 1 | -1.45124 | 0.70469 | 4.2411 | 0.0395 | 0.234 |
| IntElectric | 1 | -1.45821 | 0.81732 | 3.1831 | 0.0744 | 0.233 |

| Linear Hypotheses Testing Results | | | |
|---|---|---|---|
| Label | Wald Chi-Square | DF | Pr > ChiSq |
| Test 1 | 6.5349 | 3 | 0.0883 |

OK for burntype

# Example: infection in burns
# Testing the PH assumption

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| Race | 1 | 15.97393 | 14.61024 | 1.1954 | 0.2742 | 8657459 |
| IntRace | 1 | -4.28022 | 3.91857 | 1.1931 | 0.2747 | 0.014 |

OK for race

# Example: infection in burns
# Testing the PH assumption

But the power of these tests may be low due to small number of events in each group (which is extra important since we "want" a non-significant result)!

| Frequency | Table of Staph by Trt | | | |
|---|---|---|---|---|
| | | Trt | | |
| | Staph | 0 | 1 | Total |
| | 0 | 42 | 64 | 106 |
| | 1 | 28 | 20 | 48 |
| | Total | 70 | 84 | 154 |

| Table of Staph by BurnType | | | | | |
|---|---|---|---|---|---|
| | BurnType | | | | |
| Staph | 1 | 2 | 3 | 4 | Total |
| 0 | 8 | 12 | 5 | 81 | 106 |
| 1 | 1 | 6 | 6 | 35 | 48 |
| Total | 9 | 18 | 11 | 116 | 154 |

| Table of Staph by Race | | | |
|---|---|---|---|
| | Race | | |
| Staph | 0 | 1 | Total |
| 0 | 18 | 88 | 106 |
| 1 | 1 | 47 | 48 |
| Total | 19 | 135 | 154 |

# Program L6

- **Regression for survival data**
  - Cox's proportional hazards regression, cont'd
    - Testing the PH assumption
    - Graphical methods to investigate the PH assumption
    - What to do if the proportional hazards assumption is violated
    - Regression diagnostics

UPPSALA UNIVERSITET

# Graphical methods to assess the assumption of proportional hazards

There are a number of graphical methods available to check the proportional hazards assumption.

1) Plots based on the log cumulative baseline hazard
2) Andersen plot, based on the cum. baseline hazard
3) Arjas plot, based on the expected number of events calculated from the estimated cumulative hazard rate and the observed number of events up to time $t$
4) Plots based on score residuals
5) Plots based on scaled Schoenfeld residuals

UPPSALA UNIVERSITET

# Plots of log cumulative baseline hazard rates

$$\left.\begin{aligned} h_1(t) &= c h_2(t) \\ H_1(t) &= c H_2(t) \end{aligned}\right\}$$ Same assumption of proportionality

$$\ln H_1(t) = \ln c + \ln H_2(t)$$

$$\ln H_1(t) - \ln H_2(t) = \ln c$$

A constant

This difference can be plotted to see if it is constant over time.

# Plots of log cumulative baseline hazard rates

$$H(t) = \int_0^t h(u)du$$

$$h_1(t) = ch_2(t)$$

$$\int_0^t h_1(u)du = \int_0^t ch_2(u)du$$

$$\int_0^t h_1(u)du = c\int_0^t h_2(u)du \qquad H_1(t) = cH_2(t)$$

# Plots of log cumulative baseline hazard rates

**Procedure:**

1) The covariate $Z_1$ for which we want to check the PH assumption has to have $K$ possible values.

   Continuous covariates need to be stratified into $K$ disjoint strata to be able to use this approach.

2) Fit a Cox model, stratified on the discrete values of $Z_1$.

3) Estimate the cumulative baseline hazard rate in each of the $K$ strata, $\widehat{H}_{go}(t)$ for $g$ = 1, … $K$.

# Plots of log cumulative baseline hazard rates

4) Plot $ln\widehat{H}_{go}(t) - ln\widehat{H}_{1o}(t)$ vs. $t$ for $g$=2,….$K$.
5) If the proportional hazards assumption holds, each curve should be roughly constant over time.

NOTE! This plot should be interpreted with some care, especially in the right-hand tail of the plot, since the variances of the curves are not constant over time.

# Example: Duration of breastfeeding

Section 1.14 describes a study with the aim to evaluate the relationship between duration of breastfeeding and a number of explanatory variables.

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding

**Variables:**

*time* = duration of breastfeeding (weeks)

*complete* = indicator of completed breastfeeding (1=yes, 0=no)

*poverty* = mother in poverty (1=yes, 0=no)

*smoke* = mother smoked at birth (1=yes, 0=no)

*birthyear* = birthyear of mother

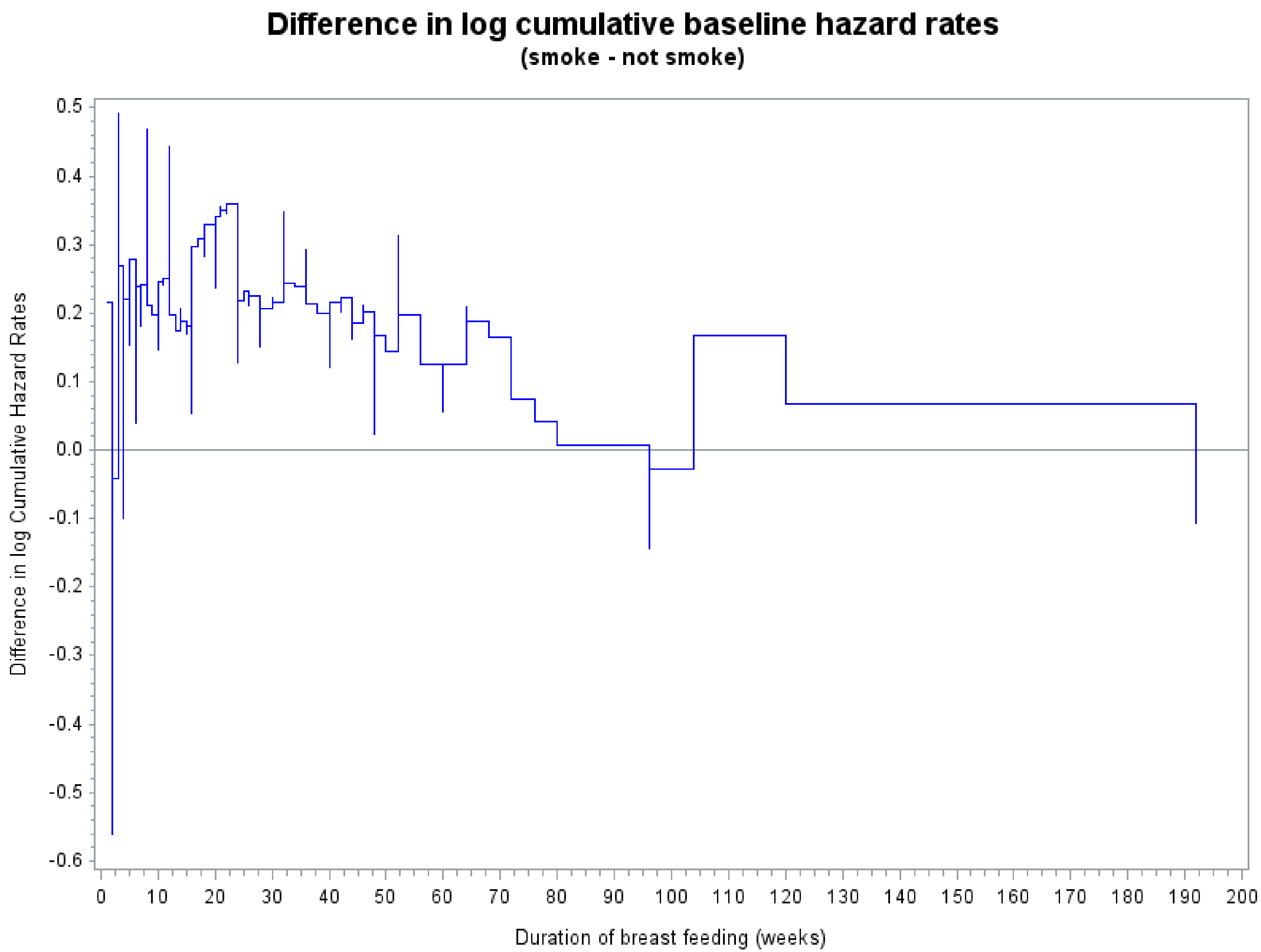
*education* = mother’s education (years of school)

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding

| Analysis of Maximum Likelihood Estimates | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio | 95% Hazard Ratio Profile Likelihood Confidence Limits | |
| poverty | 1 | -0.17178 | 0.09230 | 3.4634 | 0.0627 | 0.842 | 0.701 | 1.007 |
| smoke | 1 | 0.19272 | 0.07588 | 6.4503 | 0.0111 | 1.213 | 1.044 | 1.405 |
| birthyear | 1 | 0.07179 | 0.01786 | 16.1582 | <.0001 | 1.074 | 1.038 | 1.113 |
| education | 1 | -0.07349 | 0.01993 | 13.5932 | 0.0002 | 0.929 | 0.894 | 0.966 |

# Example: Duration of breastfeeding




Roughly constant over time?

Not that easy to interpret.

UPPSALA UNIVERSITET

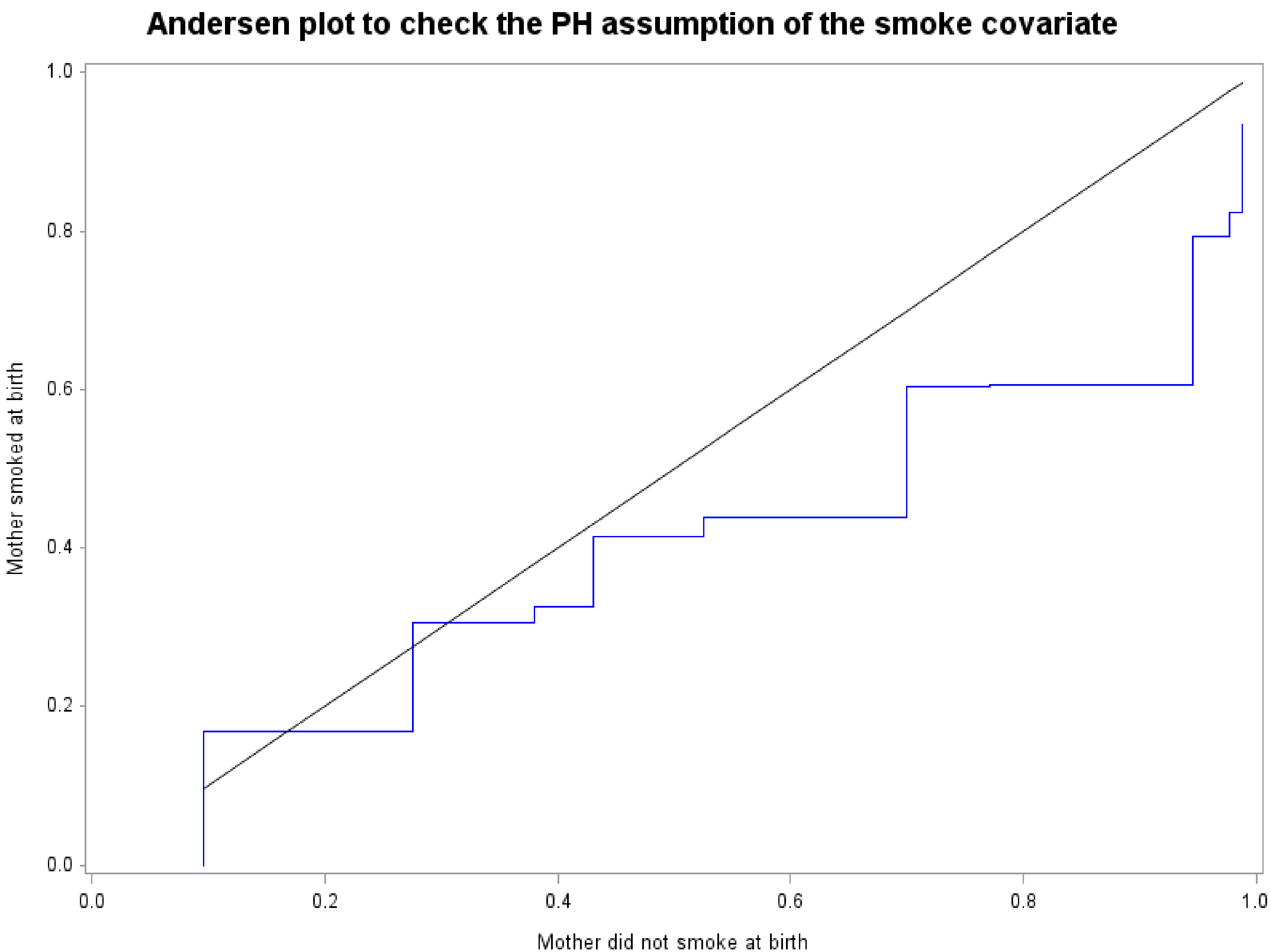
# Andersen plot (based on the cumulative baseline hazard)

**Procedure:**

1) Plot, for all $t$ (or a subset of $t$), $\widehat{H}_{1o}(t)$ versus $\widehat{H}_{go}(t)$ for $g$ = 2, … $K$.
2) If the proportional hazard assumption holds, these curves should be straight lines through the origin.

NOTE! This plot should also be interpreted with some care, especially in the right-hand tail of the plot, again since the variances of the curves are not constant over time.

# Example: Duration of breastfeeding




Roughly a straight line through the origin?

What does "roughly" mean?

UPPSALA UNIVERSITET

# Arjas plot

Arjas plots can be used both to check the proportional hazards assumption, and to check the overall fit of the Cox model.

$Z_1$ = a covariate for which we want to check the PH assumption, or check if the covariate should be included in the model.

# Arjas plot

**Procedure:**

1) Fit the Cox model without the covariate $Z_1$.
2) For each event time and for each level of $Z_1$ calculate the expected number of events and the observed no. of events up to this time.

# Arjas plot

Expected number of events up to time $t_i$:
(referred to as the "total time on test" in the book)

$$E\_TOT_g(t_i) = \sum_{Z_{1j}=g} \widehat{H}(\min(t_i, T_j)|\mathbf{Z}_j^*)$$

Stratum $g$

Sum over all individuals in stratum $g$

Cumulative hazard up to time $t_i$ (or the last time for the individual)

based on a model of all covariates except $Z_1$.

Observed number of events up to time $t_i$:

$$N_g(t_i) = \sum_{Z_{1j}=g} \delta_j (T_j \le t_i)$$

UPPSALA UNIVERSITET

# Arjas plot

**Procedure:**

1) Fit the Cox model without the covariate $Z_1$.

2) For each event time and for each level of $Z_1$ calculate the expected number of events and the observed no. of events up to this time.

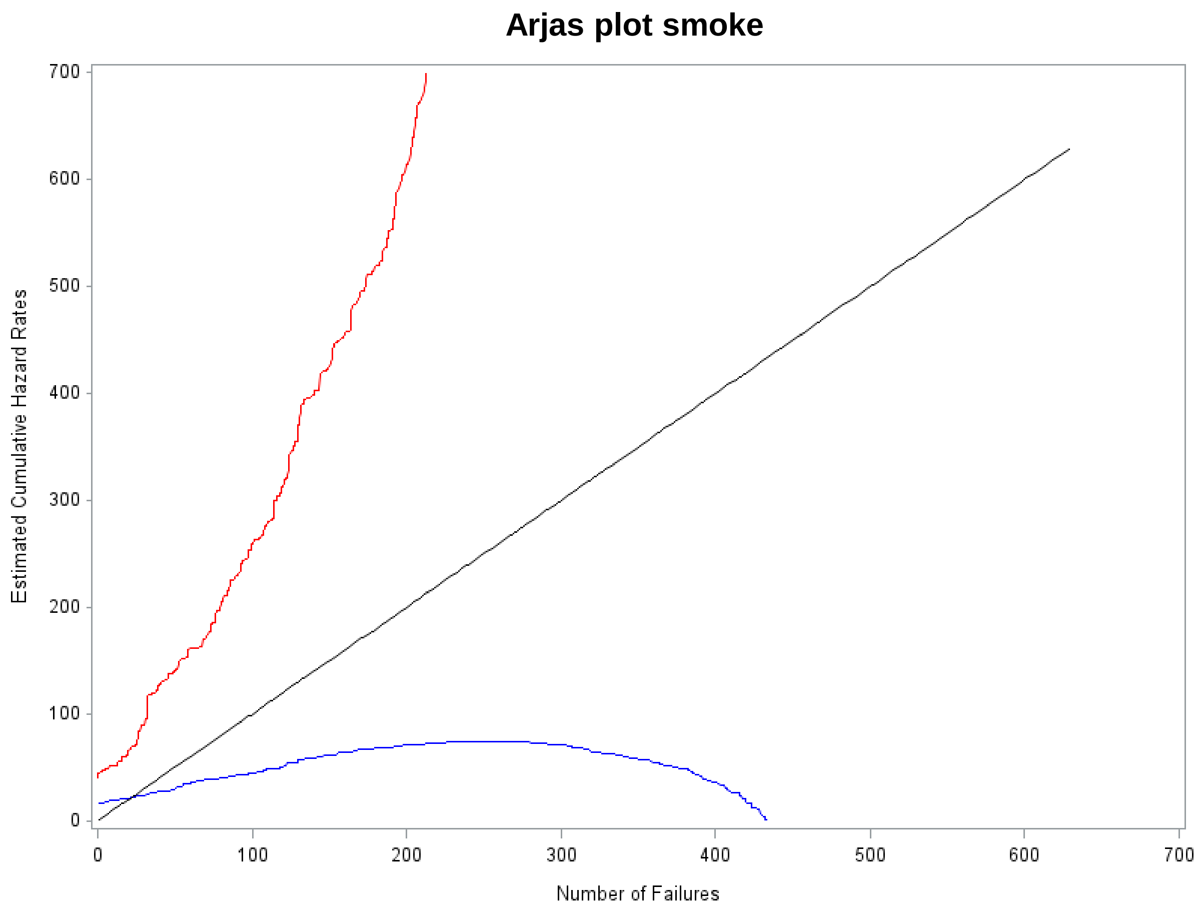
3) If the covariate $Z_1$ has a nonproportional hazards effect on the hazard rate, a plot of $N_g(t_i)$ versus $E\_TOT_g(t_i)$ provides curves that differ nonlinearly from a 45 degree line through the origin.

# Arjas plot

4) If the covariate $Z_1$ is redundant (does not need to be included in the model), the plot of $N_g(t_i)$ versus $E\_TOT_g(t_i)$ provides curves that should roughly follow a 45 degree line through the origin.

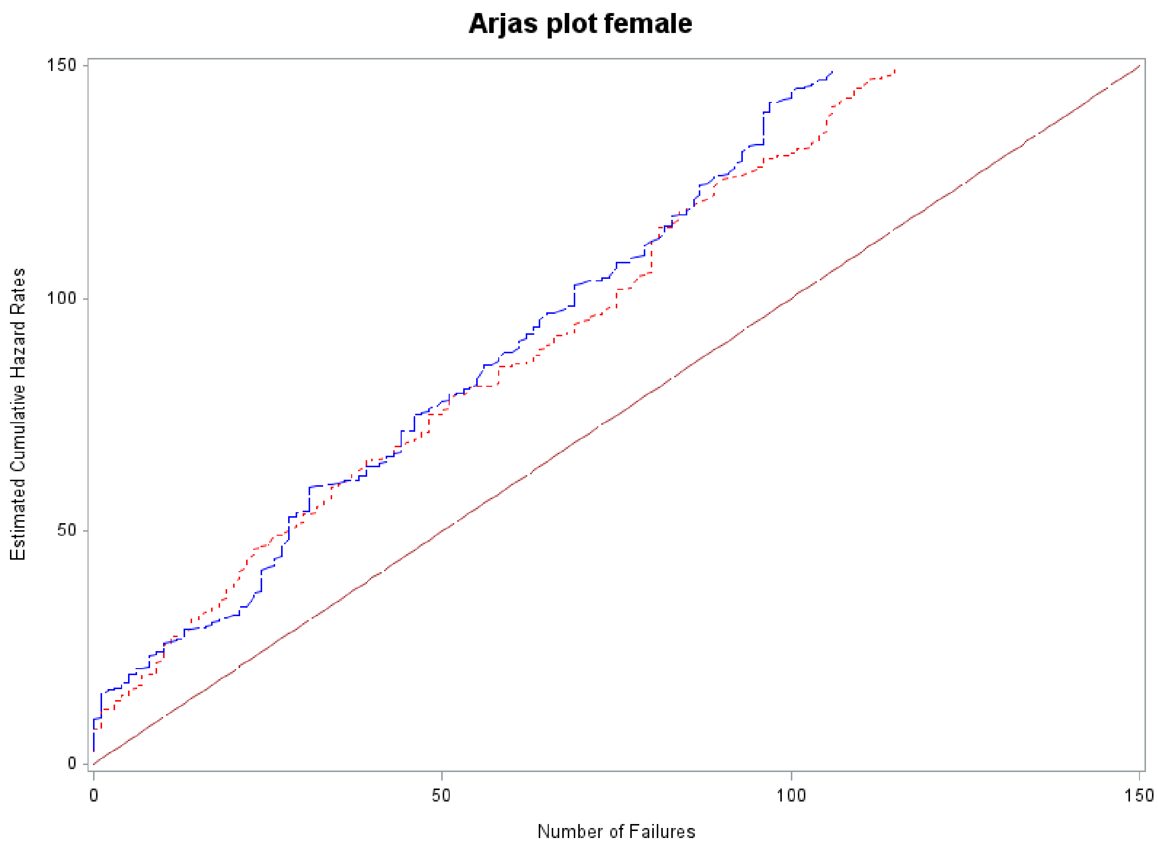
   If the covariate $Z_1$ should be included in the model, the plot provides curves that differ linearly from a 45 degree line through the origin (with slopes different from 1).

# Example: Duration of breastfeeding



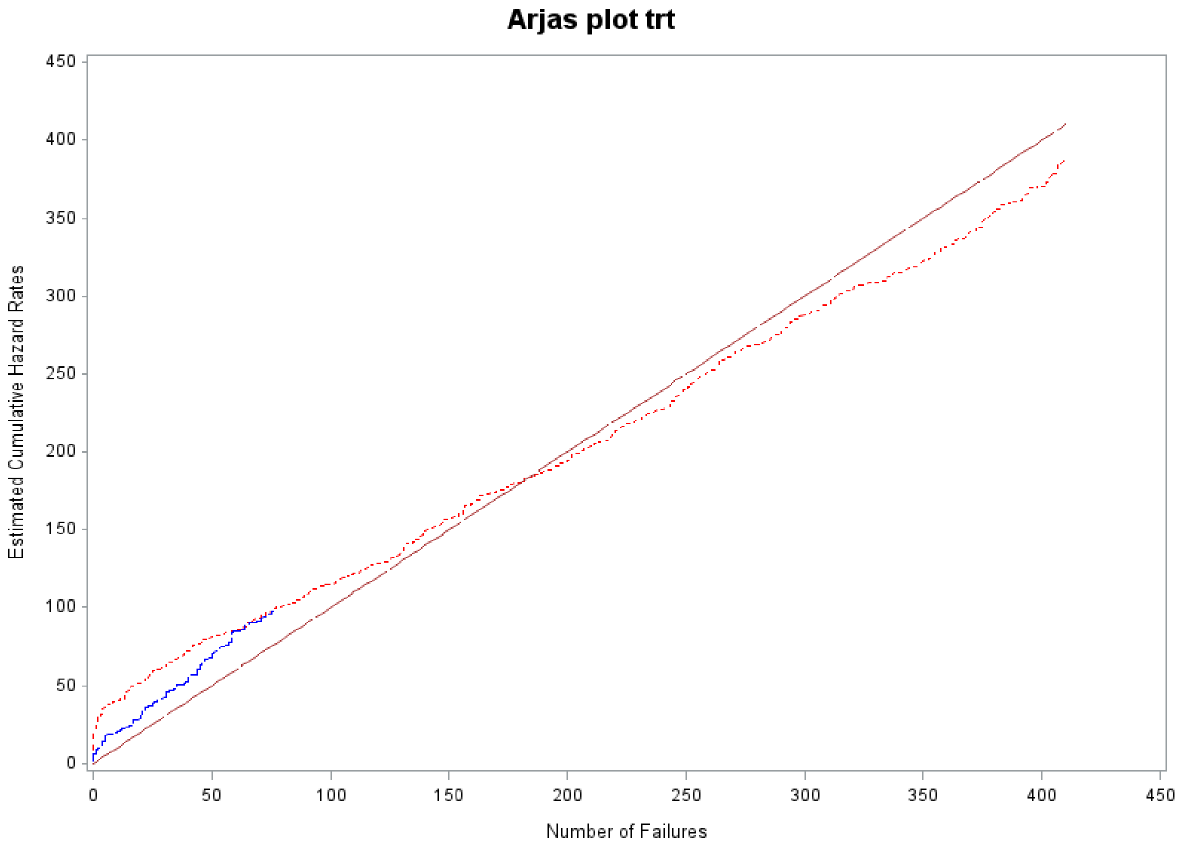

The curves (especially the blue one) differ non-linearly from the line, indicating nonproportional hazards

# Example: Exam 2019-01-18 Task 3



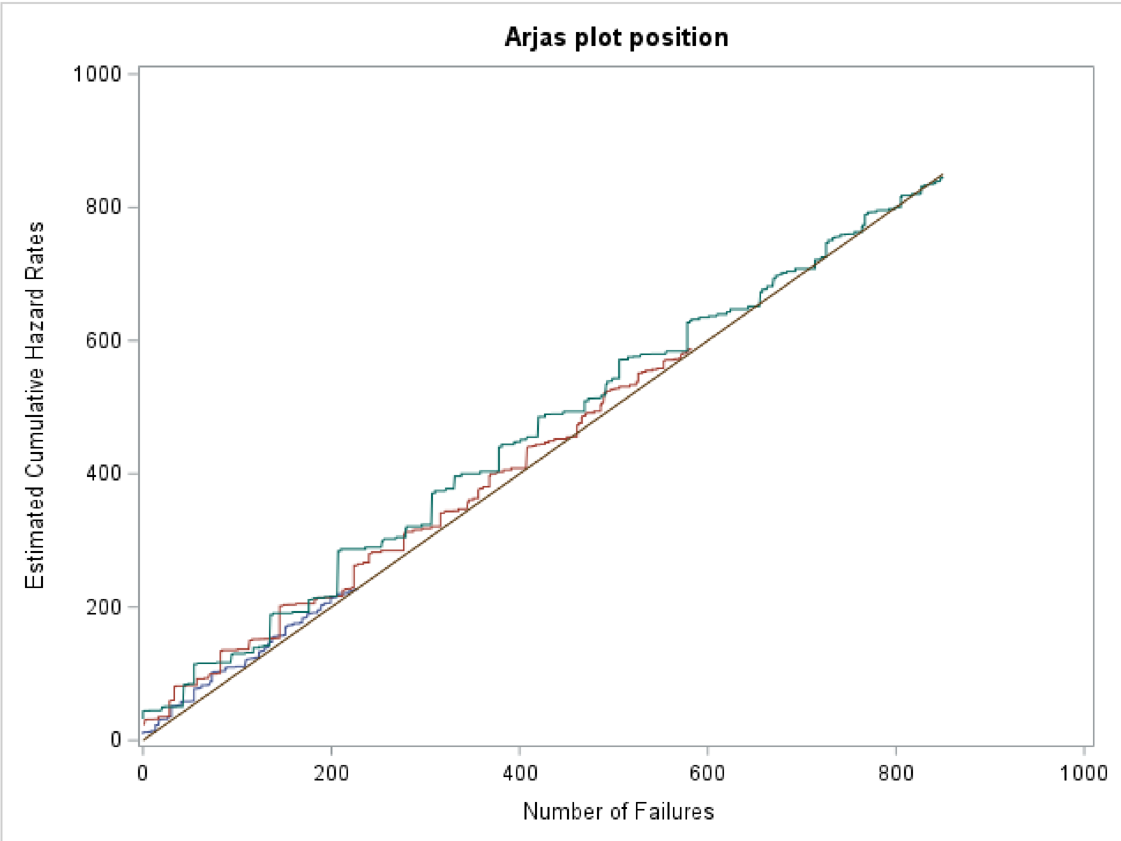

The curves differ linearly from the line, not contradicting proportional hazards

# Example: Exam 2024-01-11 Task 3




The red curve differs non-linearly from the line, indicating nonproportional hazards

# Example: Exam 2025-01-17 Task 3C




The curves are quite linear, not contradicting proportional hazards

But they roughly follow the line, which means that this covariate doesn't contribute much to the model.

# Arjas plot

One advantage of the Arjas plot is that you can change the way the investigated covariate is stratified without refitting the basic Cox model.

One disadvantage is that continuous covariates cannot be investigated as they are but have to be categorized.

# Using Score residuals for graphical checks of the PH assumption

The score residuals are the first partial derivative with respect to $\beta_k$ of the contribution to the log likelihood for the fitted Cox model, using only information of an individual $j$ accumulated up to time $t$.

Score residuals are useful in assessing the influence of each subject on individual parameter estimates.

Interpreted as a weighted difference between the value of a given covariate for a given subject and the average value of this covariate in a risk set.

UPPSALA UNIVERSITET

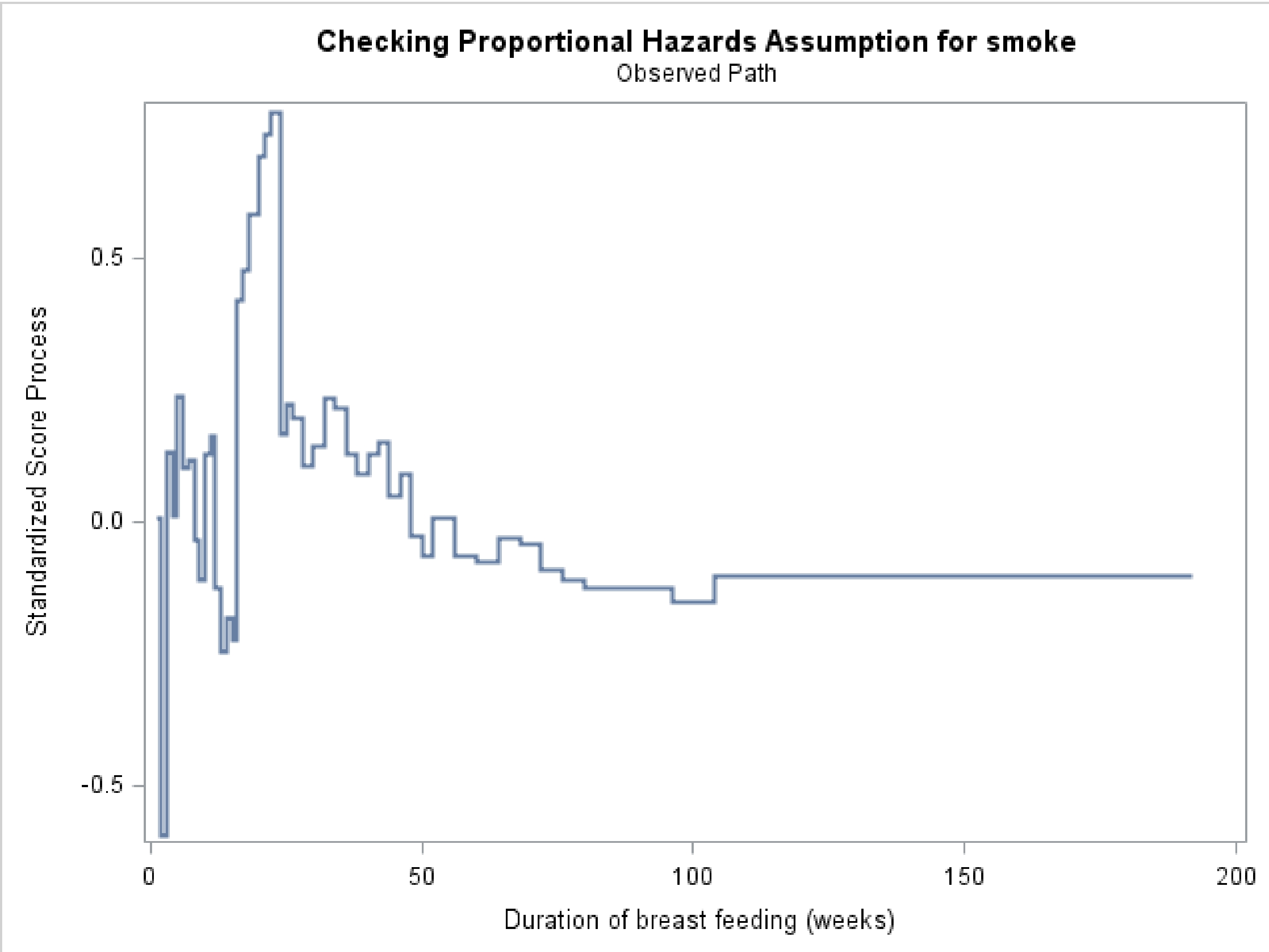
# Using Score residuals for graphical checks of the PH assumption

Standardized score residuals can be plotted against time to check for nonproportional hazards.

If the model fits properly, the standardized residuals should behave like a Brownian Bridge. If the residuals get unusually large at any time point, we don't have PH.

One advantage of the score process is that a continuous covariate doesn't have to be discretized to make the plot.

# Example: Duration of breastfeeding




No evidence of nonproportional hazards here

Not that easy to interpret

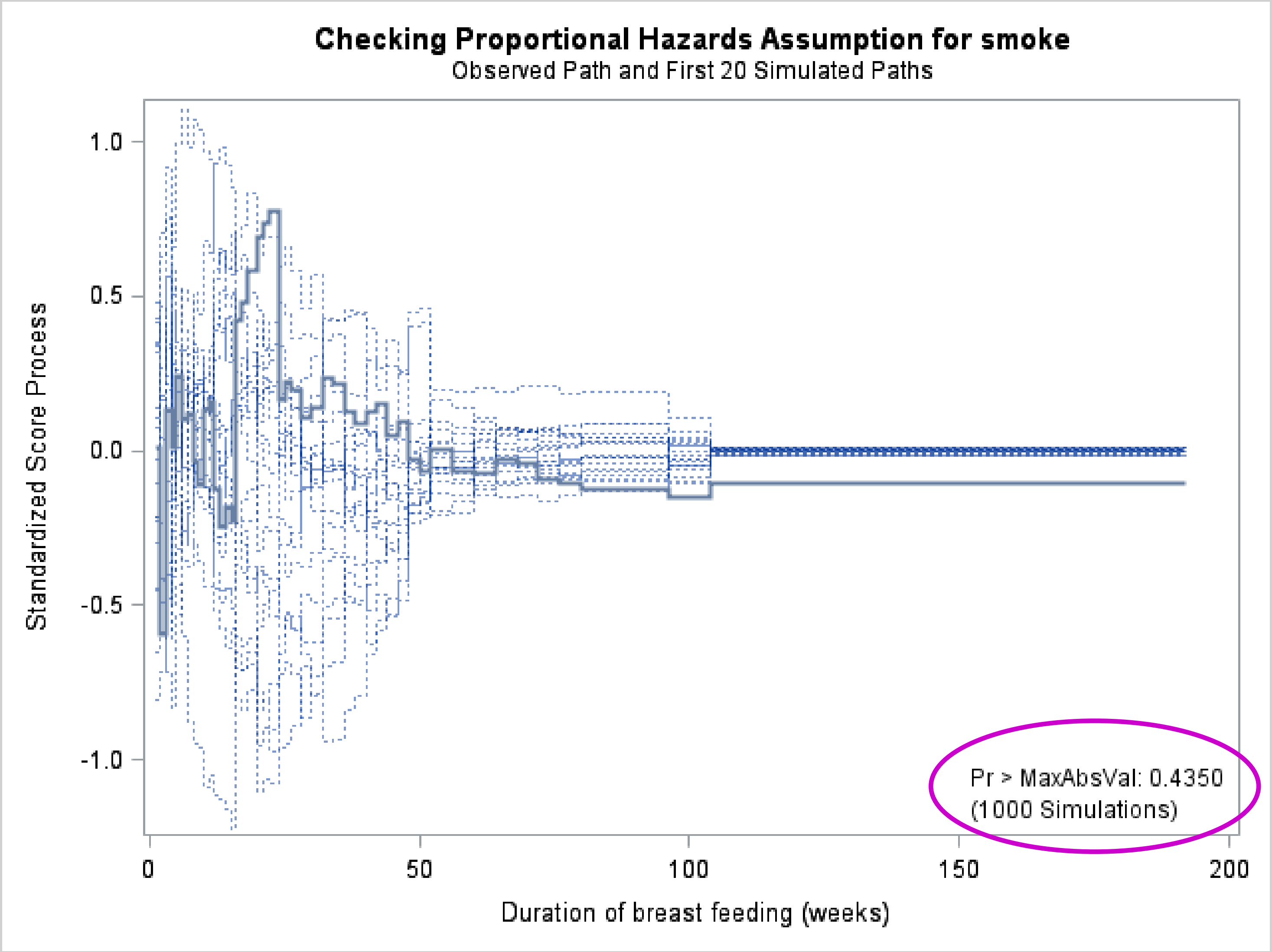
# Using Score residuals for graphical checks of the PH assumption

The standardized score residuals can be used to provide a p-value of testing the PH assumption, based on simulations*.

This is included in proc phreg, option **assess ph**.

*Lin, Wei, Ying (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika*, 80, 3, pages 557-572

# Example: Duration of breastfeeding



# Score residuals plot

One advantage of the score residuals plot is that continuous covariates can be investigated as they are (no need for categorization).

One disadvantage is that this method is computationally demanding and often cannot be used with the free online software SAS OnDemand for Academics.

# Schoenfeld's partial residuals

Schoenfeld's partial residuals are a part of the score residuals and defined as

$$s_{jk} = \delta_j \{ Z_{jk} - \bar{Z}_k(T_j) \}$$

Defined for events only

Covariate value for individual *j*

Expected Covariate value for individual *j*

This describes the difference between a covariate value and the expected value of this covariate at this time.

# Scaled Schoenfeld residuals

Grambsch and Therneau (1994) showed that a scaled version of the Schoenfeld residual at time *T* for a particular covariate will approximate the change in the regression coefficient at this time point:

$$E(s_{jk}^{*}) \approx \hat{\beta}_k + \beta_j(T_j)$$

Scaled Schoenfeld residual for individual *j*, covariate *k.*

Time-invariant coefficient

Time-dependent coefficient

$s_{jk}^{*} = s_{jk}/\hat{V}(\hat{\beta}_k)$

Grambsch and Therneau (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, Vol. 81, No. 3, 1994, pp. 515-526

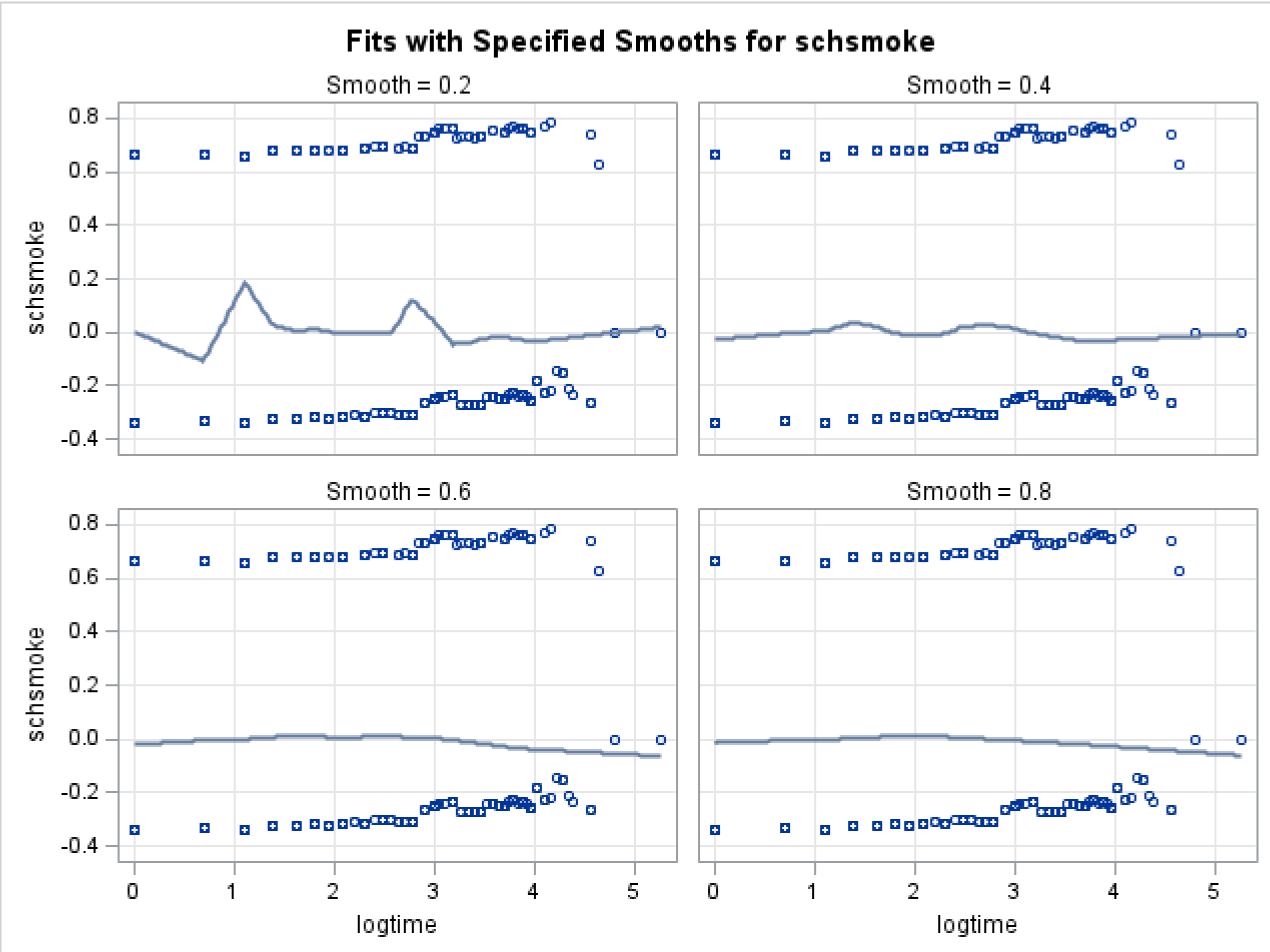
UPPSALA UNIVERSITET

# Plots based on scaled Schoenfeld residuals

Smoothed scatterplots are used to explore the scaled Schoenfeld residuals’ relationship with time (preferably log(time) or rank(time) to find non-linear relationships)

If the average of the scaled Schoenfeld residuals is zero across time (flat smoothed lines at 0), this suggests that the coefficient does not vary over time (the assumption of proportional hazards holds for this covariate).

# Example: Duration of breastfeeding




No evidence of nonproportional hazards here

UPPSALA UNIVERSITET

# Choice of method to check the assumption of proportional hazards

**Tests:**

A number of statistical tests, including the Cox model with time-dependent covariates (with different functions of time) have been evaluated by Persson and Khamis (2008).

The standard test, including a time-dependent covariate in the Cox model ($g(t) = \ln t$), is shown to work well in most situations (not good power if <~50 events/group).

Inger Persson, Harry Khamis (2008). A Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption in the Cox Model. *Journal of Statistics and Applications*, Vol. 3, No. 1-2, 2008, pages 135-154

# Choice of method to check the assumption of proportional hazards

**Graphical methods:**
A number of graphical methods have been evaluated by Persson and Khamis (2007).

The Arjas plot is shown to work well in most situations, and it is the easiest graphical method to evaluate visually.

Also, the method based on standardized score residuals (phreg option **assess ph**) seems to have good properties (not included in the study by Persson and Khamis).

Inger Persson, Harry Khamis (2007). A comparison of graphical methods for assessing the proportional hazards assumption in the Cox model. *Journal of Statistics and Applications*, Vol 2, No. 1-2, 2007, 1-32.

# Choice of method to check the assumption of proportional hazards

**Recommendations:**

Combine test+graphical method.

If any of the chosen methods finds nonproportionality, reject the null hypothesis of proportional hazards (the power of any of these methods is not very high).

# Example: Duration of breastfeeding Testing the PH assumption

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| smoke | 1 | 0.25264 | 0.15330 | 2.7160 | 0.0993 | 1.287 |
| IntSmoke | 1 | -0.01172 | 0.06365 | 0.0339 | 0.8538 | 0.988 |

OK for smoke

Time-dependent covariate test and scaled Schoenfeld residual plot show no deviation from proportionality. But the Arjas plot suggests non-proportionality, which means that the assumption of proportional hazards is violated.

UPPSALA UNIVERSITET

# Program L6

- **Regression for survival data**
  - Cox's proportional hazards regression, cont'd
    - Testing the PH assumption
    - Graphical methods to investigate the PH assumption
    - What to do if the proportional hazards assumption is violated
    - Regression diagnostics

UPPSALA UNIVERSITET

# If the PH assumption doesn’t hold

In many instances the model diagnostics reveal only a small to moderate deviation from the proportional hazards assumption.

In such cases the hazard ratio estimated from the Cox model is quite close to a geometric average of the hazard ratio*. This does not exactly describe the truth, since the hazard rate changes over time when the hazards are not proportional. It can however provide an approximate estimate.

*Inger Persson, Harry Khamis (2005). Bias of the Cox Model Hazard Ratio. *Journal of Modern Applied Statistical Methods*, May, 2005, Vol. 4, No. 1, 90-99

# “Rescuing” a violated proportional hazards assumption

Two possible solutions:

1) Stratify the model on the covariate that violates the PH assumption (might lead to the hazards being proportional within each stratum, for the other covariates).

2) Include the time-dependent covariate(s) in the model (may be harder to interpret, especially for continuous covariates)

# Stratified proportional hazards models

With a stratified proportional hazards model, the regression coefficients are assumed to be the same in each stratum, but the baseline hazard function might differ between strata.

$$\underbrace{h_j}_{\text{Stratum } j}(t \mid \mathbf{Z}) = h_{0j}(t)\exp(\boldsymbol{\beta}^t \mathbf{Z})$$

# Stratified proportional hazards models

**Assumption of stratified models**: covariates are acting similarly on the baseline hazard function in each stratum.

This can be tested, e.g. using the Likelihood ratio test.

$$H_0 : \text{All } \beta\text{'s are the same for all } s \text{ strata}$$

$$H_a : \text{At least one of the } \beta\text{'}s \text{ is/are different}$$

# Likelihood ratio test for the assumption of stratified models

1) Fit the stratified model (with common $\beta$'s in each stratum) and obtain the log partial likelihood $LL(\mathbf{b})$.

2) Use only data from the $j$th stratum, and fit a Cox model to estimate $\boldsymbol{\beta}_j$ and corresponding $LL_j(\mathbf{b}_j)$. Do this for all strata.

3) Calculate the log likelihood under the model with distinct covariates for each of the $s$ strata:

$$\sum_{j=1}^{s} LL_j(\mathbf{b}_j)$$

# Likelihood ratio test for the assumption of stratified models

4) Calculate the likelihood ratio chi square test for $H_0$: All β's are the same in the different strata

$$-2\left[LL(\mathbf{b}) - \sum_{j=1}^{s} LL_j(\mathbf{b}_j)\right] \sim \chi^2_{(s-1)p}$$

$s$ = no. of strata

$p$ = no. of covariates

$$\text{Equivalently:} \left[-2LL(\mathbf{b}) - \sum_{j=1}^{s} -2LL_j(\mathbf{b}_j)\right]$$

# Example: Duration of breastfeeding

The **strata** statement is used to estimate a stratified proportional hazards model

```
/* Stratify on smoke */
proc phreg data=bf;
  model time*complete(0)=poverty education birthyear
   /ties=exact;
  strata smoke;
run;
```

# Example: Duration of breastfeeding

Stratified by smoke (same parameter estimates, different baseline hazards)

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Without Covariates | With Covariates |
| -2 LOG L | 5320.374 | 5296.306 |
| AIC | 5320.374 | 5302.306 |
| SBC | 5320.374 | 5316.686 |

-2LogL = 5296.306

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| poverty | 1 | -0.16198 | 0.09260 | 3.0603 | 0.0802 | 0.850 |
| education | 1 | -0.07309 | 0.01997 | 13.3963 | 0.0003 | 0.930 |
| birthyear | 1 | 0.07257 | 0.01794 | 16.3595 | <.0001 | 1.075 |

# Example: Duration of breastfeeding

The **by** statement is used to estimate separate models for each stratum

```
/* estimate model for each stratum separately */
proc phreg data=bf;
  model time*complete(0)= poverty education birthyear
   /ties=exact;
  by smoke;
run;
```

# Example: Duration of breastfeeding

Smoked at birth

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Without Covariates | With Covariates |
| -2 LOG L | 1498.413 | 1495.547 |
| AIC | 1498.413 | 1501.547 |
| SBC | 1498.413 | 1512.264 |

Did not smoke

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Without Covariates | With Covariates |
| -2 LOG L | 3821.962 | 3797.824 |
| AIC | 3821.962 | 3803.824 |
| SBC | 3821.962 | 3817.157 |

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding

$$\sum_{j=1}^{2} -2LL_j(\mathbf{b}_j) = 1495.547 + 3797.824$$

$$= 5293.371$$

$$-2\left[LL(\mathbf{b}) - \sum_{j=1}^{2} LL_j(\mathbf{b}_j)\right] = 5296.306 - 5293.371]$$

$$= 2.9 \sim \chi^2_{(2-1)3=3\,df}$$

# Example: Duration of breastfeeding

The corresponding *p*-value is 0.407 (table C.2 only shows that it is larger than 0.1)

This means that the effects of the covariates are not significantly different for all strata.

A stratified model is thus appropriate in this case.

UPPSALA UNIVERSITET

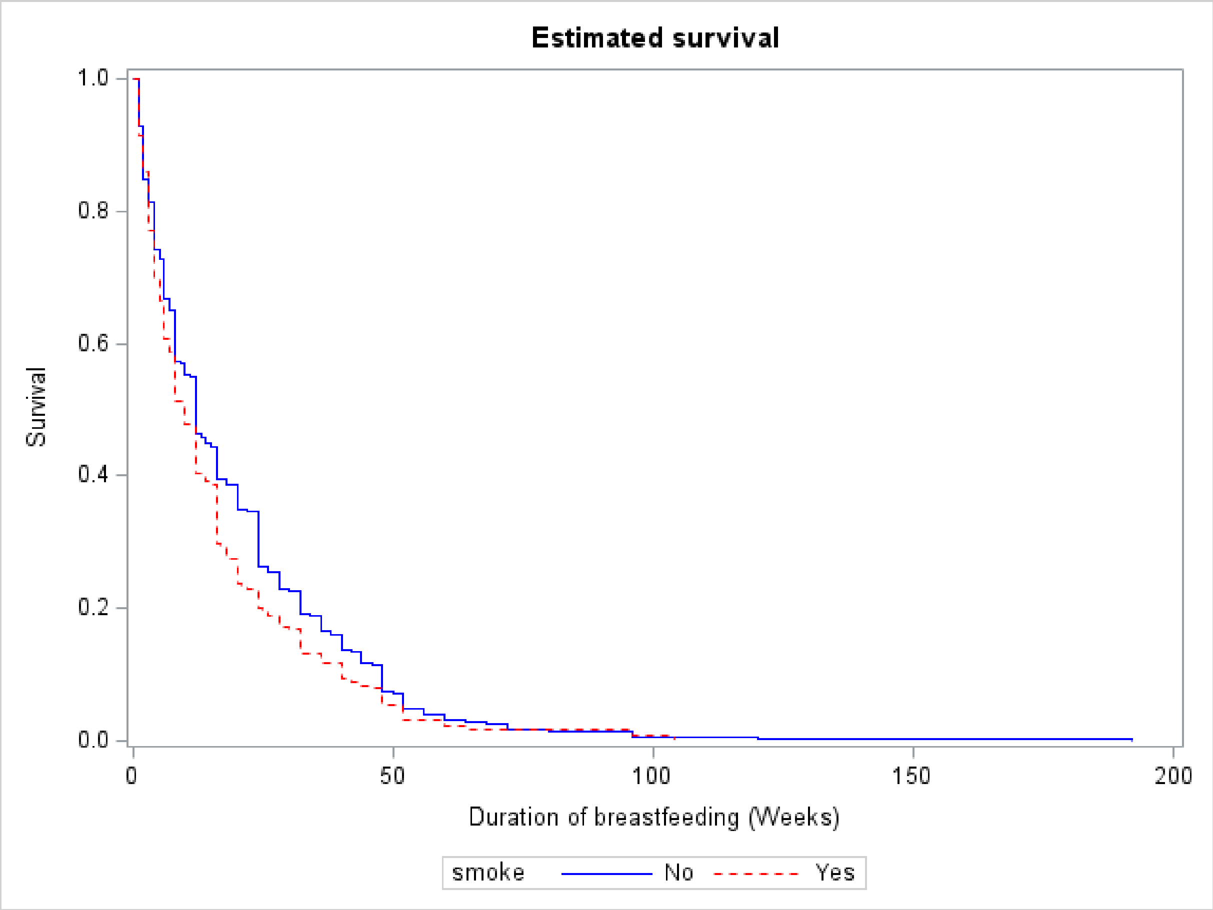
# Example: Duration of breastfeeding

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| poverty | 1 | -0.16198 | 0.09260 | 3.0603 | 0.0802 | 0.850 |
| education | 1 | -0.07309 | 0.01997 | 13.3963 | 0.0003 | 0.930 |
| birthyear | 1 | 0.07257 | 0.01794 | 16.3595 | <.0001 | 1.075 |

The stratified model doesn't contain any information on the stratification covariate.

Information on this covariate can be provided by plots of the estimated survival for the different strata, for some values of the other covariates.

# Example: Duration of breastfeeding




Survivor functions can be estimated by the Cox model for different values of the stratification variable, for selected values of the other covariates (most common values chosen here)

# “Rescuing” a violated proportional hazards assumption

Two possible solutions:

1) Stratify the model on the covariate that violates the PH assumption (might lead to the hazards being proportional within each stratum, for the other covariates).

2) Include the time-dependent covariate(s) in the model (may be harder to interpret, especially for continuous covariates)

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
| poverty | 1 | -0.13921 | 0.09225 | 2.2771 | 0.1313 | 0.870 |
| education | 1 | -0.17740 | 0.03843 | 21.3128 | <.0001 | 0.837 |
| IntEducation | 1 | 0.04855 | 0.01536 | 9.9834 | 0.0016 | 1.050 |
| birthyear | 1 | 0.07415 | 0.01794 | 17.0928 | <.0001 | 1.077 |

The hazard ratio for education is not constant over time, needs to be interpreted separately for different time points.

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding

HR for education:

$$\exp\{b_1 \cdot education + b_2 \cdot \ln(t) \cdot education\}$$

$$= \exp\{(b_1 + b_2 \cdot \ln(t))education\}$$

HR for a 1 unit's increase in education:

$$\exp\{b_1 + b_2 \cdot \ln(t)\}$$

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding

HR for education at 3 months (13 weeks):

$$\exp\left\{-0.17740 + 0.04855 \cdot \ln(13)\right\}$$

$$= 0.949$$

At 6 months (26 weeks) the HR = 0.981

At 1 year (52 weeks) the HR = 1.01

UPPSALA UNIVERSITET

# Program L6

- **Regression for survival data**
  - Cox's proportional hazards regression, cont'd
    - Testing the PH assumption
    - Graphical methods to investigate the PH assumption
    - What to do if the proportional hazards assumption is violated
    - Regression diagnostics

# Regression diagnostics

Four aspects of the Cox proportional hazards regression model are of interest to examine:

1) Which is the best functional form of a covariate to explain the effect on survival, adjusting for other covariates ($Z$, ln $Z$, etc)?
2) Does the assumption of proportional hazards hold?
3) What is the accuracy of the model for predicting the survival of a given subject?
4) What influence does each individual have on the model fit?

# Residuals and Cox regression

| | Observed value, $y$ | Estimated value, $\hat{y}$ | Residual $y - \hat{y}$ |
|---|---|---|---|
| **Linear regression** | 132 | 128 | 4 |
| **Cox regression** | $t$ | $h(t)$ | ? |

In Cox regression there is no residual in the model (as there is for linear regression)

# Residuals and Cox regression

There are a number of different residuals for the Cox model:

- **Cox-Snell residuals**, used to check the overall fit of a Cox model
- **Martingale residuals**, used to determine the functional form of a covariate to be included in a Cox model
- **Score residuals**, used for graphical checks of the proportional hazards assumption
- **Deviance residuals**, used to examine model accuracy for each individual
- **Schoenfeld residuals**, used to determine the influence of individual observations

UPPSALA UNIVERSITET

# Cox-Snell residuals

True hazard for an individual $j$:

$$H_j(t) = -\ln S_j(t) = \underbrace{H_0(t)}_{\text{Baseline cumulative hazard}} \underbrace{\exp(\boldsymbol{\beta}^t \mathbf{Z}_j)}_{\text{adjusted for values of } \mathbf{Z}}$$

If the model is correct, the estimated cumulative hazard for each individual at the time of their event or censoring should be like a censored sample from a unit exponential distribution*.

One cumulative hazard can be estimated for each individual $j$, and this is used as a residual.

*Cox, D. R., and E. J. Snell (1968). A General Definition of Residuals. Journal of the Royal Statistical Society. Series B, Vol. 30, No. 2, pp. 248-275*

# Cox-Snell residuals

**Cox-Snell residual:**

$$r_j = \hat{H}_0(t)\exp(\mathbf{b}^t \mathbf{Z}_j) \quad = -\ln(\underset{\substack{\uparrow \\ \text{estimated by} \\ \text{the Cox model}}}{\hat{S}_j(t)})$$

This can be thought of as the expected number of events for individual $j$.

If the model is correct, and the estimated $\boldsymbol{\beta}$'s are close to the true $\boldsymbol{\beta}$'s, then the $r_j$'s should behave as they come from a unit exponential distribution.

# Using the Cox-Snell residuals to check the overall fit of the model

**Procedure:**

1) Calculate the Nelson-Aalen estimator of the cumulative hazard *of the residuals.*
2) Plot the estimated cumulative hazard vs. $r_j$
3) If the model fits, this plot should be a straight line through the origin with a 45 degree slope.

UPPSALA UNIVERSITET

# Using the Cox-Snell residuals to check the overall fit of the model

**Drawbacks/warnings:**

- If this plot is not linear, this method does not indicate the type of departure from the model
- This method is not appropriate for small samples
- The exponential distribution for the residuals holds only for the true parameter values. When using estimates (as we always do), departures from the exponential distribution may partly be due to the uncertainty in estimating $\beta$ and H. This uncertainty is largest in the right of the plot and for small samples.

# Example: Duration of breastfeeding

```
proc phreg data = bf noprint;
  model time*complete(0)= poverty smoke birthyear/ties=exact;
  output out = coxsnell LOGSURV = h;
run;
```

-logsurv gives the
Cox-Snell residuals

```
data coxsnell; set coxsnell;
  r=-h;
run;
```

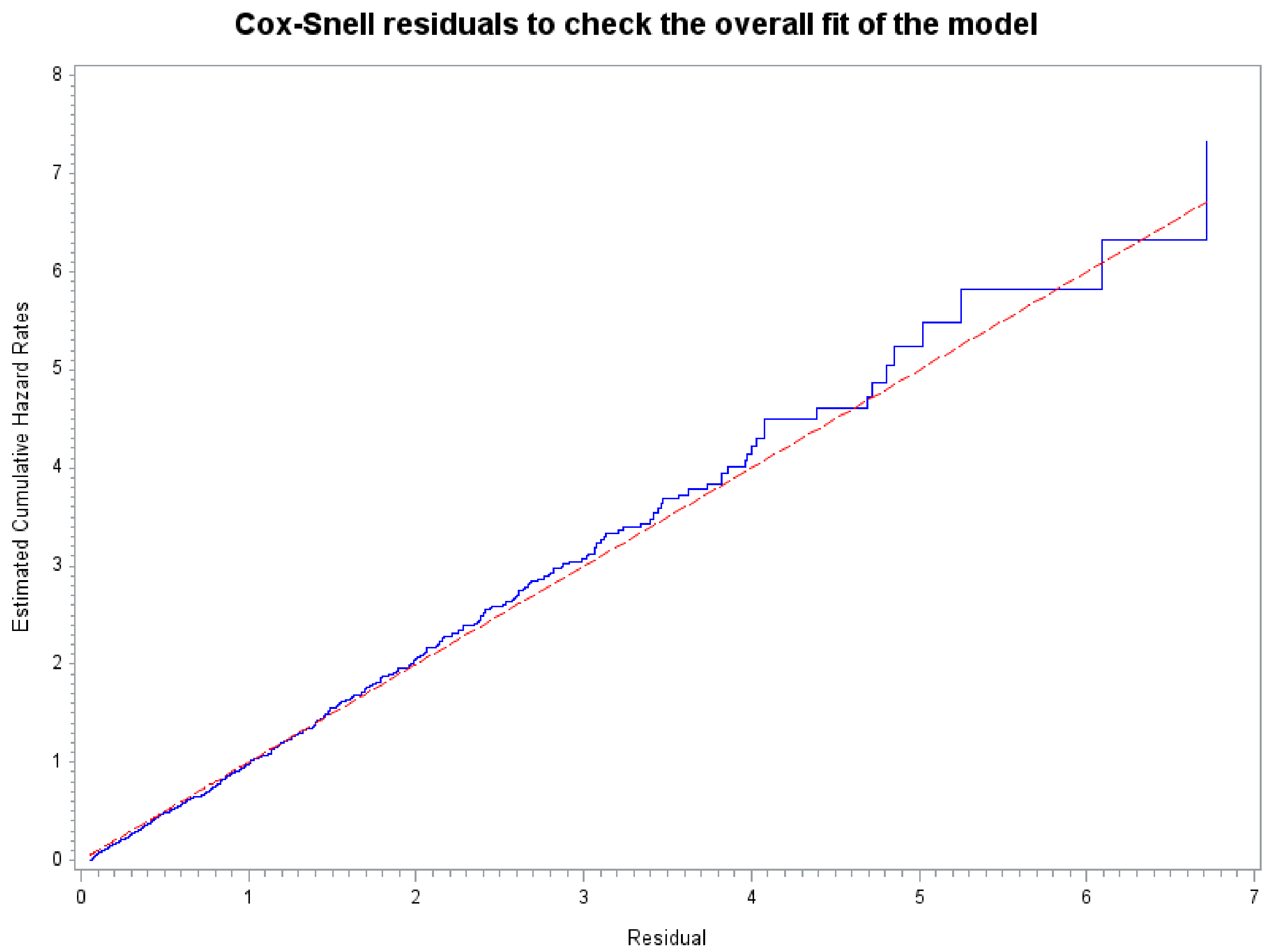
# Example: Duration of breastfeeding

```
ods output ProductLimitEstimates=figure;
proc lifetest data=coxsnell nelson plots=none;
 time r*complete(0);
run;
```

Nelson-Aalen estimates

Based on the Cox-Snell residuals

# Example: Duration of breastfeeding




The model fits the data pretty well

UPPSALA UNIVERSITET

# Generalized $R^2$ (Cox and Snell 1989)

$$R^2 = 1 - e^{-(LRT/n)}$$

where LRT= Likelihood ratio test statistic:

$$LRT = \text{-2logL(0)} - [\text{-2logL(p)}]$$

Likelihood of model without covariates

Likelihood of the fitted model with p covariates

Cox, D. R., and E. J. Snell, 1989. *The Analysis of Binary Data, 2nd ed.* London: Chapman and Hall.

# Generalized $R^2$ (Cox and Snell 1989)

The generalized $R^2$ is a statistic between 0 and 1 that is larger when the covariates are more strongly associated with the dependent variable.

NOT "proportion of variation explained by the model“.

Sensitive to the degree of censoring.

# Example: Duration of breastfeeding

| Model Fit Statistics | | |
|---|---|---|
| Criterion | Without Covariates | With Covariates |
| -2 LOG L | 5485.411 | 5451.970 |
| AIC | 5485.411 | 5459.970 |
| SBC | 5485.411 | 5479.143 |

| Summary of the Number of Event and Censored Values | | | |
|---|---|---|---|
| Total | Event | Censored | Percent Censored |
| 927 | 892 | 35 | 3.78 |

$$LRT = 5485.411 - 5451.970 = 33.441$$

$$R^2 = 1 - e^{-(33.441/927)} = 0.037$$

There is a week association between the covariates and the duration of breastfeeding

# Residuals and Cox regression

There are a number of different residuals for the Cox model:

- **Cox-Snell residuals**, used to check the overall fit of a Cox model
- **Martingale residuals**, used to determine the functional form of a covariate to be included in a Cox model
- **Score residuals**, used for graphical checks of the proportional hazards assumption
- **Deviance residuals**, used to examine model accuracy for each individual
- **Schoenfeld residuals**, used to determine the influence of individual observations

# Martingale residuals

**Martingale residual** for fixed Z (not time-dependent):

$$M_j = \delta_j - r_j$$

Event indicator for individual *j* (0 or 1)

Cox-Snell residual for individual *j* (estimated cumulative hazard)

This can be thought of as the expected number of events for individual *j*.

Can be interpreted as the difference between observed and expected number of events over time, i.e. an estimate of the excess number of events seen in the data but not caught by the model.

# Using Martingale residuals to determine the functional form of a covariate

$Z_1$ = a covariate for which we are uncertain of what functional form of the covariate we should use.

$f(Z_1)$ = best function of $Z_1$ to explain its effect on the time to event.

UPPSALA UNIVERSITET

# Using Martingale residuals to determine the functional form of a covariate

**Procedure:**

1) Fit a Cox model excluding the covariate which we want to investigate ($Z_1$), and calculate the Martingale residuals.
2) Plot the residuals vs. the value of $Z_1$ using a lowess smoothing function (lowess=LOcally Weighted Scatterplot Smooting).
3) If the plot is linear, no transformation of $Z_1$ is needed. If there is a threshold, a discretized version of $Z_1$ is indicated.

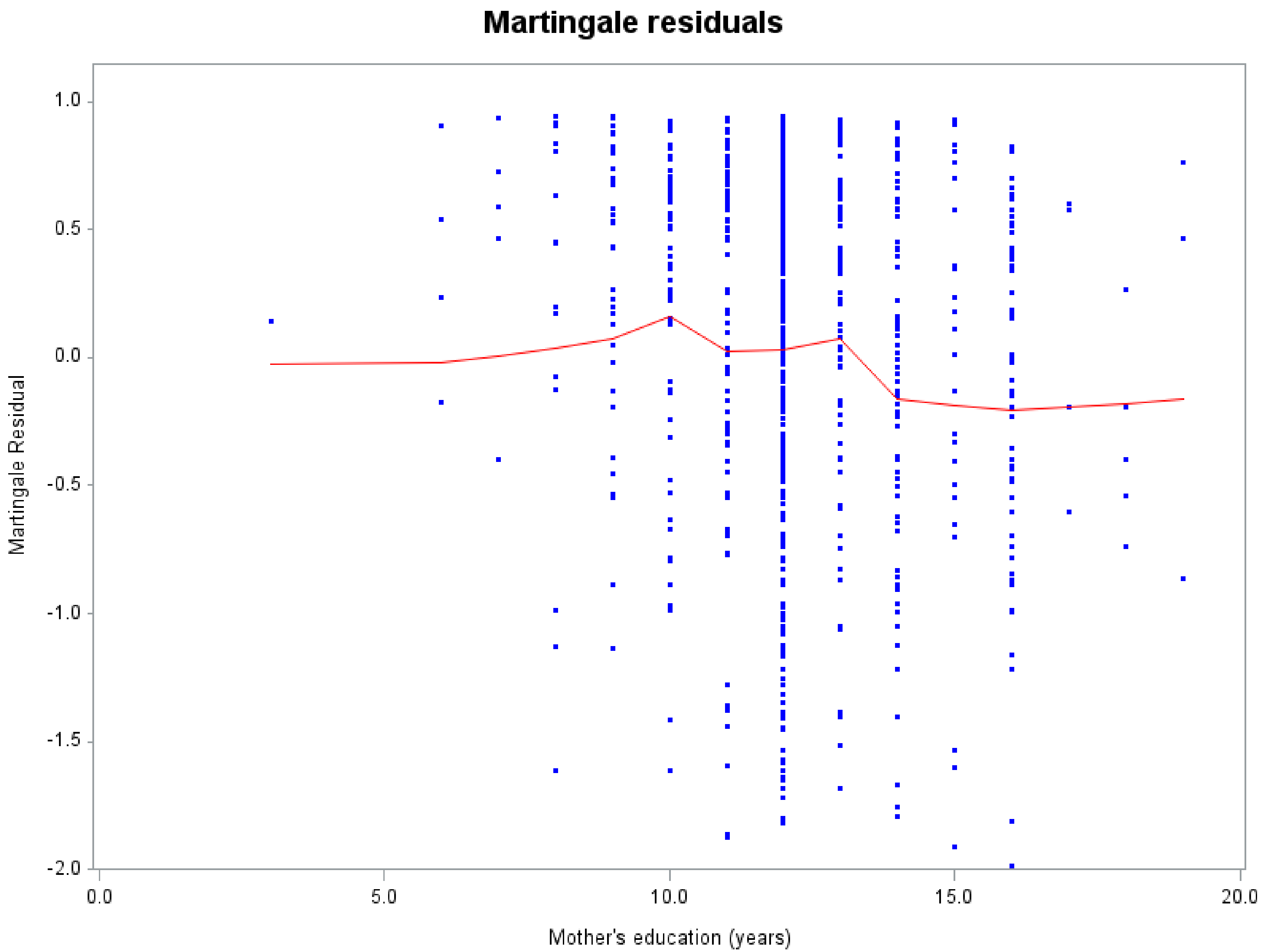
# Example: Duration of breastfeeding

```
proc phreg data = bf noprint;
  model time*complete(0)= smoke poverty birthyear/ties=exact;
  output out = martingale resmart=mgale;
run;
```

Martingale residuals

UPPSALA UNIVERSITET

# Example: Duration of breastfeeding




Roughly linear, no obvious threshold.

# Example of non-linear relationship

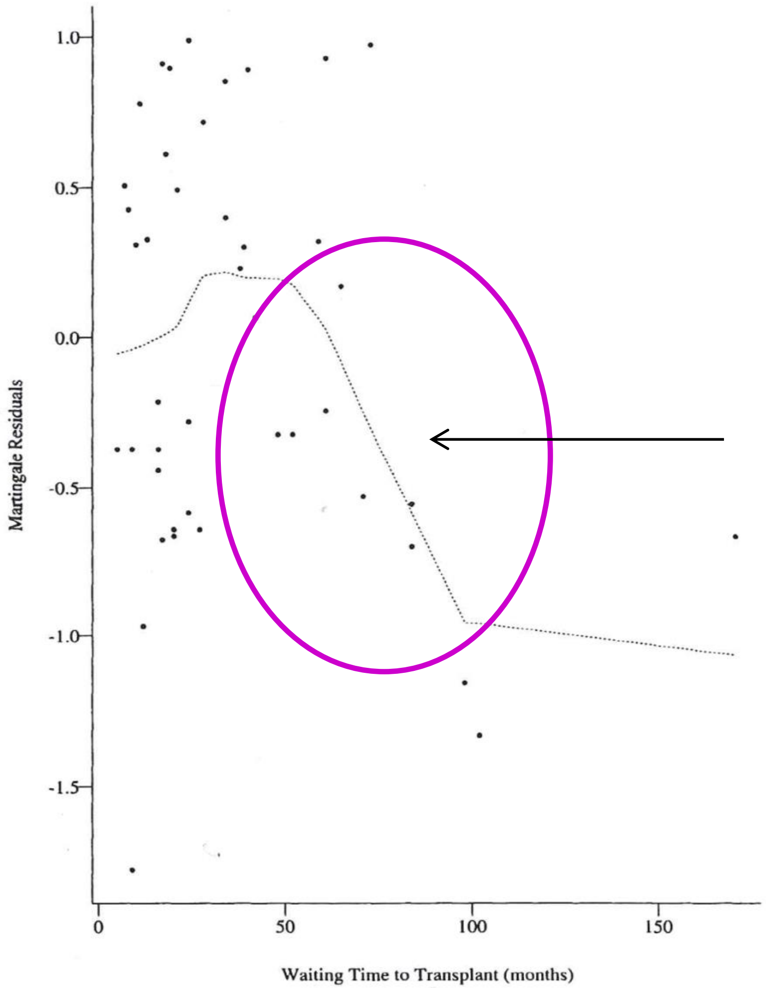



**Figure 11.4** *Plot of martingale residual verus waiting time to transplant and LOWESS smooth*

Not linear.

A drop, suggesting that this covariate could be categorized to better explain the relationship with the time variable

# Residuals and Cox regression

There are a number of different residuals for the Cox model:

- **Cox-Snell residuals**, used to check the overall fit of a Cox model
- **Martingale residuals**, used to determine the functional form of a covariate to be included in a Cox model
- **Score residuals**, used for graphical checks of the proportional hazards assumption
- **Deviance residuals**, used to examine model accuracy for each individual
- **Schoenfeld residuals**, used to determine the influence of individual observations

# Searching for outliers

It is of interest to see if any of the observations has a response not well predicted by the fitted model, i.e. to examine the model for any outliers.

The martingale residual could be used, being a measure of the difference between the observed number of events and the expected number of events the individual would experience according to the model.

But: Martingale residuals are highly skewed, can take values from $-\infty$ to $+1$.

# Deviance residuals

The deviance residuals are a transform of the martingale residuals, more symmetrically distributed around zero.

Deviance residuals have a distribution closer to the Normal distribution (asymptotically ND).

UPPSALA UNIVERSITET

# Using Deviance residuals to assess the effect of each individual (outlier check)

**Procedure:**

1) Plot the individual deviance residuals $D_j$ versus the risk scores $\sum_{k=1}^{p} b_k Z_{jk}$

2) Under light to moderate censoring, $D_j$ should look like a sample of Normally distributed noise.

3) Potential outliers will have deviance residuals with large absolute values that don't fit the Normal distribution.

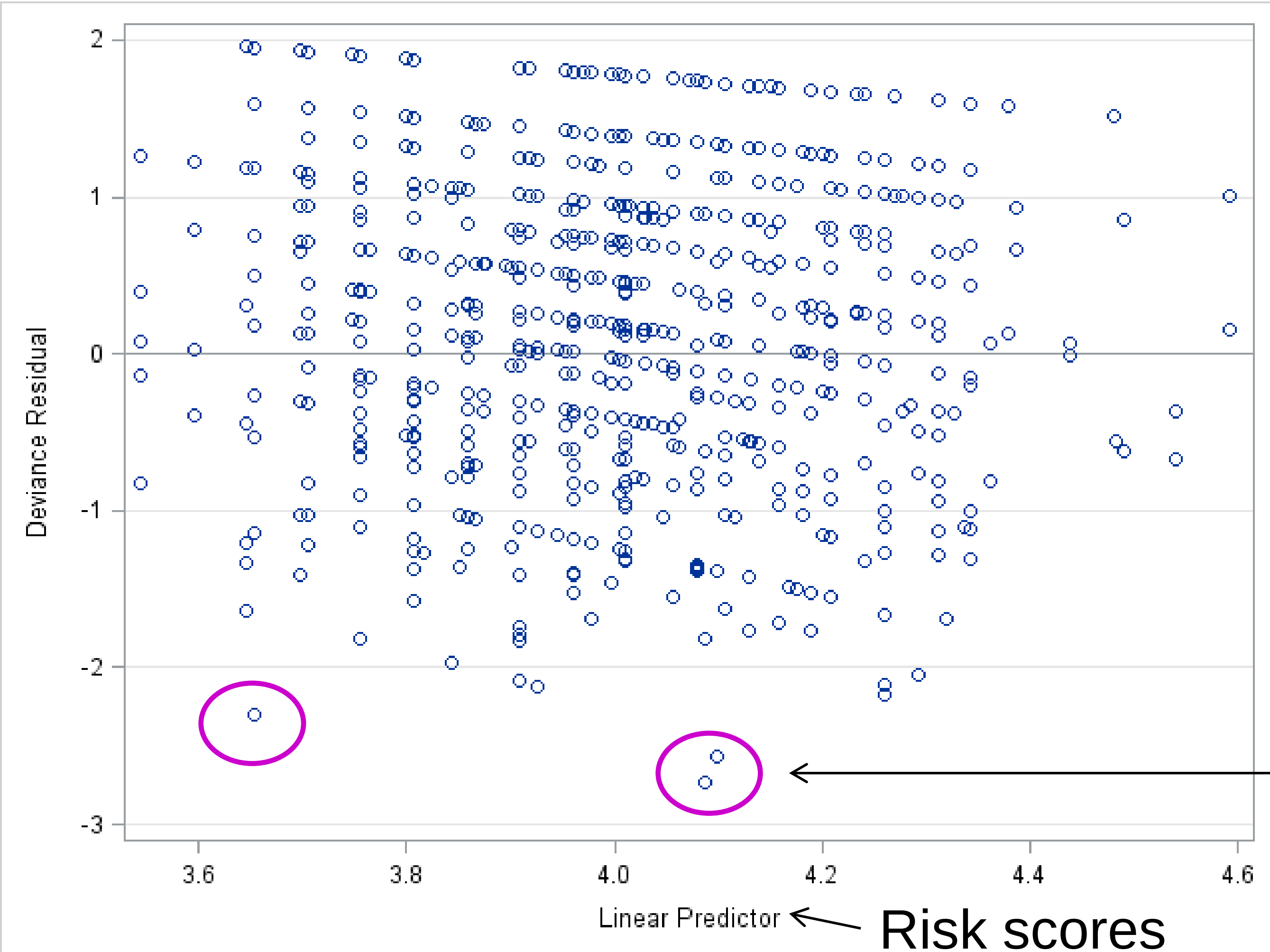
# Risk score vs. the estimated risk of experiencing the event

Risk score: $\sum_{k=1}^{p} b_k Z_{jk}$

Estimated risk of experiencing the event for an individual:

$$\exp \sum_{k=1}^{p} b_k Z_{jk}$$

# Example: Duration of breastfeeding




Potential outliers

Risk scores

# Example: Duration of breastfeeding

| | Duration of breast feeding (weeks) | Linear Predictor | Deviance Residual |
|---|---|---|---|
| 1 | 120 | 4.0877810086 | -2.737429889 |
| 2 | 104 | 4.0989420776 | -2.566037036 |
| 3 | 192 | 3.6535222193 | -2.294796572 |
| 4 | 64 | 4.2599687974 | -2.174277282 |
| 5 | 96 | 3.9267542888 | -2.118338552 |
| 6 | 60 | 4.2599687974 | -2.110382689 |

The potential outliers

Based on their risk scores they should have relatively short/medium breast-feeding duration time, but they have in fact the longest times in the study.

UPPSALA UNIVERSITET

# Checking the influence of possible outliers

The influence of possible outliers on the estimation process can be checked by the following procedure:

1) Fit the Cox model including all observations to obtain **b**, estimates of β
2) Fit the Cox model excluding one of the possible outliers, obtaining $\mathbf{b}_j$ (excluding individual *j*).
3) If $\mathbf{b}-\mathbf{b}_j$ is close to 0, the *j*th observation has little influence on the estimate.

This is feasible for a relatively small number of outliers.

# Retention or deletion of the outlier

Retain possible outliers unless demonstrable proof indicates that they are truly deviant and not representative of any observations in the population.

**If they do represent any part of the population, no matter how uncommon they are, they should be retained to ensure generalizability to the entire population.**

When deleting outliers, the analysis may improve, but the generalizability is limited. You have to redefine the population!

# Residuals and Cox regression

There are a number of different residuals for the Cox model:

- **Cox-Snell residuals**, used to check the overall fit of a Cox model
- **Martingale residuals**, used to determine the functional form of a covariate to be included in a Cox model
- **Score residuals**, used for graphical checks of the proportional hazards assumption
- **Deviance residuals**, used to examine model accuracy for each individual
- **Schoenfeld residuals**, used to determine the influence of individual observations

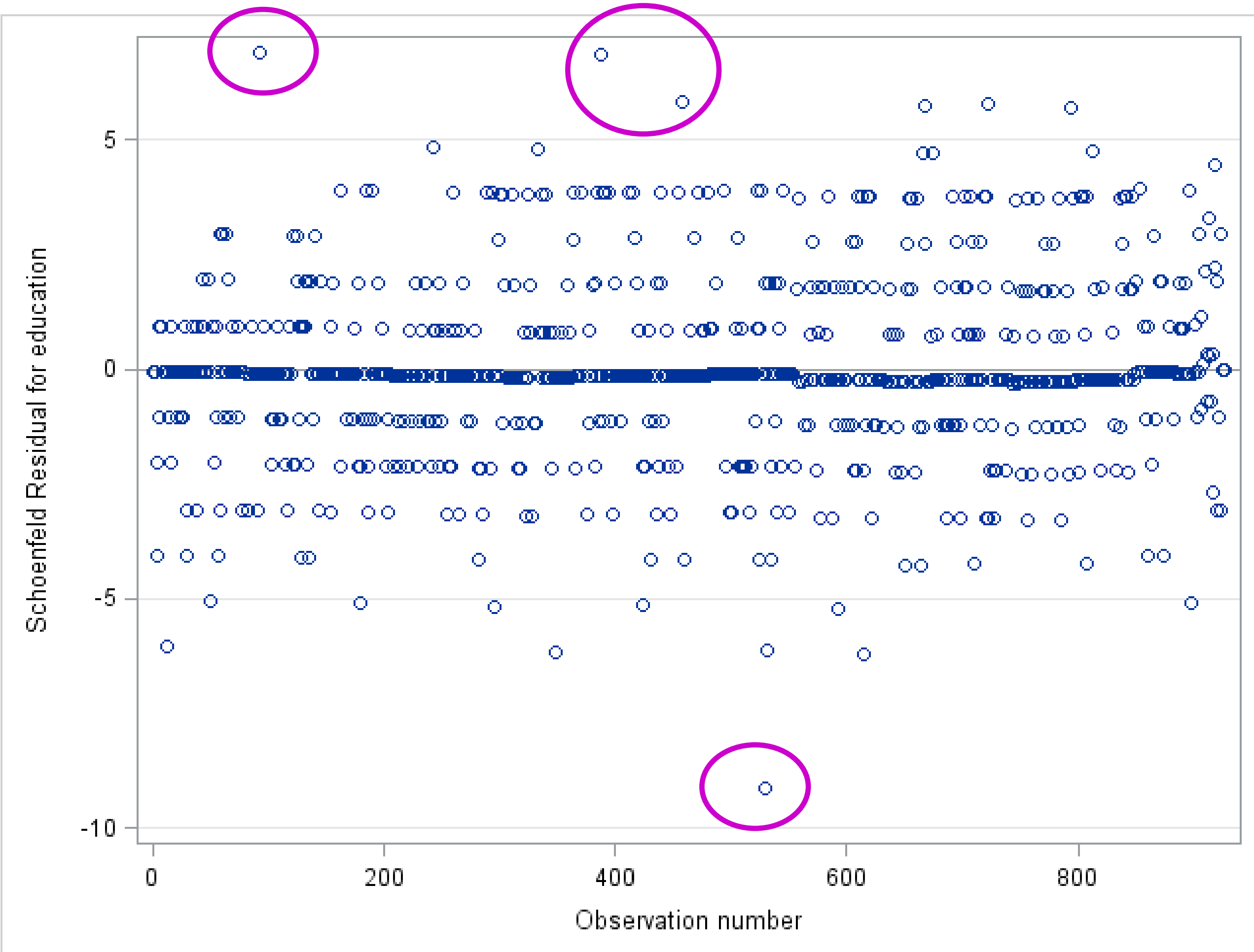
UPPSALA UNIVERSITET

# Using Schoenfeld residuals to determine the influence of individual observations

**Procedure:**

1) Fit a Cox model including all covariates, and calculate the Schoenfeld residuals.
2) Plot the residuals vs. the individual number, or vs. the covariate $Z_{jk}$ for each covariate.
3) This shows the influence of the $j$th observation for the $k$th covariate.

# Example: Duration of breastfeeding




Observations with large influence on the model estimates should be examined further