

# Practical "Multivariable survival analysis"

## P9

M. Hauptmann

**Exercise 1** Use the same dataset `bromo.sav` as in the previous practical. Variables are

- **sex**: gender (0=female, 1=male)
  - **dose**: dose of 1-bromopropane in ppm
  - **event**: event indicator (1=death, 0=censored)
  - **time**: time to death or end of study (730 days for males, 729 for females)
  - **count**: frequency of observation
- (a) Evaluate the effects of the dose groups and gender on survival in one analysis and make a statement about the direction and magnitude of the effect sizes. Is gender confounding the association between dose and death?  
SPSS tip: Tell SPSS about the frequency with which each record occurred by clicking Data – Weight Cases and selecting the appropriate variable. Then perform Cox regressions with dose in 4 categories alone and together with gender. For that, click Analyze – Survival – Cox Regression and fill in the form. Do not forget to specify dose as a 4-category variable. Click on Categorical and select variable **dose**. Select the Indicator Contrast with the first category as the reference and confirm by clicking on Change.
- (b) Use continuous dose instead of categorical dose in order to evaluate a trend in survival with increasing dose.  
SPSS tip: As above without specifying variable **dose** as a categorical variable.

### Exercise 1 – Suggested answer

- (a) Fit a Cox regression model with dose in 4 categories. Use the unexposed as the reference category.

```
COXREG time
  /STATUS=event(1)
  /CONTRAST (dose)=Indicator(1)
  /METHOD=ENTER dose
  /PRINT=CI(95)
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

In order to specify **dose** as a categorical variable, click "Categorical" and move **dose** from the left to the right panel. In the "Change Contrast" panel, click "First" and then "Change" in order to make the first dose category the reference.

	B	SE	Wald	df	Sig.	Exp(B)
dose			12.752	3	.005	
dose(1)	-.010	.218	.002	1	.964	.990
dose(2)	.315	.206	2.331	1	.127	1.370
dose(3)	.592	.199	8.833	1	.003	1.808

Fit a Cox regression model with gender and dose in 4 categories.

```
COXREG time
  /STATUS=event(1)
  /CONTRAST (sex)=Indicator
  /CONTRAST (dose)=Indicator(1)
  /METHOD=ENTER sex dose
  /PRINT=CI(95)
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

A binary variable, e.g., gender, with values 0 and 1 is automatically treated as a categorical variable with zero as reference value.

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
sex	-.599	.146	16.739	1	.000	.549	.412	.732
dose			12.802	3	.005			
dose(1)	-.016	.218	.005	1	.942	.984	.642	1.510
dose(2)	.323	.206	2.451	1	.117	1.381	.922	2.069
dose(3)	.589	.199	8.738	1	.003	1.802	1.220	2.663

The estimated hazard ratios, column "Exp(B)" are virtually unchanged when gender is included in the model, indicating gender is not a confounder. Hazard ratios increase with increasing dose and are statistically significantly 1.8-fold elevated for the highest dose category compared with the unexposed.

(b) Without specifying *dose* as a categorical variable, SPSS uses it as a continuous variable.

```
COXREG time
  /STATUS=event(1)
  /METHOD=ENTER dose sex
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

**Variables in the Equation**

	B	SE	Wald	df	Sig.	Exp(B)
dose	.001	.000	11.934	1	.001	1.001
sex	.597	.146	16.639	1	.000	1.817

The trend is highly significant ( $p=.001$ ).

**Exercise 2** The data *anderson.sav* consists of remission time data for two groups of leukemia patients with 21 patients in each group. Variables are

- *rx*: treatment (0=treatment, 1=placebo)
- *sex*: gender
- *logWBC*: log white blood cell count, a well-known prognostic indicator of survival for leukemia patients
- *lwbc3*: *logWBC* divided into low, medium and high values
- *status*: event indicator (1=relapse, 0=censored)
- *survt*: time to relapse or end of study in weeks

- (a) Check the proportional hazards assumption for the covariates. Use `lwb3` to check the assumption for `logWBC`.  
 SPSS tip: For plotting the cumulative hazard function, click Analyze – Survival – Kaplan-Meier and fill in the form. Request the hazard to be plotted under Options. Do this for each covariate. Alternatively, click Analyze – Survival – Kaplan-Meier and fill in the form. Request that the hazard is saved by clicking on Save, which creates a new variable. Calculate the negative natural logarithm of the new variable by clicking Transform – Compute Variable. Then calculate the natural logarithm of survival time and plot both variables against each other by clicking Graphs – Legacy Dialogs – Scatter/Dot – Simple.
- (b) Run a Cox-PH model for the data by using stratification.  
 SPSS tip: Click Survival – Cox Regression and use the box Strata.
- (c) Compare the stratified model with separate models.  
 SPSS tip: Create groups by clicking Data – Split File. Then, click Compare Groups.

### Exercise 2 – Suggested answer

- (a) First, we have to check the proportionality assumption for all three covariates `rx`, `sex`, `logWBC`. We can do this in two ways, (i) plot the cumulative hazard functions for the covariates, or (ii) plot the negative natural logarithm of the cumulative hazard functions for the covariate against the natural logarithm of survival time. The second option is a "safer" alternative but often the first alternative is "good enough". If we want to check the proportionality assumption for a continuous covariate, we have to categorize it and that is done for covariate `logwbc` in `lwb3`.

For alternative (i), click the following in SPSS:

Analyze – Survival – Kaplan-Meier

Time: survival time

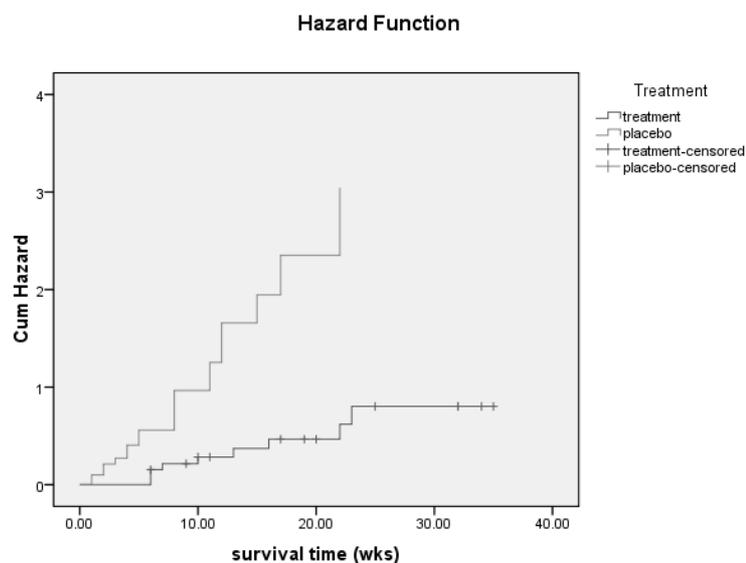
Status: status

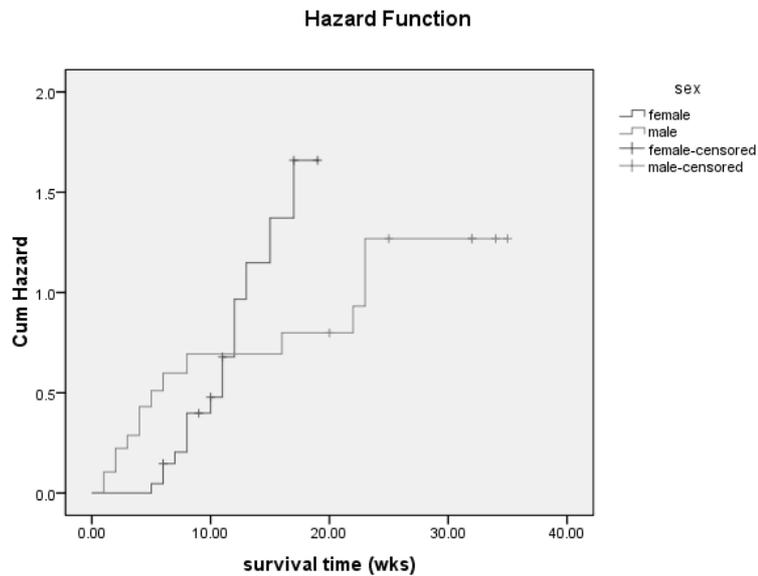
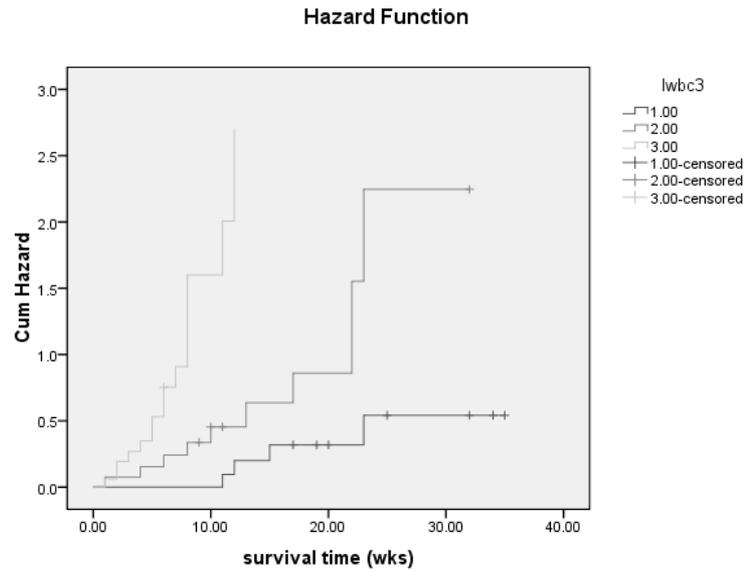
Click "Define event" and write 1 for single value.

Factor: Treatment

Click Options/Plots – Hazard

Do the same for `sex` and `lwb3`, which produces the following plots.





From this we can see that the covariate `sex` is non-proportional because the lines cross.

For alternative (ii), click the following in SPSS:

Analyze – Survival – Kaplan-Meier

Time: `survt`

Status: `status`

Click "Define event" and write 1 for single value.

Factor: `sex`

Click Save – Hazard

Now you will have a new variable in your worksheet called `HAZ_1`.

To take the negative natural logarithm of the variable `HAZ_1`:

Transform – Compute

Target variable: `minloghaz`

Numeric expression:  $-\text{LN}(\text{HAZ}_1)$

Now you will have a variable in your worksheet called `minloghaz` and this variable is the negative natural logarithm of the variable `HAZ_1`.

To take the natural logarithm of survival time `survt`:

Transform – Compute

Target variable: `logtime`

Numeric expression: `LN(survt)`

Now you will have a variable in your worksheet called `logtime` and this variable is the natural logarithm of survival time `survt`.

To make a plot of `minloghaz` against `logtime`:

Graphs – (Legacy Dialogs) – Scatter – Simple

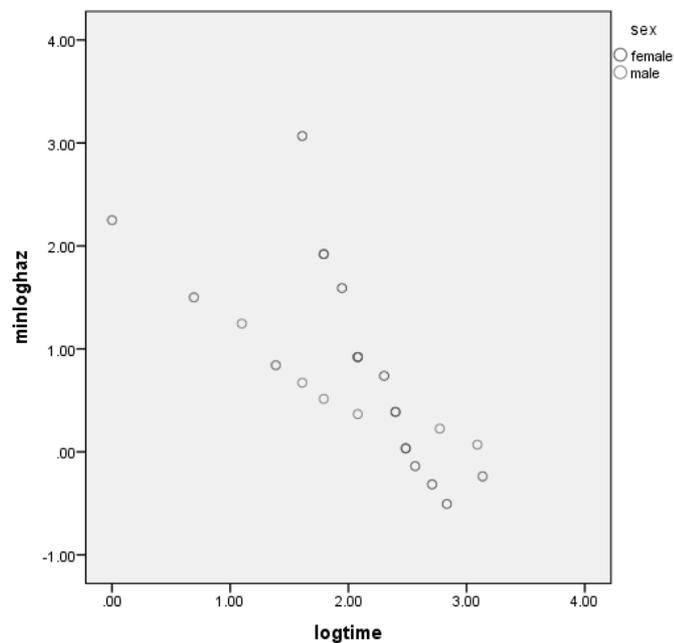
Y-axis: `minloghaz`

X-axis: `logtime`

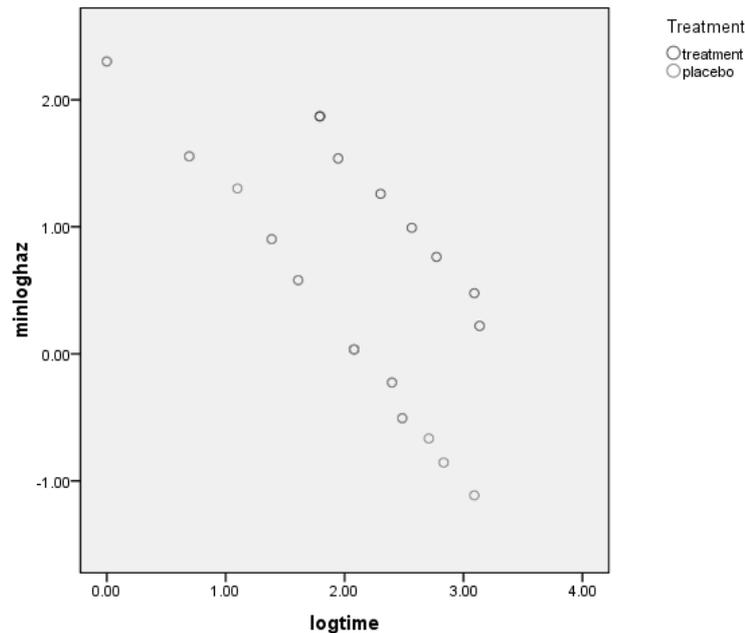
Put `sex` in the "Set Markers By" box

You will have a plot with points in different colors corresponding to the different groups. Parallel curves support the PH assumption for treatment.

Do the same for the other covariates `treatment` and `WBC`. The plot for `sex` is below and shows clearly that `sex` is non-proportional (because the lines are not parallel).



The plot for `treatment` is as follows.



(b) We can't use gender in the model because the variable is non-proportional. If we use gender as a stratification variable, the analysis is still adjusted for gender because the model uses separate baseline hazard functions for males and females.

In SPSS, click: Analyze – Survival – Cox Regression

Time: **survt**

Status: **status**

Click "Define event" and write 1 for single value.

Covariates: Treatment (**rx**)

**logwbc** (In the model you can use **logwbc** as a covariate, **lwbc3** is only used for testing the proportional hazards assumption.)

Mark Treatment **rx** and click categorical.

Choose reference category (last or first), remember to click change.

Strata: **sex**

Choosing the placebo group as the reference category (reference category is last for the covariate treatment), the output is as follows.

Variables in the Equation									
	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)		
							Lower	Upper	
rx	-.931	.472	3.894	1	.048	.394	.156	.994	
logwbc	1.390	.338	16.956	1	.000	4.016	2.072	7.783	

Interpretation: In this model, it is assumed that females and males have different baseline hazard functions for the risk of failure. Both covariates are significant. **logWBC** is significant at the .01 level and treatment is significant at the .048 level. For each unit increase of **logWBC**, the risk of failure increases four times. The treatment group has 40% of the risk of failure compared with the placebo group, i.e., risk is reduced by 60%.

(c) One reason for non-proportional hazards can be that the covariates have different effects for the different groups. In this example it means that the covariates may have different effects on males and females. To investigate this we can perform separate analyses for males and females.

In SPSS, click:

Data – Split file – Compare groups

Groups based on sex.

Follow the instructions in (b) but do not include a stratification variable.

**Variables in the Equation**

sex		B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
								Lower	Upper
female	rx	-.267	.566	.222	1	.637	.766	.253	2.322
	logwbc	1.170	.499	5.508	1	.019	3.222	1.213	8.562
male	rx	-1.859	.729	6.501	1	.011	.156	.037	.650
	logwbc	1.639	.519	9.970	1	.002	5.150	1.862	14.242

Interpretation: It appears that the treatment effect differs between males and females. Treatment is only significant for males ( $p=.011$ ). In males, treatment reduces the risk of failure by 84%, whereas there is no significant treatment effect for females ( $p=.637$ ). For males, it is somewhat more dangerous to have high values of  $\log\text{wbc}$ : the risk increases 5 times for each unit increase. For females, the risk increases about 3 times per unit increase in  $\log\text{wbc}$ .