Actuarial Research Centre

Institute and Faculty of Actuaries

Cox regression, the proportional hazards assumption and time-varying covariates

Dr Ilyas Bakbergenuly (UEA)

The **'Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks'** research programme is being funded by the Actuarial Research Centre.

30/10/2019

Actuarial Research Centre
Institute and Faculty of Actuaries

Disclaimer

The views expressed in this presentation are those of invited contributors and not necessarily those of the Institute and Faculty of Actuaries. The Institute and Faculty of Actuaries does not endorse any of the views stated, nor any claims or representations made in this presentation and accept no responsibility or liability to any person for loss or damage suffered as a consequence of their placing reliance upon any view, claim or representation made in this presentation. The information and expressions of opinion contained in this presentation are not intended to be a comprehensive study, nor to provide actuarial advice or advice of any nature and should not be treated as a substitute for specific advice concerning individual situations. On no account may any part of this presentation be reproduced without the written permission of the Institute and Faculty of Actuaries .

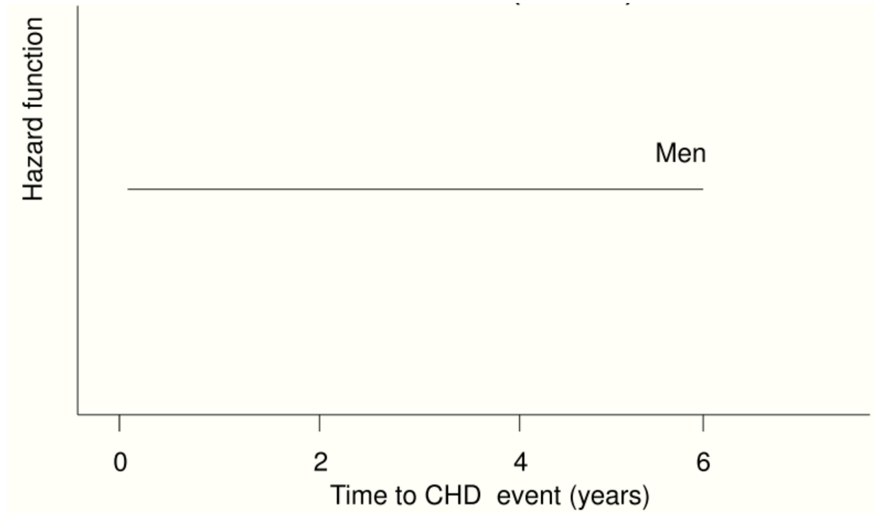
Objectives

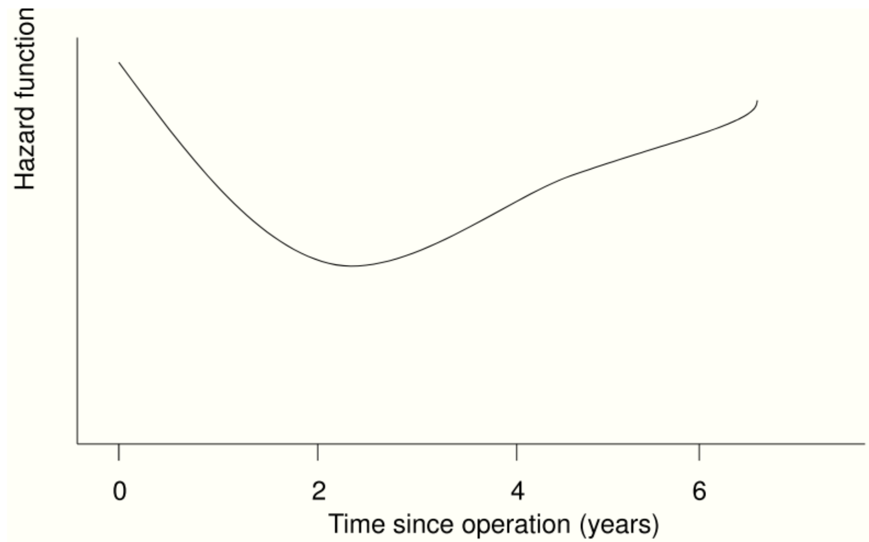
- What is a hazard ratio and Cox proportional hazards model.
- Describe methods to check the assumption of proportional hazards in the Cox model
- Describe methods how to deal with non-proportional hazards in the Cox model

Actuarial Research Centre
Institute and Faculty of Actuaries

Hazard aka "force of mortality" and "mortality intensity"

- Hazard is an instantaneous failure rate at time *t*
  - Probability that an individual will experience the event at time *t* given that the event has not yet occurred.





Cox proportional hazards regression

- The type of regression model typically used in survival analysis in medicine is the Cox proportional hazards regression model.
- The Cox model estimates the hazard $\mu_i(t)$ for subject i for time t by multiplying the baseline hazard function $\mu_0(t)$ by the subject's risk score r_i as

$$\mu_i(t, \beta, Z_i) = \mu_0(t)\, r_i(\beta, Z_i) = \mu_0(t) e^{\beta\, Z_i}$$

- The risk factors Z have a log-linear contribution to the force of mortality which does not depend on time t.

Actuarial Research Centre
Institute and Faculty of Actuaries

Hazard ratio (HR)

- Taking a ratio of the hazard functions for two subjects *i* and *j* who differ in one risk factor *z* (with the values z_0 and z_1 , respectively) but not in the other risk factors,

$$\mathrm{HR}(t,\beta,Z) = \frac{\mu_i(t,\beta,Z_i)}{\mu_j(t,\beta,Z_j)} = \frac{\mu_0(t)e^{\beta\, Z_i}}{\mu_0(t)e^{\beta\, Z_j}} = \frac{e^{\beta_Z\, z_1}}{e^{\beta_Z\, z_0}} = e^{\beta_Z\,(z_0 - z_1)}.$$

- This means that the baseline hazard $\mu_0(t)$ does not have to be specified and the hazard ratio $\mathrm{e}^{\beta_Z\,(z_0-z_1)}$ is constant with respect to time *t*.

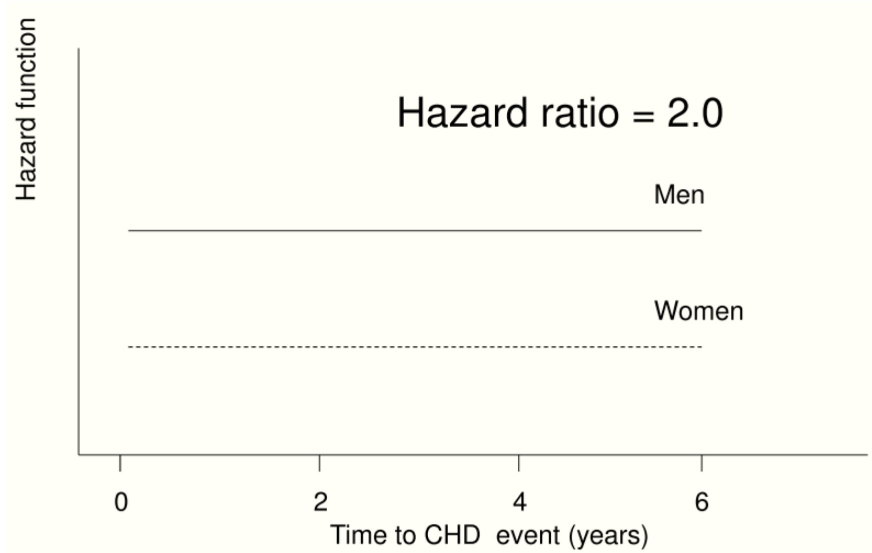
- Because of this, the Cox model does not make any assumptions about the shape of the baseline hazard.

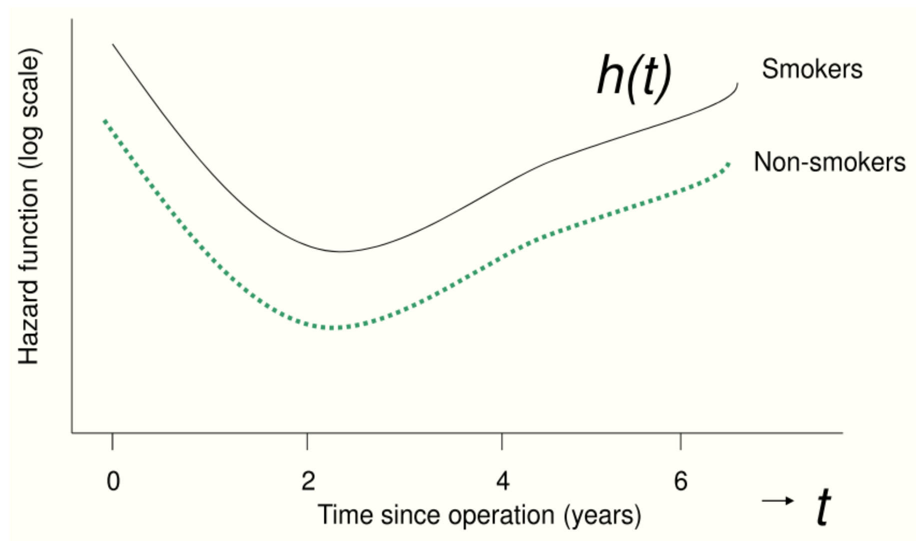
- $\mathrm{e}^{\beta_Z\,(z_0-z_1)}$ is an adjusted HR, i.e. all other risks are already accounted for by the model.

Actuarial Research Centre
Institute and Faculty of Actuaries

Hazard ratio

- Comparison of two hazard functions
- Cox model assumes constant hazard ratio over time





Proportional hazards assumption

- Graphical methods:
  - Comparison of Kaplan-Meier estimates by group
  - Plot (minus the log cumulative baseline hazard) for each group against (log survival time)

- Formal tests:
  - Grambsch and Therneau’s test based on Schoenfeld residuals
  - Include interaction between covariate and a function of time
    - Log(time) often used but could be any function of time

Actuarial Research Centre
Institute and Faculty of Actuaries

Example: Cox model for death from Parkinson's disease

- Data: parkison disease
  - Sample of 520 patients
  - Study period of 17 years
- Outcome: time to death (266 events)
- Exposure: new vs standard treatment
- Covariates:

  Sex (baseline male / female)

  Age (baseline 25-59 / 60-69 / 70-92)

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.216	0.8221	0.9549	1.55

Concordance= 0.527 (se = 0.016)

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.216	0.8224	0.9545	1.549
sex	1.031	0.9701	0.8099	1.312

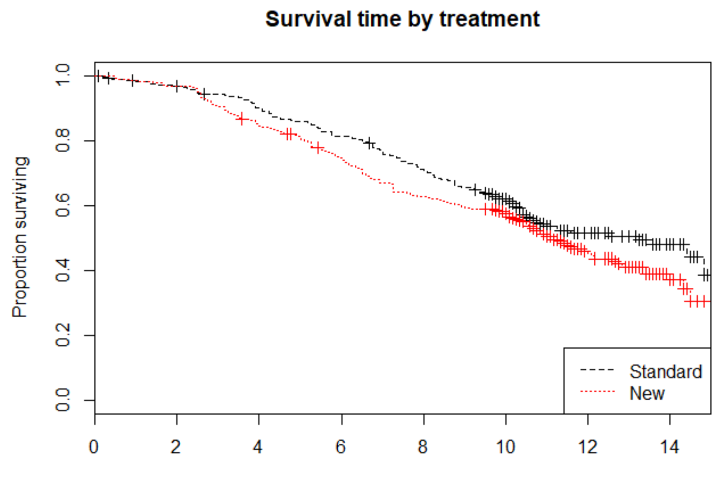
Concordance= 0.522 (se = 0.018)

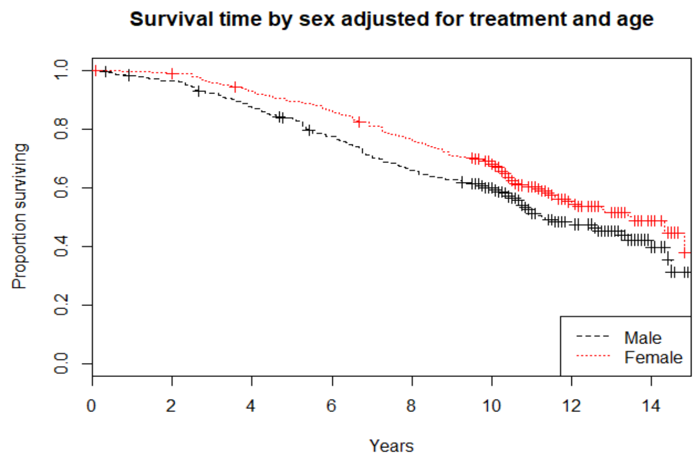
	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.1615	0.8610	0.9101	1.4822
sex	0.7412	1.3491	0.5774	0.9516
agegrp2	3.4363	0.2910	2.3853	4.9504
agegrp3	7.7408	0.1292	5.3363	11.2286

Concordance= 0.706 (se = 0.019)

Kaplan-Meier plots by levels of a factor

- Estimated survival function
  - Does not adjust for other covariates!
  - Crossing of hazard lines indicates non-proportional hazards
  - Otherwise, can be difficult to judge





Complementary log-log plot of S(t;Z)

- From the hazard function of the PH model, we obtain the survivor function

$$S(t; Z) = \exp\{ -M_0(t)\, e^{\beta Z}\}$$

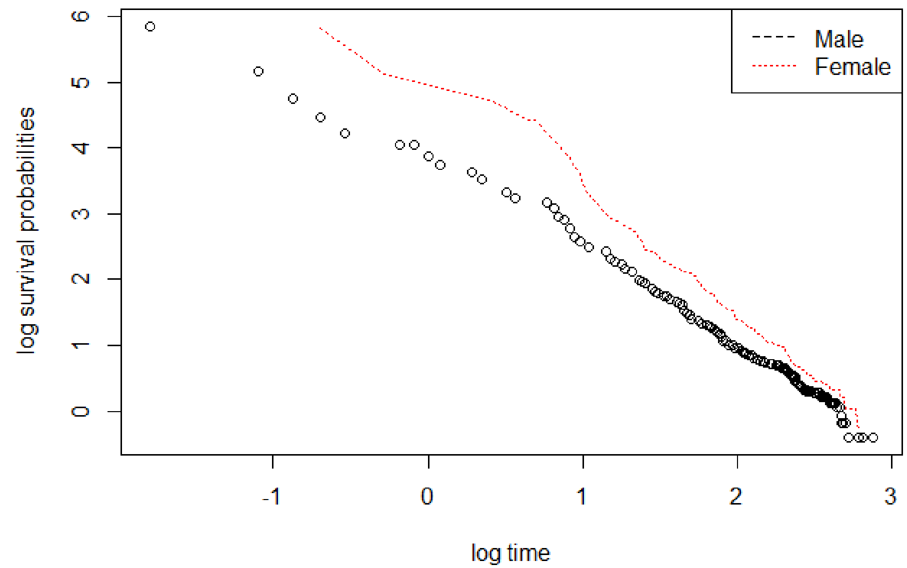
where $M_0(t)$ is the cumulative hazard corresponding to $\mu_0(t)$.

- Hence $\ln\{ -\ln S(t; Z)\} = \ln(M_0(t)) + \beta Z$.
- Hence any two such functions, S(t;z1) and S(t; z2) for different values of the covariate vector z, will be parallel.
- Plot ln{ -ln S(t; Z)} vs t or a function of t.

Actuarial Research Centre
Institute and Faculty of Actuaries

Complementary log-log plot for Parkinson's data

- Can be unadjusted or adjusted (here adjusted for treatment and age group)
- Proportional hazards assumption violated if curves are not parallel to each other
- Plot vs log(t) shows straight lines for Weibull distribution.



This only works if there are few covariates and few distinct values, only then S(t;Z) is reliably estimated for each Z value.

Actuarial Research Centre
Institute and Faculty of Actuaries

Residuals

- Residual is the difference between an observed value and a predicted value.
- Due to censoring, this is not straightforward in survival analysis
- Therefore, there are many types of residuals
- Here we are going to concentrate on **Cox-Snell residuals** and **Schoenfeld residuals**

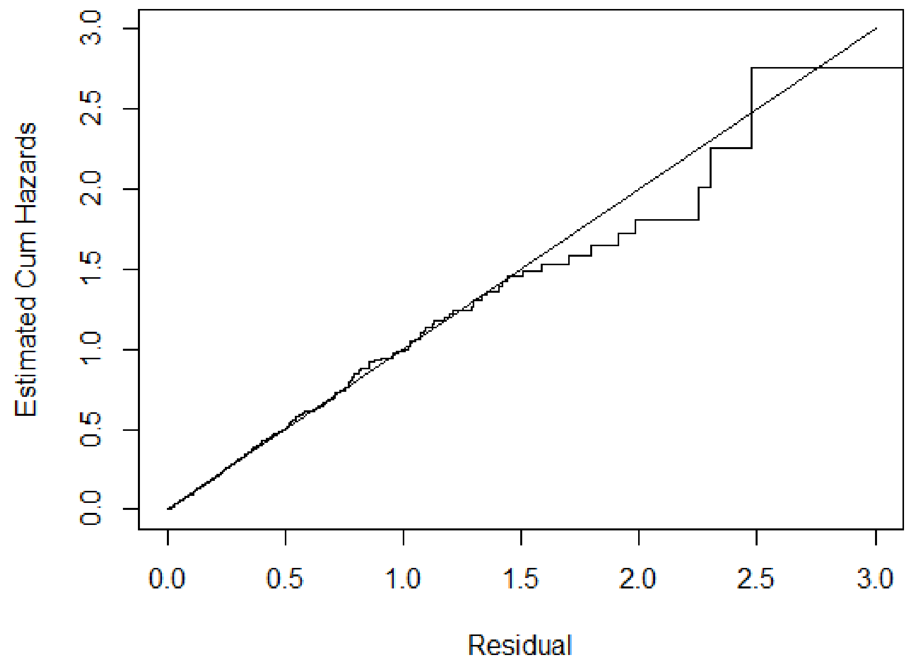
Cox-Snell residuals

- In order to assess an overall goodness of fit of a model, we use Cox-Snell residuals
- Cox-Snell residuals are $-\log(\hat{S}(t; Z))$, i.e. estimated cumulative hazards at the time of death or censoring
- If the model is correct, Cox-Snell residuals should have exponential distribution exp(1)

Actuarial Research Centre
Institute and Faculty of Actuaries

Cox-Snell residuals

- Overall goodness-of-fit
  - The first survival model for Parkinson’s data with treatment, sex, and age group. Graph indicates good fit.
- Plot of Cox-Snell residuals is just a QQ-plot for exponential distribution



Actuarial Research Centre
Institute and Faculty of Actuaries

Schoenfeld residuals

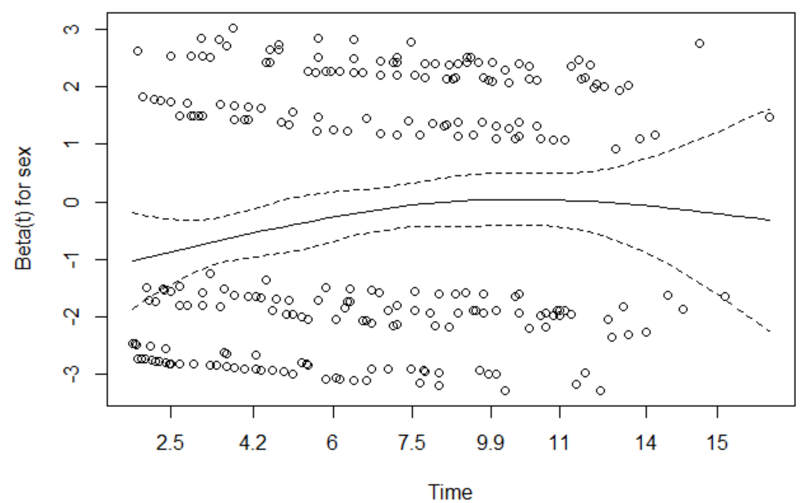
- Schoenfeld residuals are the differences between the covariate value Z_i of subject i who experienced an event at time t_i and the weighted average of all covariate values across all subjects at risk at t_i
- Schoenfeld residuals are used for testing the proportionality of hazards assumption using Grambsh and Therneau’s test

Actuarial Research Centre
Institute and Faculty of Actuaries

Grambsch and Therneau test

- Testing correlation between Schoenfeld residuals and survival time
- Significant correlation indicates non-proportional hazards

	rho	chisq	p
treat	-0.06626	1.156192	0.2823
sex	0.13732	5.459043	0.0195
agegrp2	0.00114	0.000344	0.9852
agegrp3	-0.07785	1.658215	0.1978
GLOBAL	NA	8.123383	0.0872



Cox model with time-varying coefficients

$$\mu(t, \beta, Z) = \mu_0(t)e^{\beta(t)\,Z}$$

Write the time-varying coefficients as

$$\beta_j(t)=\beta_j+\theta_j g_j(t), \quad j=1,\ldots,p$$

where g_j(t) is known. A standard choice is g_j(t) = log(t).

Test H_0: θ=0 (as a vector and for each component.).

Actuarial Research Centre
Institute and Faculty of Actuaries

Testing interaction of covariate with time

- Significant correlation indicates non-proportional hazards
- NB: very sensitive

```
  n= 520, number of events= 266

                        coef  exp(coef)   se(coef)      z Pr(>|z|)
treat              4.078e+00  5.905e+01  5.989e-01  6.810 9.76e-12 ***
sex                3.109e+00  2.239e+01  4.942e-01  6.290 3.17e-10 ***
agegrp2            1.814e+01  7.566e+07  2.184e+00  8.307  < 2e-16 ***
agegrp3            1.903e+01  1.843e+08  2.149e+00  8.857  < 2e-16 ***
treat:log(time)   -2.054e+00  1.283e-01  2.900e-01 -7.081 1.43e-12 ***
sex:log(time)     -1.590e+00  2.040e-01  2.493e-01 -6.377 1.81e-10 ***
agegrp2:log(time) -7.222e+00  7.300e-04  8.989e-01 -8.035 8.88e-16 ***
agegrp3:log(time) -7.597e+00  5.019e-04  8.938e-01 -8.500  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Actuarial Research Centre
Institute and Faculty of Actuaries

What if the proportional hazards assumption is not met?

- Stratify the analysis on violating variable: $\mu_s(t,\beta,Z') = \mu_{0s}(t)e^{\beta Z'}$ for Z' being all covariates but that one.
  - Fit one model: allow baseline hazards to vary by group but assume covariate effects are the same across strata. Only if the variable is of no direct interest. (There should be no significant interactions between covariates and stratum variable)
  - Fit separate models: allow both baseline hazards and hazard ratios to vary by group

What if the proportional hazards assumption is not met?

- Include time-dependent effect
  - Split follow-up time such that the hazards are proportional within these time bands
  - Continuous (could be any function of time)

Actuarial Research Centre
Institute and Faculty of Actuaries

Stratified analysis

- Check for interactions
- Fit one stratified Cox model (n=520, events=266)
- Fit separate models
  - Male (n=283, events=141)
  - Female (n=237, events=125)
- Easy procedure but comes at the cost of no estimate for the effect of the violated variable associated with the outcome

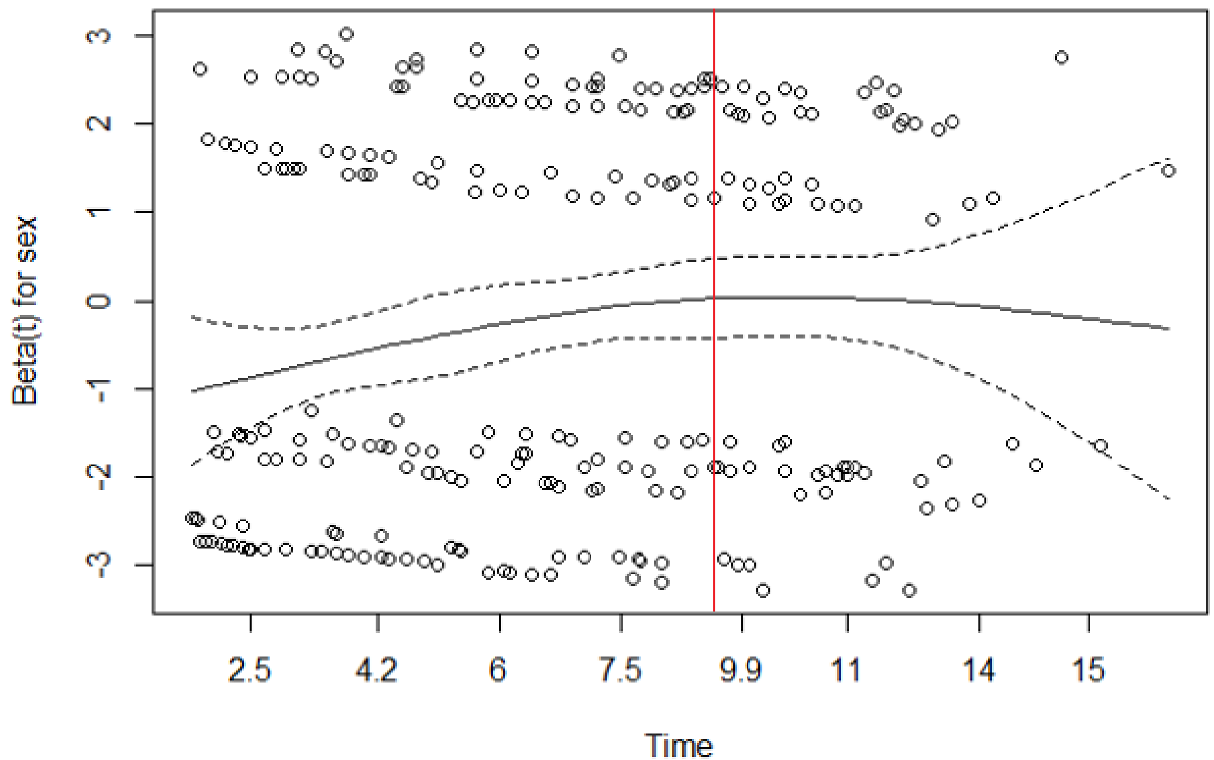
	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.0395	0.9620	0.4877	2.216
sex	0.7895	1.2666	0.2964	2.103
agegrp2	2.1979	0.4550	0.7261	6.653
agegrp3	23.7526	0.0421	7.8517	71.855
treat:sex	1.0938	0.9142	0.6695	1.787
sex:agegrp2	1.3487	0.7415	0.6348	2.865
sex:agegrp3	0.4752	2.1045	0.2243	1.007

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.162	0.8605	0.9099	1.484
agegrp2	3.392	0.2948	2.3549	4.887
agegrp3	7.389	0.1353	5.0865	10.735

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.126	0.88817	0.8067	1.571
agegrp2	2.918	0.34272	1.8230	4.670
agegrp3	10.540	0.09488	6.5101	17.063

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.258	0.7950	0.8737	1.811
agegrp2	4.005	0.2497	2.2151	7.242
agegrp3	5.335	0.1875	2.9578	9.622

Schoenfeld Residuals plot of effect of sex over time



Actuarial Research Centre
Institute and Faculty of Actuaries

Step-wise time-dependent hazards

- Split follow-up time in intervals in which the proportional hazards assumpation is no longer violated
- Create time dependent effect
  - Here: 0 = male's hazard (baseline),
    1=female's hazard 0-9 years,
    2=female's hazard 9+ years
- Fit mode with time dependent effect
- More time consuming procedure due to creating the most effective time intervals

	exp(coef)	exp(-coef)	lower .95	upper .95
treat	1.1636	0.8594	0.9117	1.4852
t_sex1	0.6887	1.4520	0.5162	0.9188
t_sex2	0.9118	1.0968	0.5678	1.4641
agegrp2	3.4215	0.2923	2.3750	4.9291
agegrp3	7.5985	0.1316	5.2343	11.0305

	rho	chisq	p
treat	-0.06453	1.101428	0.2940
t_sex1	0.09955	2.762015	0.0965
t_sex2	0.02402	0.158846	0.6902
agegrp2	0.00158	0.000657	0.9795
agegrp3	-0.07938	1.708571	0.1912
GLOBAL	NA	6.446771	0.2651

Actuarial Research Centre
Institute and Faculty of Actuaries

References

- Therneau, T.M. and Grambsch, P.M., 2013. *Modeling survival data: extending the Cox model*. Springer Science & Business Media.
- Martinussen T, Scheike TH. *Dynamic Regression Models for Survival Data*. Springer: New York, 2006.

Actuarial
Research Centre

Institute and Faculty
of Actuaries

The Actuarial Research Centre (ARC)

A gateway to global actuarial research

The Actuarial Research Centre (ARC) is the Institute and Faculty of Actuaries' (IFoA) network of actuarial researchers around the world.

The ARC seeks to deliver cutting-edge research programmes that address some of the significant, global challenges in actuarial science, through a partnership of the actuarial profession, the academic community and practitioners.

The '**Use of Big Health and Actuarial Data for understanding Longevity and Morbidity Risks**' research programme is being funded by the ARC.

www.actuaries.org.uk/arc

Questions

Comments

The views expressed in this presentation are those of the presenter.