Survival analysis and regression – Part 3

Michael Otterstatter

BCCDC Biostats Session

November 22, 2019

Session overview

- In this session we will
  - continue discussing the structure and assumptions of regression models for survival data
  - consider extension of these regression models using time-varying covariates

Reminder: survival analysis

1. Define event of interest, time zero, time scale and how participants exit
   - Consideration of censoring
2. Descriptive analysis: univariate modeling
   - KM curves and descriptive statistics
3. Inferential analysis: multivariate modeling
   - Cox regression (semi-parametric)

Reminder: Proportional hazards models

- In the case of survival (time-to-event) analysis, we model the **hazard**
- log of the hazard ratio is the link used connect to the linear predictors

$$\log(HR) = \log\left(\frac{h(t)}{h_0(t)}\right) = \beta_1 x_1 + \beta_2 x_2 \ldots$$

$$\log h(t) = \log(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 \ldots$$

Intercept slopes

Reminder: Proportional hazards models

- The most common proportional hazards model is the **Cox regression**
- covariate set (linear predictors) is parametric, but no assumptions are made about baseline hazard (often written as $\lambda_0(t)$)

$$h(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2 \dots}$$

An example

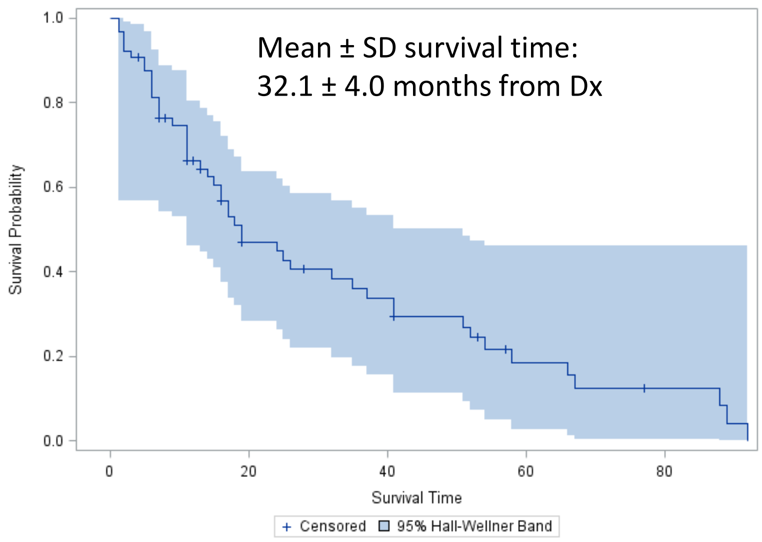
- Multiple myeloma study (see Krall et al, 1975)
  - 65 patients undergoing treatment (48 died during study)
  - Analysis of survival time from diagnosis
  - Identifying factors associated with survival

Time (months)
Alive/Dead
Blood urea nitrogen
Hemoglobin
White blood cells
Plasma cells in marrow
Protein in urine
Serum calcium

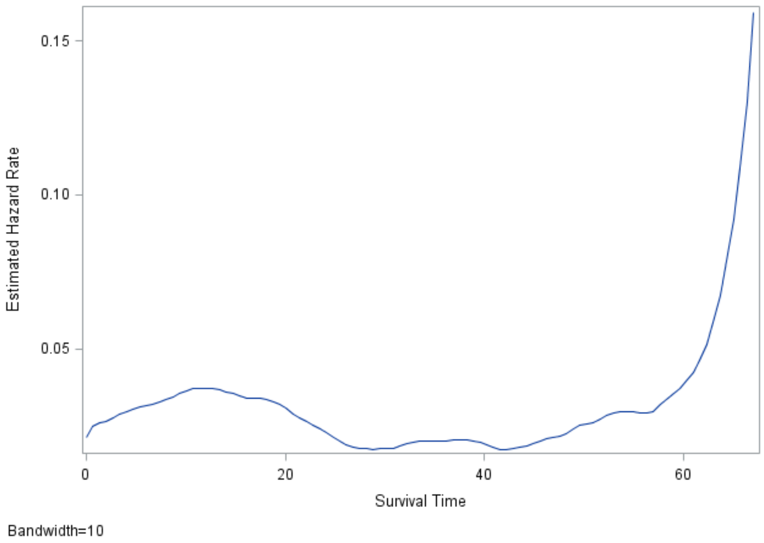
Time	Status	LogBUN	HGB	Platelet	Age	LogWBC	Frac	LogPBM	Protein	SCalc
1.25	1	2.2175	9.4	1	67	3.6628	1	1.9542	12	10
1.25	1	1.9395	12	1	38	3.9868	1	1.9542	20	18
2.00	1	1.5185	9.8	1	81	3.8751	1	2	2	15
2.00	1	1.7482	11.3	0	75	3.8062	1	1.2553	0	12
2.00	1	1.301	5.1	0	57	3.7243	1	2	3	9
3.00	1	1.5441	6.7	1	46	4.4757	0	1.9345	12	10
4.00	0	1.9542	10.2	1	59	4.0453	0	0.7782	12	10
4.00	0	1.9243	10	1	49	3.959	0	1.6232	0	13
5.00	1	2.2355	10.1	1	50	4.9542	1	1.6628	4	9
5.00	1	1.6812	6.5	1	74	3.7324	0	1.7324	5	9
6.00	1	1.3617	9	1	77	3.5441	0	1.4624	1	8

Descriptive analysis

Kaplan-Meier survival curve



Estimated hazard function



Inferential analysis: the model

- Cox proportional hazards regression

$$h(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2 \dots}$$

```
proc phreg data=Myeloma;
   model Time*VStatus(0)=LogBUN HGB Platelet Age LogWBC
                         Frac LogPBM Protein SCalc;
run;
```

Inferential analysis: the model

- Cox proportional hazards regression

$$h(t) = \lambda_0(t)e^{1.798*LogBUN-0.126*HGB+\cdots}$$

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
65	48	17	26.15

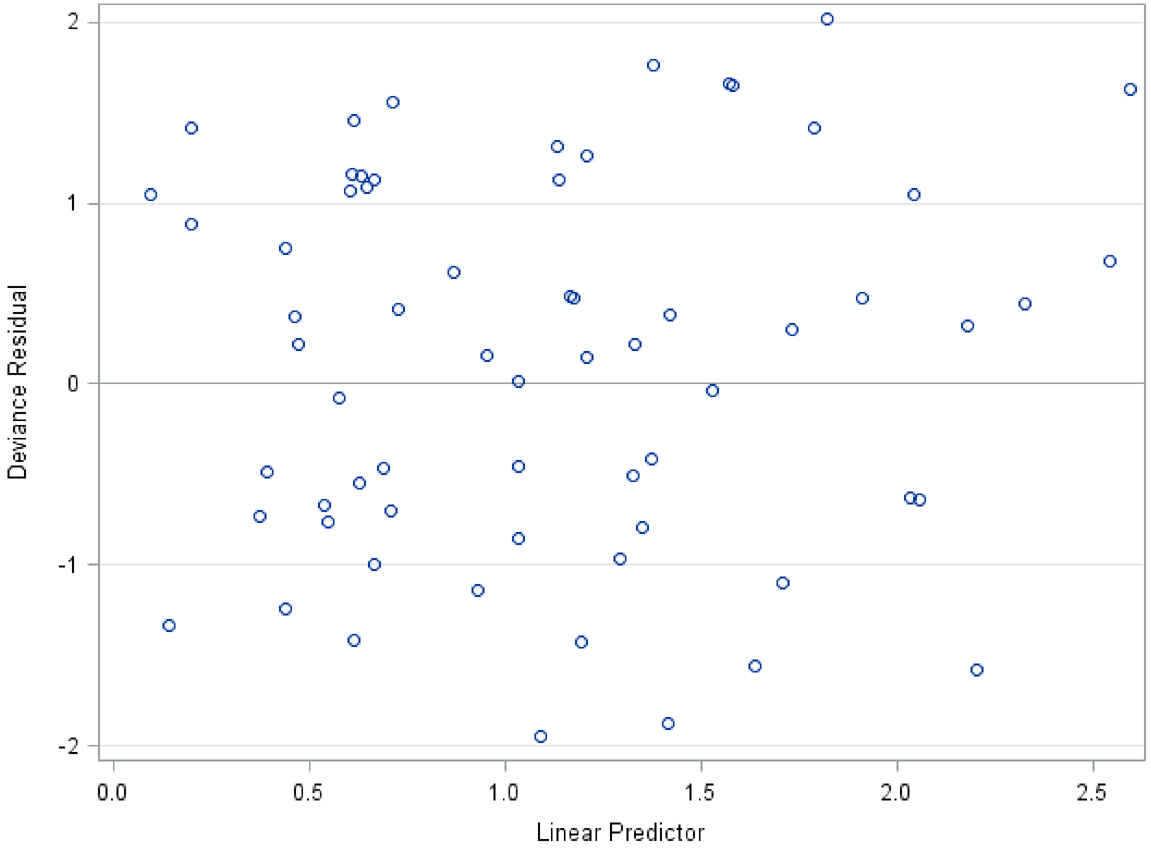
Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	309.716	292.588
AIC	309.716	310.588
SBC	309.716	327.429

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
LogBUN	1	1.79836	0.64833	7.6942	0.0055	6.040
HGB	1	-0.12631	0.07183	3.0920	0.0787	0.881
Platelet	1	-0.25059	0.50747	0.2438	0.6214	0.778
Age	1	-0.01279	0.01948	0.4316	0.5112	0.987
LogWBC	1	0.35371	0.71319	0.2460	0.6199	1.424
Frac	1	0.33788	0.40728	0.6883	0.4068	1.402
LogPBM	1	0.35893	0.48603	0.5454	0.4602	1.432
Protein	1	0.01307	0.02617	0.2494	0.6175	1.013
SCalc	1	0.12595	0.10340	1.4837	0.2232	1.134

Inferential analysis: model fit

- Assessing model fit (as usual, with residuals)



Inferential analysis: assumptions

- Assessing model assumptions: **log-linearity**
  - *covariates assumed to have linear relation with log of hazard*

$$h(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2 \dots}$$

$$\log h(t) = \log(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 \dots$$

Inferential analysis: assumptions

- Assessing model assumptions: **log-linearity**
  - *covariates assumed to have linear relation with log of hazard*
  - **simple test**: add quadratic terms and test for significant non-linearity; and,
  - bin continuous covariates and examine estimates for each stratum

Inferential analysis: assumptions

- Assessing model assumptions: **log-linearity**
  - *covariates assumed to have linear relation with log of hazard*
  - <u>add quadratic terms and test for significant non-linearity</u>

```
proc phreg data=Myeloma;
   model Time*Vstatus(0)=LogBUN HGB LogBUN_2 HGB_2;
   LogBUN_2 = LogBUN * LogBUN;
   HGB_2 = HGB * HGB;
run;
```

quadratic terms

Inferential analysis: assumptions

- Assessing model assumptions: **log-linearity**
  - *covariates assumed to have linear relation with log of hazard*
  - <u>add quadratic terms and test for significant non-linearity</u>

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
LogBUN	1	-8.22578	3.54128	5.3955	0.0202	0.000
HGB	1	-0.05381	0.41569	0.0168	0.8970	0.948
LogBUN_2	1	3.30650	1.16915	7.9983	0.0047	27.289
HGB_2	1	-0.00370	0.02146	0.0297	0.8631	0.996

quadratic terms

Inferential analysis: assumptions

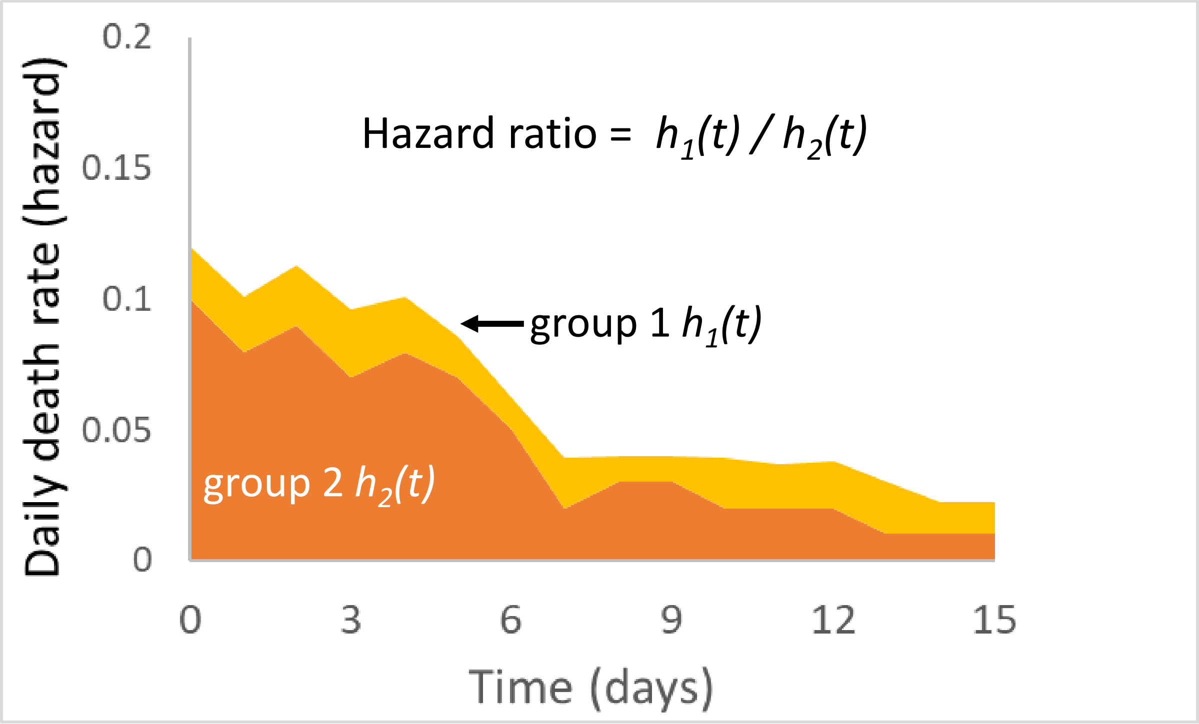
- Assessing model assumptions: **log-linearity**
  - *covariates assumed to have linear relation with log of hazard*
  - <u>bin continuous covariates and examine estimates for each stratum</u>

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
LogBUN_cat	2	1	-0.11559	0.43748	0.0698	0.7916	0.891	LogBUN_cat 2
LogBUN_cat	3	1	-0.20499	0.45563	0.2024	0.6528	0.815	LogBUN_cat 3
LogBUN_cat	4	1	0.76252	0.45333	2.8292	0.0926	2.144	LogBUN_cat 4
HGB_cat	2	1	0.52388	0.42464	1.5220	0.2173	1.689	HGB_cat 2
HGB_cat	3	1	-0.45328	0.42443	1.1406	0.2855	0.636	HGB_cat 3
HGB_cat	4	1	-0.73908	0.42852	2.9747	0.0846	0.478	HGB_cat 4

original variable binned into quartiles

Inferential analysis: assumptions

- Assessing model assumptions: **proportional hazards**
  - *hazard ratios assumed to be constant over time*



Inferential analysis: assumptions

- Assessing model assumptions: **proportional hazards**
  - *hazard ratios assumed to be constant over time*
  - **simple test**: create new variables as

    [original variable] x [log(time)]

    and test for significance (change in effect over time); and,
  - examine survival curves by strata and look for obvious non-proportionality (non-parallel lines)

Inferential analysis: assumptions

- Assessing model assumptions: **proportional hazards**
  - *hazard ratios assumed to be constant over time*
  - create new variables as [original variable] x [log(time)] and test for significance (change in effect over time)

```
proc phreg data=Myeloma;
   model Time*VStatus(0) = LogBUN HGB Platelet Age LogWBC Frac LogPBM Protein
                           SCalc LogBUN_t HGB_t Platelet_t Age_t LogWBC_t Frac_t
                           LogPBM_t Protein_t SCalc_t;
    LogBUN_t = LogBUN*log(Time);
    HGB_t = HGB*log(Time);
    Platelet_t = Platelet*log(Time);
    Age_t = Age*log(Time);
    LogWBC_t = LogWBC*log(Time);
    Frac_t = Frac*log(Time);
    LogPBM_t = LogPBM*log(Time);
    Protein_t = Protein*log(Time);
    SCalc_t = SCalc*log(Time);

    proportionality_test: test LogBUN_t, HGB_t, Platelet_t, Age_t,
                               LogWBC_t, Frac_t, LogPBM_t, Protein_t, SCalc_t;
run;
```

New variables: original variable x log(time)

Inferential analysis: assumptions

- Assessing model assumptions: **proportional hazards**
  - *hazard ratios assumed to be constant over time*
  - create new variables as [original variable] x [log(time)] and test for significance (change in effect over time)

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
LogBUN_t	1	-1.80037	0.78370	5.2774	0.0216	0.165
HGB_t	1	-0.03453	0.07914	0.1904	0.6626	0.966
Platelet_t	1	0.31026	0.60659	0.2616	0.6090	1.364
Age_t	1	-0.02107	0.01843	1.3066	0.2530	0.979
LogWBC_t	1	-0.25876	0.67118	0.1486	0.6998	0.772
Frac_t	1	0.12365	0.40442	0.0935	0.7598	1.132
LogPBM_t	1	-1.43320	0.67690	4.4830	0.0342	0.239
Protein_t	1	-0.00316	0.03127	0.0102	0.9196	0.997
SCalc_t	1	-0.18544	0.08950	4.2926	0.0383	0.831

Linear Hypotheses Testing Results			
Label	Wald Chi-Square	DF	Pr > ChiSq
proportionality_test	14.6747	9	0.1003

Non-proportional hazards

- What does it mean when hazards are not proportional?
  - The effect of that variable changes over time
  - Note in the Cox proportional hazards model, only the baseline hazard can vary with time, not the covariates

$$h(t) = \lambda_0(t)e^{\beta_1 x_1 + \beta_2 x_2 \ldots}$$

Baseline hazard

covariates

Extended Cox regression model

- The Cox proportional hazards regression utilizes fixed covariates (e.g., measured at baseline and do not change over time)
- However, the ‘extended Cox regression model’ can be used to incorporate time-dependent covariates (covariates whose values change over time)

Extended Cox regression model

- As a simple example, we might have covariate that changes value from 0 to 1 at time *z* for each patient
- In this case the simple model of the hazard would be:

$$h(t) = \lambda_0(t)e^{\beta_1 x_1}$$

$$\begin{cases} \lambda_0(t) & \text{before time } z \\ \lambda_0(t)e^{\beta_1 x_1} & \text{at or after time } z \end{cases}$$

References

- Columbia University Mailman School of Public Health. Population Health Methods. Time to event data analysis. https://www.mailman.columbia.edu/research/population-health-methods/time-event-data-analysis

- George H. Dunteman & Moon-Ho R. Ho. 2011. Survival Analysis. *In*, An Introduction to Generalized Linear Models. SAGE Publications, Inc.

- Krall, J. M., Uthoff, V. A., and Harley, J. B. 1975. A Step-up Procedure for Selecting Variables Associated with Survival. *Biometrics* 31: 49–57.

- McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. Chapman & Hall.

- O'Quigley, J., 2008. *Proportional hazards regression* (Vol. 542). New York: Springer.

- Sainani, K.L. Introduction to Survival Analysis. Stanford University Department of Health Research and Policy. https://web.stanford.edu/~kcobb/index.html