Stratification in the Cox model

Patrick Breheny

November 13

Introduction

- Today's topic is the use of stratification in Cox regression
- There are two main purposes of stratification:
  - It is useful as a diagnostic for checking the proportional hazards assumption
  - It offers a way of extending the Cox model to allow for non-proportionality with respect to some covariates

VA Lung Cancer data

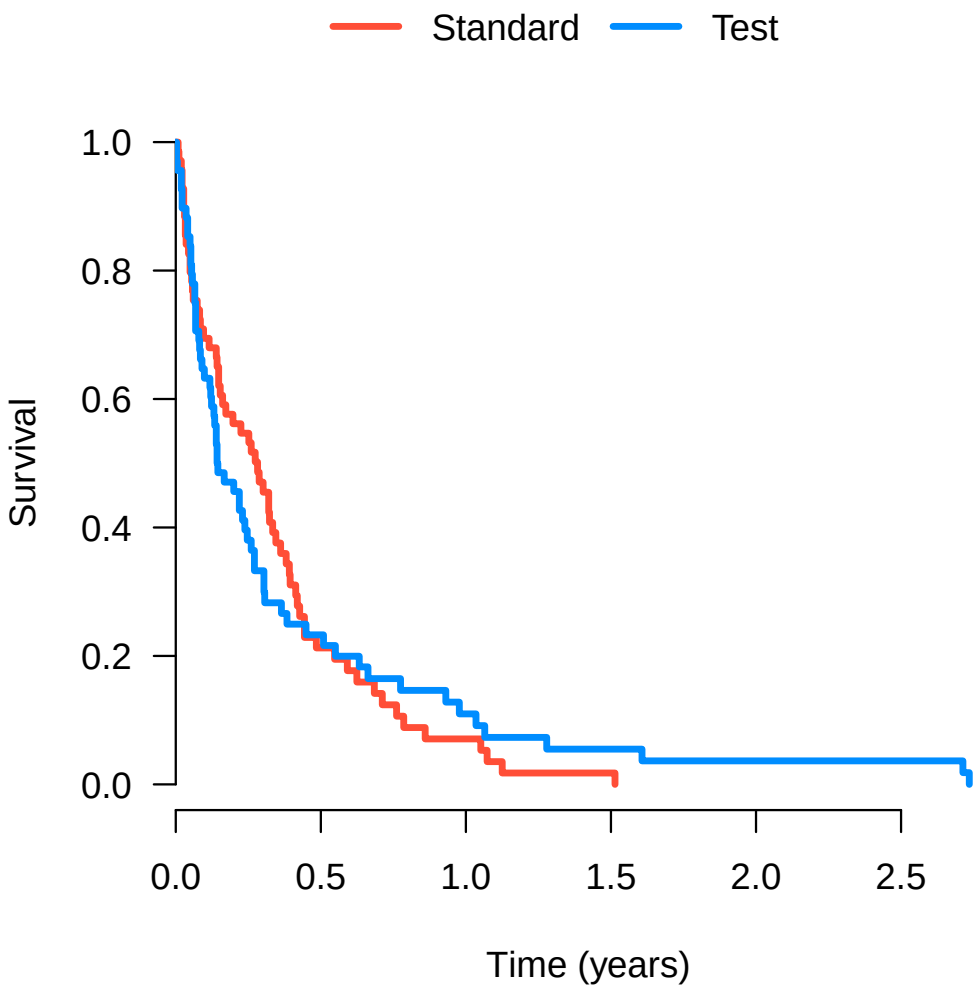
- To illustrate these concepts, we will look at a classic survival data set, the VA lung cancer data (`veteran` in the `survival` package)
- The data comes from a clinical trial carried out by the Veterans' Administration on male veterans with advanced, inoperable lung cancer
- In the trial, patients were randomized to receive either a standard chemotherapy or an experimental chemotherapy, and the primary endpoint was the time until death

Covariates

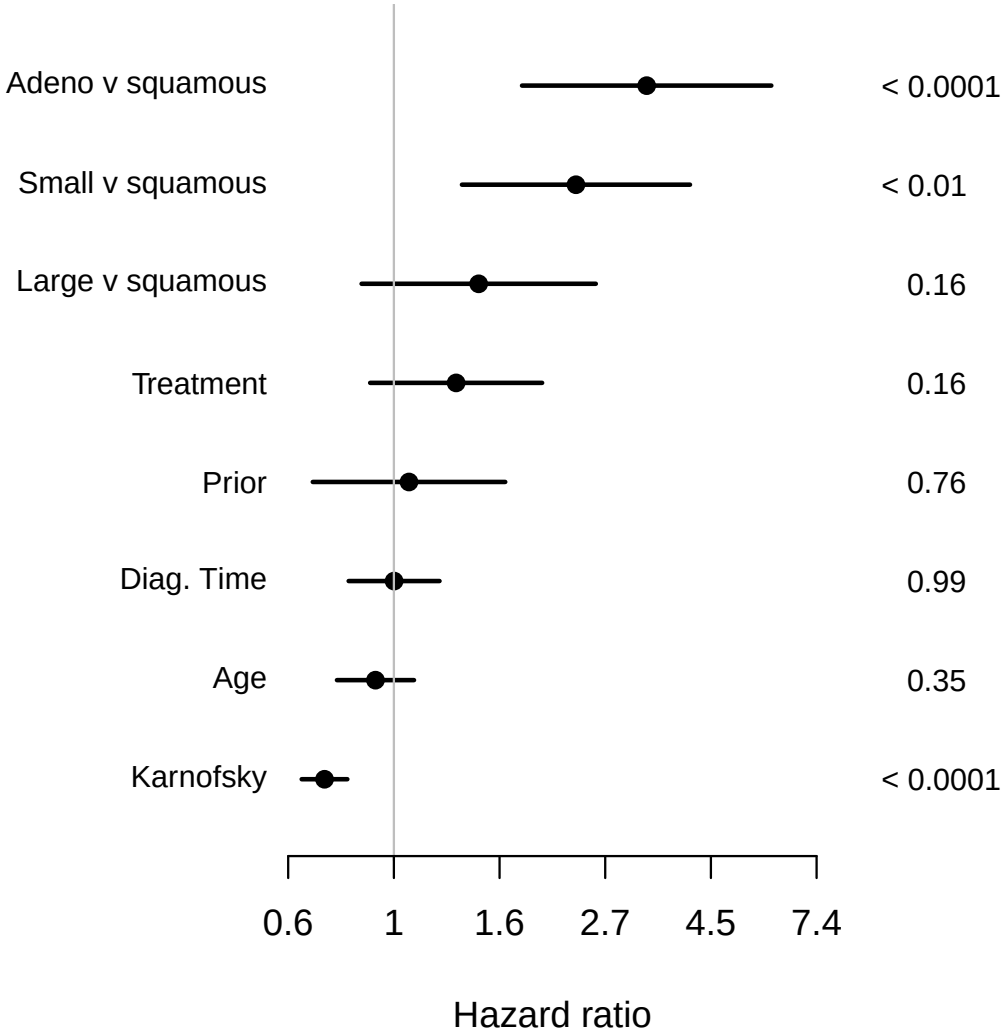
A number of covariates which potentially affect survival were also recorded:

- `karno`: The Karnofsky score, a way of quantifying the patient's overall baseline status, with ≥ 70 denoting that the patient is able to care for themselves, $40 - 60$ meaning that the patient requires assistance and regular medical care, and $10 - 30$ meaning that the patient is hospitalized
- `diagtime`: Time in months from diagnosis to randomization
- `age`: Age in years at randomization
- `prior`: Indicator for whether the patient had received prior therapy
- `celltype`: Type of tumor (small cell, large cell, squamous, adenocarcinoma)

Kaplan-Meier



Cox results


Adeno v squamous
Small v squamous
Large v squamous
Treatment
Prior
Diag. Time
Age
Karnofsky
< 0.0001
< 0.01
0.16
0.16
0.76
0.99
0.35
< 0.0001
0.6
1
1.6
2.7
4.5
7.4
Hazard ratio

Diagnostics for proportional hazards

- Consider the following as a way to assess the proportional hazards assumption: rather than including a term in the model as a covariate, we will estimate separate baseline hazards $\hat{\Lambda}_{01}$, $\hat{\Lambda}_{02}$, ..., for each level of the covariate
- If the baseline hazards appear proportional, then it is reasonable to model the term in the regular manner

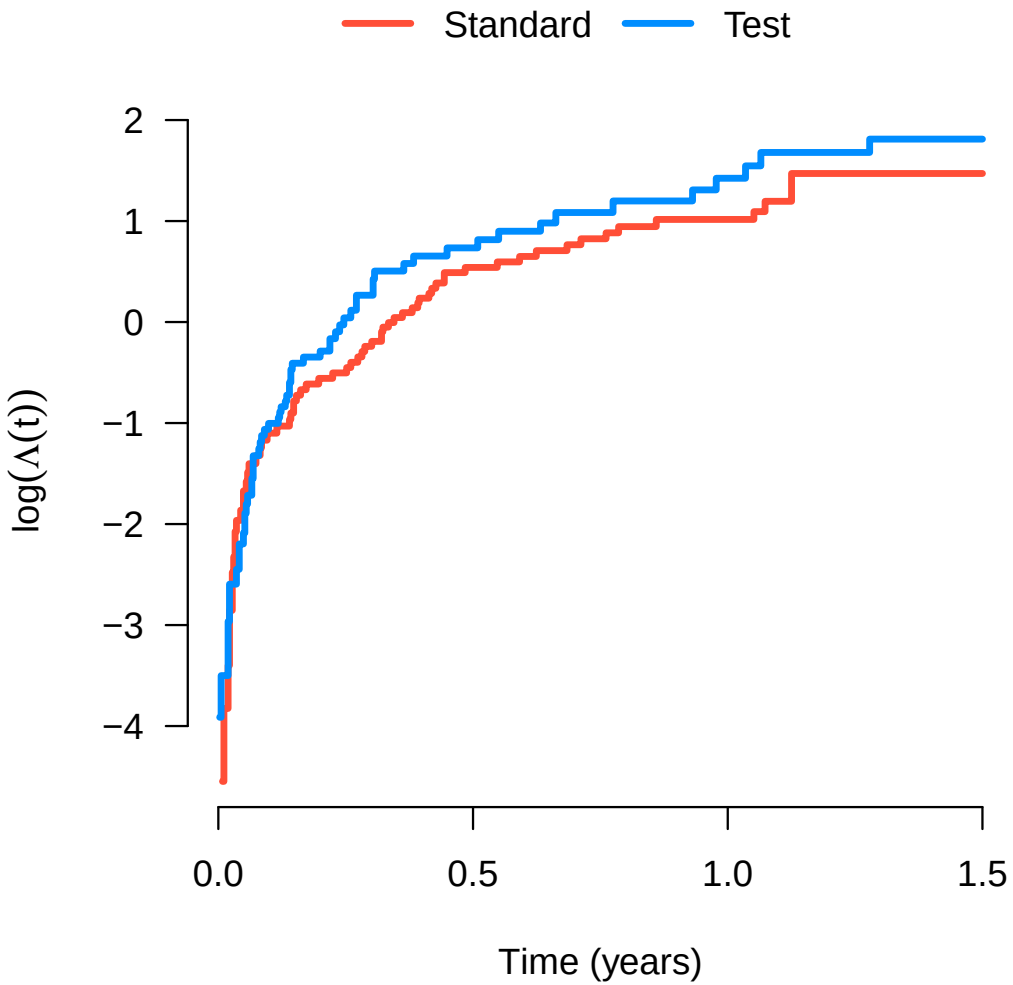
Diagnostic plot types

- Because proportionality is difficult to assess by visual inspection, it is common to plot $\log \hat{\Lambda}_0$:

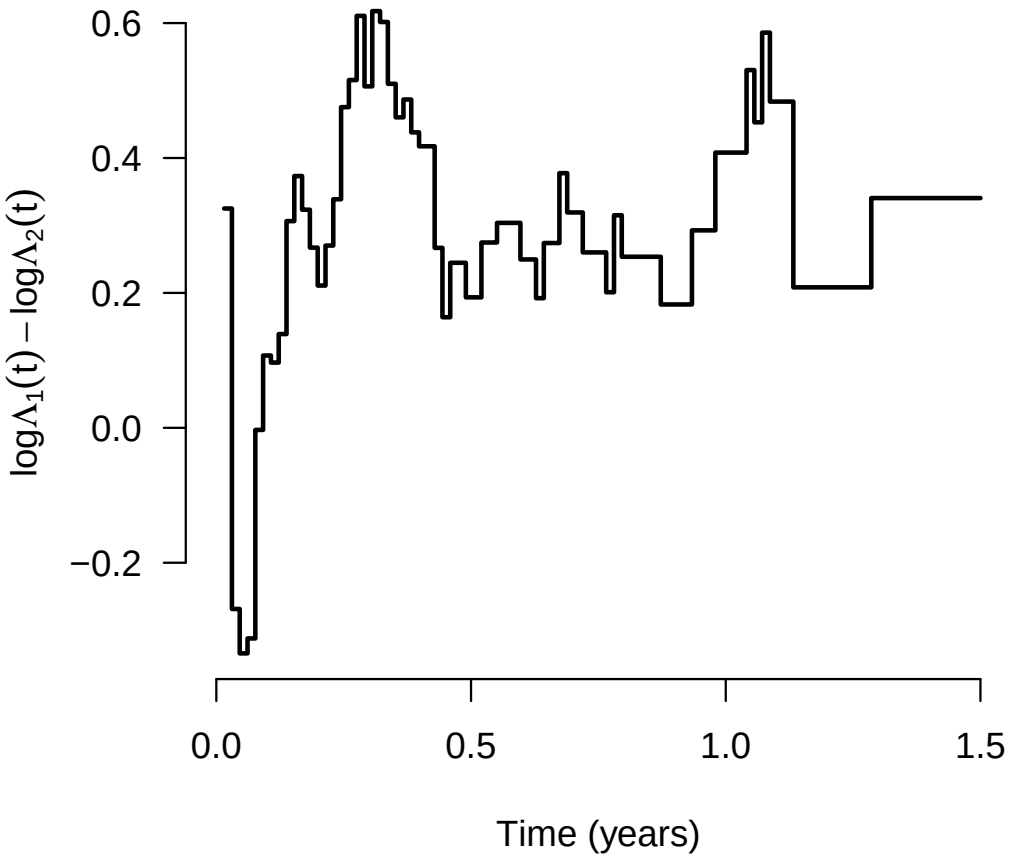
$$\begin{aligned} \Lambda_i(t) &= \Lambda_0(t) \exp(\eta_i) \\ \implies \log \Lambda_i(t) &= \log \Lambda_0(t) + \eta_i \end{aligned}$$

- An alternative, known as the *Andersen plot*, is to plot $\hat{\Lambda}_{01}$ versus $\hat{\Lambda}_{02}$; under proportional hazards this should be a straight line with slope $\exp(\eta)$
- The most common approach, however, for assessing proportional hazards is using Schoenfeld residuals, which we will discuss after Thanksgiving

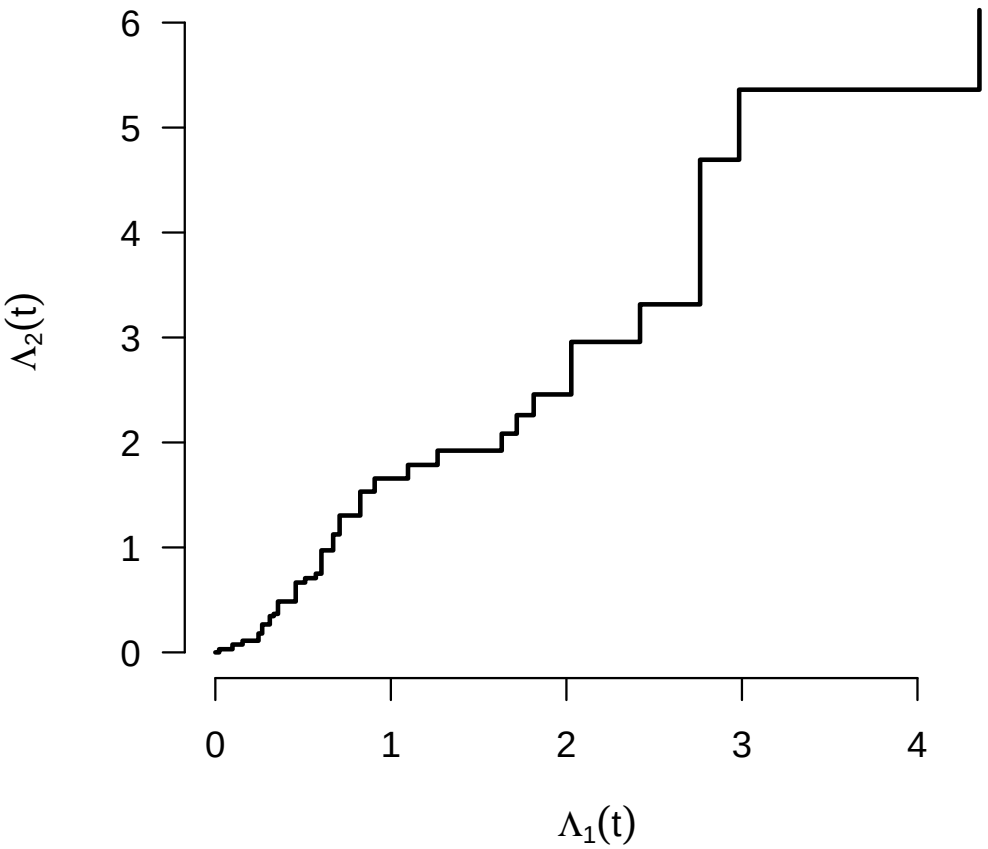
Treatment (Version 1)



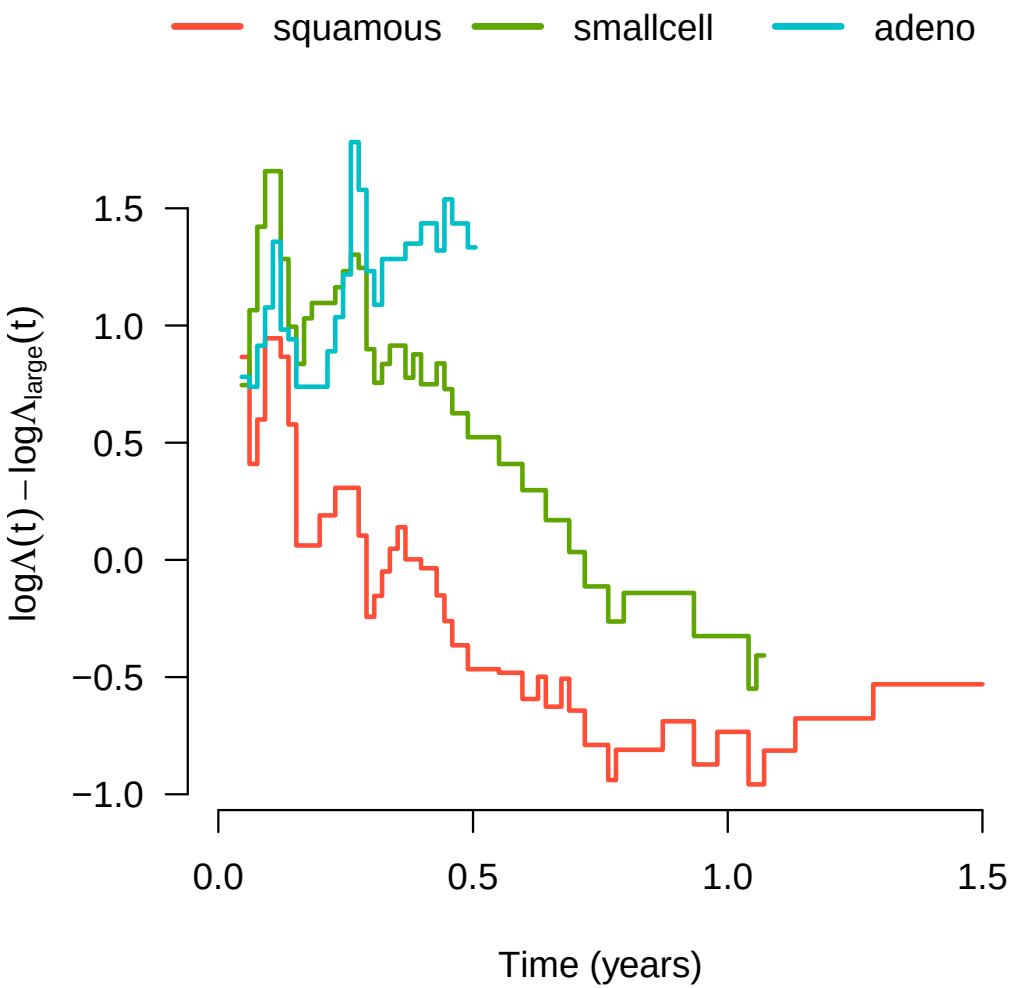
Treatment (Version 2) ($\widehat{\beta} = 0.29$)



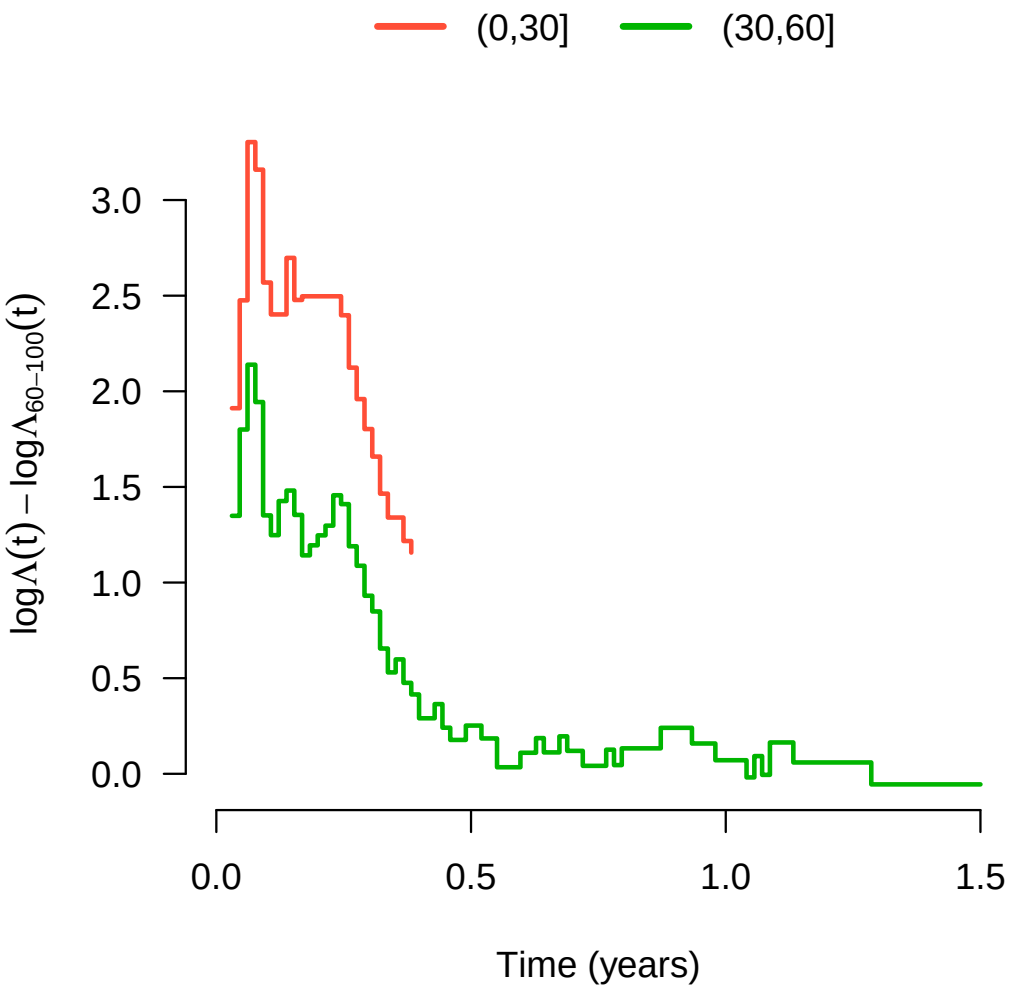
Treatment (Version 3, the Andersen plot)



Cell type


squamous
smallcell
adeno
1.5
1.0
0.5
0.0
−0.5
−1.0
log Λ(t) − log Λ_large(t)
0.0
0.5
1.0
1.5
Time (years)

Karnofsky


(0,30]
(30,60]
3.0
2.5
2.0
1.5
1.0
0.5
0.0
$\log\Lambda(t) - \log\Lambda_{60-100}(t)$
0.0
0.5
1.0
1.5
Time (years)

Remarks

- Treatment appears broadly proportional except for very short-term survival
- Proportional hazards appears questionable with respect to cell type
- Karnofsky status also appears non-proportional, with the variable losing relevance over time (which makes sense)

The stratified Cox model

- What should we do in the presence of variables with non-proportional effects?
- One remedy is to allow for different baseline hazards for each level of the variable:

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta}),$$

  where $\lambda_{ij}(t)$ is the hazard function for the ith subject, who belongs to the jth stratum
- The model may seem complex, but is entirely straightforward in the likelihood framework, as we can simply combine likelihoods across strata:

$$L(\boldsymbol{\beta}) = \prod_j L_j(\boldsymbol{\beta})$$

Stratified Cox model: Details

Furthermore,

$$\ell(\boldsymbol{\beta}) = \sum_j \ell_j(\boldsymbol{\beta})$$
$$\mathbf{u}(\boldsymbol{\beta}) = \sum_j \mathbf{u}_j(\boldsymbol{\beta})$$
$$\mathbf{I}(\boldsymbol{\beta}) = \sum_j \mathbf{I}_j(\boldsymbol{\beta}),$$

so estimation, the Newton-Raphson algorithm, and inference are all straightforward as well: we simply have to sum the contributions from each stratum

R code

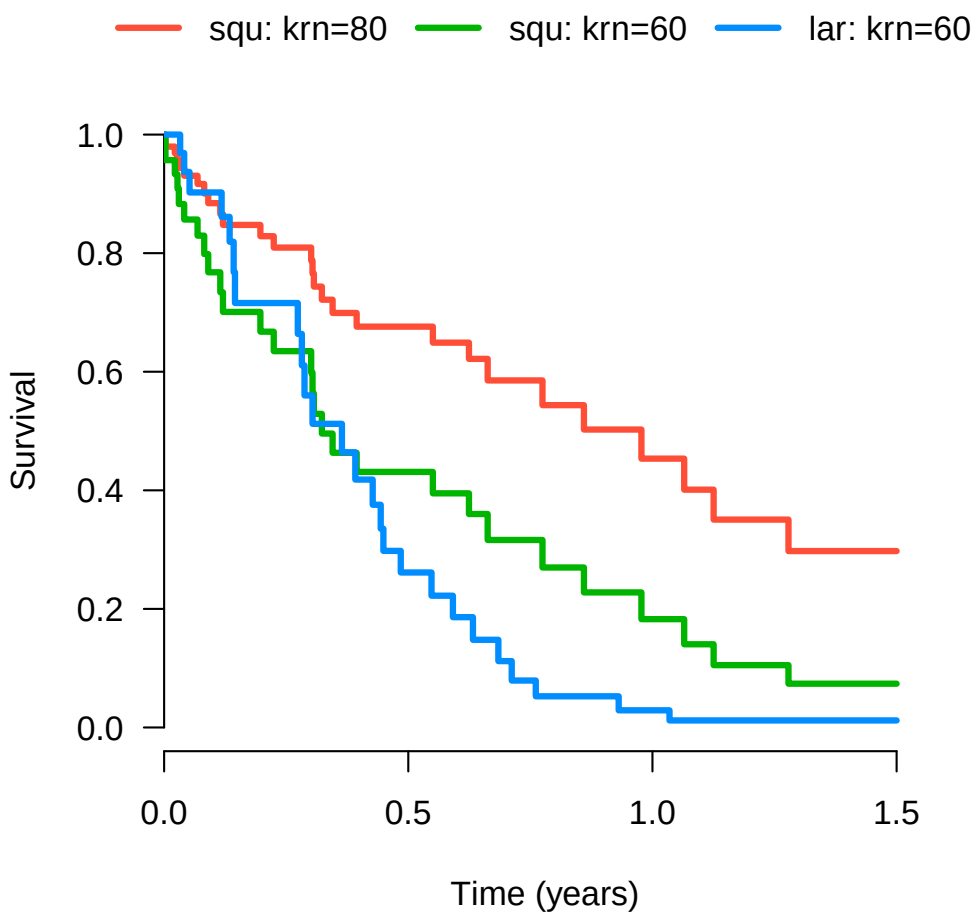
- The `survival` package makes it easy to fit stratified Cox models through the use of the `strata` function:

```
fit <- coxph(S ~ trt + karno + ... + strata(celltype))
```

- `summary(fit)` will then provide a summary for all the parametric terms (`trt`, `karno`, ...), but not `celltype`
- `survfit(fit)` will estimate K different baseline hazard functions, one for each stratum (here, $K = 4$)

Predictions

Standard treatment, wait 12 months, age 40, no prior treatment



Final remarks

- Stratified Cox models are a useful extension of the standard Cox models to allow for covariates with non-proportional hazards
- A minor drawback is that stratifying unnecessarily (i.e., even though the PH assumption is met) reduces estimation efficiency, although the loss is typically small
- A larger limitation of stratification is that it becomes messy with continuous variables and with multiple stratification variables, as there is no way to impose an additive structure

Final remarks (cont'd)

- Stratification is most commonly used when one is not interested in carrying out inference concerning the stratification variables; for example, when aggregating results across multi-center studies, because comparing these sites is typically not of interest
- Stratification is less useful in dealing with non-proportionality with respect to treatment – we are definitely interested in estimating the effect of treatment, and although we can estimate baseline coefficients, inference is not straightforward
- An alternative, which we will discuss next week, is to directly model the changing effect of the predictor over time