Flexible parametric alternatives to the Cox model

Paul Lambert[1,2], Patrick Royston[3]

[1]Department of Health Sciences, University of Leicester, UK
[2]Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Sweden
[3]MRC Clinical Trials Unit, London

pr@ctu.mrc.ac.uk

11 September 2009

Motivation

- "The success of Cox regression has perhaps had the unintended side-effect that practitioners too seldom invest efforts in studying the baseline hazard. . .

Motivation

- "The success of Cox regression has perhaps had the unintended side-effect that practitioners too seldom invest efforts in studying the baseline hazard. . .
- . . . a parametric version [of the Cox model], ... if found to be adequate, would lead to more precise estimation of survival probabilities and ... concurrently contribute to a better understanding of the phenomenon under study. " (Hjort 1992)

Sir David Roxbee Cox (b. 15 July 1924)

Taken at IBC, 2008

Problems with the Cox model

- Discards information on the *survival distribution*
- Difficult to visualize the *hazard function*
- Problem of non-proportional hazards (non-PH)—modelling to deal with non-PH can be complex
- May want a *completely specified probability model*
    - e.g. for prediction, simulation or model validation

Quote from Sir David Cox (Reid, 1994)

Reid "What do you think of the cottage industry that's grown up around [the Cox model]?"

Cox "In the light of further results one knows since, I think I would normally want to tackle the problem parametrically. ... I'm not keen on non-parametric formulations normally."

Reid "So if you had a set of censored survival data today, you might rather fit a parametric model, even though there was a feeling among the medical statisticians that that wasn't quite right."

Cox "That's right, but since then various people have shown that the answers are very insensitive to the parametric formulation of the underlying distribution. And if you want to do things like predict the outcome for a particular patient, it's much more convenient to do that parametrically."

Outline (part 1—Patrick)

- Motivation for flexible parametric proportional hazards models
- Very brief introduction to Royston-Parmar models
    - Generalization of flexible parametric PH models
- Smoothing baseline distribution functions with splines

Outline (part 2—Paul)

- The `stpm2` command
- Why we need flexible models
- Applications of `stpm2`

Flexible parametric proportional hazards models I

- Start with about the simplest survival distribution: exponential

$$S(t) = \exp(-\lambda t)$$

- Transform to log cumulative hazard scale:

$$\begin{aligned} \ln H(t) &= \ln\left[-\ln S(t)\right] \\ &= \ln(\lambda t) = \ln \lambda + \ln t \end{aligned}$$

- We see that $\ln H(t)$ is a linear function of $\ln t$ with slope 1

Flexible parametric proportional hazards models II

- Generalize $\ln H(t)$ to a linear function of $\ln t$ with slope γ:

$$\ln H(t) = \ln \lambda + \gamma \ln t$$

- This is a Weibull distribution. Now add covariates, $\mathbf{x}$:

$$\ln H(t|\mathbf{x}) = \ln \lambda + \gamma \ln t + \mathbf{x}\beta$$

- Further generalize the $\ln t$ part to allow greater flexibility:

$$\ln H(t|\mathbf{x}) = \ln \lambda + s(\ln t) + \mathbf{x}\beta$$

- Here, $s(\ln t)$ is a suitable smooth function of $\ln t$

Flexible parametric proportional hazards models III

- We actually use a *restricted cubic spline* function for $s(\ln t)$
  - more coming on splines shortly
- Let's write the model more generally as "log cumul. hazard = log baseline cumul. hazard + covariate effect":

$$\ln H(t|\mathbf{x}) = \ln H_0(t) + \mathbf{x}\beta$$

- The baseline log cumulative hazard is a smooth function of $\ln t$:

$$\begin{aligned} \ln H_0(t) &= \ln H(t; \mathbf{x} = 0) \\ &= \ln \lambda + s(\ln t) \end{aligned}$$

Why model on the log cumulative hazard scale? I

- We are used to the log hazard scale—why change to the log cumulative hazard scale?
- Under proportional hazards, covariate effects are proportional on the log cumulative hazard scale AND on the log hazard scale:

$$\begin{aligned} \ln H(t|\mathbf{x}) &= \ln H_0(t) + \mathbf{x}\beta \\ \therefore H(t|\mathbf{x}) &= H_0(t)\exp(\mathbf{x}\beta) \\ \therefore h(t|\mathbf{x}) &= h_0(t)\exp(\mathbf{x}\beta) \end{aligned}$$

Why model on the log cumulative hazard scale? II

- The log cumulative hazard as a function of log time is often simple
  - in the Weibull model, it is a straight line
- It's easier to capture the shape of simple functions than complicated ones
  - e.g. the log hazard function is usually more complicated than the log cumulative hazard
  - you can too easily get 'wiggly' fitted hazard functions

Royston-Parmar models

A further generalization!

- We can generalize the model for the log cumulative hazard function by going back a few steps
  - No details given here—see Royston & Parmar (2002)
- The most important other class of Royston-Parmar models is *proportional cumulative odds* models:

$$O\left(t|\mathbf{x}\right) = O_0\left(t\right)\exp\left(\mathbf{x}\beta\right)$$

  where

$$O\left(t\right) = \left[1 - S\left(t\right)\right] / S\left(t\right)$$

  is the (cumulative) odds of an event occurring before time t
- Proportional odds models for survival data were first suggested by Bennett (1983)
- They sometimes give a better fit than proportional hazards models
- But we won't be discussing them much today

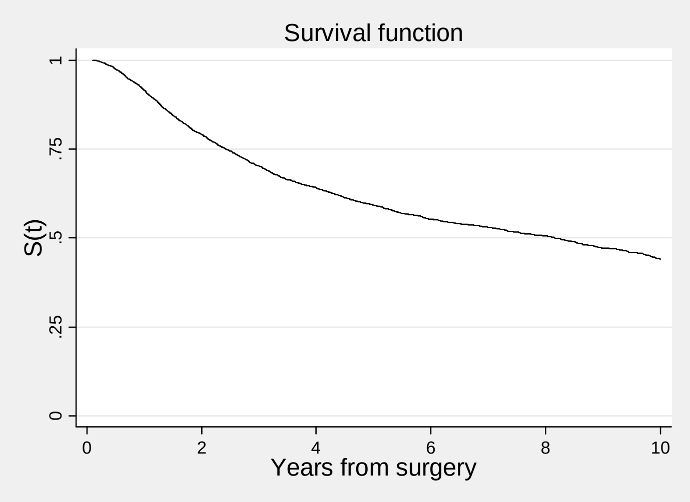
Examples of survival, cumulative hazard and cumulative odds functions in real data

The Rotterdam breast cancer dataset:

- N = 2982 patients with primary breast cancer
- 1,477 events for relapse-free survival
- Time is years since surgery to remove the primary tumour
- Data restricted to first 10 years of follow-up

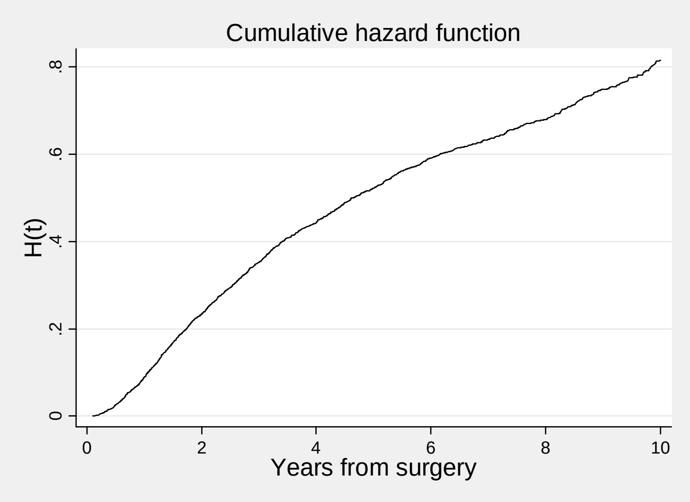
Survival function (Kaplan-Meier)

```
sts gen s = s
```



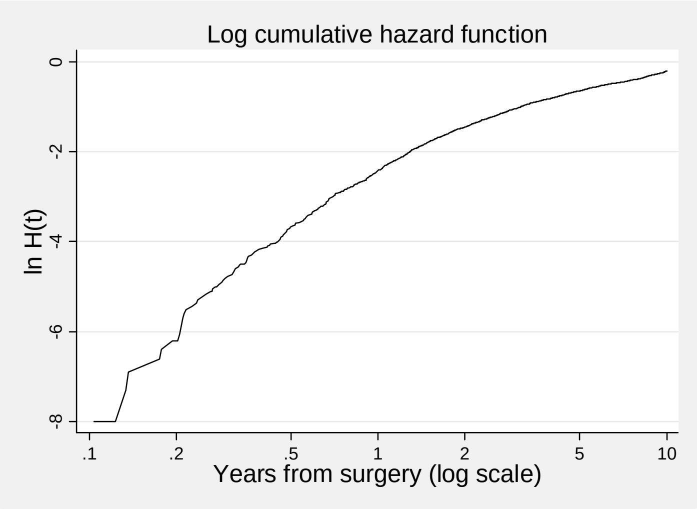
Cumulative hazard function (Nelson-Aalen)

```
sts gen H = na
```



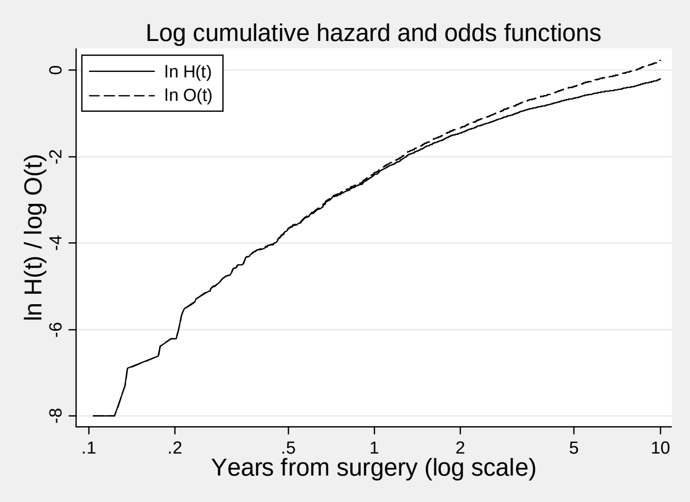
Log cumulative hazard function

```
gen lnH = ln(H)
```



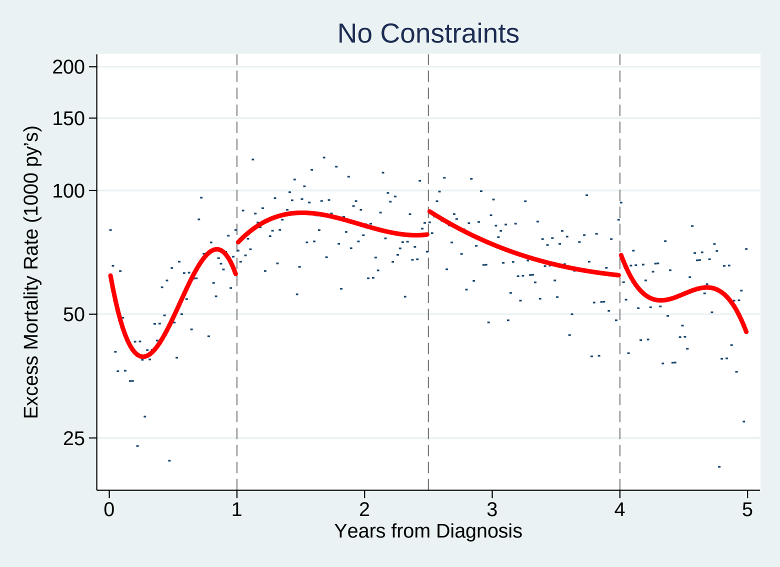
Log cumulative hazard and odds functions

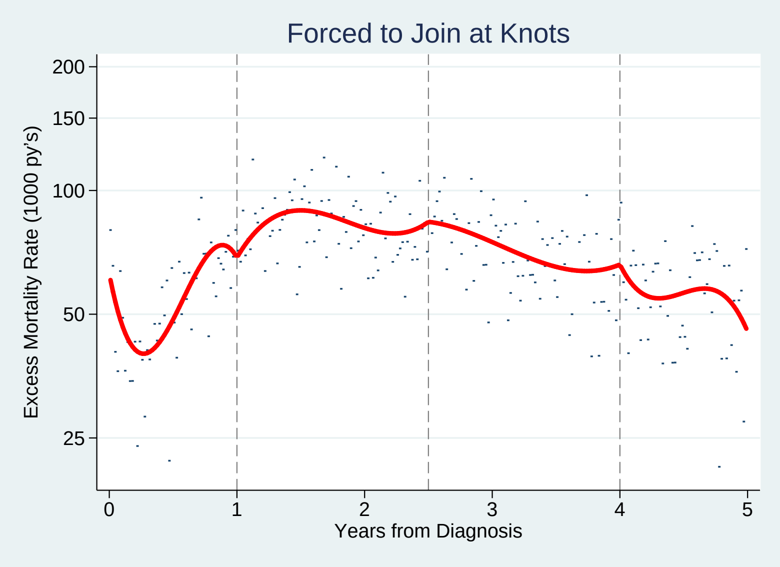
```
gen lnO = ln(1 - s) / s
```

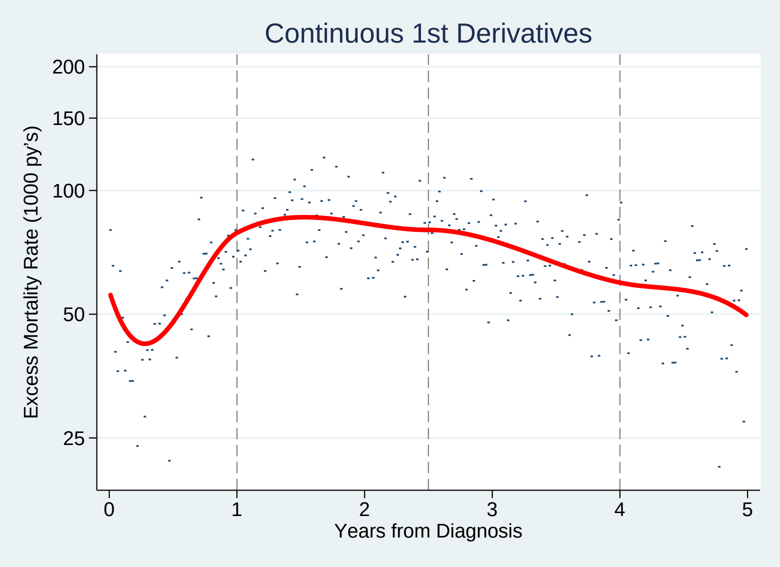


Spline functions: A brief description

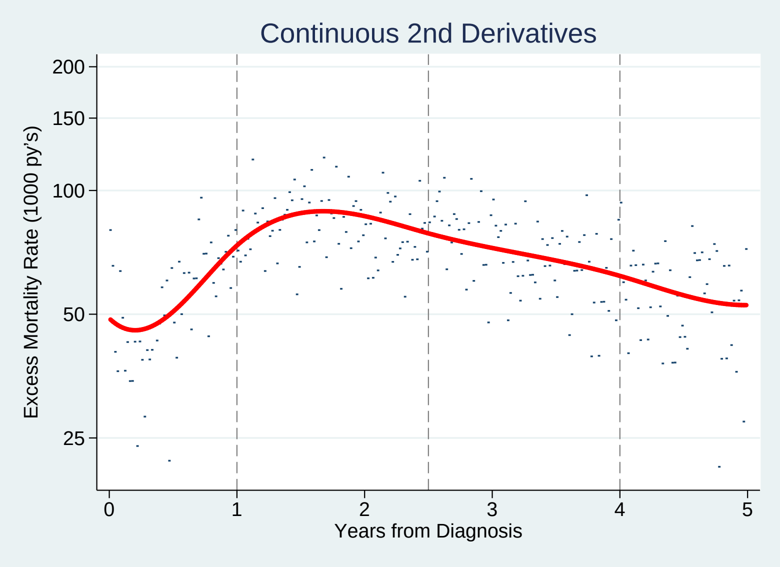
- Splines are flexible mathematical functions defined by piecewise polynomials joined at points on the x axis known as *knots*
- Regression splines are particularly useful, because
  - they can be incorporated into any regression model which has a linear predictor
  - they are relatively simple but still flexible enough for most practical data
- Cubic regression splines can be constrained in different ways to improve their smoothness
  - e.g. the fitted function is forced to have continuous 0th, first and second derivatives at the knots
- The following graph shows how smoothness improves as the curve and its derivatives are more and more constrained


No Constraints
Excess Mortality Rate (1000 py's)
200
150
100
50
25
0
1
2
3
4
5
Years from Diagnosis


Forced to Join at Knots
Excess Mortality Rate (1000 py's)
200
150
100
50
25
0
1
2
3
4
5
Years from Diagnosis



Continuous 1st Derivatives
Excess Mortality Rate (1000 py's)
200
150
100
50
25
0
1
2
3
4
5
Years from Diagnosis


Continuous 2nd Derivatives
Excess Mortality Rate (1000 py's)
200
150
100
50
25
0
1
2
3
4
5
Years from Diagnosis

Restricted cubic splines I

- A further useful constraint is to force the fitted function to be linear in the tails
  - i.e. beyond the lowest and highest knots—known as boundary knots
- This stabilizes the function in the two regions which usually have the sparsest data
- These splines are known as 'restricted' (or 'natural') cubic splines—RCS
- We use RCS in Royston-Parmar models to smooth the baseline distribution function

Restricted cubic splines II

- An RCS with K interior knots and 2 boundary knots has $K + 1$ d.f. and $K + 1$ *basis functions*:

$$s(x) = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_{K+1} z_{K+1}$$

  where z_1 equals x and the other z's are simple functions of the knots and x
- If $K = 0$ the spline simplifies to the linear function $\gamma_0 + \gamma_1 x$
  - no boundary knots, of course
- Will shortly return to the question of how many knots to choose in the Stata demonstration

But how do I do any of this in Stata?

Let's visit Stata now and see how to fit some simple flexible proportional hazards models to the Rotterdam breast cancer data

We use Paul's powerful `stpm2` program to do the model-fitting, and its `predict` function to get the outputs we need.

Notes on stpm2

- `stpm2` can fit a wide variety of types of model—Paul will demonstrate some of them shortly
- Particularly strong are its `prediction` facilities:
  - Survival, cumulative hazard, hazard, cumulative odds, ...
  - All with pointwise confidence intervals if desired (`ci` option)
  - Can handle important derived quantities such as hazard ratios, hazard differences and survival differences

Now hand over to Paul to continue.