# Survival analysis:
## Penalised Cox models

**Mark Stevenson**

Faculty of Veterinary and Agricultural Sciences

The University of Melbourne, Parkville Victoria 3010 Australia

mark.stevenson1@unimelb.edu.au

# Roadmap

- Background
- Splines
- Examples:
  - clinical mastitis at calving in New Zealand dairy cows
  - equine colic
  - culling and deaths in New Zealand dairy goats

# Background

- When modelling survival using CPH regression, we routinely check that each continuous explanatory variable is linear in its log hazard …

# Background

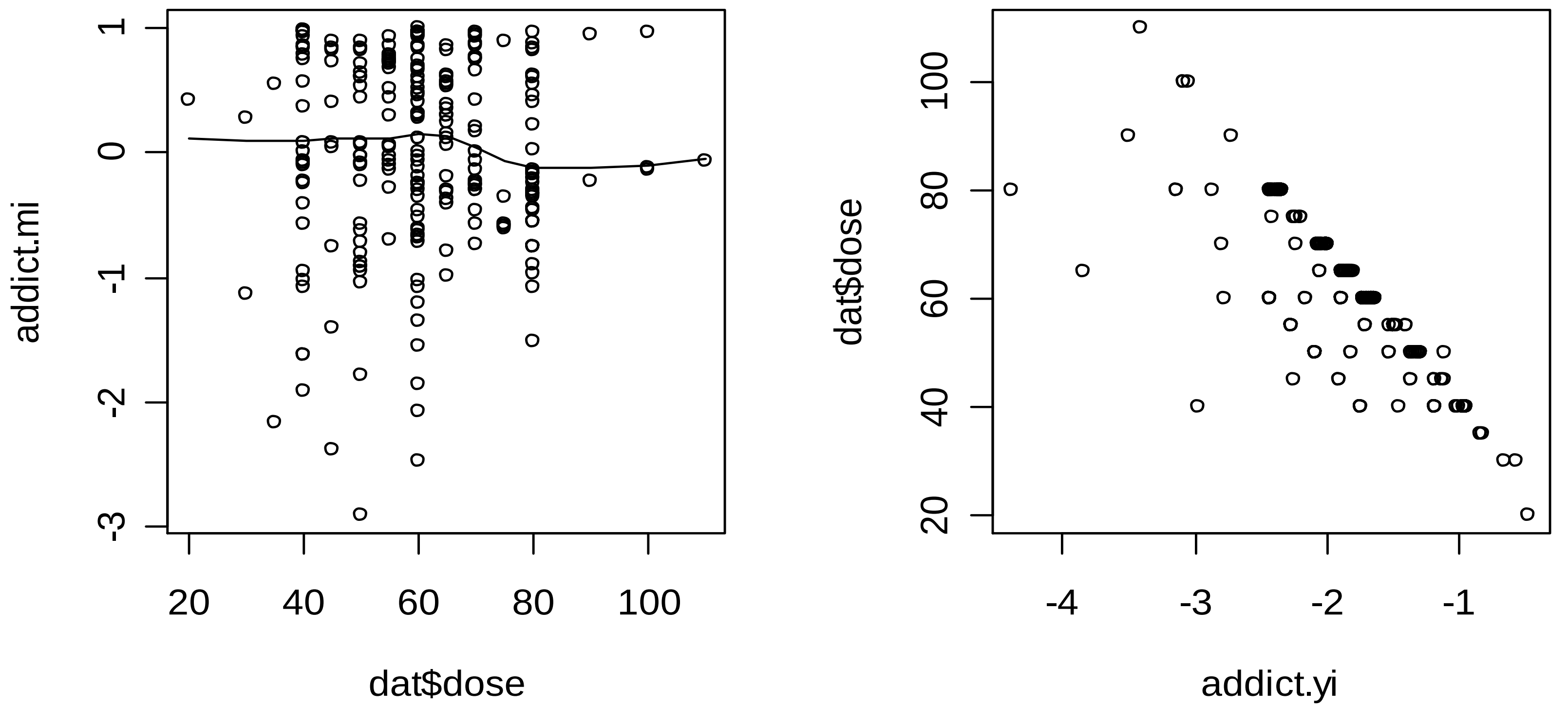
- Method 1
  - replace the continuous explanatory variable with three design variables using Q1, Q2, and Q3 as cutpoints
  - plot the estimated coefficients for the design variables versus the midpoint of the group: a fourth point is included at zero using the midpoint of the first group
  - if the correct scale is linear, then the line connecting the four points should approximate a straight line (consider transforming the continuous explanatory variable if this is not the case)

# Model building

- Method 2
  - fit the preliminary main effects model, including explanatory variable of interest (e.g. 'age')
  - save the Martingale residuals ($M_i$) from this model and calculate $H_i = ci - M_i$, where $c_i$ is the censoring variable
  - plot $c_i$ versus the covariate of interest and calculate a lowess smooth (called $c_{lsm}$)
  - plot $H_i$ versus the covariate of interest and calculate a lowess smooth (called $H_{lsm}$)
  - calculate

$$y_i = \ln\left(\frac{c_{lsm}}{H_{lsm}}\right) + \beta_{age} \times \text{age}$$

THE UNIVERSITY OF MELBOURNE




If covariate is linear in its log hazard each of these plots will follow a straight line.

# Background

- What do we do when our continuous explanatory variable is <u>not</u> linear in its log hazard?

- Two options
  - re-code the variable into categories and treat as a factor
  - parameterise the explanatory variable using smoothing functions

## Background

- Polynomials are one of the easiest smooth functions to fit: simply add $x$, $x^2$, $x^3$, ... to the right-hand side of the model equation

- Polynomials have a number of flaws however:
  - the data fits are not local
  - the fitting process for polynomials can be numerically ill-conditioned

# Background

- Penalised Cox models offer a means for dealing with non linearity in log hazards by fitting non-parametric functions (for example, spline smoothers) to account for relationships between explanatory and outcome variables

- A nice feature of this technique is that results can be displayed graphically to illustrate the multivariable functional form of these relationships (e.g. linear, quadratic or cubic)

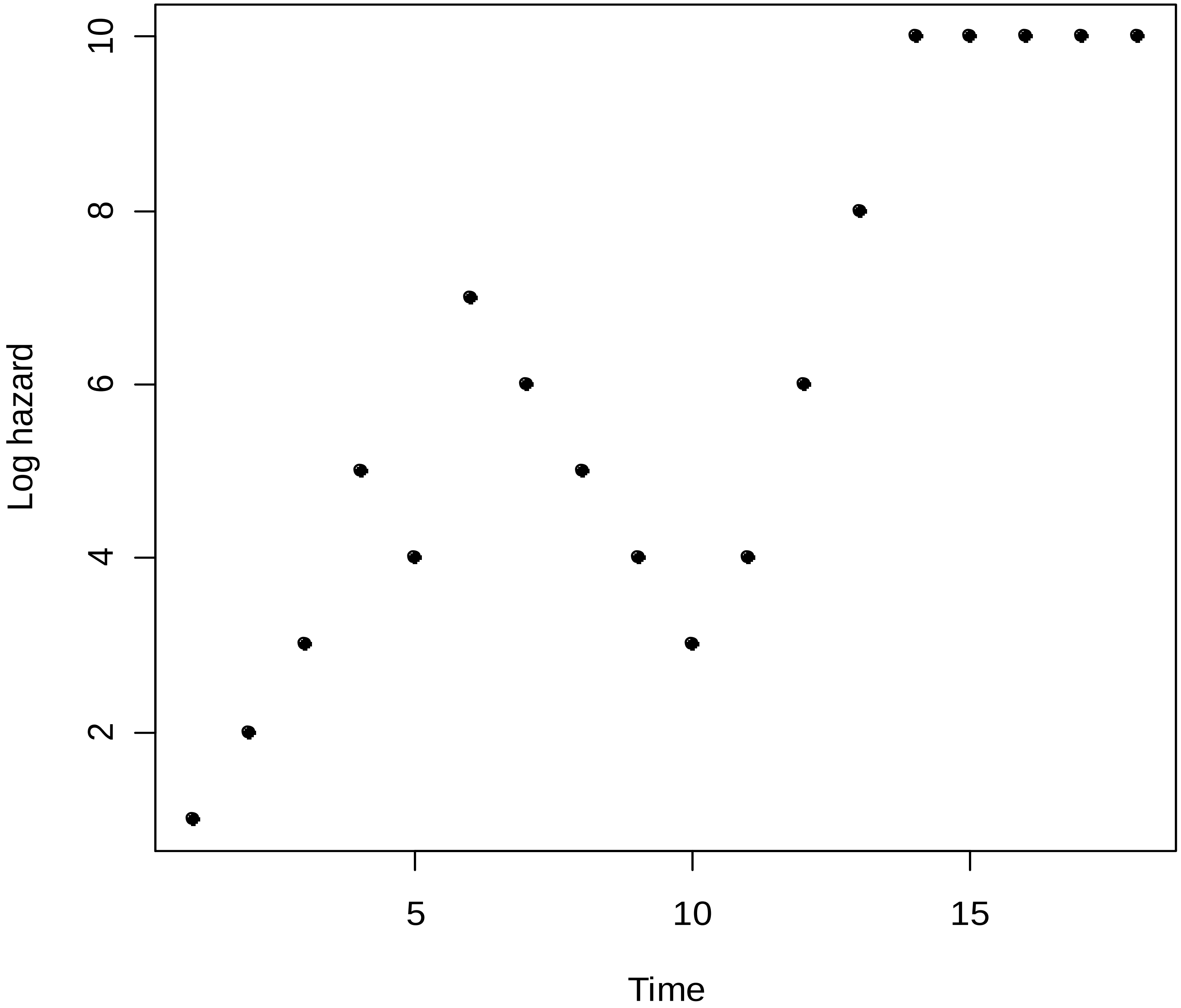
# Roadmap

- Background
- <u>Splines</u>
- Examples:
  - clinical mastitis at calving in New Zealand dairy cows
  - equine colic
  - culling and deaths in New Zealand dairy goats

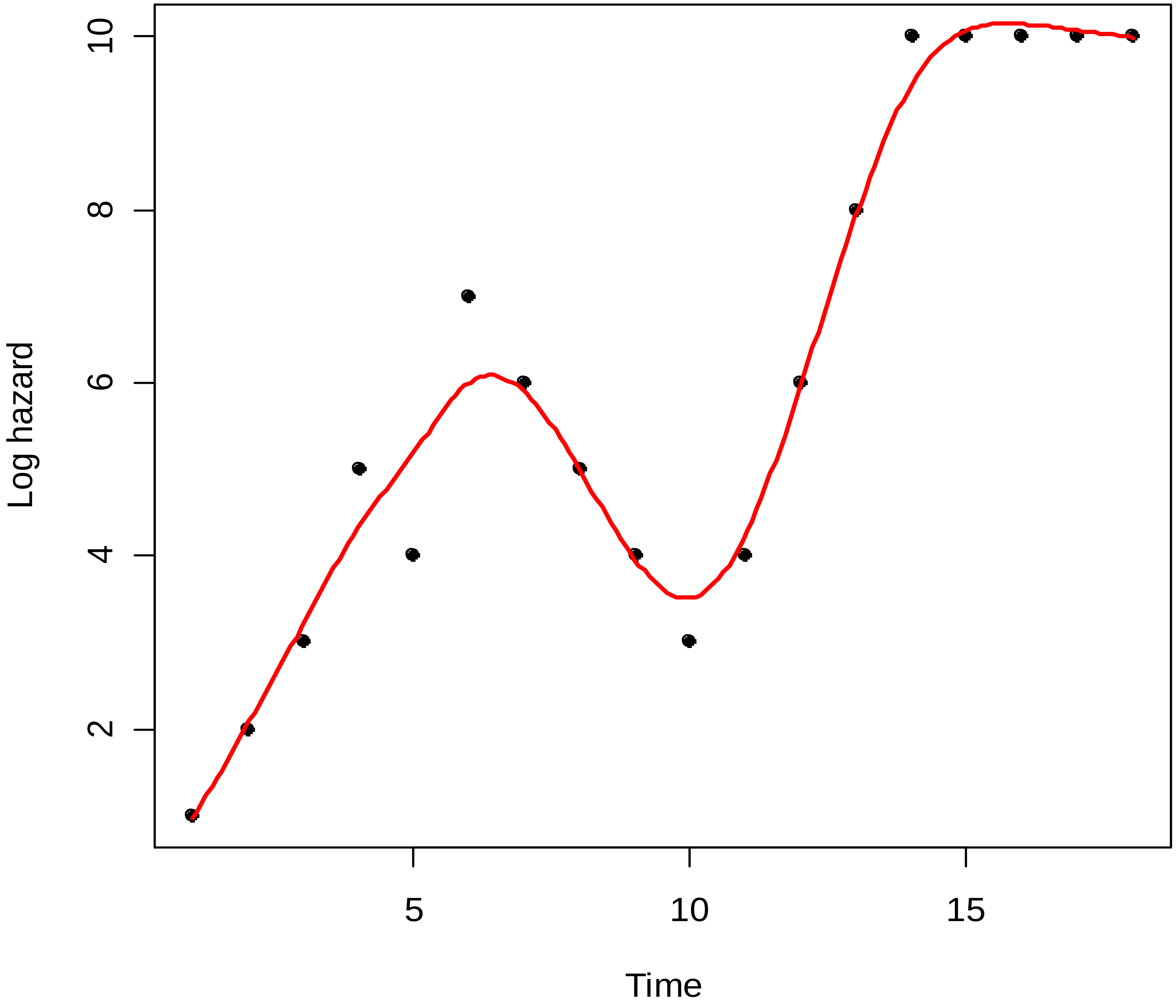
# Splines

- Think of the continuous variable as a number of nails in a wooden board
  - interpolate the values of the variable using a thin, flexible metal strip
  - the metal strip is your spline

Scatterplot of log hazard estimates for various levels of a continuously distributed variable as a function of follow up time.



Scatterplot of log hazard estimates for various levels of a continuously distributed variable as a function of follow up time. A smoothing spline has been fit, and superimposed.



# Splines

- Properties of splines
  - locality of influence
  - the curve can be constrained to be linear beyond the last control point
- Splines can be fit by creating an appropriate set of dummy variables, based on the pre-specified locations of the control points (the nails)
  - these control points are called knots and the dummy variables are known as basis functions
  - you can then regress the data on the basis of the dummy variables

# Roadmap

- Background
- Splines
- Examples:
  - clinical mastitis at calving in New Zealand dairy cows
  - equine colic
  - culling and deaths in New Zealand dairy goats