Improving Dynamic Predictions from Joint Models using Time-Varying Effects

Eleni-Rosalina Andrinopoulou & **Dimitris Rizopoulos**

July 31, 2017

Learning-Health System Prostate Biopsies

- Screening has resulted in an increase in the number of newly diagnosed prostate cancers

- Up to 80% of men with PSA screen-detected prostate cancer are over-diagnosed

- Current treatments have a number of side effects
    - intervention should be restricted to those who need it

PRIAS Study

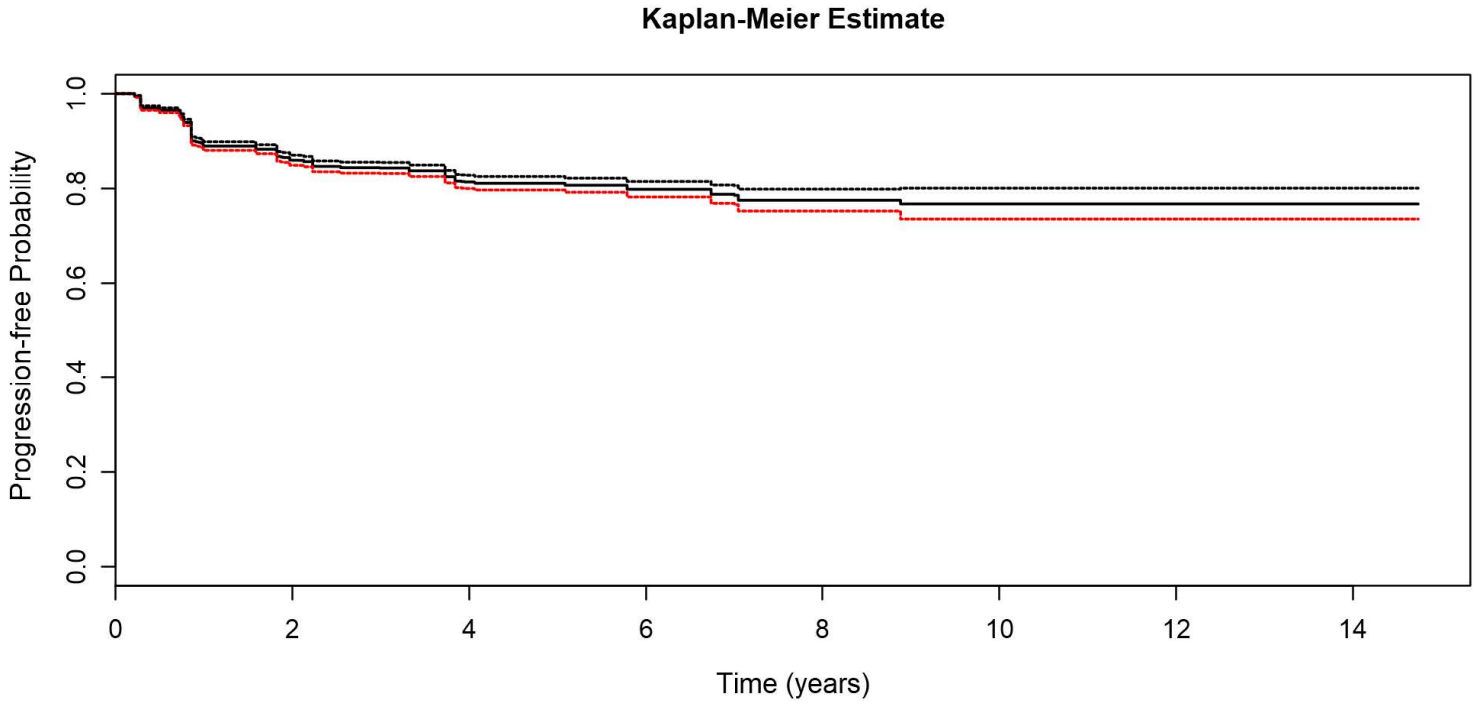
- A program in which men with early prostate cancer are managed by a follow-up strategy
  - biopsies at baseline, 1, 4, 7 and 10 years
  - or yearly after PSA doubling within a year

- Outcomes of interest:
  - time to Gleason score reclassification (from 6 to ≥ 7)
  - longitudinal PSA measurements

PRIAS Study (cont'd)

Outcome:

survival



PRIAS Study (cont'd)

- Research Questions:
  - *How the longitudinal PSA profiles are related to Gleason Score reclassification?*
  - *How to derive dynamic predictions of progression probabilities?*
  - How to optimally plan biopsies?

Time-varying Covariates

- To answer these questions we need to link
  - the time to progression (survival outcome)
  - the PSA measurements (longitudinal outcome)

- Biomarkers are *endogenous* time-varying covariates
  - their future path depends on previous events
  - standard time-varying Cox model not appropriate

Time-varying Covariates (cont'd)

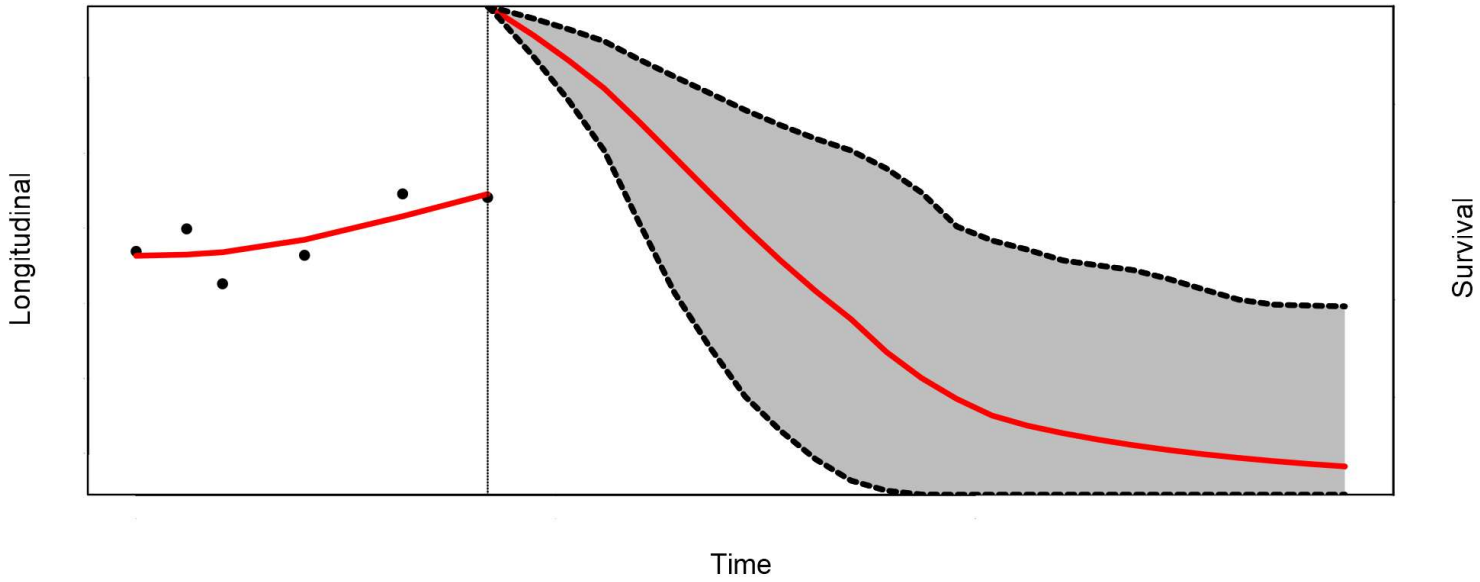
To account for endogeneity we use the framework of

Joint Models for Longitudinal & Survival Data

The Basic Joint Model

Measurements:

1 6 10



The Basic Joint Model (cont'd)

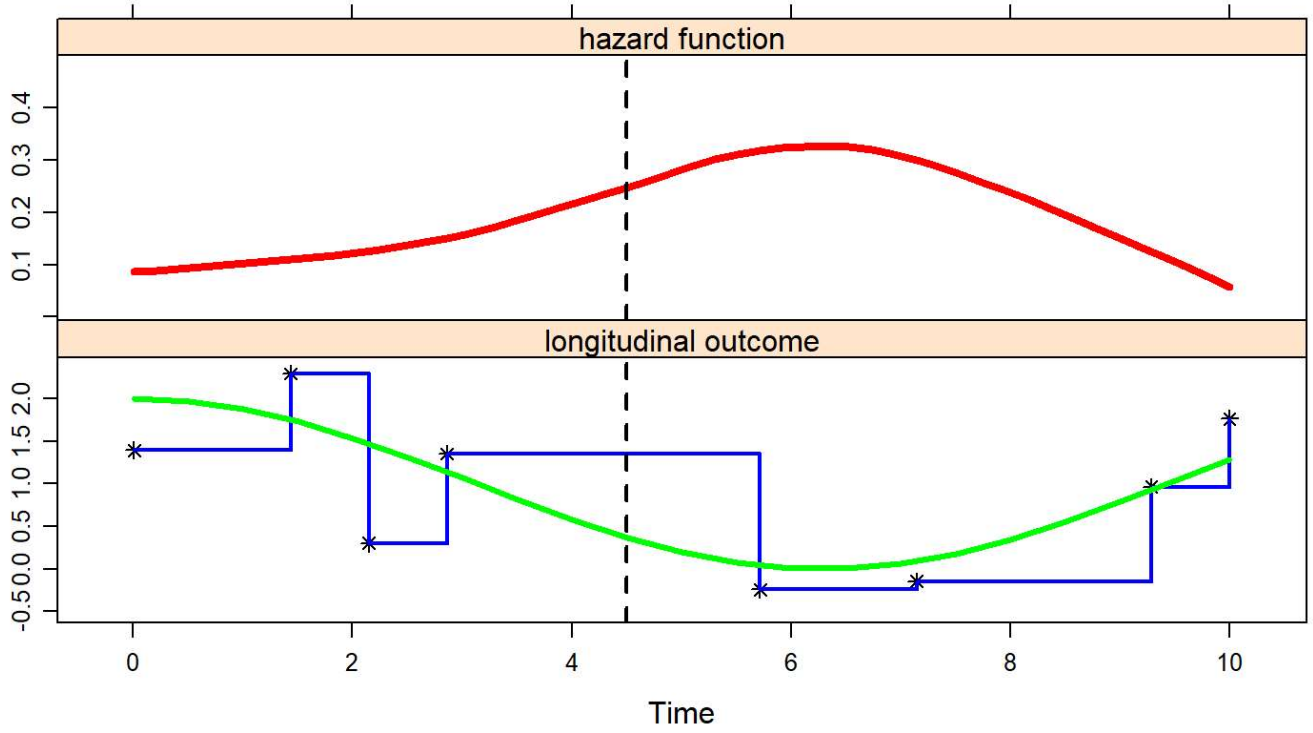
- We need some notation
  - T_i^* the true progression time
  - T_i^L last biopsy time point Gleason Score was < 7
  - T_i^R first biopsy time point Gleason Score was ≥ 7
  - $T_i^R = \infty$ for patients who haven't progressed yet
  - $\mathbf{y}_i$ vector of longitudinal PSA measurements
  - $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$

The Basic Joint Model (cont'd)

- Formally, we have

$$
\begin{cases}
h_i(t) &= h_0(t)\exp\{\gamma^\top \mathbf{w}_i + \alpha\eta_i(t)\} \\
\\
y_i(t) &= \eta_i(t) + \varepsilon_i(t) \\
&= \mathbf{x}_i^\top(t)\beta + \mathbf{z}_i^\top(t)\mathbf{b}_i + \varepsilon_i(t) \\
\\
& \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2)
\end{cases}
$$

The Basic Joint Model (cont'd)



The Basic Joint Model (cont'd)

- The longitudinal and survival outcomes are jointly modeled

$$p(y_i, T_i^L, T_i^R) = \int p(y_i \mid b_i) \times \left\{ S(T_i^L \mid b_i) - S(T_i^R \mid b_i) \right\} \times p(b_i)\, db_i$$

  - the random effects b_i explain the interdependencies

The Basic Joint Model (cont'd)

- Estimation of joint models is based on either
  - Maximum likelihood (requires numerical integration)
  - Bayesian approaches (e.g., MCMC or HMC)

- Here, we follow a Bayesian approach
  - more on this later…

Functional Form

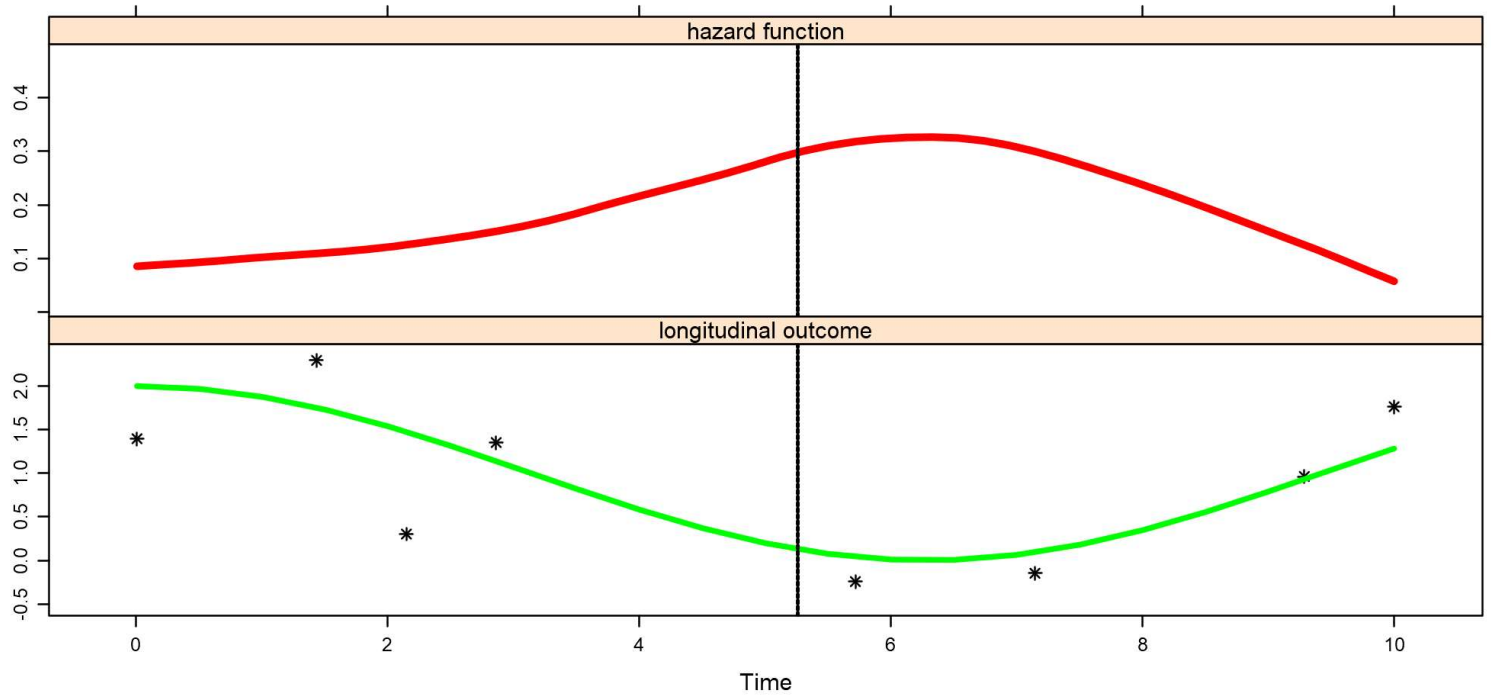
- The link between the two processes
  - the basic joint model assumes

$$\begin{cases} h_i(t) &= h_0(t)\exp\{\gamma^\top \mathbf{w}_i + \alpha\eta_i(t)\} \\ \\ y_i(t) &= \eta_i(t) + \varepsilon_i(t) \\ &= \mathbf{x}_i^\top(t)\beta + \mathbf{z}_i^\top(t)\mathbf{b}_i + \varepsilon_i(t) \end{cases}$$

Functional Form (cont'd)

times:

1 11 20



Functional Form (cont'd)

Is this the only option?

- Especially when interest
  - in studying the association structure
  - predictions

- Let's see some possibilities…

Functional Form (cont'd)

- Some options: Biomarker's rate of change
  - In prostate cancer, fast increasing PSA indicative of cancer

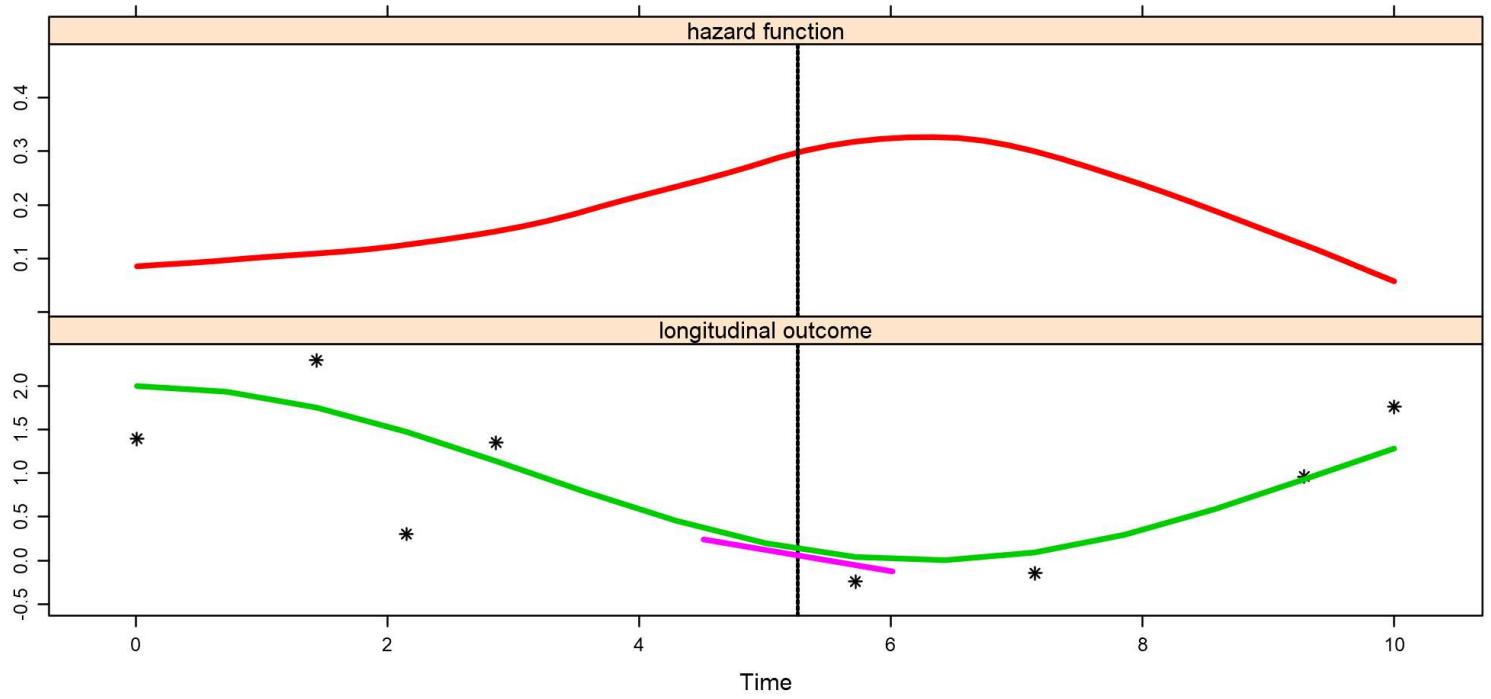
$$h_i(t) = h_0(t)\exp\{\gamma^\top \mathbf{w}_i + \alpha_1 \eta_i(t) + \alpha_2 \eta_i'(t)\}$$

where $\eta_i'(t) = \frac{d}{dt}\eta_i(t)$

Functional Form (cont'd)

times:

1 11 20



Functional Form (cont'd)

- Some options: Biomarker's cumulative effect
  - In diabetes, the accumulated HbA1c levels are related to the risk of side effects

$$h_i(t) = h_0(t) \exp\Big\{\gamma^\top \mathbf{w}_i + \alpha \int_0^t \eta_i(s) ds\Big\}$$

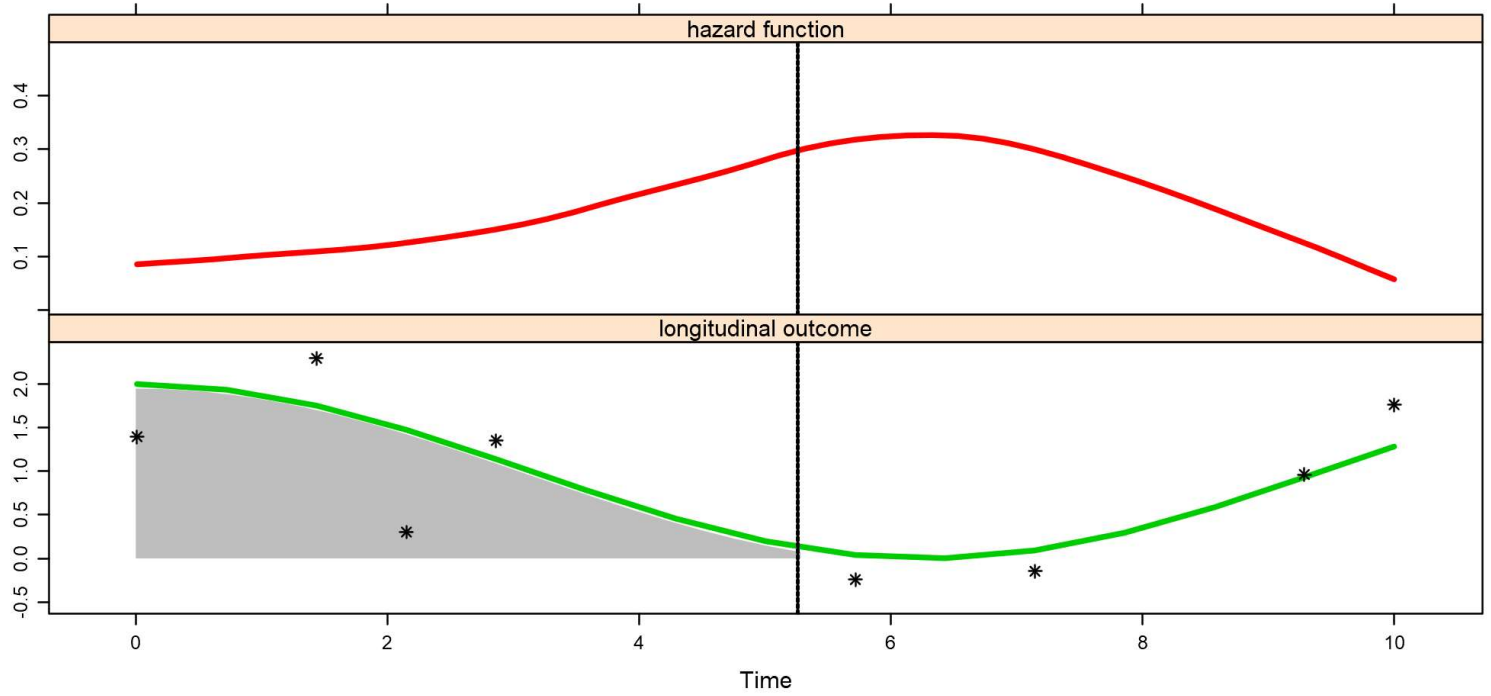
  - or even weighted cumulative effects

$$h_i(t) = h_0(t) \exp\Big\{\gamma^\top \mathbf{w}_i + \alpha \int_0^t \varpi(t-s)\eta_i(s) ds\Big\}$$

Functional Form (cont'd)

times:

1 11 20



20/32

Functional Form (cont'd)

- Previous functional forms: *Which features of the longitudinal profiles relate to the risk of Gleason reclasification?*

But is the strength of the association constant over time?

Functional Form (cont'd)

- Allowing association parameters to be time-varying

$$h_i(t) = h_0(t) \exp\left\{\gamma^\top \mathbf{w}_i + \sum_{l=1}^{L} f_l(\mathcal{H}_i(t), \alpha_l(t))\right\}$$

- $f(\cdot)$ specifies which features of the longitudinal profile enter in the linear predictor
  - value
  - slope
  - area
  - …

Functional Form (cont'd)

- The time-varying functions $\alpha_l(t)$ are approximated using B-splines

$$\sum_{k=1}^{K} \lambda_k \mathcal{B}_k(t, \mathbf{v})$$

  where

  - $\mathcal{B}_k(t, \mathbf{v})$ denotes the k-th basis function of B-spline with vector of knots $\mathbf{v}$

Functional Form (cont'd)

- To appropriately control for smoothness we use the following hierarchical prior specification

$$\lambda \sim \mathcal{N}(0, \tau_\lambda \mathbf{M})$$

$$\tau_\lambda \sim \text{inv-Gamma}(1, 0.005)$$

where

- $\mathbf{M} = \mathcal{D}_r^\top \mathcal{D}_r + 10^{-6}\mathbf{I}$
- $\mathcal{D}_r$ denotes the r-th order differences matrix

PRIAS Study Analysis

- PSA growth

$$\begin{cases} \log_2(PSA) &= \eta_i(t) + \varepsilon_i(t) \\ &= \beta_0 + \sum_{k=1}^{3} \beta_k \text{NS}_k(t, \nu) + \beta_4 \text{Age} + \beta_5 \text{Age}^2 \\ &\quad + b_{i0} + \sum_{k=1}^{2} b_{ik} \text{NS}_k(t, \nu) + \varepsilon_i(t) \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), & \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2) \end{cases}$$

PRIAS Study Analysis (cont'd)

- Risk of reclassification

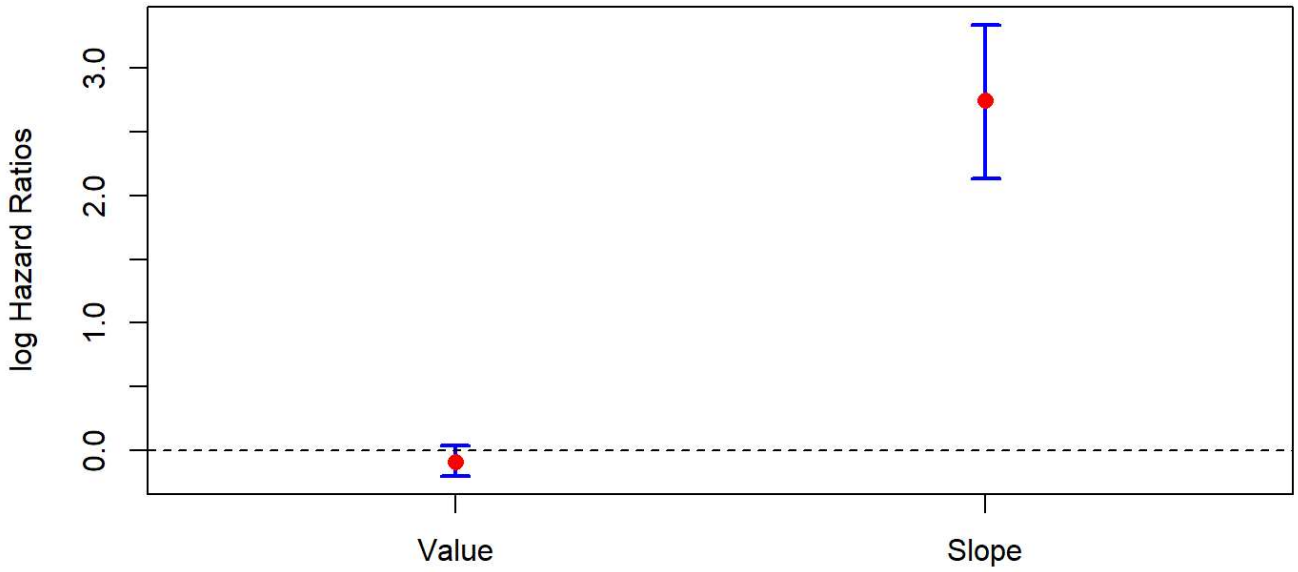
$$h_i(t) = h_0(t) \exp\left\{ \gamma_1 \text{Age} + \gamma_2 \text{Age}^2 + \alpha_1 \eta_i(t) + \alpha_2 \frac{d\eta_i(t)}{dt} \right\}$$

where

- $\eta_i(t)$ log2(PSA) current value
- $\frac{d\eta_i(t)}{dt}$ log2(PSA) velocity

PRIAS Study Analysis (cont'd)

- Results



PRIAS Study Analysis (cont'd)

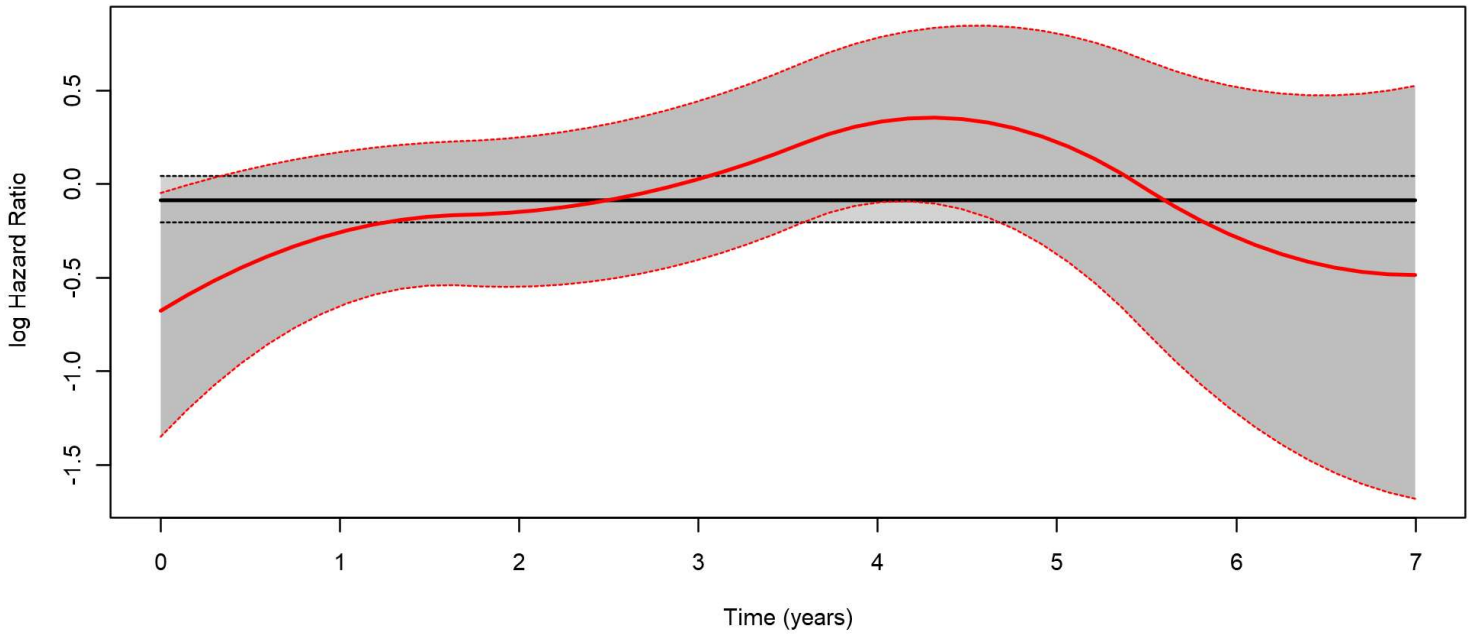
- We allow for time-varying coefficients

$$h_i(t) = h_0(t) \exp\Big\{\gamma_1 \text{Age} + \gamma_2 \text{Age}^2 + \alpha_1(t)\eta_i(t) + \alpha_2(t)\frac{d\eta_i(t)}{dt}\Big\}$$

PRIAS Study Analysis (cont'd)

Effect Type: Both

Parameter: Value



Simulations

- Different scenarios with time-constant and time-varying effects

- *When true association time-constant*
  - assuming a time-varying coefficient did not affected predictive ability

- *When true association time-varying*
  - assuming a time-constant coefficient resulted in diminished predictive ability

Discussion

- The P-splines approach provides a flexible framework for estimating time-varying association parameters

- Software: available in the development version of **JMbayes** on GitHub (https://github.com/drizopoulos/JMbayes (https://github.com/drizopoulos/JMbayes))

- More info on current status of the project at http://www.drizopoulos.com/ (http://www.drizopoulos.com/)

Thank you for your attention!

http://www.drizopoulos.com/ (http://www.drizopoulos.com/)