

# Dynamic predictions from Joint Models using Super Learning

**Dimitris Rizopoulos<sup>1</sup> and Jeremy M.G. Taylor<sup>2</sup>**

<sup>1</sup>Department of Biostatistics, Erasmus Medical Center Rotterdam

<sup>2</sup>Department of Biostatistics, University of Michigan



d.rizopoulos@erasmusmc.nl  
jmgt@umich.edu



@drizopoulos

# 1 Background & Motivation

---

## Setting: Follow-up studies

- ▷ multiple longitudinal outcomes
  - \* biomarkers
  - \* patient parameters
  - \* patient reported outcome scores
  
- ▷ one or multiple endpoints
  - \* relapse of disease
  - \* requirement for intervention
  - \* death

# 1 Background & Motivation (cont'd)

---

**Obtain accurate predictions for the (cumulative) risk of an event to guide decision making**

**Using the available longitudinal information**

# 1 Background & Motivation (cont'd)

---

## University of Michigan Prostatectomy Data

- ▷ 3634 PCa patients followed-up in 1996–2013
  - \* aged 40 to 84 years with clinically localized cT1 to cT3 disease
  - \* received radical prostatectomy

# 1 Background & Aim (cont'd)

---

University of Michigan Prostatectomy Data

**Patients remain at risk of metastasis**

▷ Follow-up

- \* PSA levels at frequent intervals
- \* when PSA increases, physicians consider Salvage Therapy (ST)
- \* ST androgen deprivation therapy, radiation therapy, chemotherapy, and combinations

# 1 Background & Motivation (cont'd)

---

University of Michigan Prostatectomy Data

**Use the longitudinal PSA & baseline covariates to predict  
the risk of metastasis**

# 1 Background & Motivation (cont'd)

---

- Two main frameworks to obtain such predictions
  - ▷ *Landmarking*
    - \* a series of Cox models at different landmark times
    - \* biomarker last value as a baseline covariate or a mixed model
    - \* Breslow estimator of survival probabilities
  - ▷ *Joint Models*
    - \* complete specification of the joint distribution of the outcomes
    - \* direct derivation of conditional risk probabilities

# 1 Background & Motivation (cont'd)

---

## Landmarking

### ▷ *Advantages*

- \* easier to use, available in standard software
- \* can generalize to multiple biomarkers without (much) extra computational cost

### ▷ *Disadvantages*

- \* predictions not consistent
- \* not plausible LOCF for biomarkers
- \* does not account for measurement error and endogeneity
- \* not valid causal interpretation



# 1 Background & Motivation (cont'd)

---

## Joint Models

### ▷ *Advantages*

- \* consistent predictions
- \* accounts for measurement error and endogeneity
- \* biomarkers follow a trajectory
- \* valid causal interpretation

### ▷ *Disadvantages*

- \* computationally intensive
- \* *sensitive to modeling assumptions*

# 1 Background & Motivation (cont'd)

---

- *Sensitive to modeling assumptions*
  - ▷ *Longitudinal profiles shape*
    - \* non-linear subject-specific trajectories
  - ▷ *Functional form*
    - \* how to link the hazard of the event with the longitudinal outcome

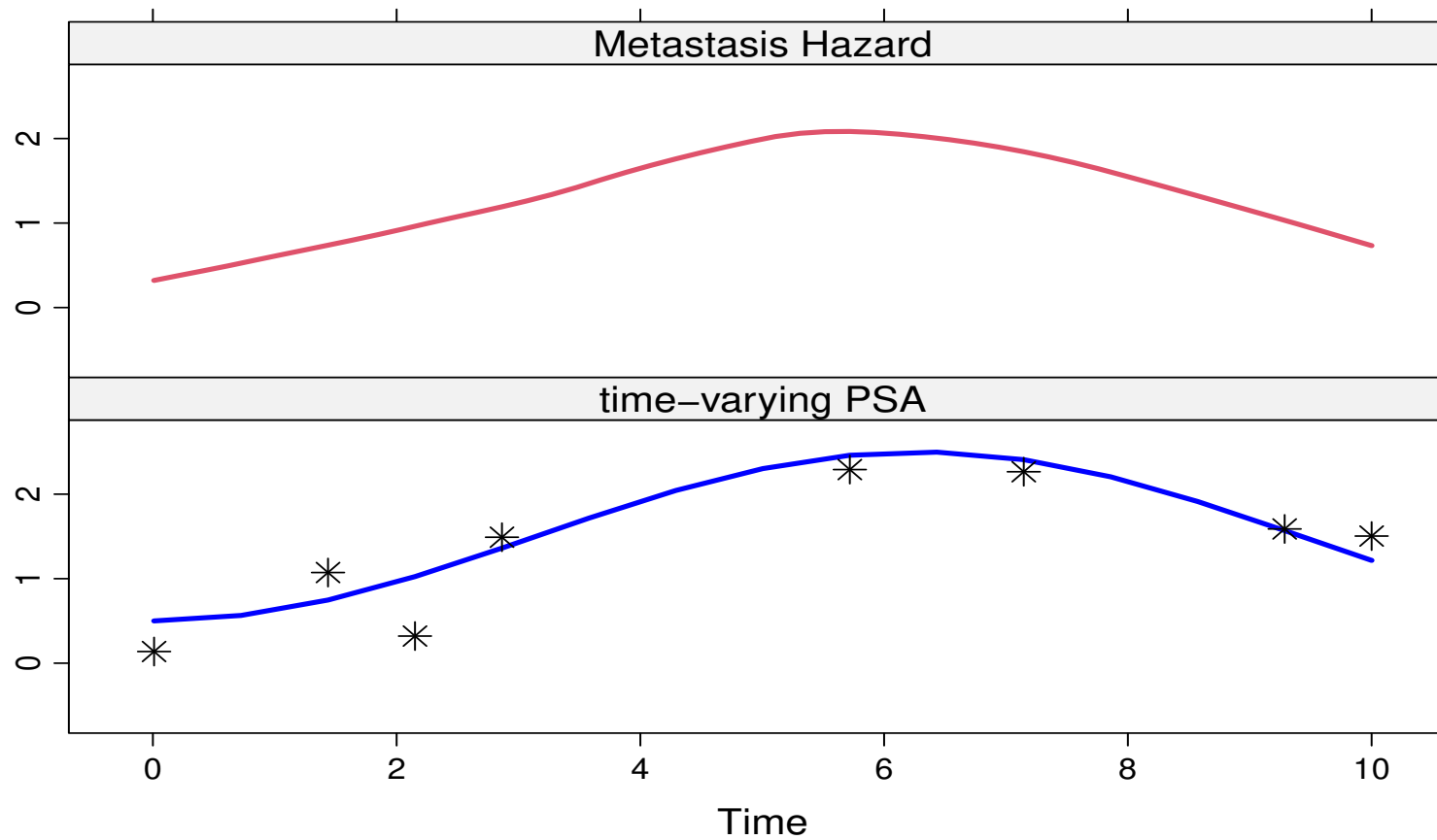
## 2 Joint Models

---

### Joint Models Framework - Basic Idea

- ▷ Use a model to describe the subject-specific longitudinal trajectories
- ▷ Use these trajectories in a hazard model for the event
- ▷ Random effects explain the association

## 2 Joint Models (cont'd)



## 2 Joint Models (cont'd)

---

Some notation

- ▷  $T_i^*$ : True event time for patient  $i$
- ▷  $T_i$ : Observed event time for patient  $i$
- ▷  $\delta_i$ : Event indicator, i.e., equals 1 for true events
- ▷  $y_i$ : Longitudinal covariate

## 2 Joint Models (cont'd)

More formally

$$\left\{ \begin{array}{l} h_i(t \mid \mathcal{H}_i(t, \mathbf{b}_i)) = h_0(t) \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + f(\alpha, \mathcal{H}_i(t, \mathbf{b}_i))\}, \\ \quad \mathcal{H}_i(t, \mathbf{b}_i) = \{\eta_i(s, \mathbf{b}_i); 0 \leq s \leq t\} \\ \\ y_i(t) = \eta_i(t, \mathbf{b}_i) + \varepsilon_i(t) \\ \quad = \mathbf{x}_i^\top(t) \boldsymbol{\beta} + \mathbf{z}_i^\top(t) \mathbf{b}_i + \varepsilon_i(t), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \\ \\ \mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \end{array} \right.$$

## 2 Joint Models (cont'd)

---

- We follow a Bayesian estimation paradigm treating  $\theta$  and  $\{b_i, i = 1, \dots, n\}$  are regarded as parameters
- Inference is based on the full posterior distribution

$$p(\theta, b \mid T, \delta, y) = \frac{\prod_i p(T_i, \delta_i \mid b_i, \theta) p(y_i \mid b_i, \theta) p(b_i, \theta) p(\theta)}{\prod_i p(T_i, \delta_i, y_i)}$$

$$\propto \prod_{i=1}^n \left\{ p(T_i, \delta_i \mid b_i, \theta) p(y_i \mid b_i, \theta) p(b_i, \theta) \right\} p(\theta)$$

## 2 Joint Models (cont'd)

---

- Dynamic predictions from joint models

$$\pi_i(u \mid t) = \Pr\{T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n\}, \quad u > t,$$

where

- ▷  $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s \leq t\}$  available measurements up to  $t$
- ▷  $\mathcal{D}_n$  the sample used to fit the model



## 2 Joint Models (cont'd)

---

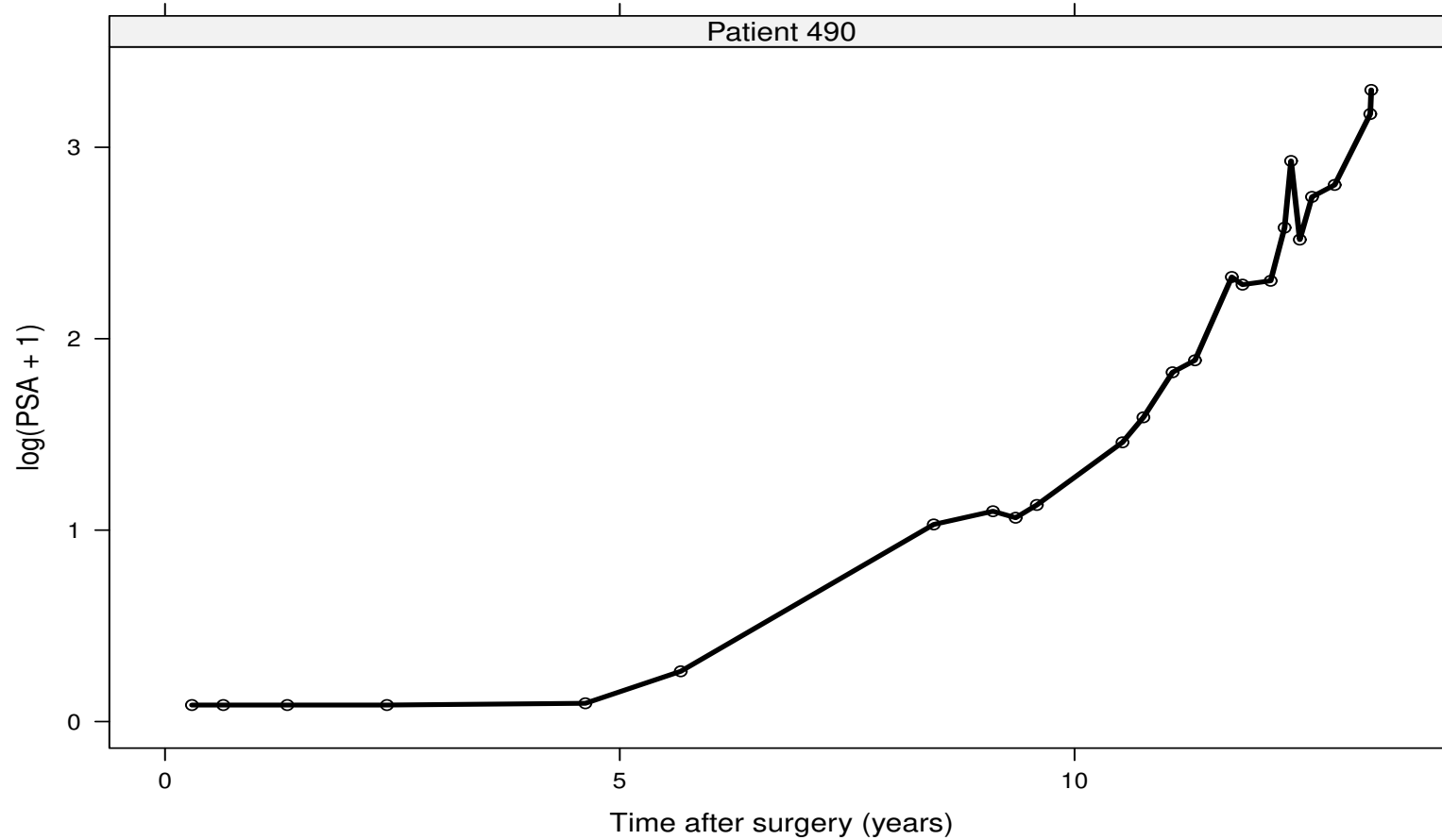
- Under the Bayesian formulation  $\pi_i(u | t)$  is written as

$$\Pr\{T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n\} = \int \Pr\{T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), \theta\} p(\theta \mid \mathcal{D}_n) d\theta$$

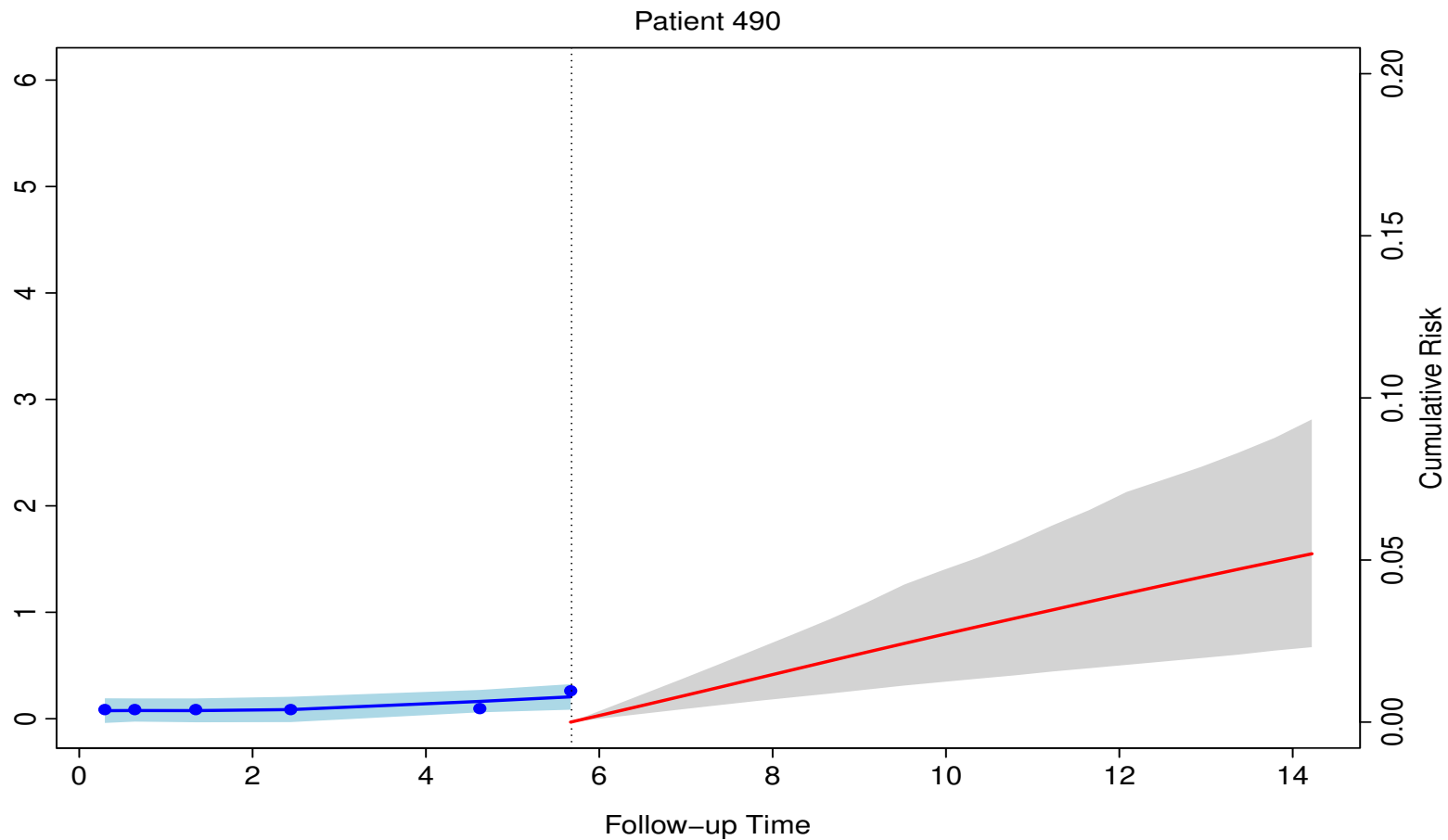
- With the first term taking the form

$$\begin{aligned} \Pr\{T_i^* \geq u \mid T_i^* > t, \mathcal{Y}_i(t), \theta\} &= \\ &= \int \frac{S_i\{u \mid \mathcal{H}_i(u, b_i, \theta), \theta\}}{S_i\{t \mid \mathcal{H}_i(t, b_i, \theta), \theta\}} p(b_i \mid T_i^* > t, \mathcal{Y}_i(t), \theta) db_i \end{aligned}$$

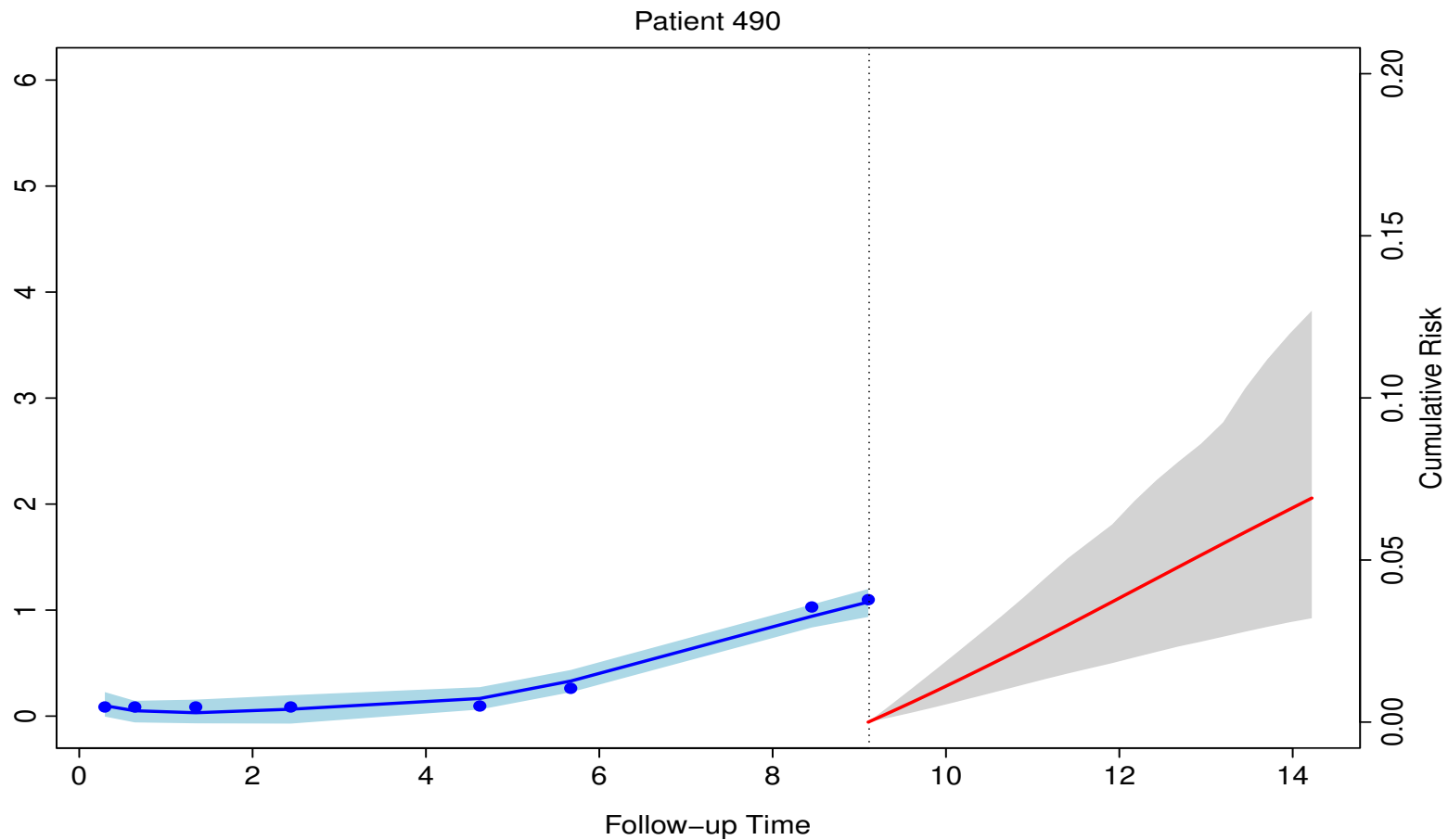
## 2 Joint Models (cont'd)



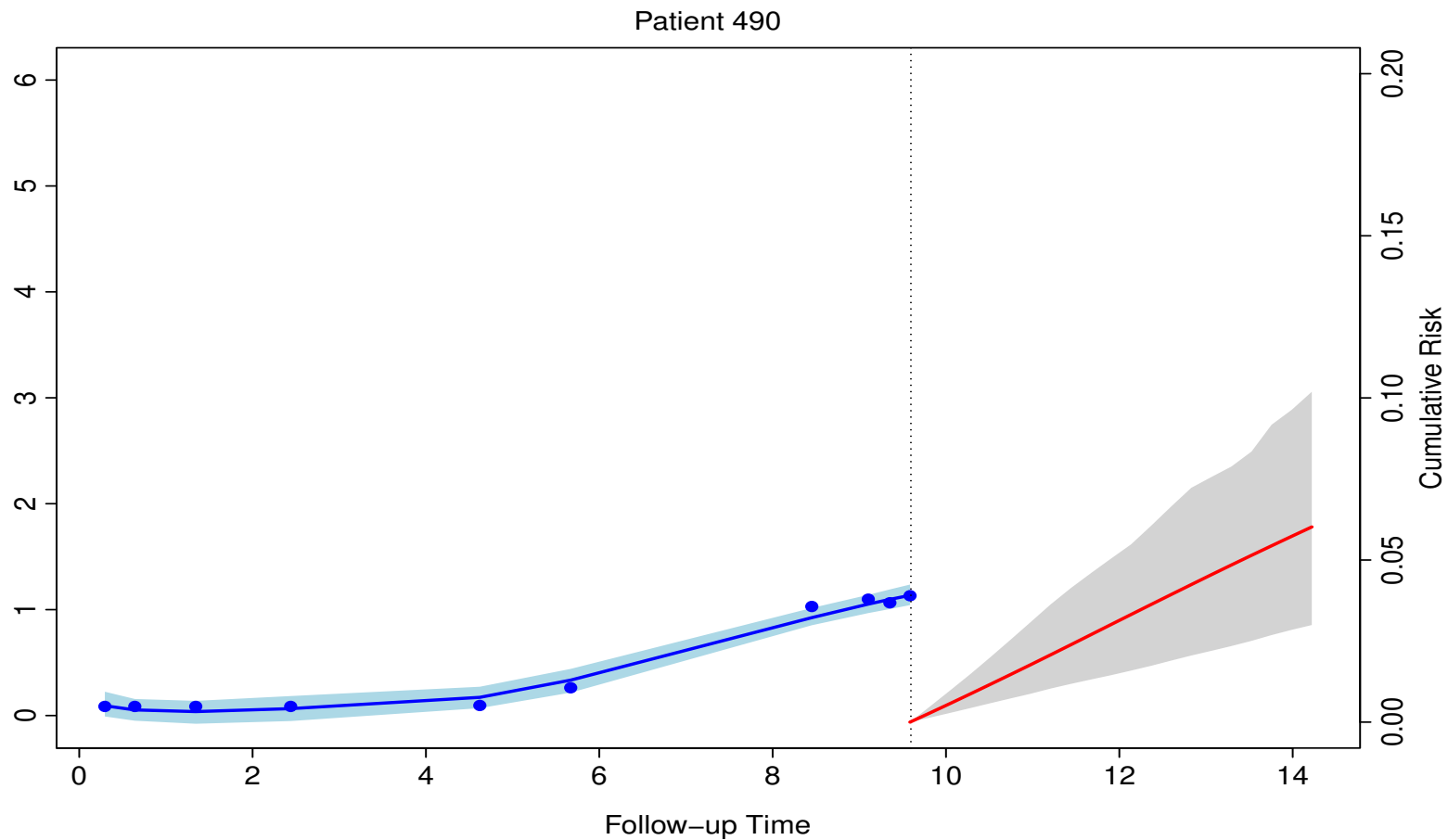
## 2 Joint Models (cont'd)



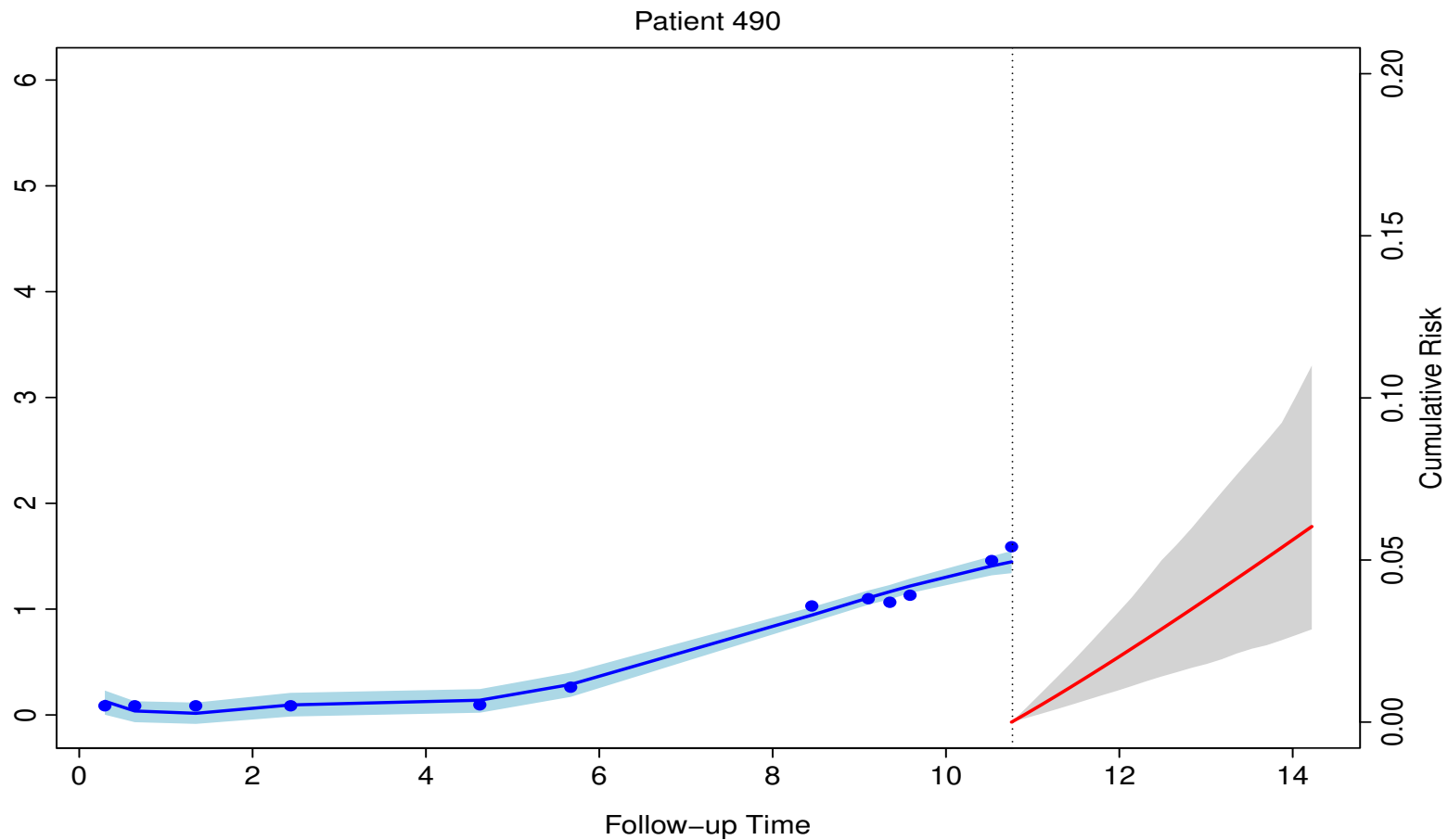
## 2 Joint Models (cont'd)



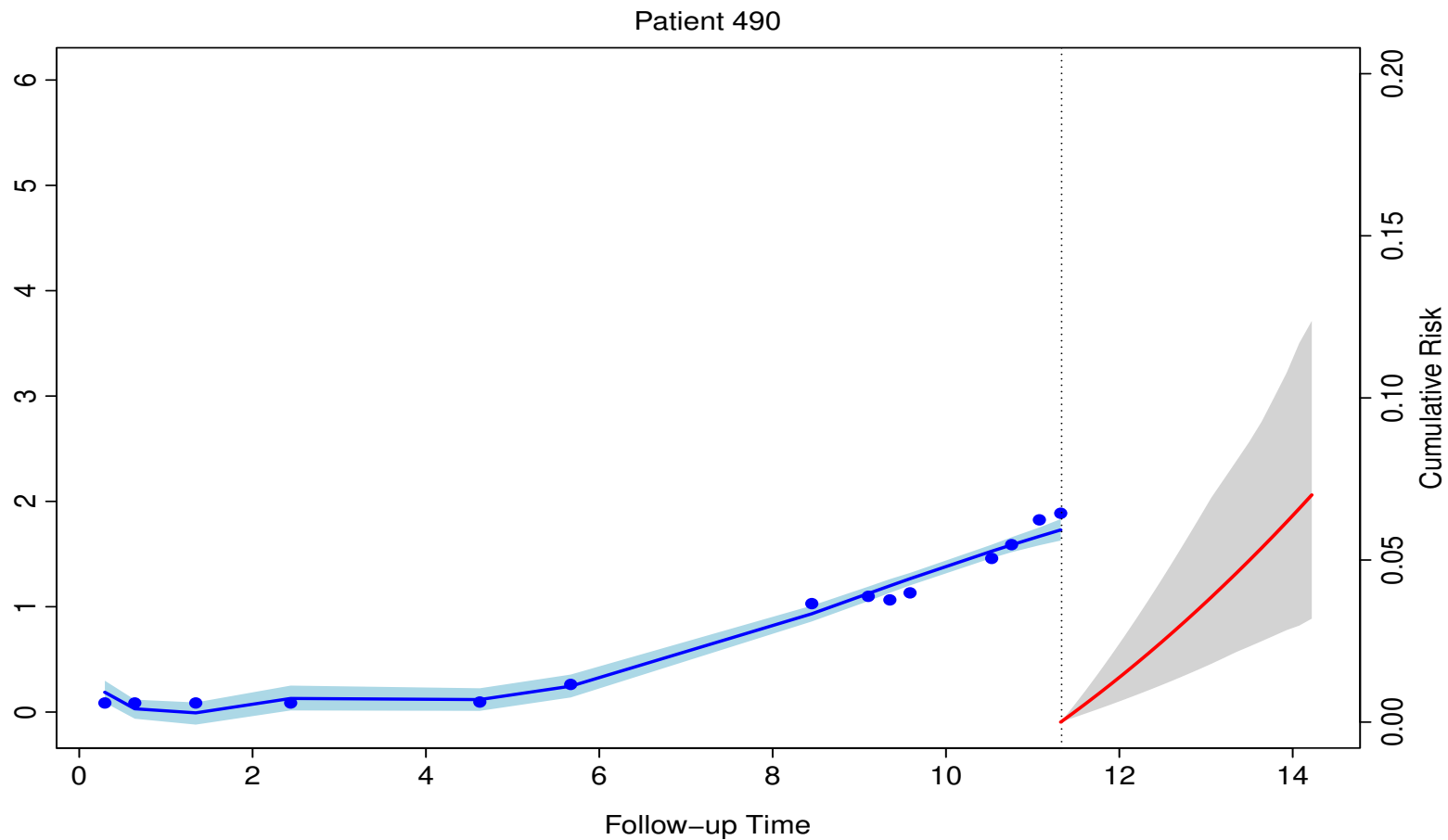
## 2 Joint Models (cont'd)



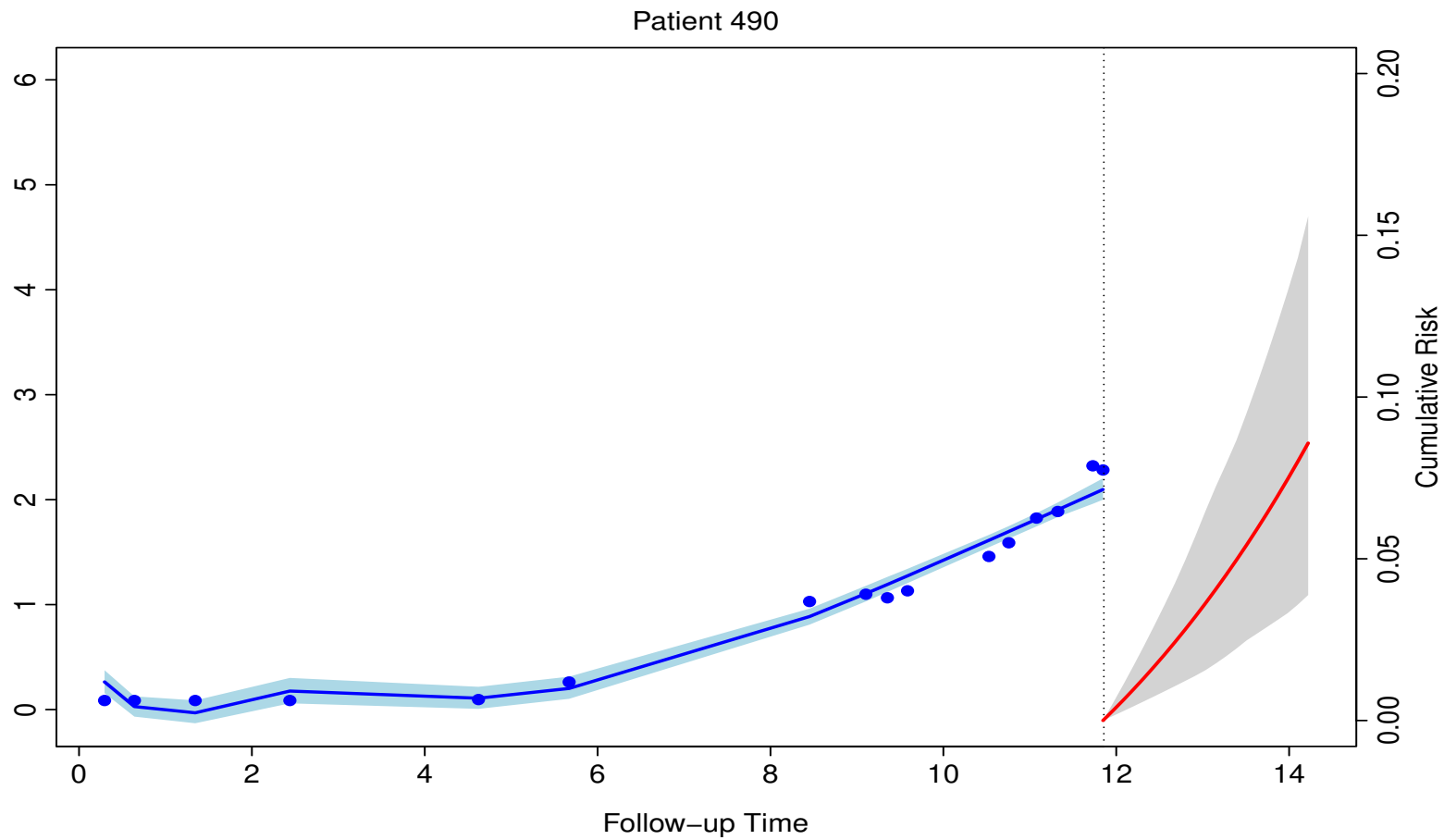
## 2 Joint Models (cont'd)



## 2 Joint Models (cont'd)

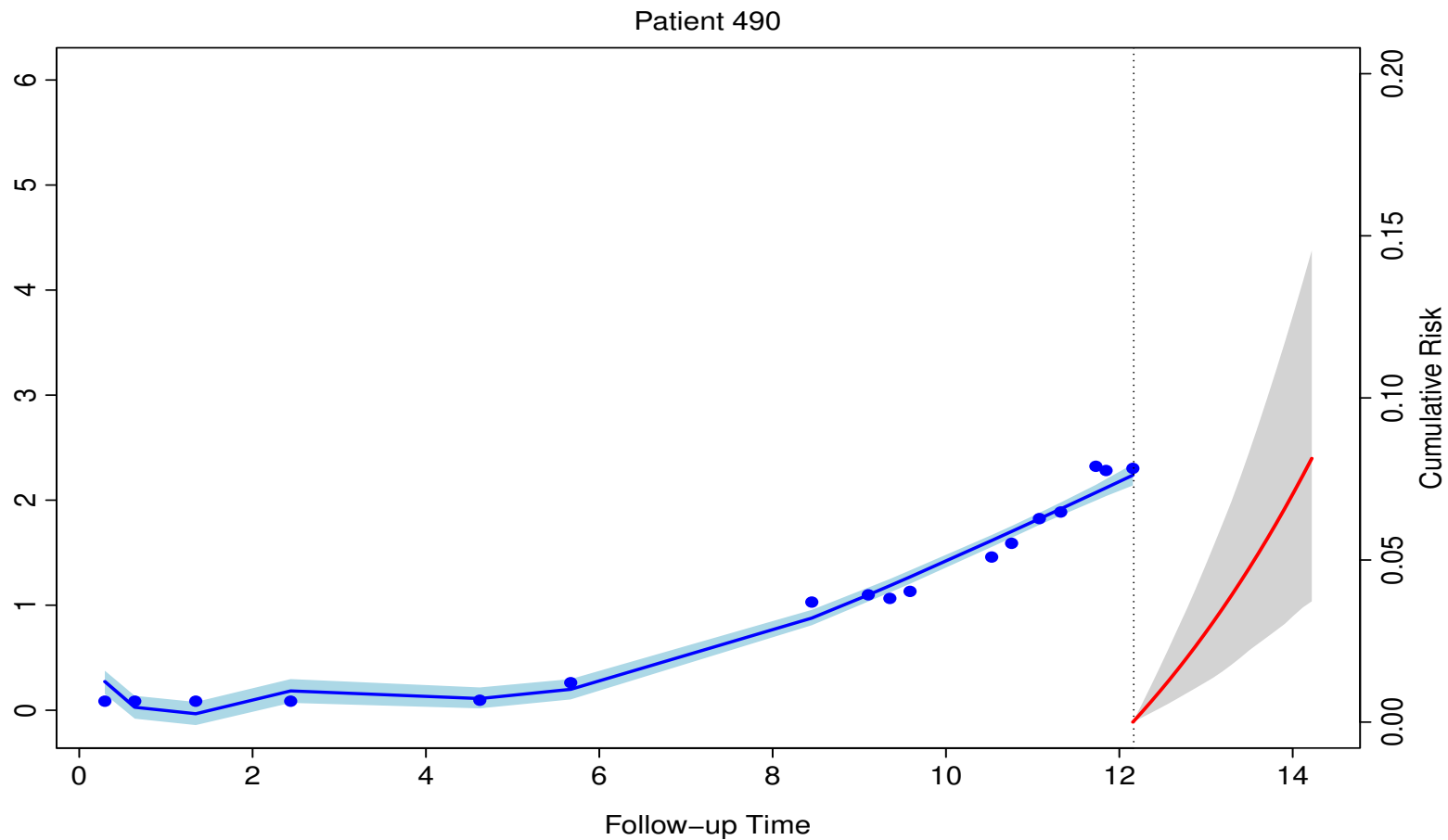


## 2 Joint Models (cont'd)

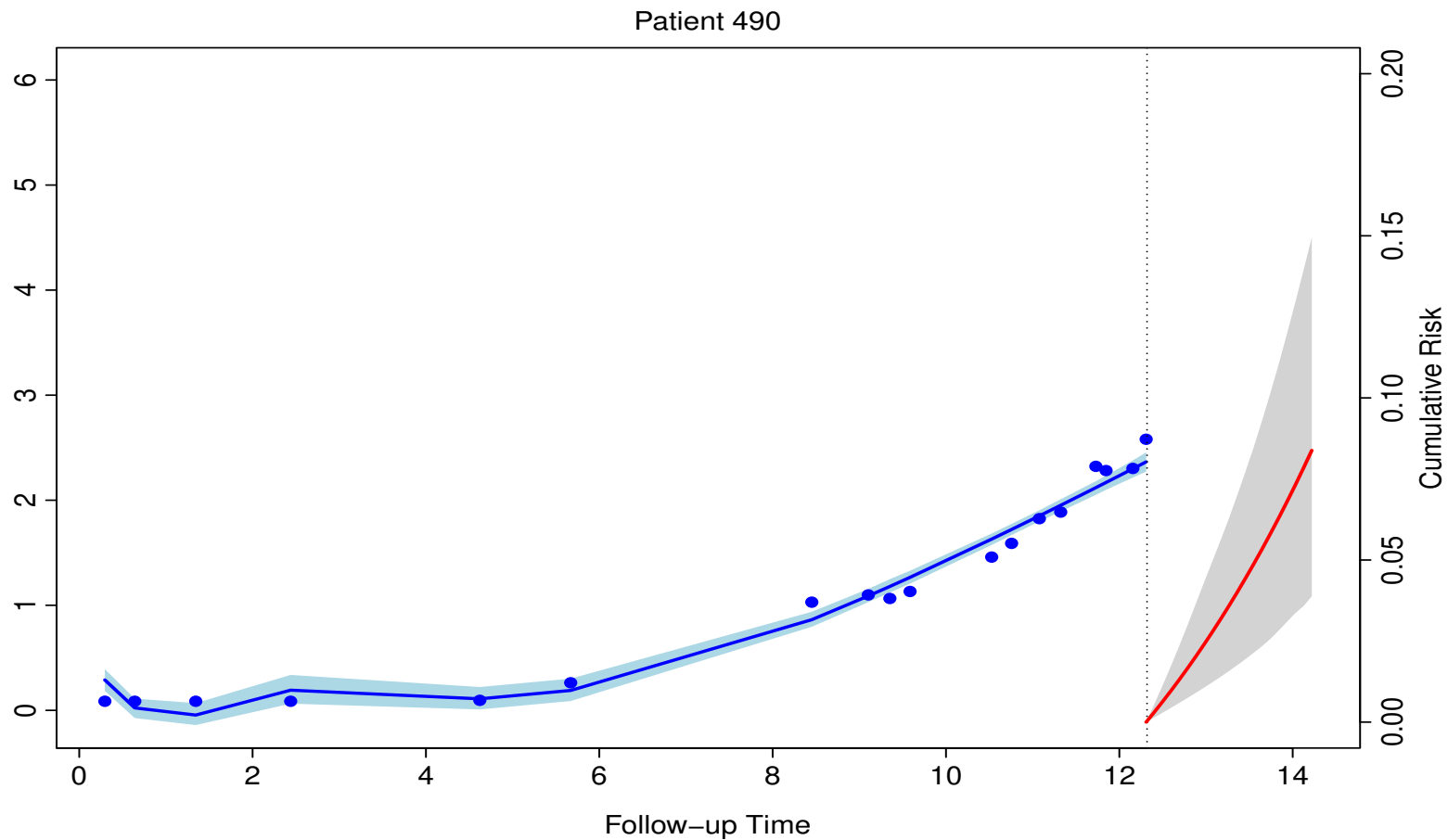




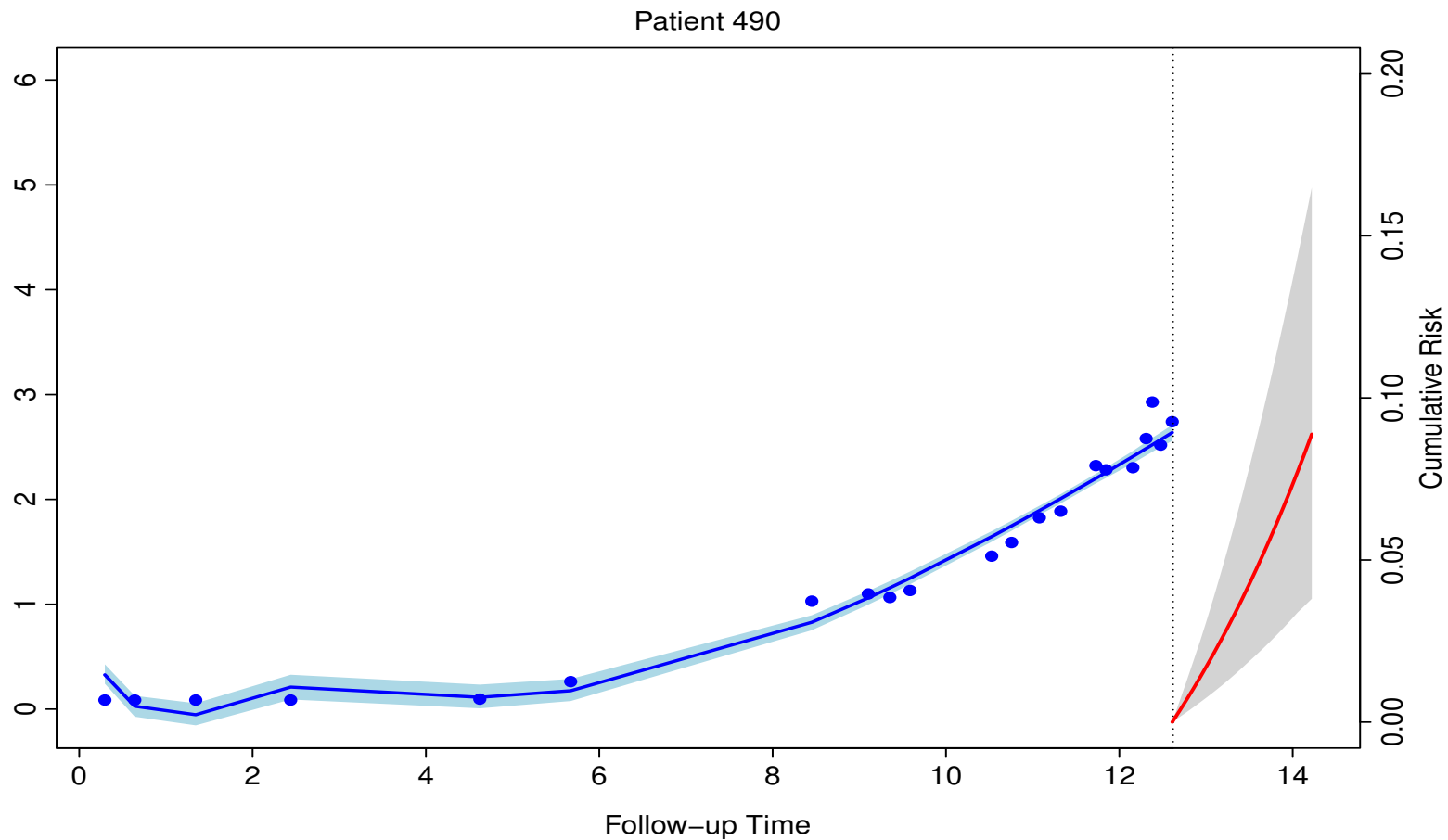
## 2 Joint Models (cont'd)



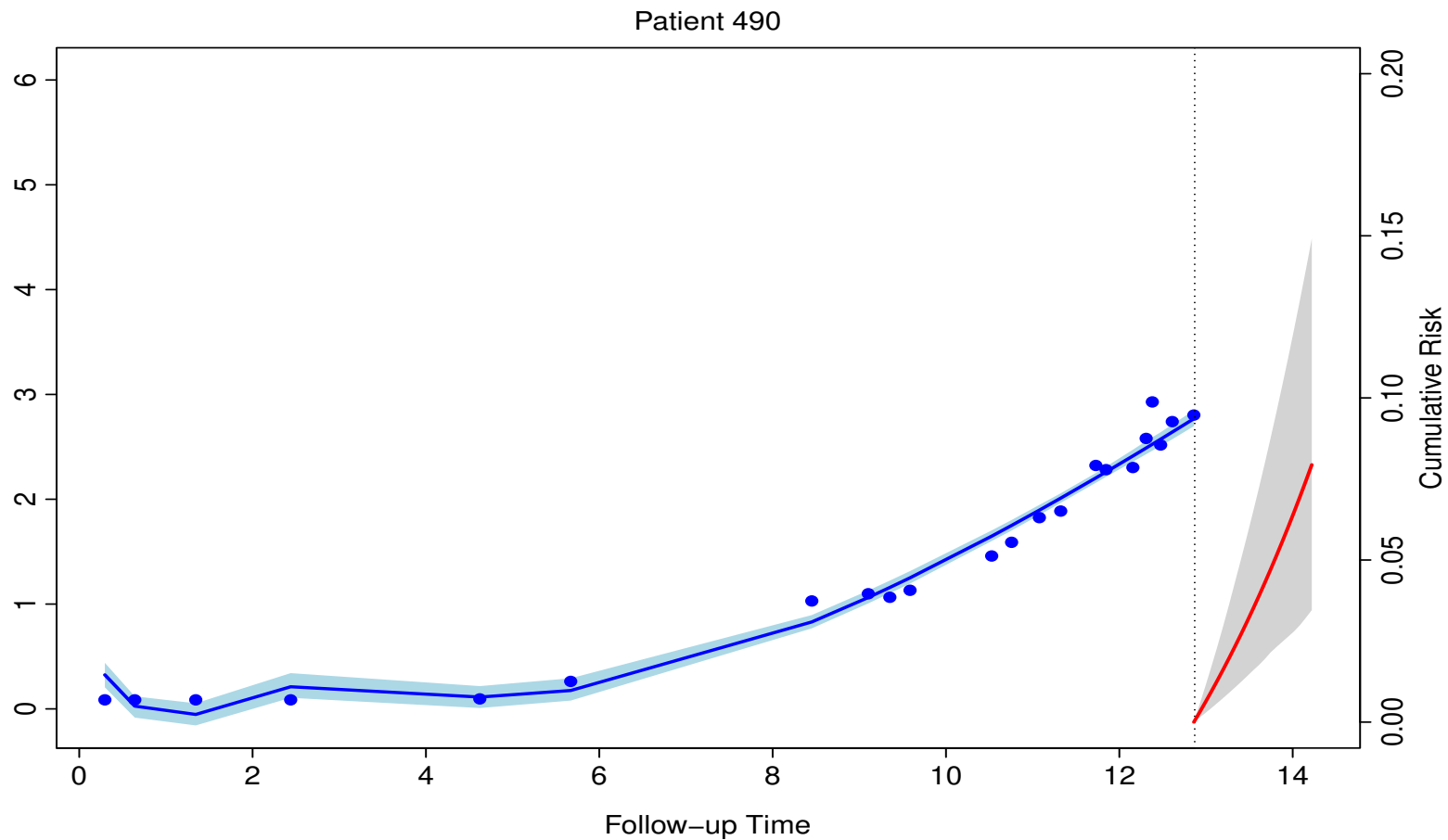
## 2 Joint Models (cont'd)



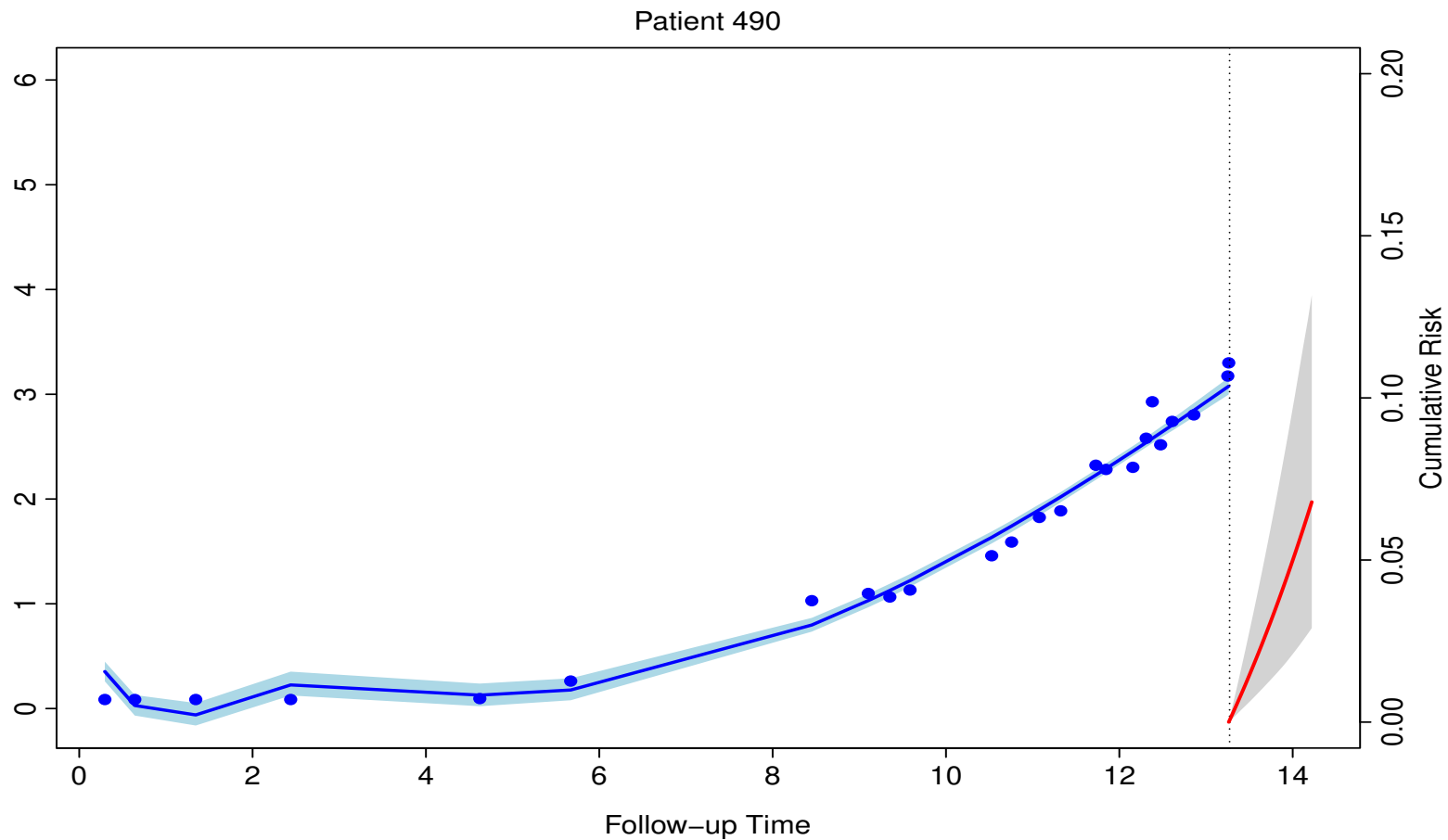
## 2 Joint Models (cont'd)



## 2 Joint Models (cont'd)



## 2 Joint Models (cont'd)



## 2 Joint Models (cont'd)

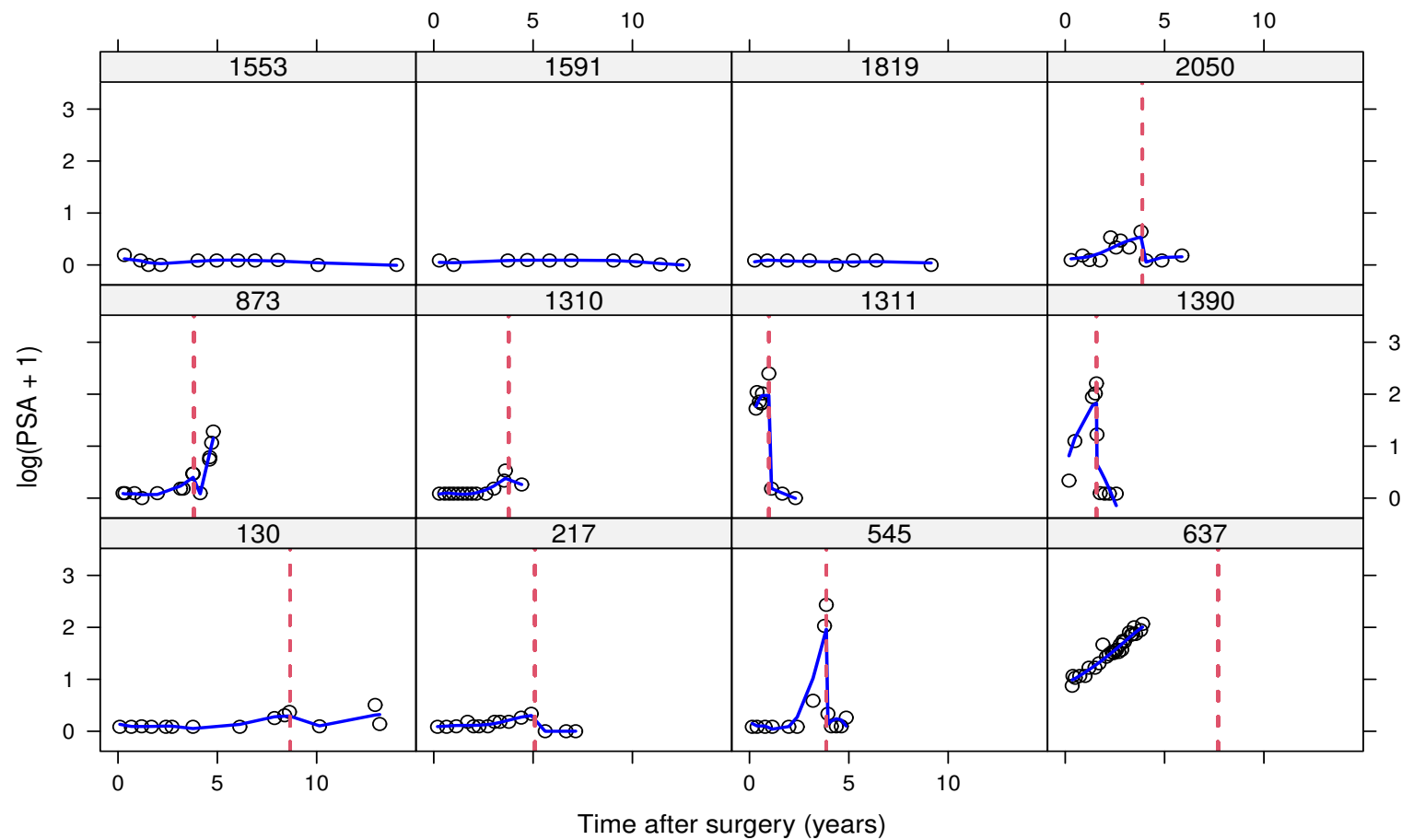
---

- In the context of dynamic predictions,
  - ▷ previous research has shown that predictive accuracy is compromised
  - ▷ when the model does not *adequately* capture the subject-specific trajectories shape

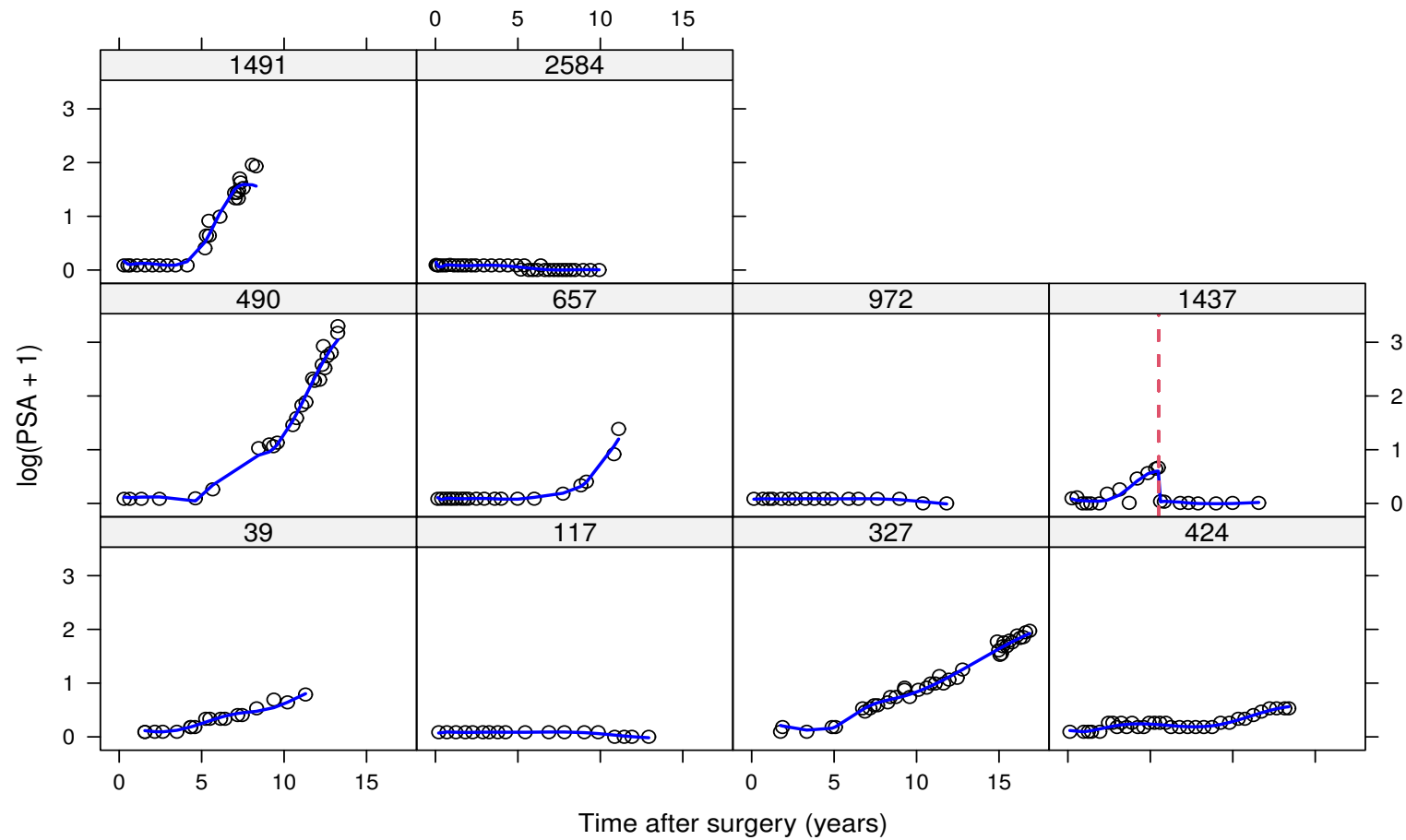
### Advice

- ▷ use flexible models, e.g., splines in both fixed- and random-effects parts
- ▷ *increased computational burden*

## 2 Joint Models (cont'd)



## 2 Joint Models (cont'd)





### 3 Functional Forms

---

**There are different ways to link the longitudinal trajectories  
to the risk of an event**

▷ Some standard options are ...

### 3 Functional Forms (cont'd)

---

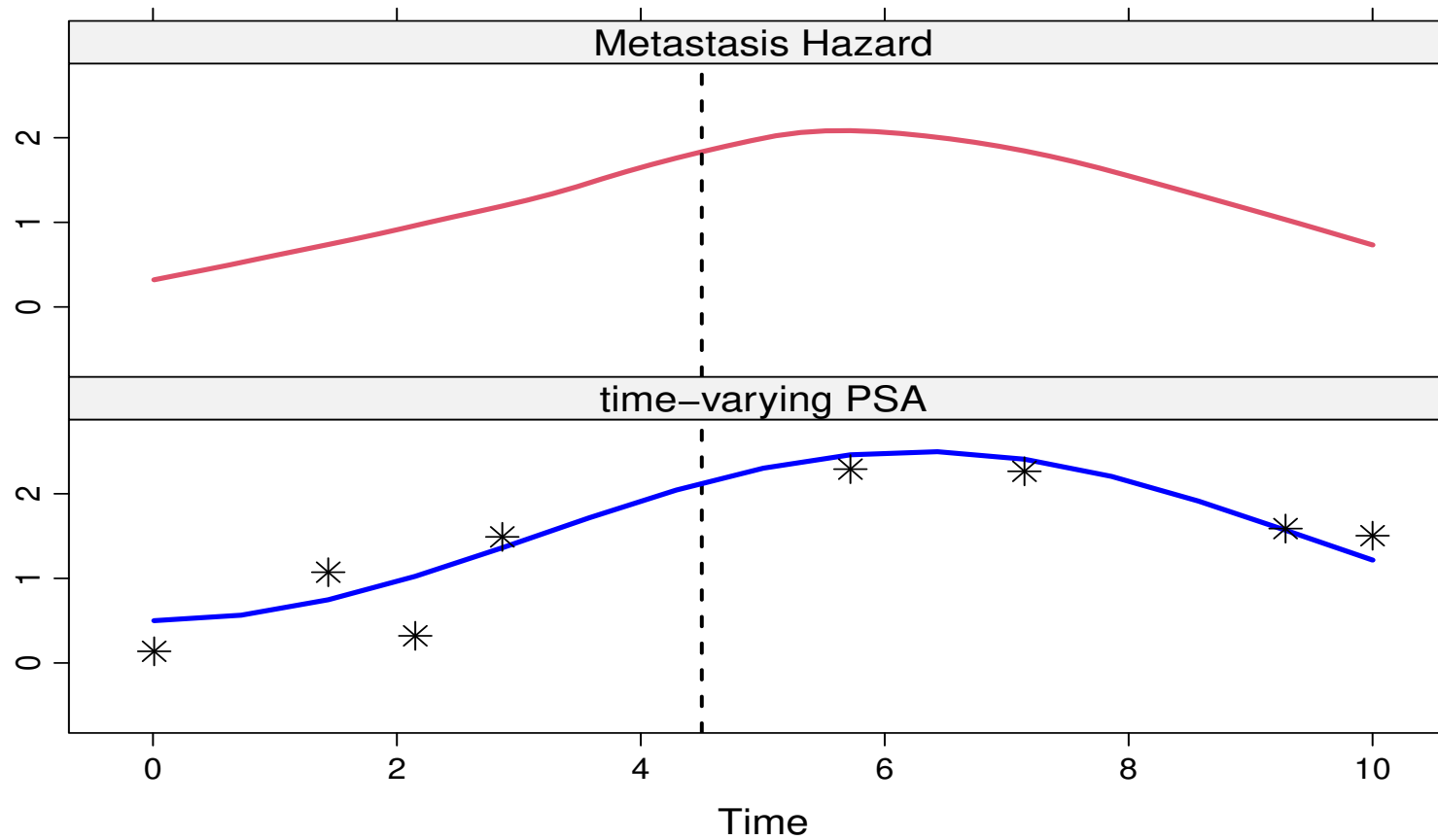
**Value:** The hazard of metastasis at  $t$  is associated with the level of PSA at  $t$ :

$$h_i(t \mid \mathcal{H}_i(t, \mathbf{b}_i)) = h_0(t) \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha \eta_i(t, \mathbf{b}_i)\}$$

where

$$\eta_i(t, \mathbf{b}_i) = \mathbf{x}_i^\top(t) \boldsymbol{\beta} + \mathbf{z}_i^\top(t) \mathbf{b}_i$$

### 3 Functional Forms (cont'd)



### 3 Functional Forms (cont'd)

---

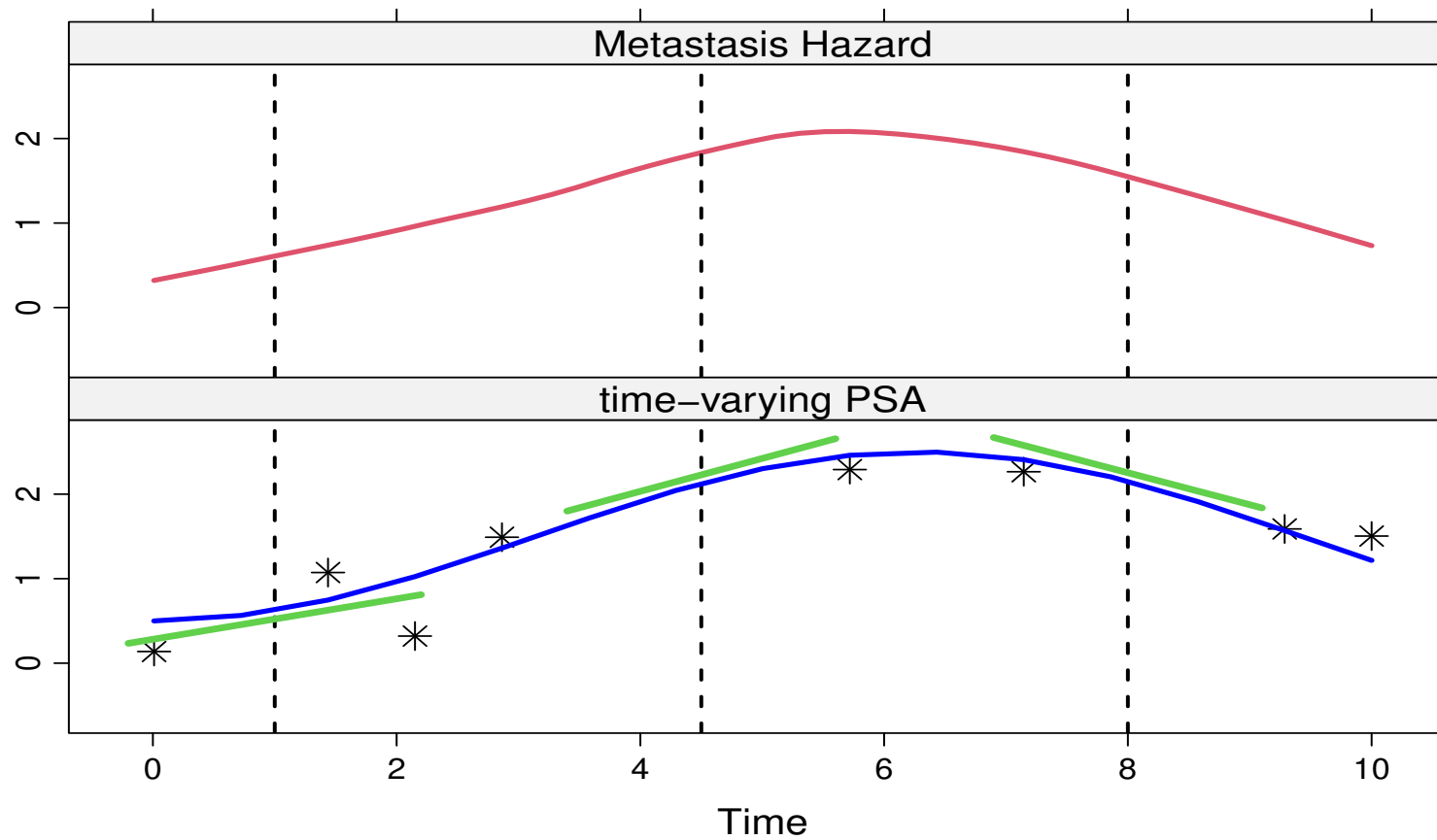
**Velocity:** The hazard of metastasis at  $t$  is associated with the slope of the PSA trajectory at  $t$ :

$$h_i(t \mid \mathcal{H}_i(t, \mathbf{b}_i)) = h_0(t) \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha \eta'_i(t, \mathbf{b}_i)\},$$

where

$$\eta'_i(t, \mathbf{b}_i) = \frac{d}{dt} \{\mathbf{x}_i^\top(t) \boldsymbol{\beta} + \mathbf{z}_i^\top(t) \mathbf{b}_i\}$$

### 3 Functional Forms (cont'd)



### 3 Functional Forms (cont'd)

---

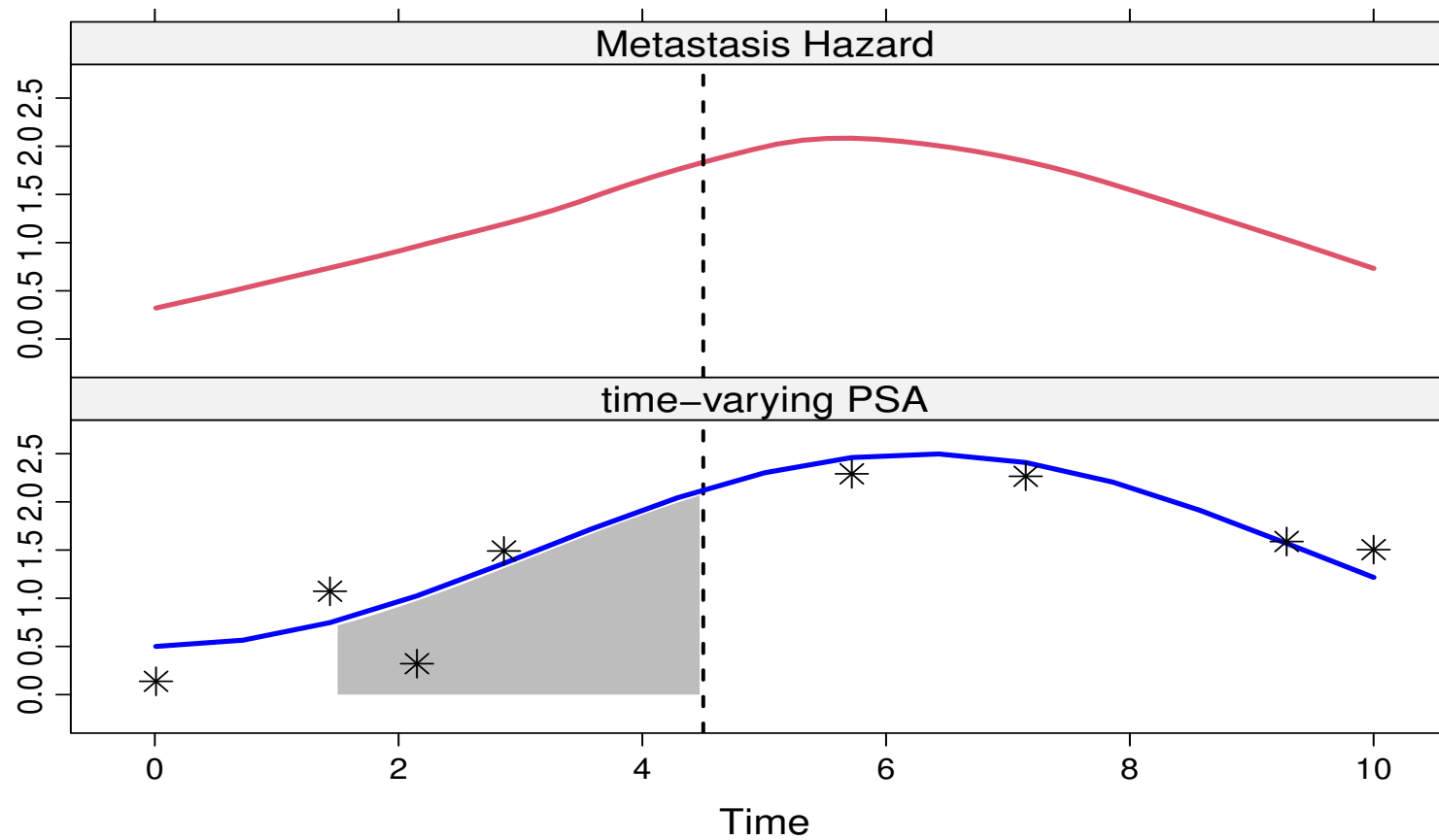
**Average Effects:**

The hazard of metastasis at  $t$  is associated with the average PSA in the interval  $(t - \Delta t, t)$ :

$$h_i(t \mid \mathcal{H}_i(t, \mathbf{b}_i)) = h_0(t) \exp \left\{ \boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha \frac{1}{\Delta t} \int_{t-\Delta t}^t \eta_i(s, \mathbf{b}_i) \, ds \right\}$$

We account for the observation period

### 3 Functional Forms (cont'd)



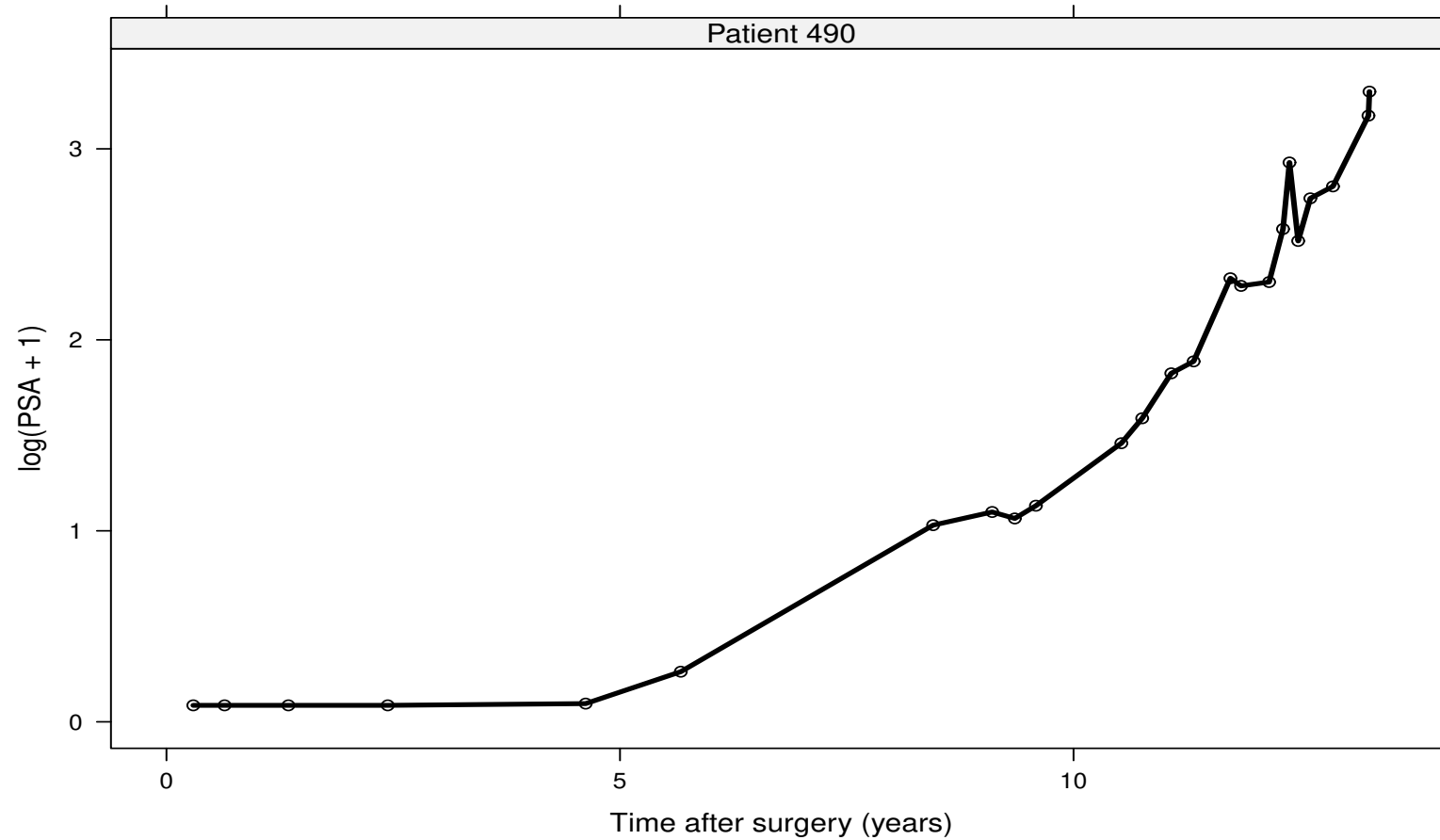
### 3 Functional Forms (cont'd)

---

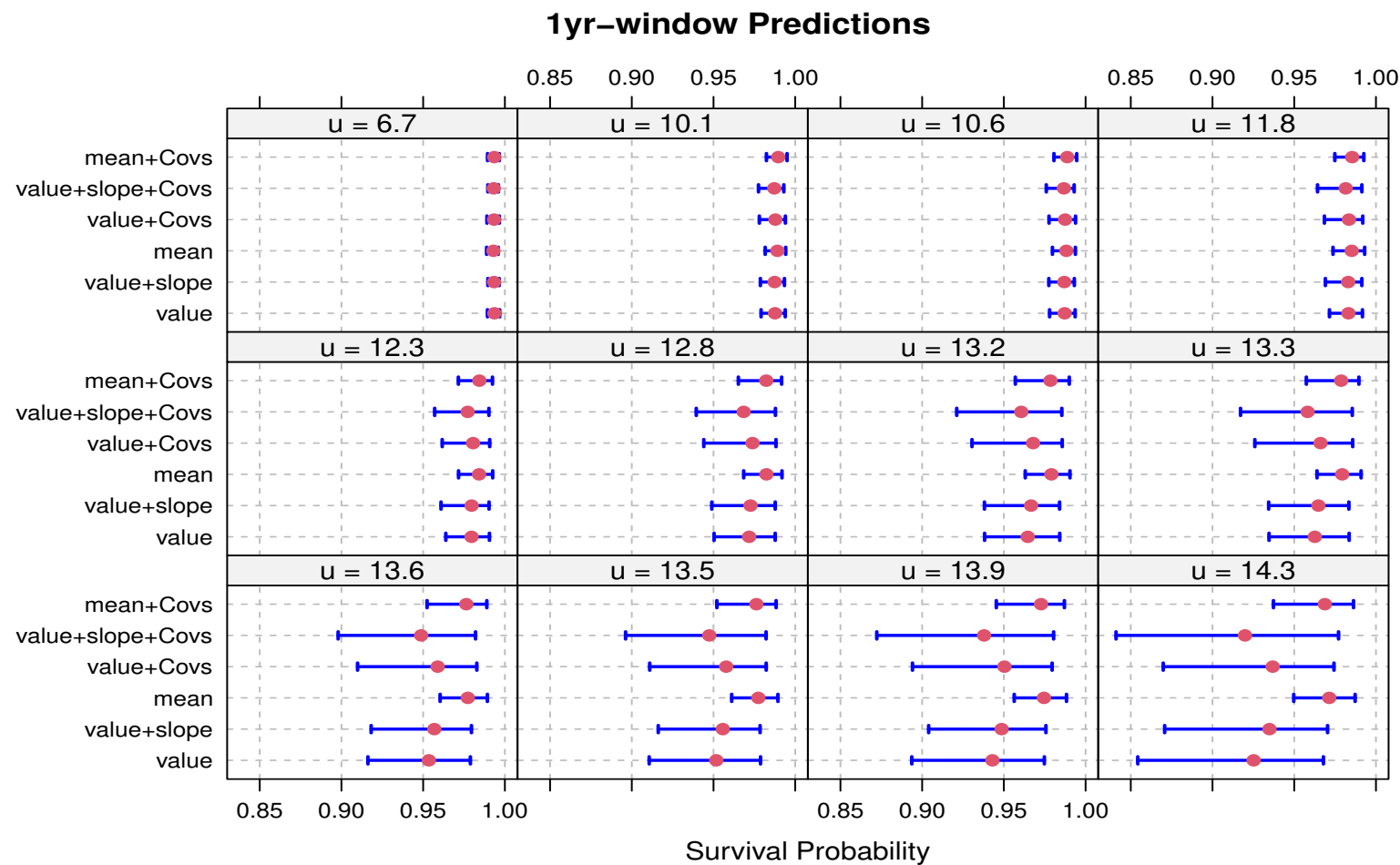
**How significant is the choice of the functional form for dynamic predictions?**



### 3 Functional Forms (cont'd)



### 3 Functional Forms (cont'd)



## 4 Super Learning

---

- The selected functional form and time effect for the longitudinal outcome can influence the derived predictions
  - ▷ especially for the survival outcome

**How to select between the different functional forms and trajectory shapes?**

## 4 Super Learning (cont'd)

---

- The standard answer is to employ information criteria, e.g., DIC, WAIC, ...
- However, the longitudinal information dominates the joint likelihood  
⇒ will not be sensitive enough wrt predicting survival probabilities
- In addition, will *a single model* be the most appropriate
  - ▷ for all follow-up times?

## 4 Super Learning (cont'd)

---

### **Solution**

- ▷ *Consider multiple plausible models with different*
  - \* longitudinal outcomes
  - \* assumptions for the longitudinal profiles
  - \* functional forms
  - \* baseline covariates, interaction terms
  - \* ...
  
- ▷ *Obtain the desired predictions from these models*
  
- ▷ *Combine predictions using weights*
  - \* *how to select the weights?*

## 4 Super Learning (cont'd)

---

- Previous research: *Bayesian Model Averaging*
  - ▷ Assume we have a library of  $L$  models  $\mathcal{L} = \{M_1, \dots, M_L\}$
  - ▷ Weights: Posterior probability of a model given the data

$$p(M_l \mid \mathcal{D}_n), \quad l = 1, \dots, L$$

where

$$^* \mathcal{D}_n = \{T_i, \delta_i, \mathbf{y}_i; i = 1, \dots, n\}$$

## 4 Super Learning (cont'd)

---

- *Issues with BMA weights*

- ▷ Requires calculating the marginal likelihood

$$p(\mathcal{D}_n \mid M_l) = \int \underbrace{p(\mathcal{D}_n \mid \boldsymbol{\theta}, M_l)}_{\text{Likelihood}} \underbrace{p(\boldsymbol{\theta} \mid M_l)}_{\text{Prior}} d\boldsymbol{\theta}$$

⇒ *Computationally demanding*

- ▷ Weights not designed to optimize predictions
- ▷ Not clear if we account for over-fitting

## 4 Super Learning (cont'd)

---

- *Issues with BMA weights*
  - ▷ The likelihood of a model that fits the data a bit better can have a likelihood value that is several units larger compared to the other models
  - ▷ Often one model dominates the weights over the others



## 4 Super Learning (cont'd)

---

Alternative Solution: *Super Learning*

- ▷ Select weights to optimize prediction metric *of your choice*
- ▷ Account for over-fitting using cross-validation

## 4 Super Learning (cont'd)

---

### How it works:

- ▷ Assume we have a library of  $L$  *base-learners* (models)  $\mathcal{L} = \{M_1, \dots, M_L\}$
- ▷ Specify the landmark time  $t$ , and a relevant future time  $u$ ,  $u > t$
- ▷ Split  $\mathcal{D}_n$  in  $V$ -folds
- ▷ For  $v \in \{1, \dots, V\}$ , train the learners in library  $\mathcal{L}$  using  $\mathcal{D}_n^{(-v)}$

## 4 Super Learning (cont'd)

---

### How it works:

- ▷ For the subjects in  $\mathcal{D}_n^{(v)}$ , not used when training the learner, calculate the predictions

$$\hat{\pi}_i^{(v)}(u \mid t, M_l) = \Pr\{T_i^* < u \mid T_i^* > t, \mathcal{Y}_i(t), M_l, \mathcal{D}_n^{(-v)}\}$$

do this for all  $v = 1, \dots, V$  to get the *cross-validated predictions*

## 4 Super Learning (cont'd)

---

How it works:

- ▷ We define the ensemble of *cross-validated predictions*

$$\hat{\pi}_i^v(u \mid t) = \sum_{l=1}^L \varpi_l(t) \hat{\pi}_i^{(v)}(u \mid t, M_l), \quad v = 1, \dots, V$$

\* the weights depend on  $t \Rightarrow$  *different weights at different follow-up times*

## 4 Super Learning (cont'd)

---

### How it works:

- ▷ Select  $\varpi_l(t)$  to optimize your *meta-learner* (predictive accuracy metric), e.g.,
  - \* Brier Score (*Proper scoring rule*)
  - \* Expected Predictive Cross-Entropy (*Proper scoring rule*)
  - \* AUC (*Not a proper scoring rule*)
  - \* ...
  
- ▷ Under the constraints
  - \*  $\hat{\varpi}_l(t) > 0$  for all  $l = 1, \dots, L$
  - \*  $\sum_{l=1}^L \hat{\varpi}_l(t) = 1$

## 5 UM Data Analysis

---

A library  $\mathcal{L}$  with twelve joint models

- PSA models
  - ▷  $M_{l1}$ : *linear* subject-specific time trends that change after salvage
  - ▷  $M_{l2}$ : the same as  $M_{l1}$  + covariates
  - ▷  $M_{l3}$ : *nonlinear* subject-specific time trends that change after salvage
  - ▷  $M_{l4}$ : the same as  $M_{l3}$  + covariates
- Baseline covariates: age at surgery, Charlson's index, Gleason score, and baseline PSA

## 5 UM Data Analysis (cont'd)

---

A library  $\mathcal{L}$  with twelve joint models

- Metastasis models
  - ▷  $M_{s1}$ : value of  $\log(\text{PSA} + 1)$
  - ▷  $M_{s2}$ : velocity of  $\log(\text{PSA} + 1)$
  - ▷  $M_{s3}$ : average  $\log(\text{PSA} + 1)$
- Time varying salvage therapy
- Baseline covariates: the same as in the PSA models

## 5 UM Data Analysis (cont'd)

---

- We evaluated predictive accuracy in two time intervals
  - ▷  $(4, 7]$ : 2514 patients at risk; 28 metastasis
  - ▷  $(6, 9]$ : 1914 patients at risk; 16 metastasis
- Metrics - meta learners
  - ▷ Integrated Brier Score
  - ▷ Expected Predictive Cross-Entropy



## 5 UM Data Analysis (cont'd)

---

### Meta-learners

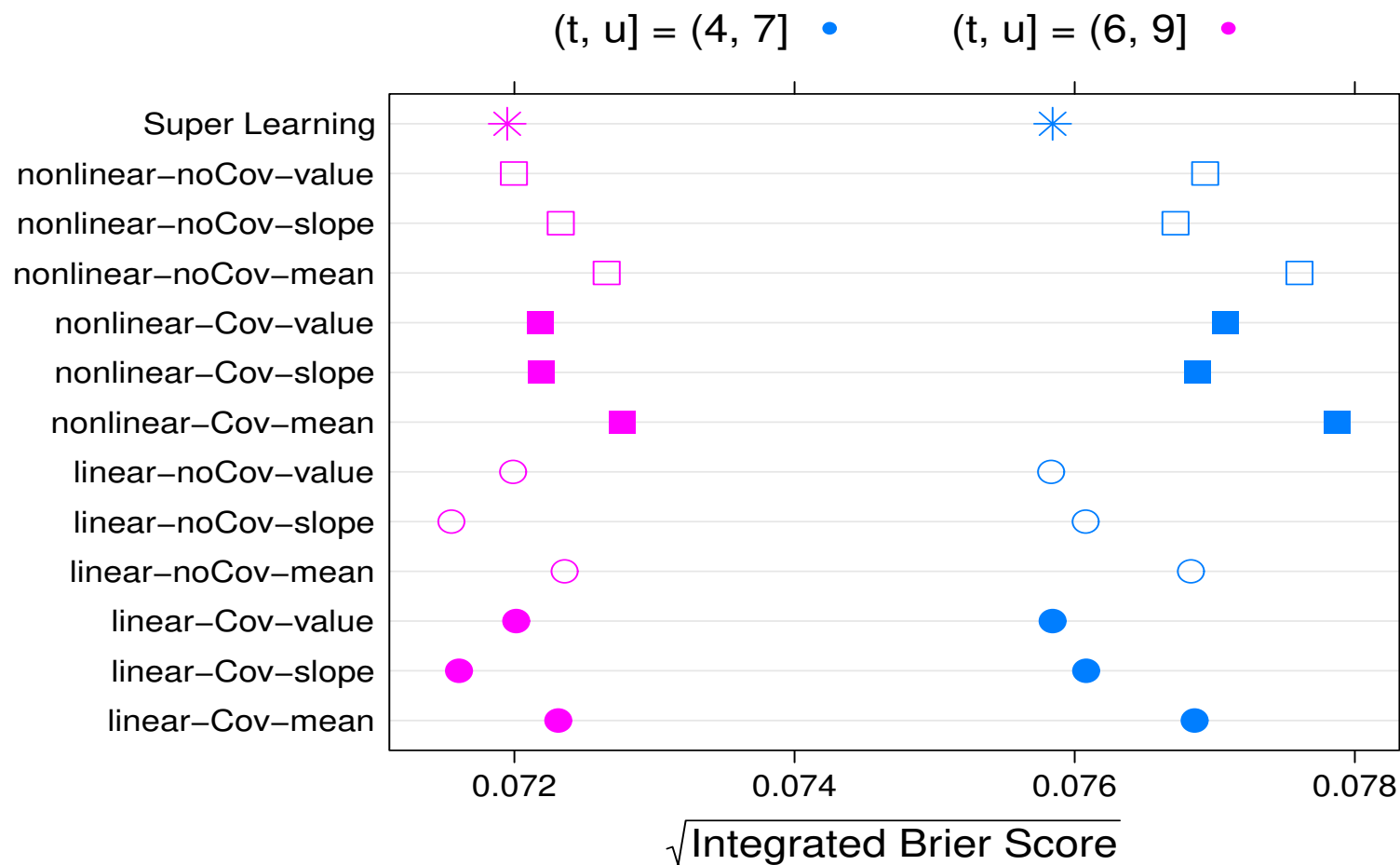
▷ *Integrated Brier Score*

$$\text{IBS}(u, t) = \frac{1}{u - t} \int_t^u E \left\{ \mathbb{I}(t < T_i^* \leq s) - \pi_i(s | t) \right\}^2 ds$$

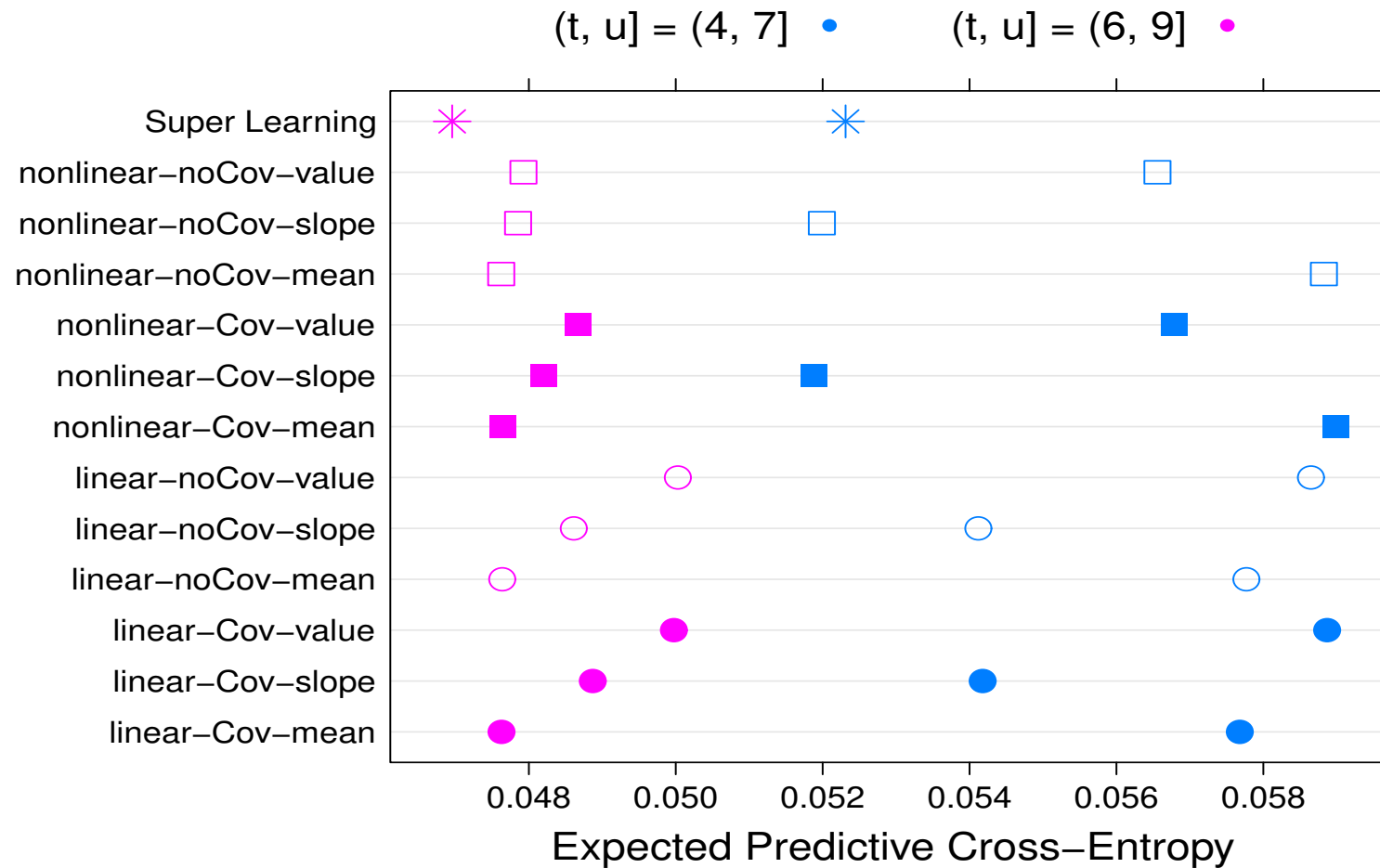
▷ *Expected Predictive Cross-Entropy*

$$\text{EPCE}(u, t) = E \left\{ -\log \left[ p \{ T_i^* | t < T_i^* \leq u, \mathcal{Y}_i(t) \} \right] \right\}$$

## 5 UM Data Analysis (cont'd)



## 5 UM Data Analysis (cont'd)



## 5 UM Data Analysis (cont'd)

---

Observations (also from the simulation study)

- ▷ ensemble Super Learning (eSL) often, *but not always*, outperforms the individual models
- ▷ In some datasets and intervals  $(t, u]$ , the discrete Super Learner (dSL) beats the eSL

## 5 UM Data Analysis (cont'd)

---

### Recommendation

**Regard eSL as an extra member of the library  $\mathcal{L}$  and use CV to select the optimal strategy**

- Available in **JMbayes2**

- ▷ cross-validated fitting of models
- ▷ combination of dynamic predictions

[https://drizopoulos.github.io/JMbayes2/articles/Super\\_Learning.html](https://drizopoulos.github.io/JMbayes2/articles/Super_Learning.html)

**Thank for your attention!**

<https://www.drizopoulos.com/>

## 7 Choice of the Meta-Learner

We focus on two meta-learners

▷ *Integrated Brier Score*

$$\text{IBS}(t + \Delta t, t) = \frac{1}{\Delta t} \int_t^{t+\Delta t} E \left[ \left\{ \mathbb{I}(T_i^* \leq s) - \pi_i(s | t) \right\}^2 \mid T_i^* > t \right] ds$$

▷ *Expected Predictive Cross-Entropy*

$$\text{EPCE}(t + \Delta t, t) = E \left\{ -\log \left[ p \{ T_i^* \mid t < T_i^* \leq t + \Delta t, \mathcal{Y}_i(t) \} \right] \right\}$$



## 7 Choice of the Meta-Learner (cont'd)

---

- For the estimation of the Brier score, we need to account for censoring in  $[t, t + \Delta t)$ 
  - \* inverse probability of censoring weighting
  - \* model-based weights

## 7 Choice of the Meta-Learner (cont'd)

---

- Brier Score with IPCW

$$\widehat{\text{BS}}_{IPCW}(t + \Delta t, t) = \frac{1}{n} \sum_{i=1}^n \widehat{W}_i(t + \Delta t, t) \left\{ \mathbb{I}(T_i \leq t + \Delta t) - \hat{\pi}_i^v(t + \Delta t | t) \right\}^2$$

where

$$\widehat{W}_i(t + \Delta t, t) = \frac{\mathbb{I}(t < T_i \leq t + \Delta t) \delta_i}{\hat{G}(T_i | t)} + \frac{\mathbb{I}(T_i > t + \Delta t)}{\hat{G}(t + \Delta t | t)},$$

with  $\hat{G}(\cdot)$  denoting Kaplan-Meier estimate of the censoring distribution  $\Pr(C_i > t)$

## 7 Choice of the Meta-Learner (cont'd)

- Brier Score with model-weights

$$\begin{aligned}\widehat{\text{BS}}_{\text{model}}(t + \Delta t, t) &= \frac{1}{n_t} \sum_{i: T_i > t} \delta_i \mathbb{I}(T_i \leq t + \Delta t) \left\{ 1 - \hat{\pi}_i^v(t + \Delta t \mid t) \right\}^2 \\ &\quad + \mathbb{I}(T_i > t + \Delta t) \left\{ \hat{\pi}_i^v(t + \Delta t \mid t) \right\}^2 \\ &\quad + (1 - \delta_i) \mathbb{I}(T_i \leq t + \Delta t) \left[ \hat{\pi}_i^v(t + \Delta t \mid T_i) \left\{ 1 - \hat{\pi}_i^v(t + \Delta t \mid t) \right\}^2 \right. \\ &\quad \left. + \left\{ 1 - \hat{\pi}_i^v(t + \Delta t \mid T_i) \right\} \left\{ \hat{\pi}_i^v(t + \Delta t \mid t) \right\}^2 \right]\end{aligned}$$

## 7 Choice of the Meta-Learner (cont'd)

---

- IPCW

- ▷ *Advantage*: it provides unbiased estimates even when the model is misspecified
- ▷ *Disadvantage*: it requires that the model for the weights is correct
  - \* challenging because censoring may depend on the longitudinal outcomes in a complex manner
  - \* sensitive to (unobserved) instrument by confounder interactions

## 7 Choice of the Meta-Learner (cont'd)

---

- Model-based Weights
  - ▷ *Advantage*: it allows censoring to depend on the longitudinal history (in any possible manner)
  - ▷ *Disadvantage*: it requires that the model is well-specified

## 7 Choice of the Meta-Learner (cont'd)

---

- An estimate of  $\text{EPCE}(t + \Delta t, t)$  that accounts for censoring

$$\widehat{\text{EPCE}}(t + \Delta t, t) = \frac{1}{n_t} \sum_{i: T_i > t} -\log \left[ p\{\tilde{T}_i, \tilde{\delta}_i \mid T_i > t, \mathcal{Y}_i(t), \mathcal{D}_n\} \right]$$

with

- ▷  $\tilde{T}_i = \min(T_i, t + \Delta t)$
- ▷  $\tilde{\delta}_i = \delta_i \mathbb{I}(t < T_i \leq t + \Delta t)$

- Features

- ▷ it allows censoring to depend on the longitudinal history
- ▷ *problem*: it is not written as a function of the predictions

## 7 Choice of the Meta-Learner (cont'd)

- The conditional predictive log-likelihood

$$\log \left[ p \{ \tilde{T}_i, \tilde{\delta}_i \mid T_i > t, \mathcal{Y}_i(t), \mathcal{D}_n \} \right] =$$

$$\tilde{\delta}_i \log [h_i \{ \tilde{T}_i \mid \mathcal{Y}_i(t), \mathcal{D}_n \}] + \log \frac{\Pr \{ T_i^* > \tilde{T}_i \mid \mathcal{Y}_i(t), \mathcal{D}_n \}}{\Pr \{ T_i^* > t \mid \mathcal{Y}_i(t), \mathcal{D}_n \}}$$

▷ the second term is  $\log \{ \pi_i(\tilde{T}_i \mid t) \}$

▷ for the first term, we write the hazard function as

$$h_i \{ \tilde{T}_i \mid \mathcal{Y}_i(t), \mathcal{D}_n \} = \frac{p(\tilde{T}_i)}{S(\tilde{T}_i)} = - \frac{\frac{d}{dt} \Pr \{ T_i^* > t \mid \mathcal{Y}_i(t), \mathcal{D}_n \} \Big|_{t=\tilde{T}_i}}{\Pr \{ T_i^* > \tilde{T}_i \mid \mathcal{Y}_i(t), \mathcal{D}_n \}}$$

## 7 Choice of the Meta-Learner (cont'd)

---

- We approximate the derivative with a forward difference and we get

$$\widehat{\text{EPCE}}(t + \Delta t, t) =$$

$$-\frac{1}{n_t} \sum_{i: T_i > t} \tilde{\delta}_i [\log\{1 - \hat{\pi}_i^v(\tilde{T}_i + \epsilon \mid \tilde{T}_i)\} - \log(\epsilon)] + \log\{\hat{\pi}_i^v(\tilde{T}_i \mid t)\}$$

that can be used to optimize  $\varpi_l(t)$



## 7 UM Data Analysis (cont'd)

	$(t, t + \Delta t] = (4, 7]$		$(t, t + \Delta t] = (6, 9]$	
	IBS	weights	IBS	weights
SL	0.07584		0.07195	
linear-noCov-value	0.07583	0.00000	0.07199	0.08333
linear-noCov-slope	0.07608	0.00000	0.07155	0.08340
linear-noCov-mean	0.07683	0.00000	0.07236	0.08332
linear-Cov-value	0.07584	1.00000	0.07201	0.08335
linear-Cov-slope	0.07608	0.00000	0.07160	0.08339
linear-Cov-mean	0.07686	0.00000	0.07231	0.08332
nonlinear-noCov-value	0.07693	0.00000	0.07200	0.08334
nonlinear-noCov-slope	0.07672	0.00000	0.07233	0.08331
nonlinear-noCov-mean	0.07760	0.00000	0.07266	0.08329
nonlinear-Cov-value	0.07708	0.00000	0.07218	0.08332
nonlinear-Cov-slope	0.07687	0.00000	0.07219	0.08333
nonlinear-Cov-mean	0.07788	0.00000	0.07277	0.08328

## 7 UM Data Analysis (cont'd)

	$(t, t + \Delta t] = (4, 7]$		$(t, t + \Delta t] = (6, 9]$	
	EPCE	weights	EPCE	weights
SL	0.05231		0.04696	
linear-noCov-value	0.05865	0.08325	0.05003	0.00002
linear-noCov-slope	0.05412	0.08320	0.04861	0.00000
linear-noCov-mean	0.05777	0.08260	0.04764	0.39649
linear-Cov-value	0.05887	0.08215	0.04997	0.00000
linear-Cov-slope	0.05418	0.08333	0.04887	0.00000
linear-Cov-mean	0.05768	0.08270	0.04763	0.12793
nonlinear-noCov-value	0.05656	0.08337	0.04793	0.00136
nonlinear-noCov-slope	0.05199	0.08517	0.04785	0.44966
nonlinear-noCov-mean	0.05882	0.08296	0.04762	0.00961
nonlinear-Cov-value	0.05679	0.08315	0.04867	0.00000
nonlinear-Cov-slope	0.05188	0.08526	0.04820	0.01327
nonlinear-Cov-mean	0.05899	0.08288	0.04764	0.00166