Joint Models for Longitudinal and Survival Data & Dynamic Predictions

Dimitris Rizopoulos
Department of Biostatistics, Erasmus University Medical Center, the Netherlands

d.rizopoulos@erasmusmc.nl

Dynamic predictions for repeated markers and repeated events

October 11th, 2013, Bordeaux

1.1 Introduction

- Over the last 10-15 years increasing interest in joint modeling of longitudinal and time-to-event data (Tsiatis & Davidian, Stat. Sinica, 2004; Yu et al., Stat. Sinica, 2004)

- The majority of the biostatistics literature in this area has focused on:
  - ▷ several extensions of the standard joint model, new estimation approaches, ...

- Recently joint models have been utilized to provide **individualized predictions**
  - ▷ Rizopoulos (Biometrics, 2011); Proust-Lima and Taylor (Biostatistics, 2009); Yu et al. (JASA, 2008)

1.1 Introduction (cont'd)

- **Goals of this talk**:
  - ▷ Introduce joint models
  - ▷ Dynamic individualized predictions of survival probabilities;
  - ▷ Study the importance of the association structure;
  - ▷ Combine predictions from different joint models

1.2 Illustrative Case Study

- Aortic Valve study: Patients who received a human tissue valve in the aortic position
  - ▷ data collected by Erasmus MC (from 1987 to 2008);
    77 received sub-coronary implantation; 209 received root replacement

- Outcomes of interest:
  - ▷ death and re-operation $\longrightarrow$ composite event
  - ▷ aortic gradient

- Research Question:
  - ▷ Can we utilize available aortic gradient measurements to predict survival/re-operation

2.1 Joint Modeling Framework

- To answer our questions of interest we need to postulate a model that relates
  - ▷ the aortic gradient with
  - ▷ the time to death or re-operation

- Problem: Aortic gradient measurement process is an **endogenous** time-dependent covariate (Kalbfleisch and Prentice, 2002, Section 6.3)
  - ▷ Endogenous (aka internal): the future path of the covariate up to any time $t > s$ IS affected by the occurrence of an event at time point s, i.e.,

$$\Pr\{\mathcal{Y}_i(t) \mid \mathcal{Y}_i(s), T_i^* \geq s\} \neq \Pr\{\mathcal{Y}_i(t) \mid \mathcal{Y}_i(s), T_i^* = s\},$$

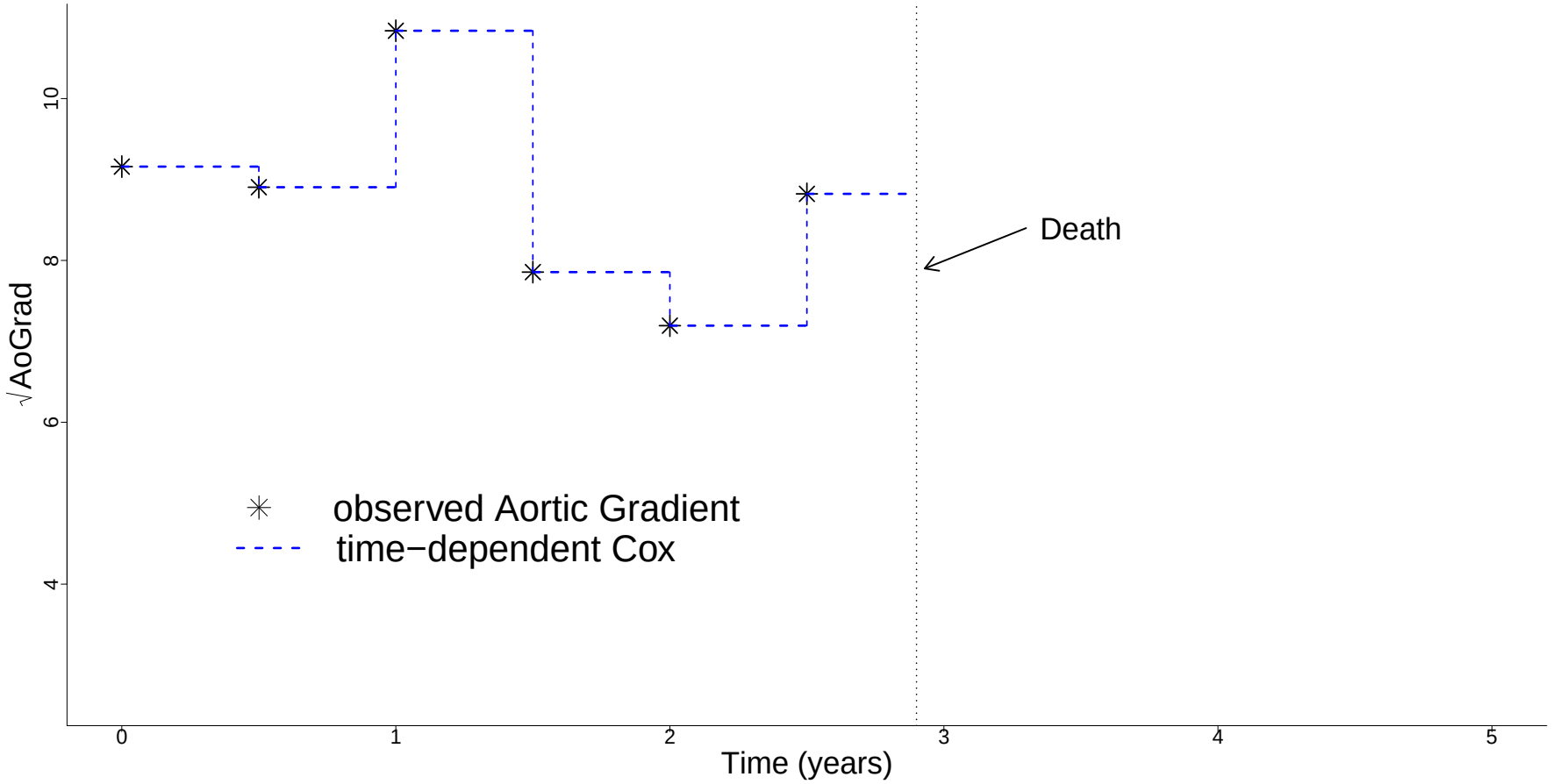
where $0 < s \leq t$ and $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$

2.1 Joint Modeling Framework (cont'd)

- What is special about endogenous time-dependent covariates
  - ▷ measured with error
  - ▷ the complete history is not available
  - ▷ existence directly related to failure status

- What if we use the Cox model?
  - ▷ the association size can be severely underestimated
  - ▷ true potential of the marker will be masked

2.1 Joint Modeling Framework (cont'd)



2.1 Joint Modeling Framework (cont'd)

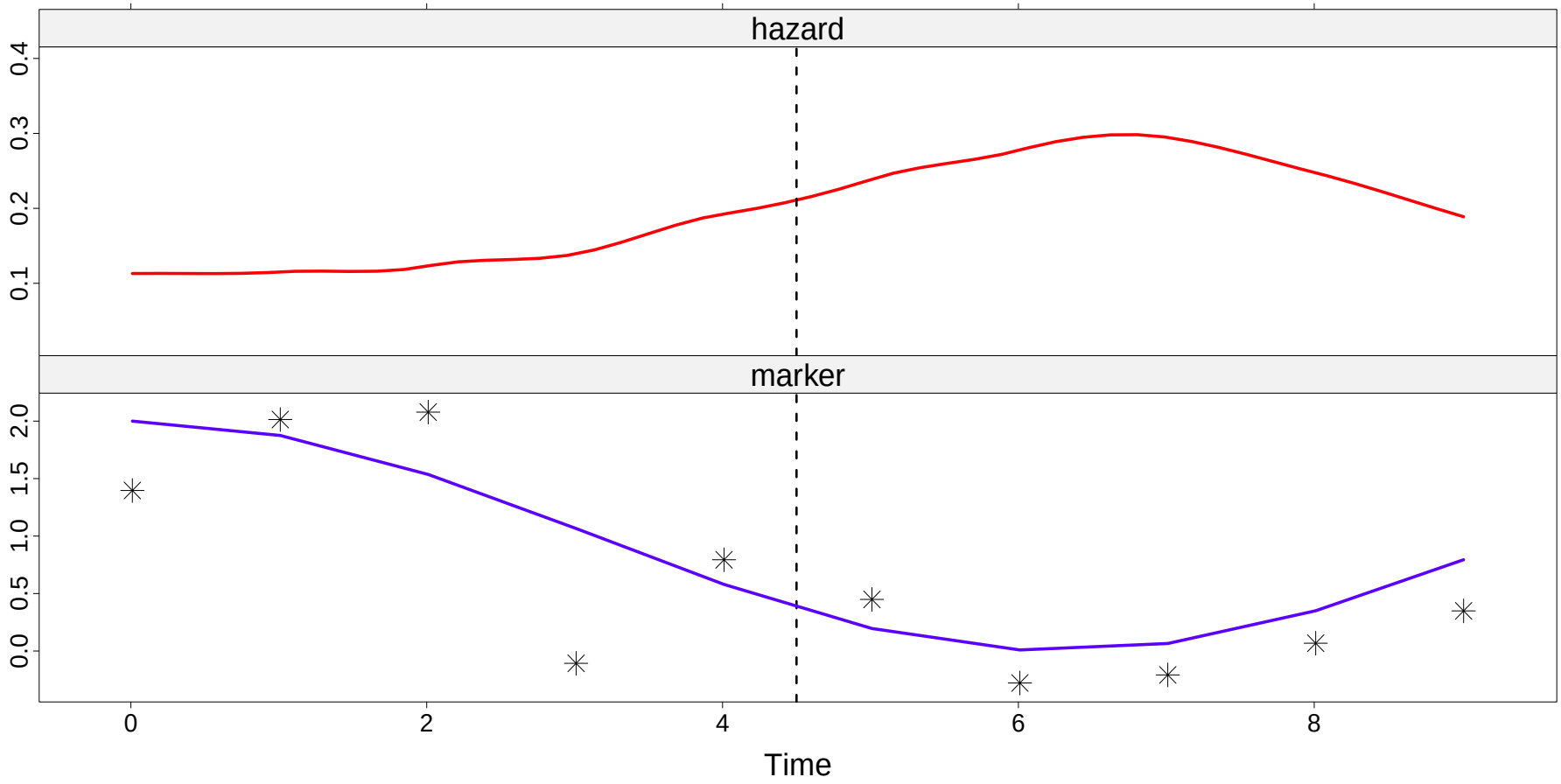
- To account for the special features of these covariates a new class of models has been developed

Joint Models for Longitudinal and Time-to-Event Data

- Intuitive idea behind these models
  1. use an appropriate model to describe the evolution of the marker in time for each patient
  2. the estimated evolutions are then used in a Cox model

- Feature: Marker level is **not** assumed constant between visits

2.1 Joint Modeling Framework (cont'd)



2.1 Joint Modeling Framework (cont'd)

- Some notation
  - ▷ T_i^*: True time-to-death for patient i
  - ▷ T_i: Observed time-to-death for patient i
  - ▷ δ_i: Event indicator, i.e., equals 1 for true events
  - ▷ y_i: Longitudinal aortic gradient measurements

2.1 Joint Modeling Framework (cont'd)

- We define a standard joint model

  ▷ Survival Part: Relative risk model

$$h_i(t \mid \mathcal{M}_i(t)) = h_0(t) \exp\{\gamma^\top w_i + \alpha m_i(t)\},$$

where

* $m_i(t)$ = the *true* & *unobserved* value of aortic gradient at time t
* $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$
* α quantifies the effect of aortic gradient on the risk for death/re-operation
* w_i baseline covariates

2.1 Joint Modeling Framework (cont'd)

▷ Longitudinal Part: Reconstruct $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ using $y_i(t)$ and a mixed effects model (we focus on continuous markers)

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t), \qquad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \end{aligned}$$

where

* $x_i(t)$ and β: Fixed-effects part
* $z_i(t)$ and b_i: Random-effects part, $b_i \sim \mathcal{N}(0, D)$

2.1 Joint Modeling Framework (cont'd)

- The two processes are associated $\Rightarrow$ define a model for their joint distribution

- Joint Models for such joint distributions are of the following form
  (Tsiatis & Davidian, Stat. Sinica, 2004)

$$p(y_i, T_i, \delta_i) = \int p(y_i \mid b_i) \left\{ h(T_i \mid b_i)^{\delta_i} S(T_i \mid b_i) \right\} p(b_i)\, db_i$$

  where

  ▷ b_i a vector of random effects that explains the interdependencies

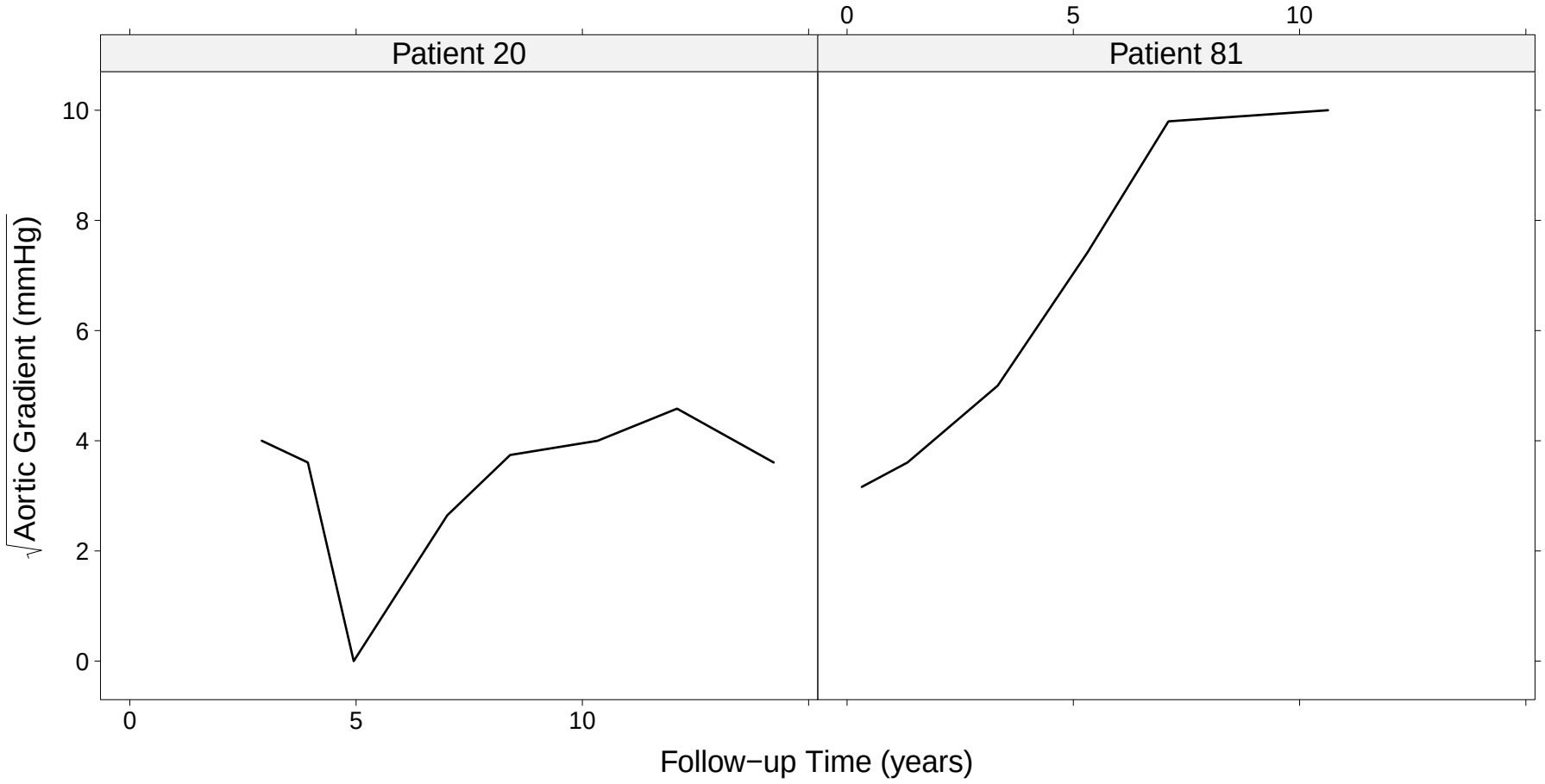
  ▷ $p(\cdot)$ density function; $S(\cdot)$ survival function

3.1 Prediction Survival – Definitions

- We are interested in predicting survival probabilities for a new patient j that has provided a set of aortic gradient measurements up to a specific time point t

- Example: We consider Patients 20 and 81 from the Aortic Valve dataset
  - ▷ **Dynamic Prediction** survival probabilities are dynamically updated as additional longitudinal information is recorded

3.1 Prediction Survival – Definitions (cont'd)



3.1 Prediction Survival – Definitions (cont'd)

- More formally, we have available measurements up to time point t

$$\mathcal{Y}_j(t) = \{y_j(s), 0 \leq s < t\}$$

and we are interested in

$$\pi_j(u \mid t) = \mathsf{Pr}\{T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n\},$$

where

- ▷ where $u > t$, and
- ▷ $\mathcal{D}_n$ denotes the sample on which the joint model was fitted

3.2 Prediction Survival – Estimation

- Joint model is estimated using MCMC or maximum likelihood

- Based on the fitted model we can estimate the conditional survival probabilities
  - ▷ Empirical Bayes
  - ▷ fully Bayes/Monte Carlo (it allows for easy calculation of s.e.)

- For more details check:
  - ▷ Proust-Lima and Taylor (2009, Biostatistics), Rizopoulos (2011, Biometrics), Taylor et al. (2013, Biometrics)

3.2 Prediction Survival – Estimation (cont'd)

- It is convenient to proceed using a Bayesian formulation of the problem $\Rightarrow$ $\pi_j(u \mid t)$ can be written as

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

- The first part of the integrand using CI

$$\Pr\{T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_j(t); \theta\} =$$

$$= \int \frac{S_j\{u \mid \mathcal{M}_j(u, b_j, \theta); \theta\}}{S_i\{t \mid \mathcal{M}_i(t, b_i, \theta); \theta\}}\, p(b_i \mid T_i^* > t, \mathcal{Y}_i(t); \theta)\, db_i$$

3.2 Prediction Survival – Estimation (cont'd)

- A Monte Carlo estimate of $\pi_i(u \mid t)$ can be obtained using the following simulation scheme:

Step 1. draw $\theta^{(\ell)} \sim [\theta \mid \mathcal{D}_n]$ or $\theta^{(\ell)} \sim \mathcal{N}(\hat{\theta}, \widehat{\mathcal{H}})$

Step 2. draw $b_i^{(\ell)} \sim \{b_i \mid T_i^* > t, \mathcal{Y}_i(t), \theta^{(\ell)}\}$

Step 3. compute $\pi_i^{(\ell)}(u \mid t) = S_i\big\{u \mid \mathcal{M}_i(u, b_i^{(\ell)}, \theta^{(\ell)}); \theta^{(\ell)}\big\} \Big/ S_i\big\{t \mid \mathcal{M}_i(t, b_i^{(\ell)}, \theta^{(\ell)}); \theta^{(\ell)}\big\}$

- Repeat Steps 1–3, $\ell = 1, \ldots, L$ times, where L denotes the number of Monte Carlo samples

3.3 Prediction Survival – Illustration

- Example: We fit a joint model to the Aortic Valve data

- Longitudinal submodel
  - ▷ fixed effects: natural cubic splines of time (d.f.$= 3$), operation type, and their interaction
  - ▷ random effects: Intercept, & natural cubic splines of time (d.f.$= 3$)

- Survival submodel
  - ▷ type of operation, age, sex + *underlying* aortic gradient level
  - ▷ log baseline hazard approximated using B-splines

3.3 Prediction Survival – Illustration (cont'd)

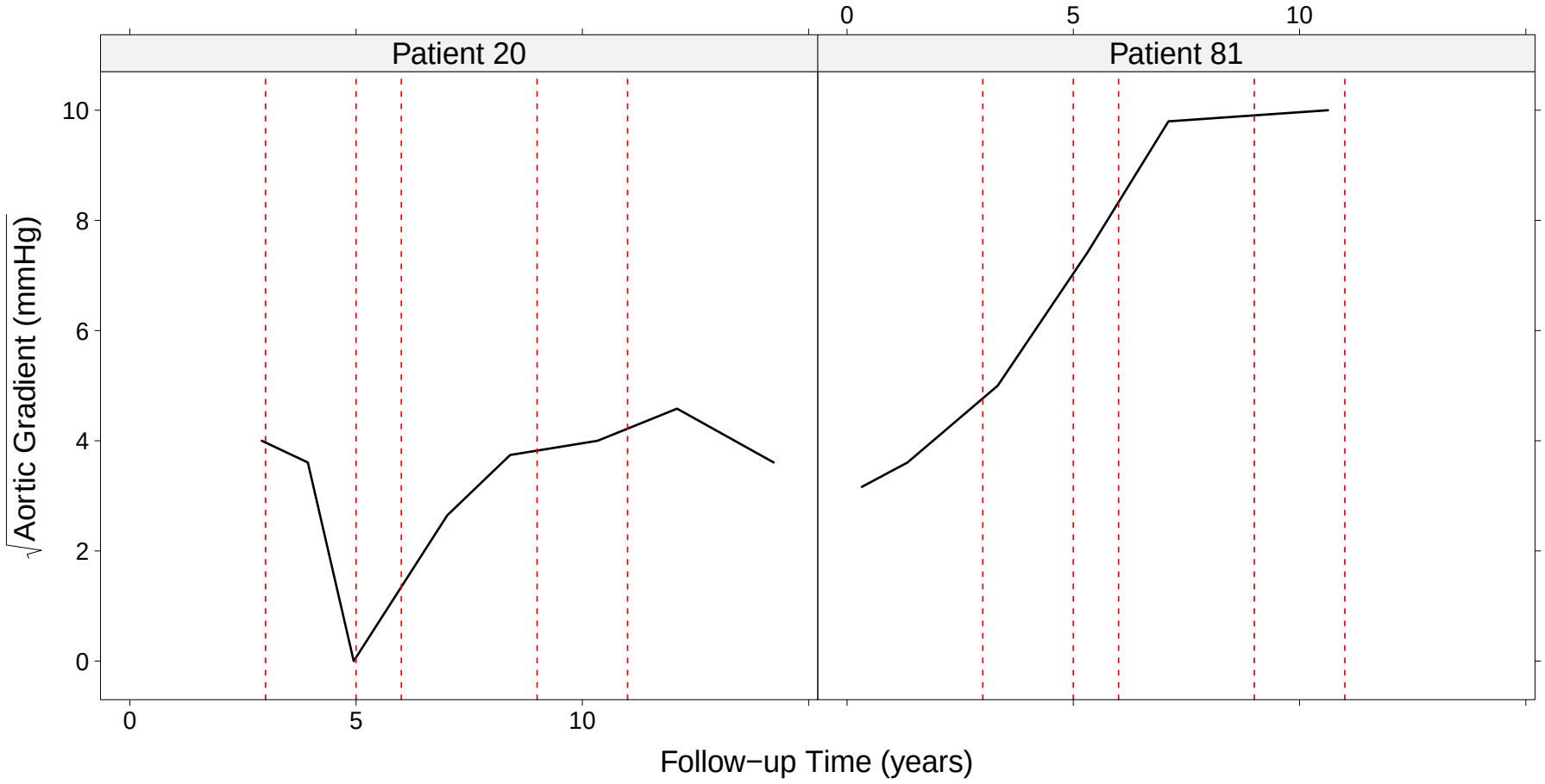
- Based on the fitted joint model we estimate $\pi_j(u \mid t)$ for Patients 20 and 81

- We used the fully Bayesian approach with 500 Monte Carlo samples, and we took as estimate

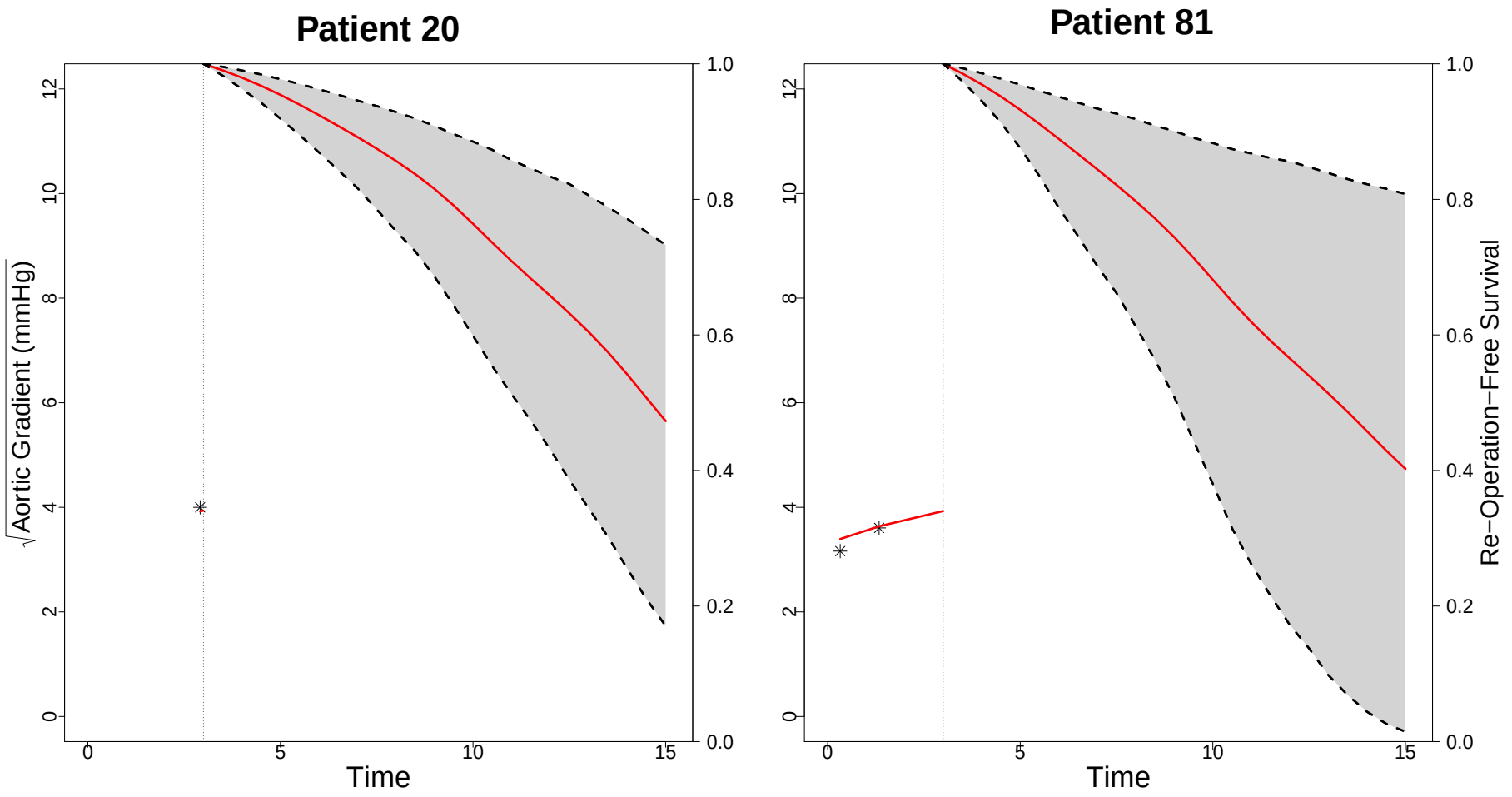
$$\hat{\pi}_j(u \mid t) = \frac{1}{L} \sum_{\ell=1}^{L} \pi_j^{(\ell)}(u \mid t)$$

and calculated the corresponding 95% pointwise CIs

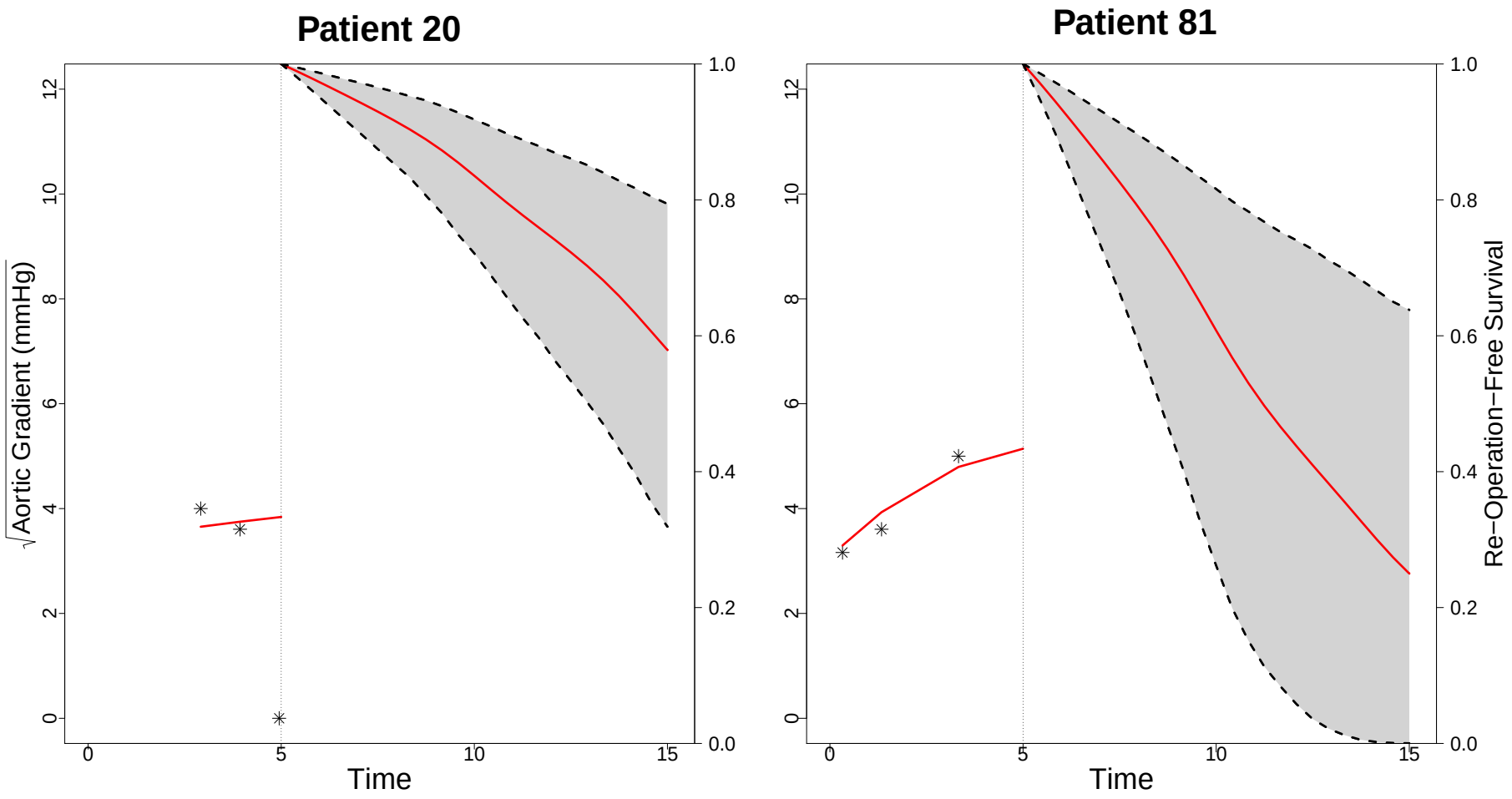
3.3 Prediction Survival – Illustration (cont'd)



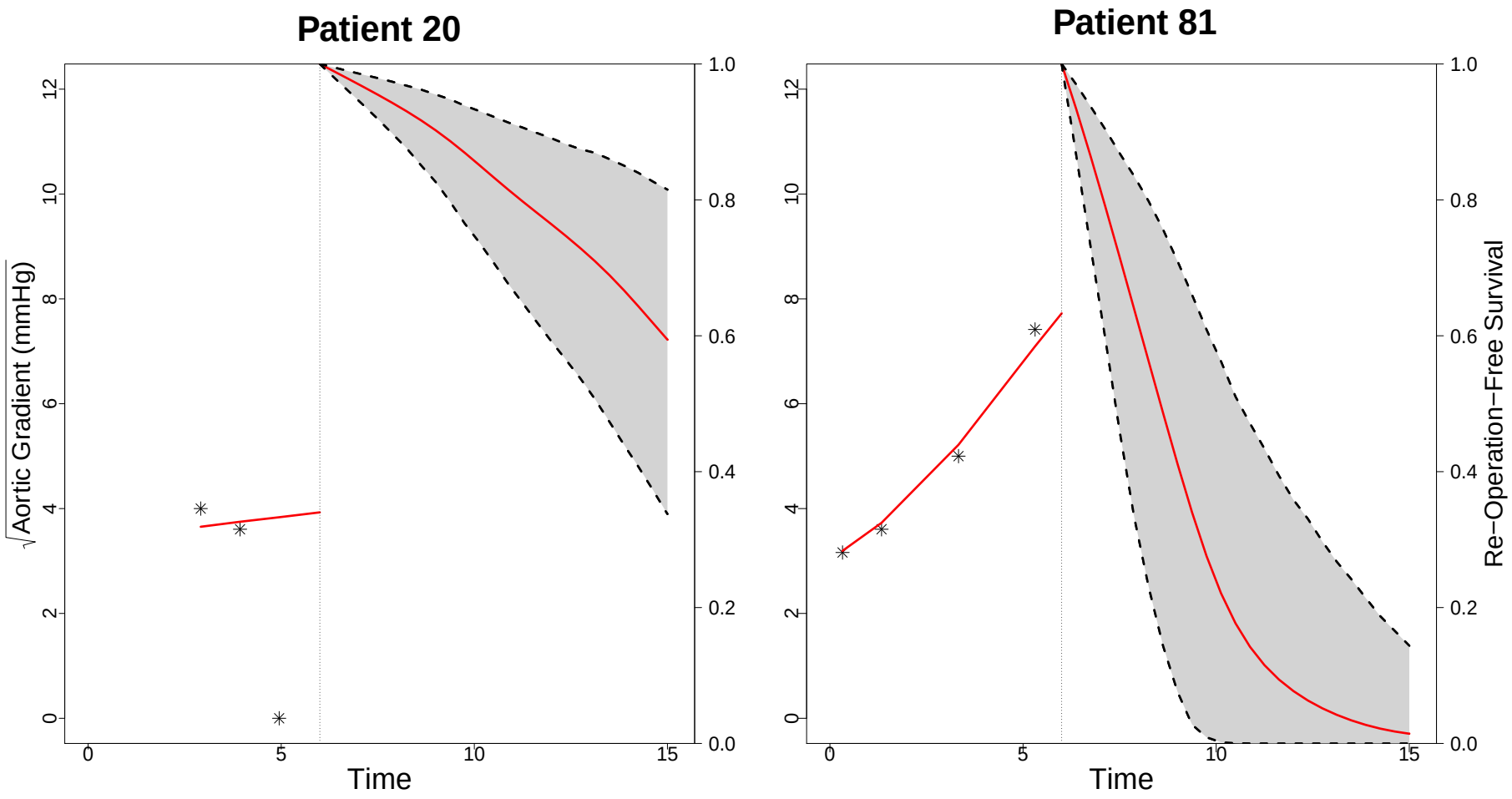
3.3 Prediction Survival – Illustration (cont'd)



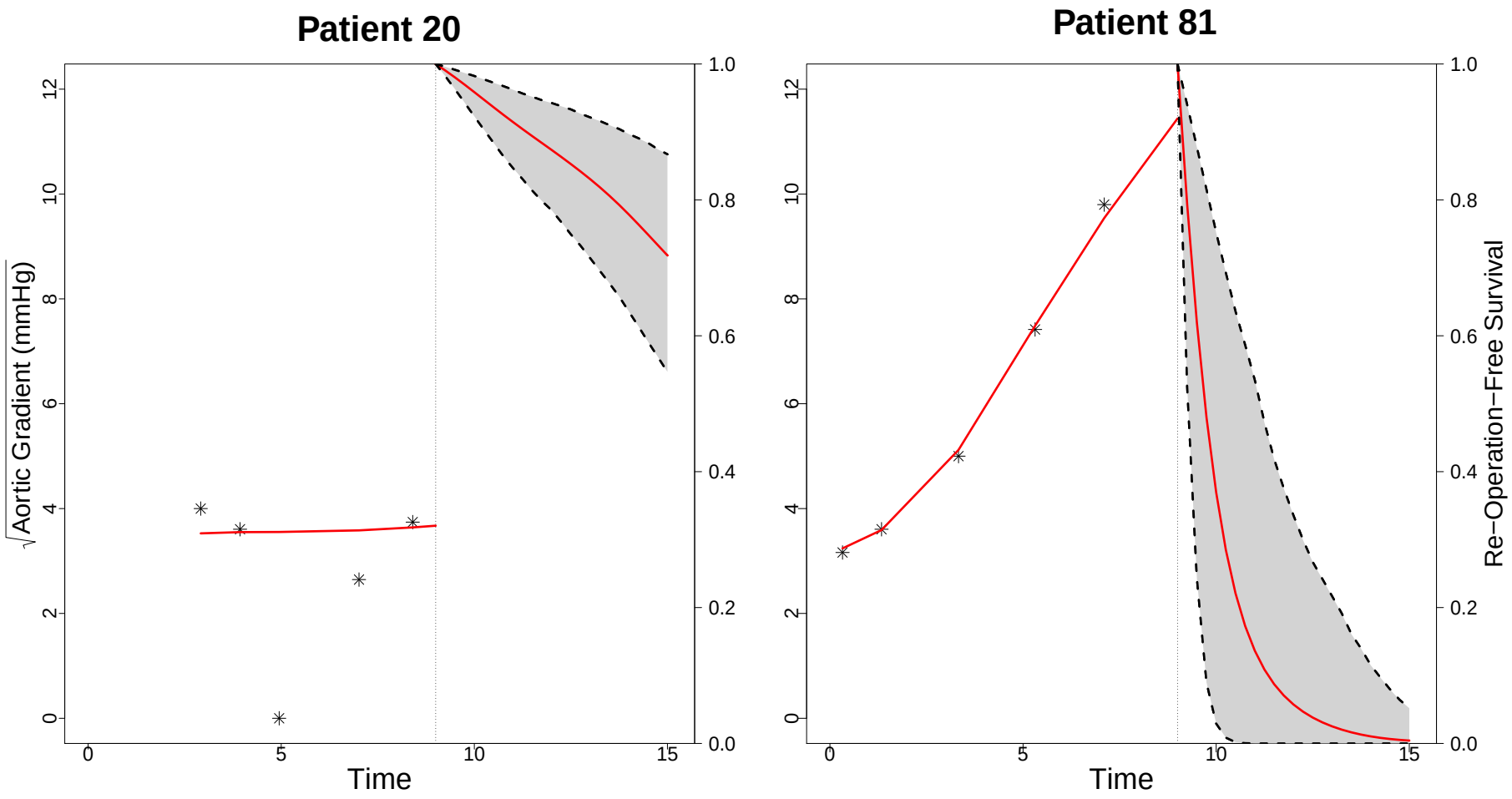
3.3 Prediction Survival – Illustration (cont'd)



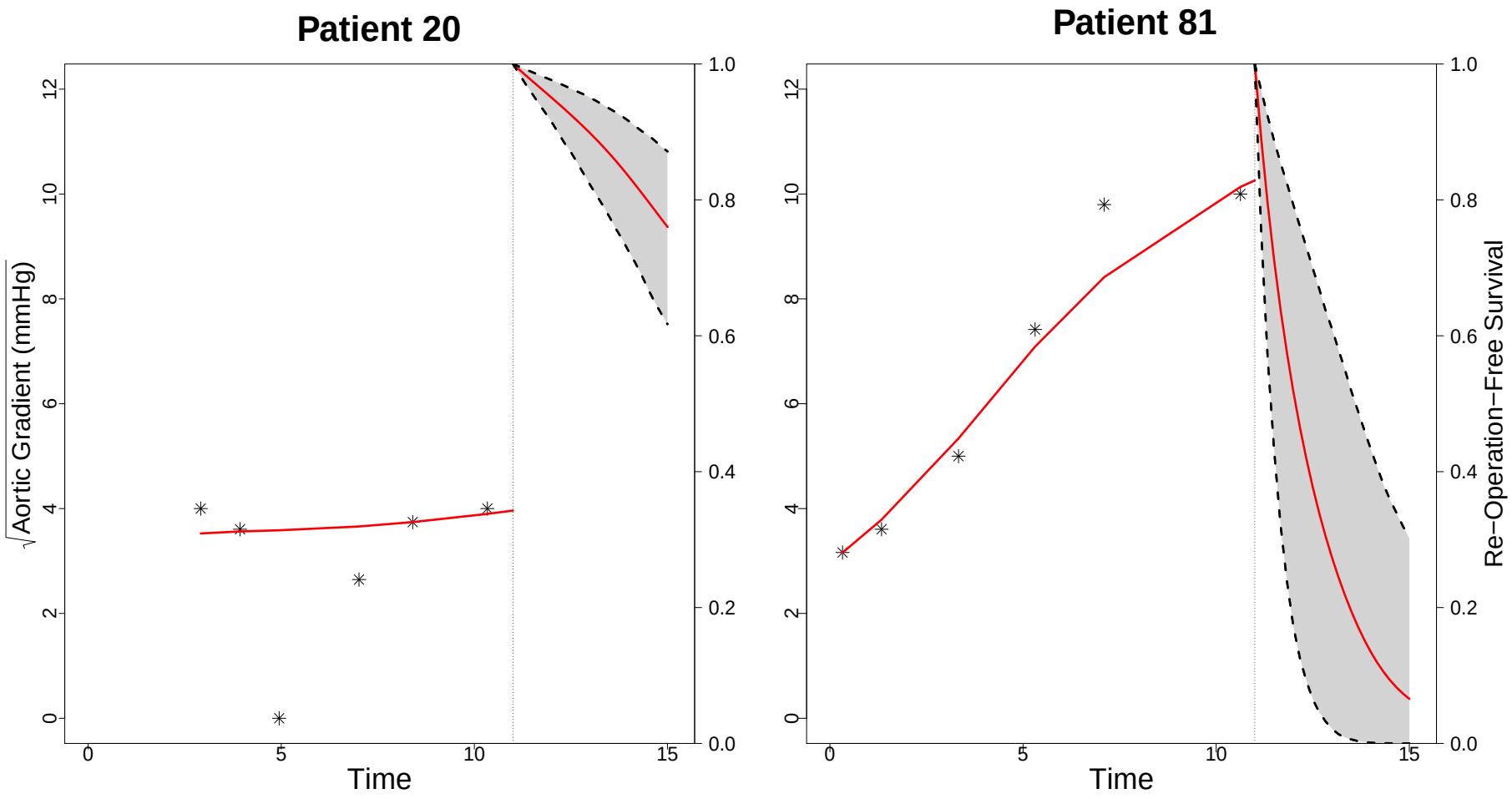
3.3 Prediction Survival – Illustration (cont'd)



3.3 Prediction Survival – Illustration (cont'd)



3.3 Prediction Survival – Illustration (cont'd)



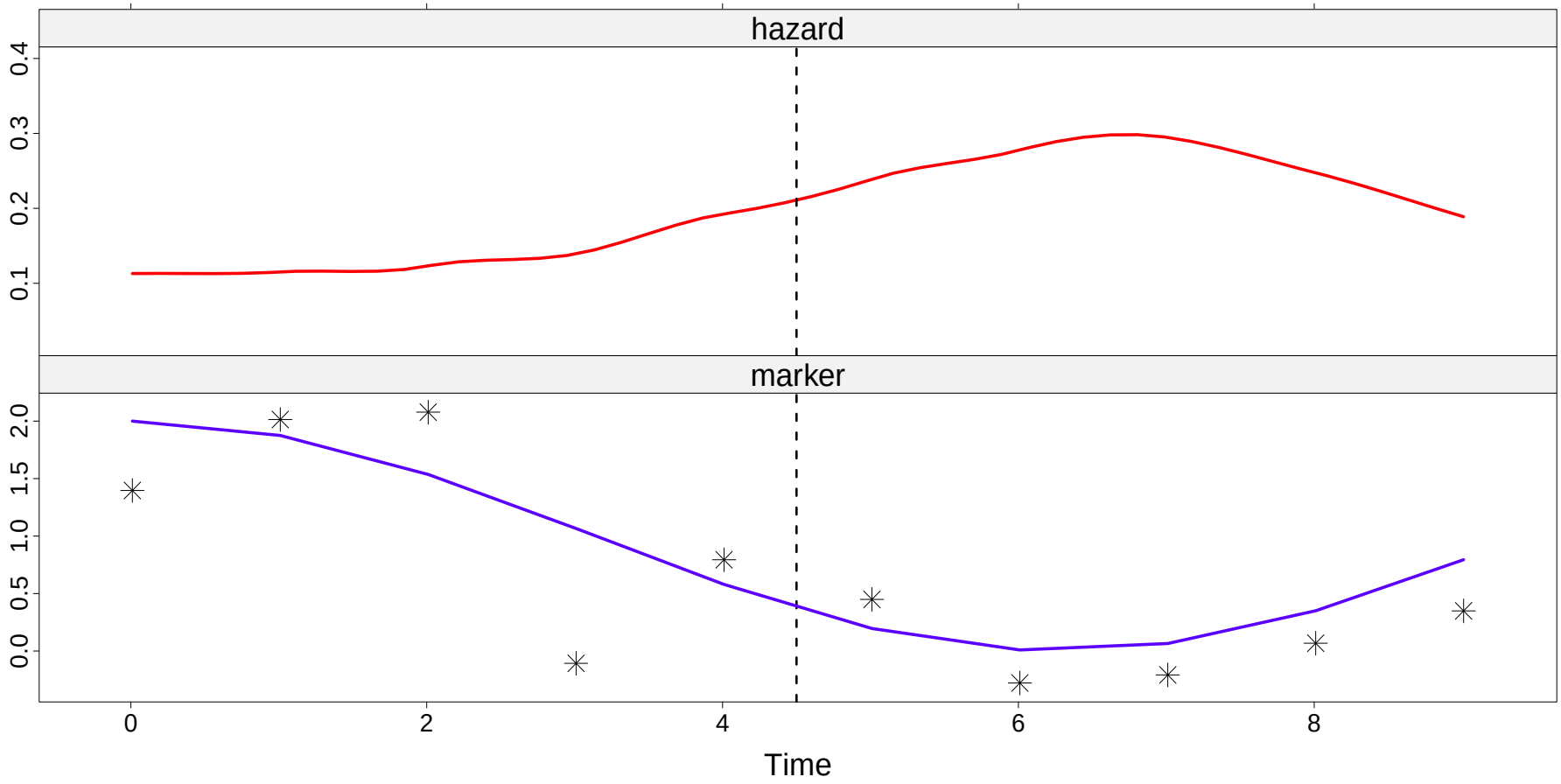
4.1 Association Structures

- The standard joint model

$$\begin{cases} h_i(t \mid \mathcal{M}_i(t)) &= h_0(t)\exp\{\gamma^\top w_i + \alpha m_i(t)\}, \\ \\ y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t), \end{cases}$$

where $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$

4.1 Association structures (cont'd)



4.1 Association Structures (cont'd)

- The standard joint model

$$\begin{cases} h_i(t \mid \mathcal{M}_i(t)) &= h_0(t)\exp\{\gamma^\top w_i + \alpha m_i(t)\}, \\ \\ y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= x_i^\top(t)\beta + z_i^\top(t)b_i + \varepsilon_i(t), \end{cases}$$

where $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$

Is this the only option? Is this the most optimal for prediction?

4.3 Time-dependent Slopes

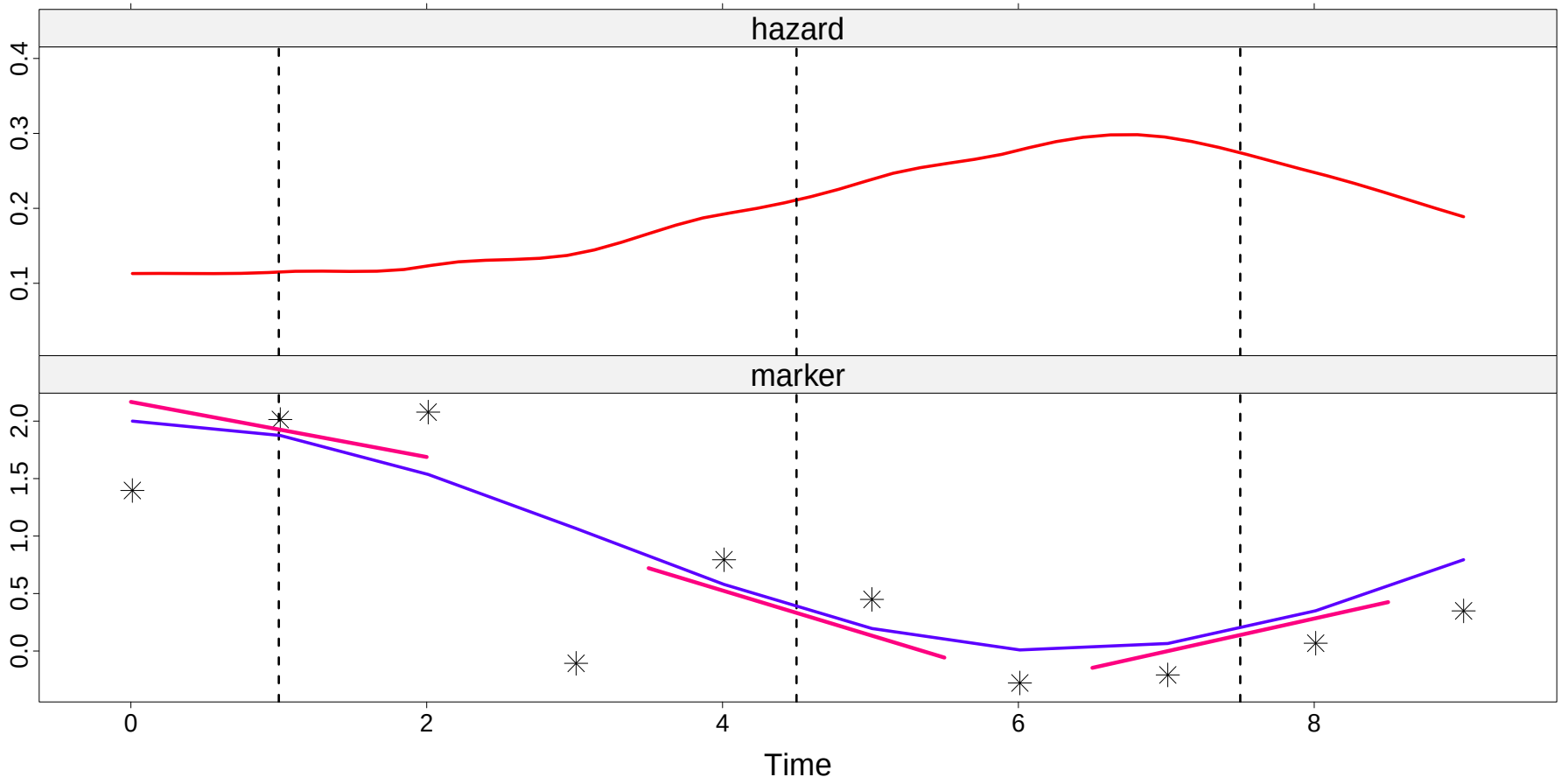
- The hazard for an event at t is associated with both the current value and the slope of the trajectory at t (Ye et al., 2008, Biometrics):

$$h_i(t \mid \mathcal{M}_i(t)) = h_0(t) \exp\{\gamma^\top w_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\},$$

where

$$m_i'(t) = \frac{d}{dt}\{x_i^\top(t)\beta + z_i^\top(t) b_i\}$$

4.3 Time-dependent Slopes (cont'd)



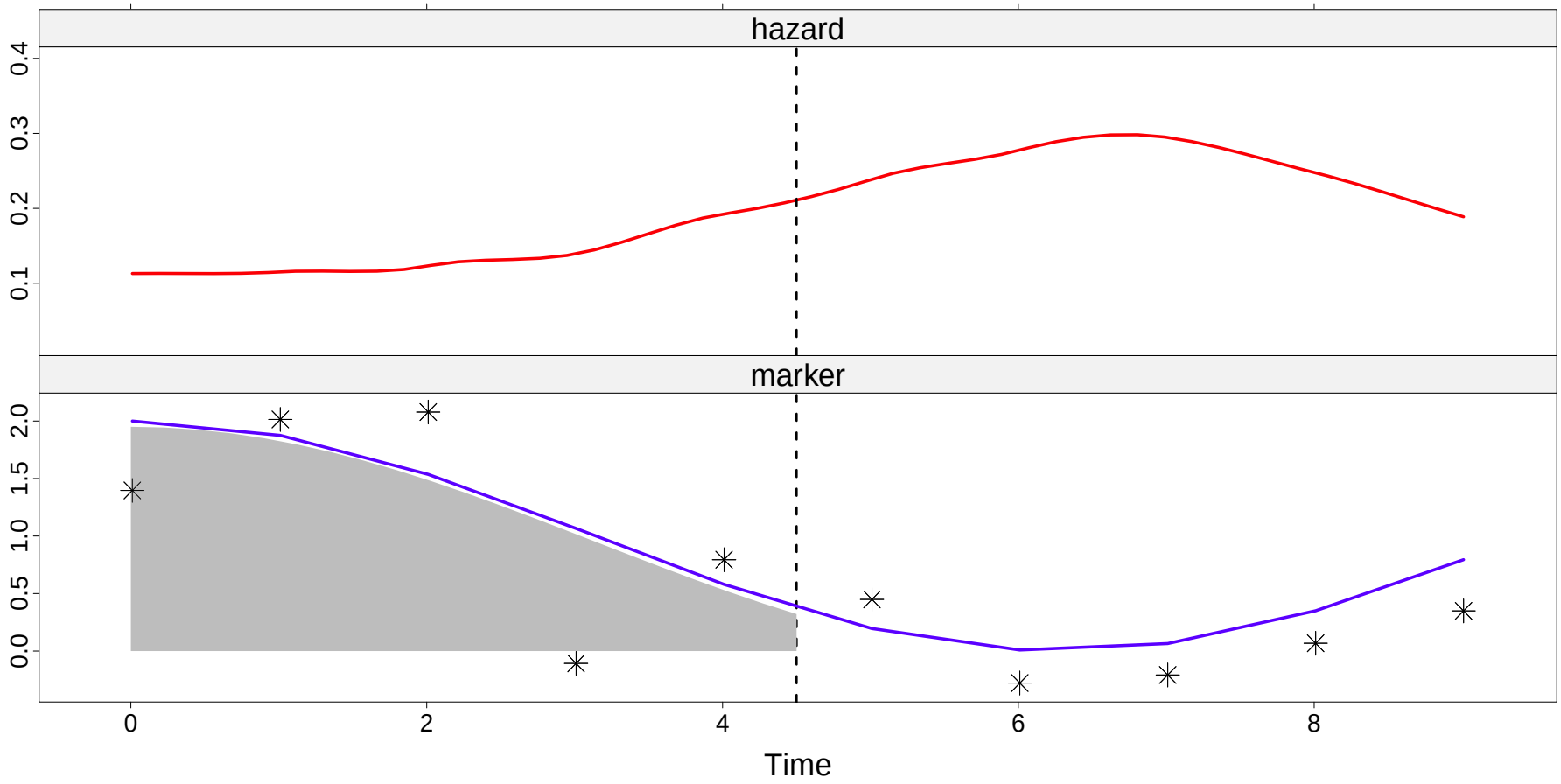
4.4 Cumulative Effects

- The hazard for an event at t is associated with area under the trajectory up to t:

$$h_i(t \mid \mathcal{M}_i(t)) = h_0(t) \exp\left\{\gamma^\top w_i + \alpha \int_0^t m_i(s)\, ds\right\}$$

- Area under the longitudinal trajectory taken as a summary of $\mathcal{M}_i(t)$

4.4 Cumulative Effects (cont'd)



4.5 Weighted Cumulative Effects

- The hazard for an event at t is associated with the area under the weighted trajectory up to t:

$$h_i(t \mid \mathcal{M}_i(t)) \;=\; h_0(t)\exp\Big\{\gamma^\top w_i + \alpha \int_0^t \varpi(t-s) m_i(s)\; ds\Big\},$$

where $\varpi(\cdot)$ appropriately chosen weight function, e.g.,

- ▷ Gaussian density
- ▷ Student's-t density
- ▷ ...

4.6 Shared Random Effects

- The hazard for an event at t is associated with the random effects of the longitudinal submodel:

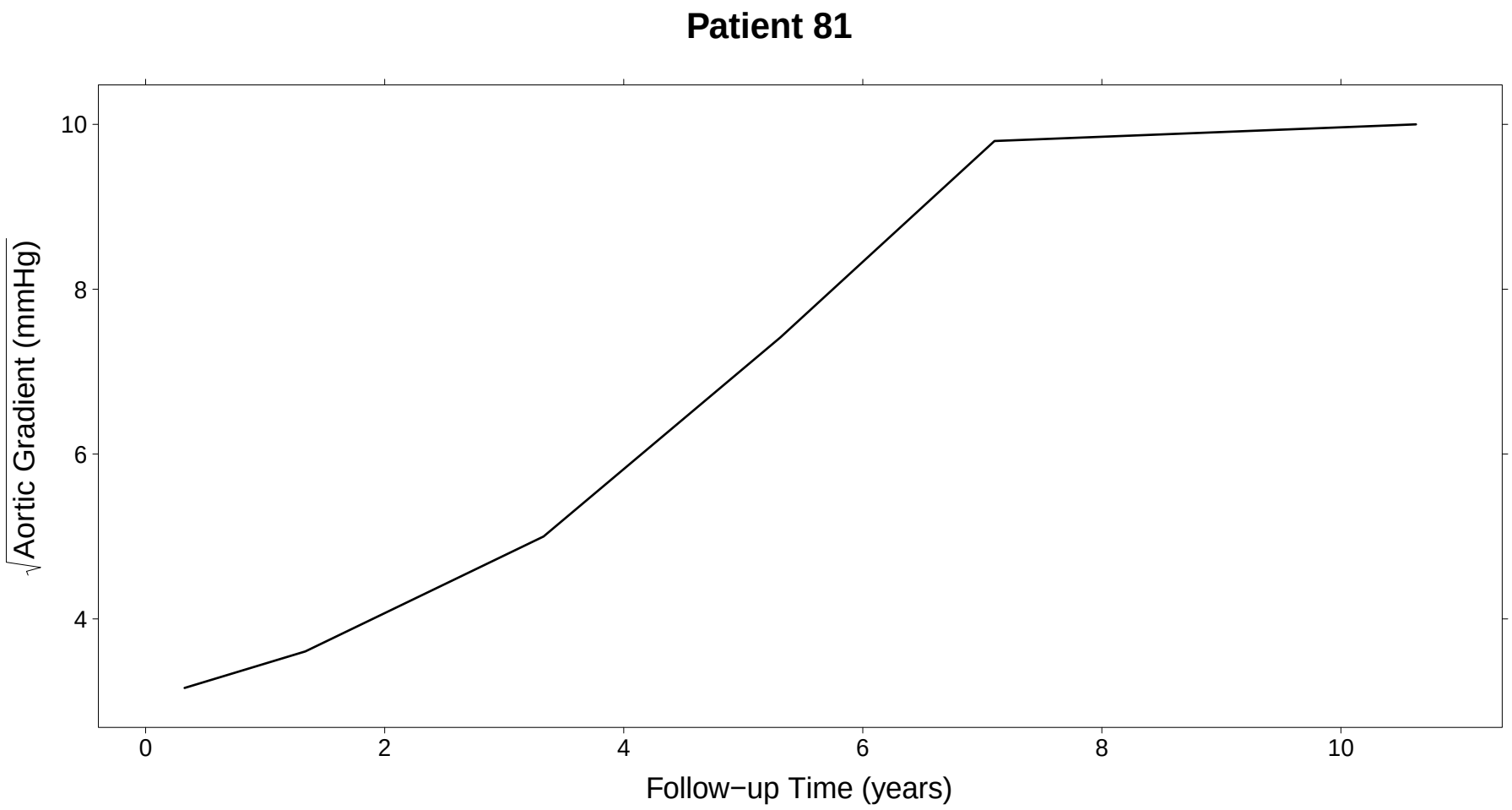
$$h_i(t \mid \mathcal{M}_i(t)) \;=\; h_0(t)\exp(\gamma^\top w_i + \alpha^\top b_i)$$

Features

▷ time-independent (no need to approximate the survival function)

▷ interpretation more difficult when we use something more than random-intercepts & random-slopes

4.7 Parameterizations & Predictions



4.7 Parameterizations & Predictions (cont'd)

- Five joint models for the Aortic Valve dataset
  - ▷ the same longitudinal submodel, and
  - ▷ relative risk submodels

$$h_i(t) = h_0(t)\exp\{\gamma_1 \texttt{TypeOP}_i + \gamma_2 \texttt{Sex}_i + \gamma_3 \texttt{Age}_i + \alpha_1 m_i(t)\},$$

$$h_i(t) = h_0(t)\exp\{\gamma_1 \texttt{TypeOP}_i + \gamma_2 \texttt{Sex}_i + \gamma_3 \texttt{Age}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\},$$

$$h_i(t) = h_0(t)\exp\left\{\gamma_1 \texttt{TypeOP}_i + \gamma_2 \texttt{Sex}_i + \gamma_3 \texttt{Age}_i + \alpha_1 \int_0^t m_i(s)ds\right\}$$

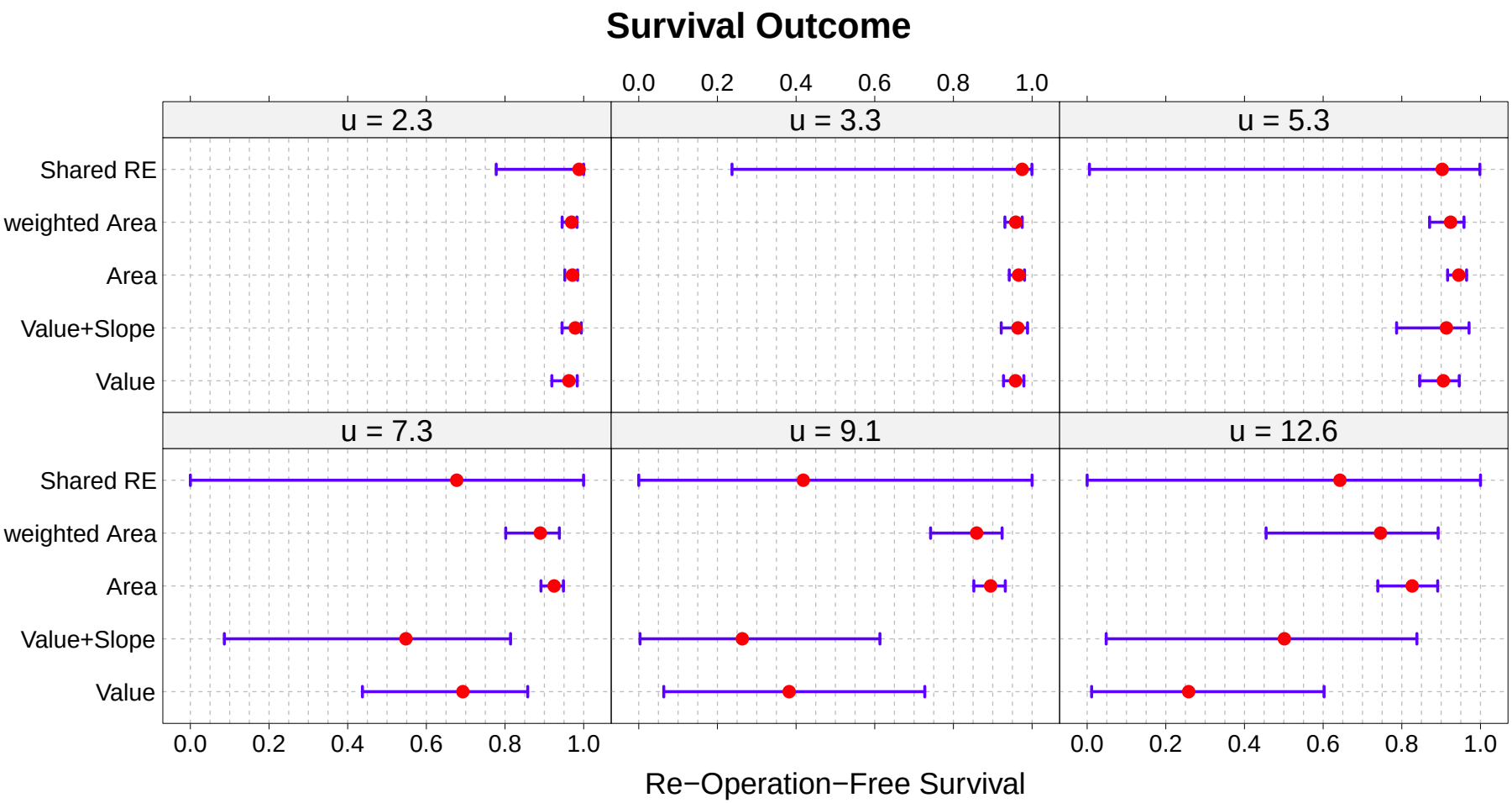
4.7 Parameterizations & Predictions (cont'd)

$$h_i(t) = h_0(t)\exp\Big\{\gamma_1 \texttt{TypeOP}_i + \gamma_2 \texttt{Sex}_i + \gamma_3 \texttt{Age}_i + \alpha_1 \int_0^t \varpi(t-s) m_i(s) ds\Big\},$$

where $\varpi(t-s) = \phi(t-s)/\{\Phi(t) - 0.5\}$, with $\phi(\cdot)$ and $\Phi(\cdot)$ the normal pdf and cdf, respectively

$$h_i(t) = h_0(t)\exp(\gamma_1 \texttt{TypeOP}_i + \gamma_2 \texttt{Sex}_i + \gamma_3 \texttt{Age}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1} + \alpha_3 b_{i2} + \alpha_4 b_{i4})$$

4.7 Parameterizations & Predictions (cont'd)



4.7 Parameterizations & Predictions (cont'd)

- The chosen parameterization can influence the derived predictions
  - ▷ especially for the survival outcome

How to choose between the competing association structures?

4.7 Parameterizations & Predictions (cont'd)

- The easy answer is to employ information criteria, e.g., AIC, BIC, DIC, ...

- However, a problem is that the longitudinal information dominates the joint likelihood
  $\Rightarrow$ will not be sensitive enough wrt predicting survival probabilities

- In addition, thinking a bit more deeply, is the same single model the most appropriate
  - ▷ for all future patients?
  - ▷ for the same patient during the whole follow-up?

The most probable answer is No

4.8 Combining Joint Models

- To address this issue we will use Bayesian Model Averaging (BMA) ideas

- In particular, we assume $M_1, \ldots, M_K$
  - ▷ different association structures
  - ▷ different baseline covariates in the survival submodel
  - ▷ different formulation of the mixed model
  - ▷ ...

- Typically, this list of models will not be exhaustive

4.8 Combining Joint Models (cont'd)

- The aim is the same as before, using the available information for a future patient j up to time t, i.e.,
  - $\triangleright\ T_j^* > t$
  - $\triangleright\ \mathcal{Y}_j(t) = \{y_j(s), 0 \leq s \leq t\}$

- We want to estimate

$$\pi_j(u \mid t) = \mathsf{Pr}\{T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_j(t), \mathcal{D}_n\},$$

  by averaging over the posited joint models

4.8 Combining Joint Models (cont'd)

- More formally we have

$$\Pr\{T_j^* \geq u \mid \mathcal{D}_j(t), \mathcal{D}_n\} = \sum_{k=1}^{K} \Pr(T_j^* > u \mid M_k, \mathcal{D}_j(t), \mathcal{D}_n)\, p(M_k \mid \mathcal{D}_j(t), \mathcal{D}_n)$$

  where

  - $\mathcal{D}_j(t) = \{T_j^* > t, y_j(s), 0 \leq s \leq t\}$
  - $\mathcal{D}_n = \{T_i, \delta_i, y_i, i = 1, \ldots, n\}$

- The first part, $\Pr(T_j^* > u \mid M_k, \mathcal{D}_j(t), \mathcal{D}_n)$, the same as before
  - i.e., model-specific conditional survival probabilities

4.8 Combining Joint Models (cont'd)

- Working out the marginal distribution of each competing model we found some very attractive features of BMA,

$$p(M_k \mid \mathcal{D}_j(t), \mathcal{D}_n) = \frac{p(\mathcal{D}_j(t) \mid M_k)\, p(\mathcal{D}_n \mid M_k)\, p(M_k)}{\sum_{\ell=1}^{K} p(\mathcal{D}_j(t) \mid M_\ell)\, p(\mathcal{D}_n \mid M_\ell)\, p(M_\ell)}$$

  - ▷ $p(\mathcal{D}_n \mid M_k)$ marginal likelihood based on the available data
  - ▷ $p(\mathcal{D}_j(t) \mid M_k)$ marginal likelihood based on the new data of patient j

Model weights are both patient- and time-dependent

4.8 Combining Joint Models (cont'd)

- For different subjects, and even for the same subject but at different times points, different models may have higher posterior probabilities

$$\Downarrow$$

Predictions better tailored to each subject than in standard prognostic models

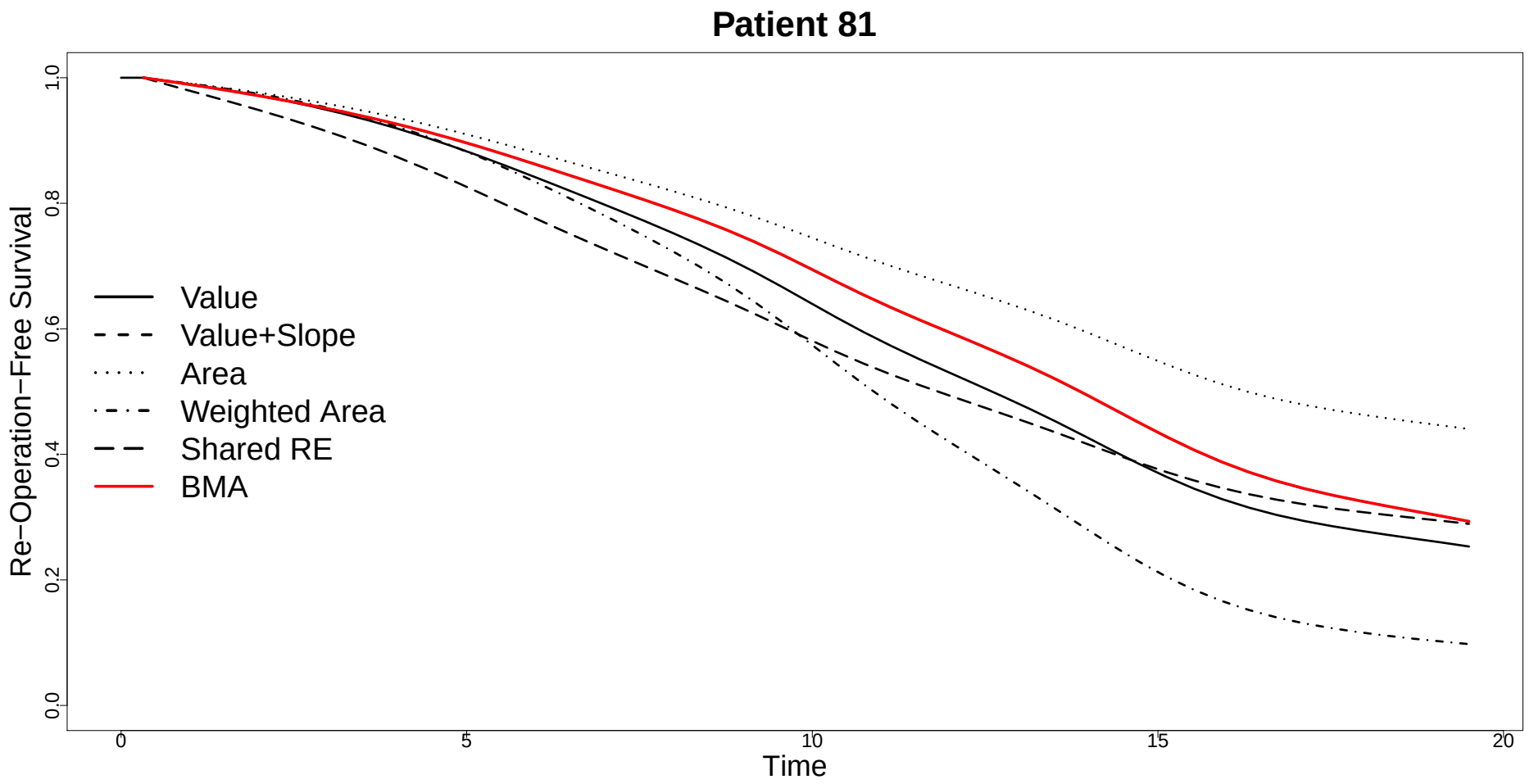
- In addition, the longitudinal model likelihood, which is
  - ▷ hidden in $p(\mathcal{D}_n \mid M_k)$, and
  - ▷ is not affected by the chosen association structure

  will cancel out because it is both in the numerator and denominator

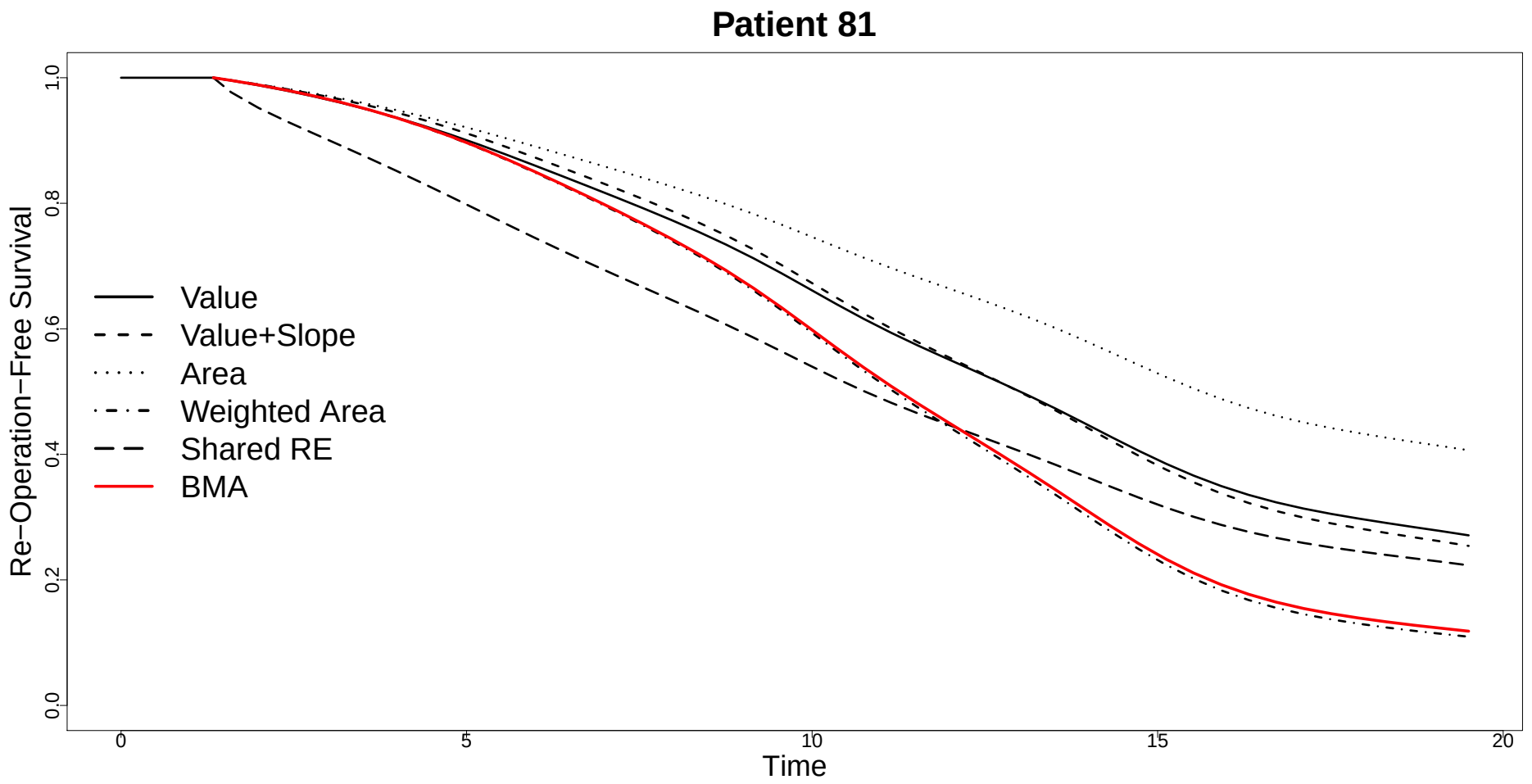
4.8 Combining Joint Models (cont'd)

- Example: Based on the five fitted joint models
  - ▷ we compute BMA predictions for Patient 81, and
  - ▷ compare with the predictions from each individual model

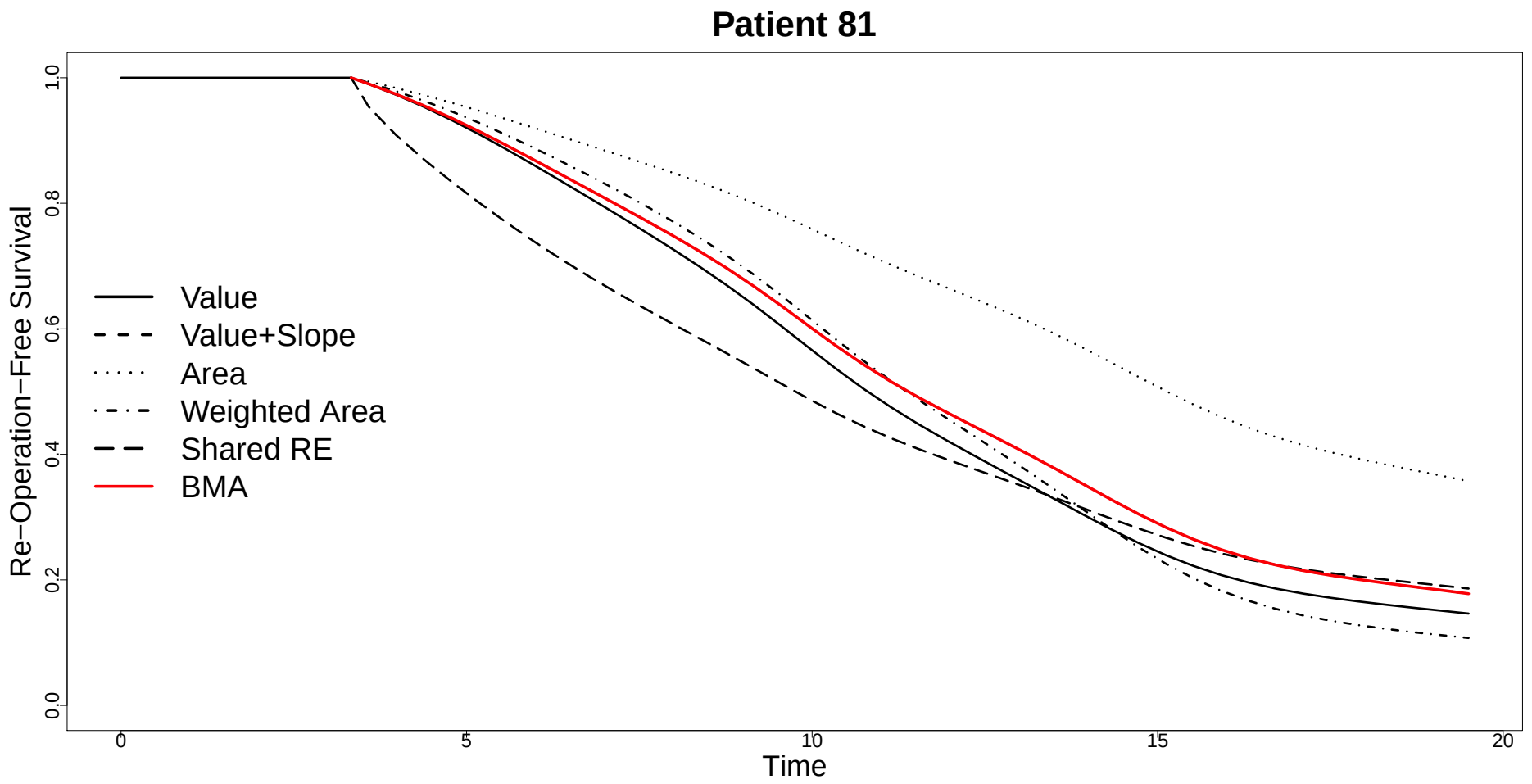
4.8 Combining Joint Models (cont'd)



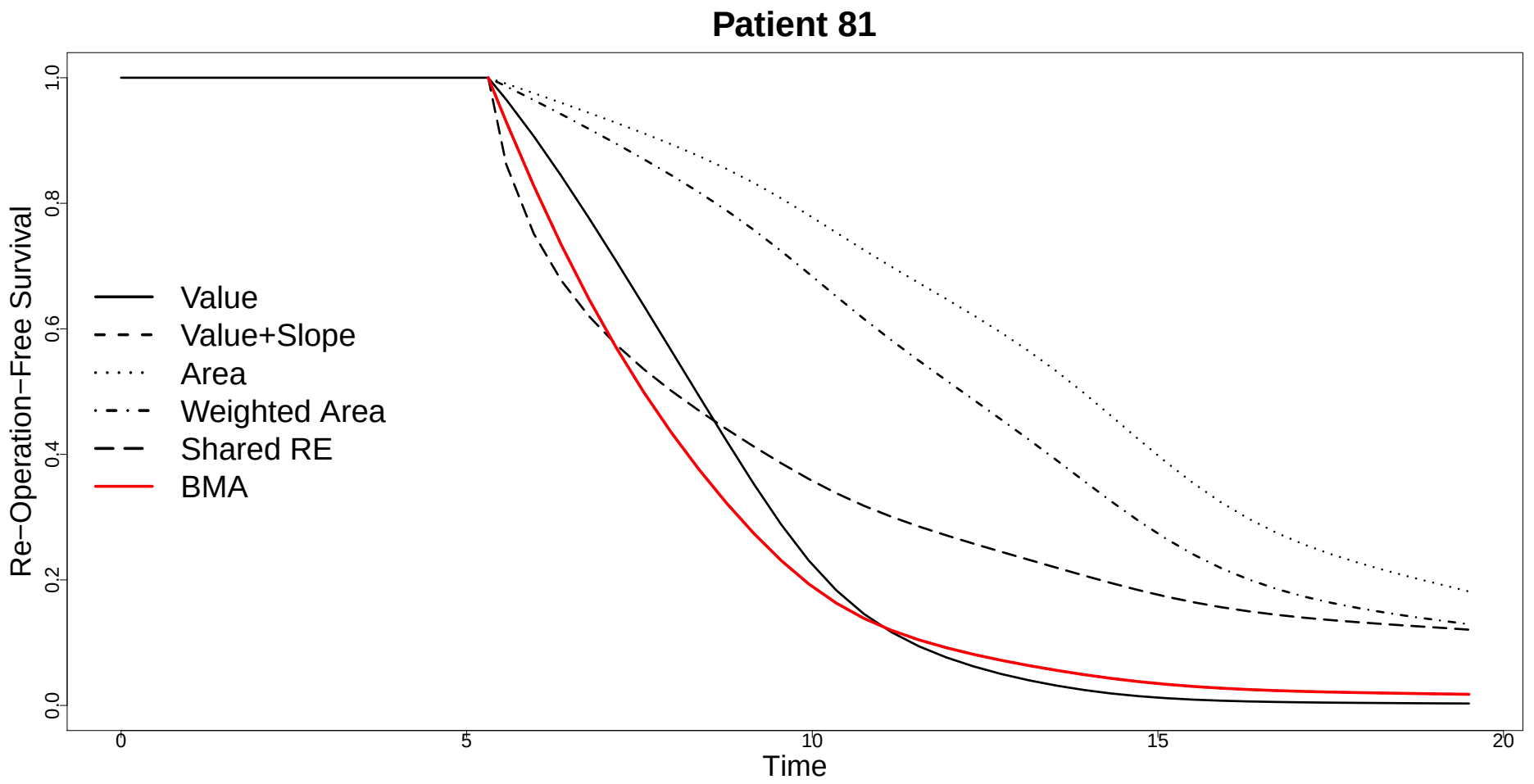
4.8 Combining Joint Models (cont'd)



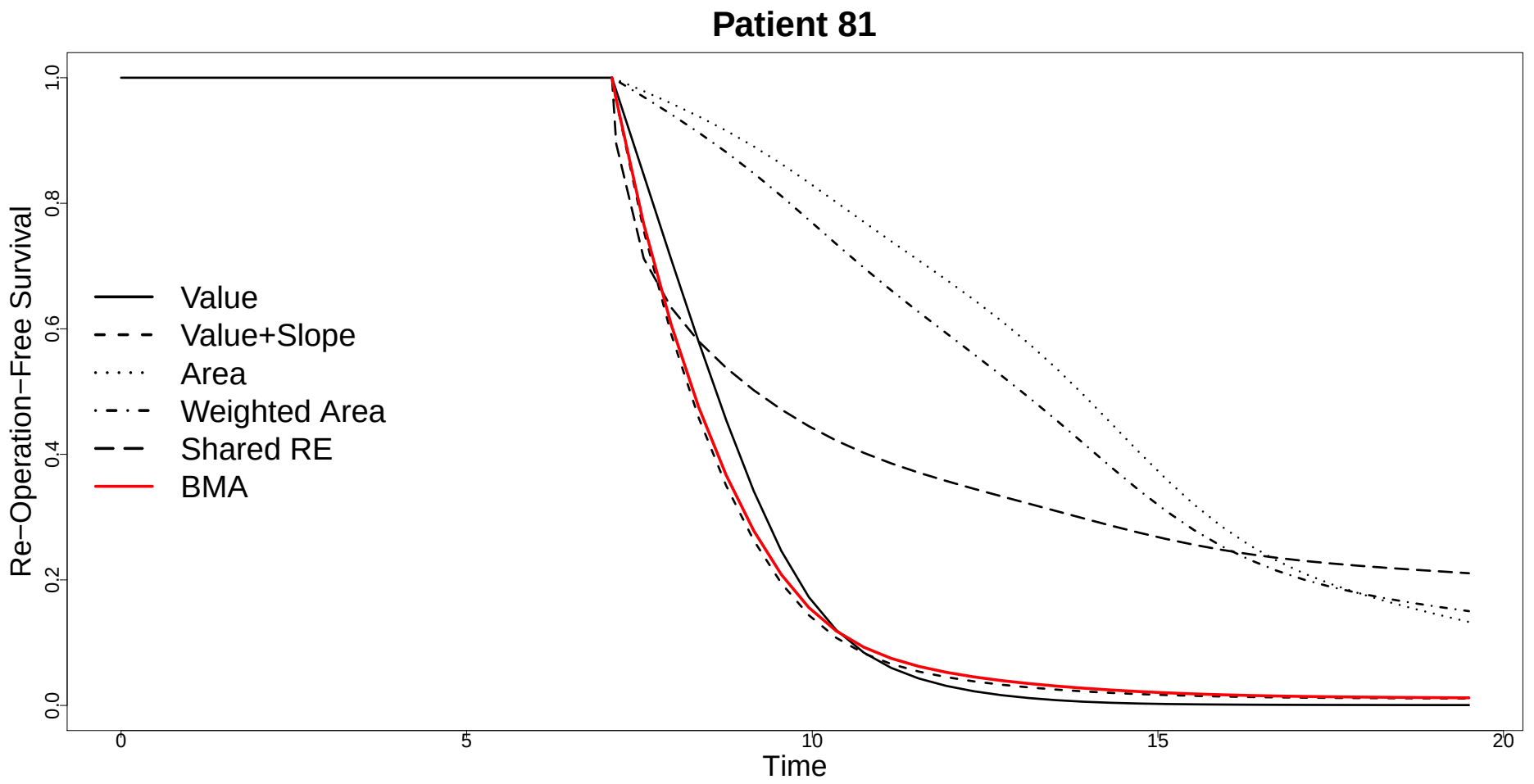
4.8 Combining Joint Models (cont'd)



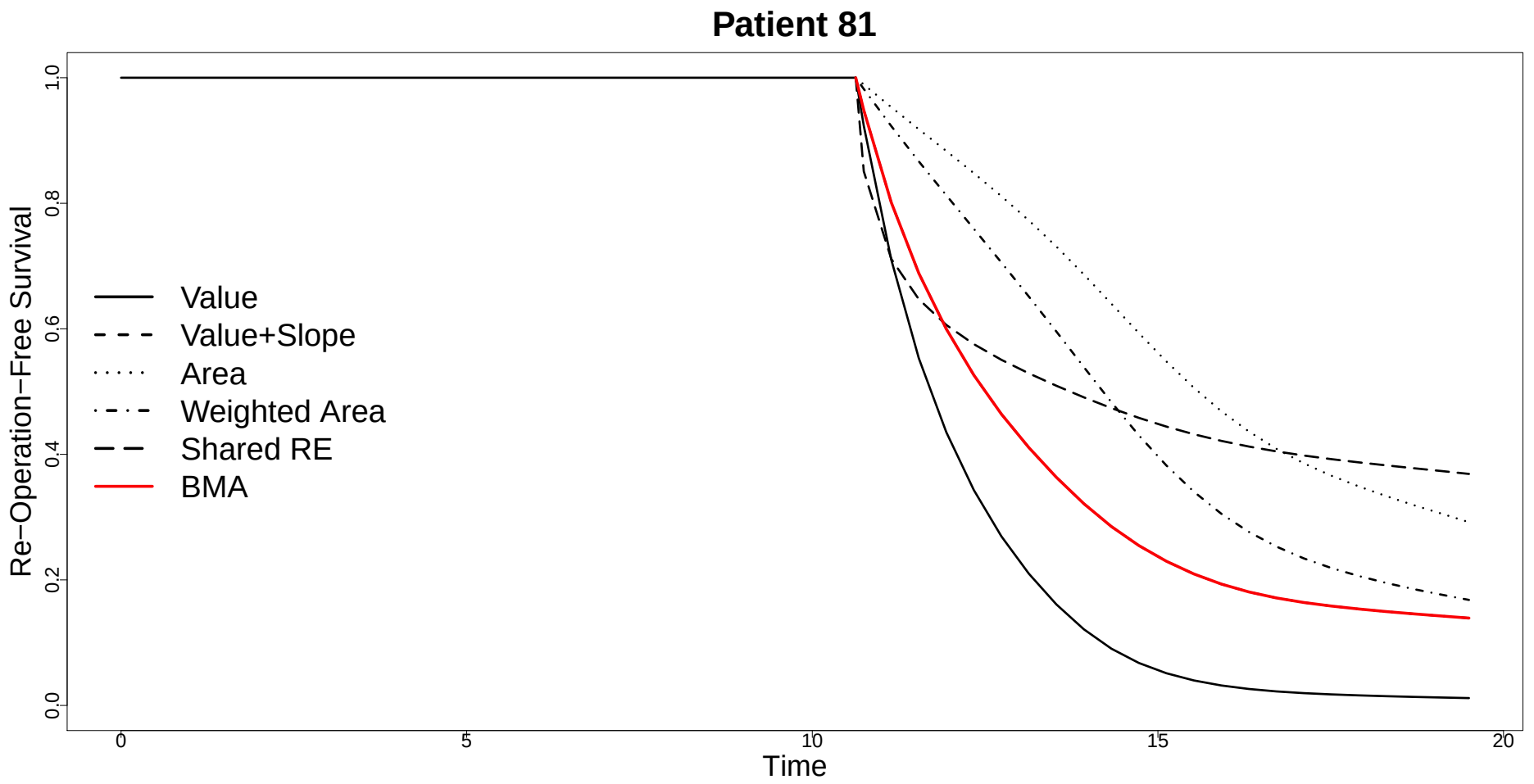
4.8 Combining Joint Models (cont'd)



4.8 Combining Joint Models (cont'd)



4.8 Combining Joint Models (cont'd)



5. Software – I

- Software: R package **JM** freely available via `http://cran.r-project.org/package=JM`
  - ▷ it can fit a variety of joint models + many other features
  - ▷ relevant to this talk: Functions `survfitJM()` and `predict()`

- More info available at:

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data, with Applications in R*. Boca Raton: Chapman & Hall/CRC.

Web site: `http://jmr.r-forge.r-project.org/`

5. Software – II

- Software: R package **JMbayes** freely available via `http://cran.r-project.org/package=JMbayes`
  - ▷ it can fit a variety of joint models + many other features
  - ▷ relevant to this talk: Functions `survfitJM()`, `predict()` and `bma.combine()`

Thank you for your attention!