Using Joint Models to Estimate Causal Effects for Salvage Therapy after Prostatectomy

Dimitris Rizopoulos[1], **Jeremy M.G. Taylor**[2] **and Grigorios Papageorgiou**[1]

[1]Department of Biostatistics, Erasmus Medical Center Rotterdam
[2]Department of Biostatistics, University of Michigan

d.rizopoulos@erasmusmc.nl

@drizopoulos

1 Background & Aim

- Setting Patients treated with surgery after diagnosis of Prostate Cancer (PCa)
  - ▷ *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 & Aim (cont'd)

- Important questions regarding Salvage Therapy
  - ▷ *who should take it?*
  - ▷ *when to start?*
  - ▷ *does it work?*

1 Background & Aim (cont'd)

Quantify the amount by which Salvage Therapy reduces the risk of metastasis

1 Background & Aim (cont'd)

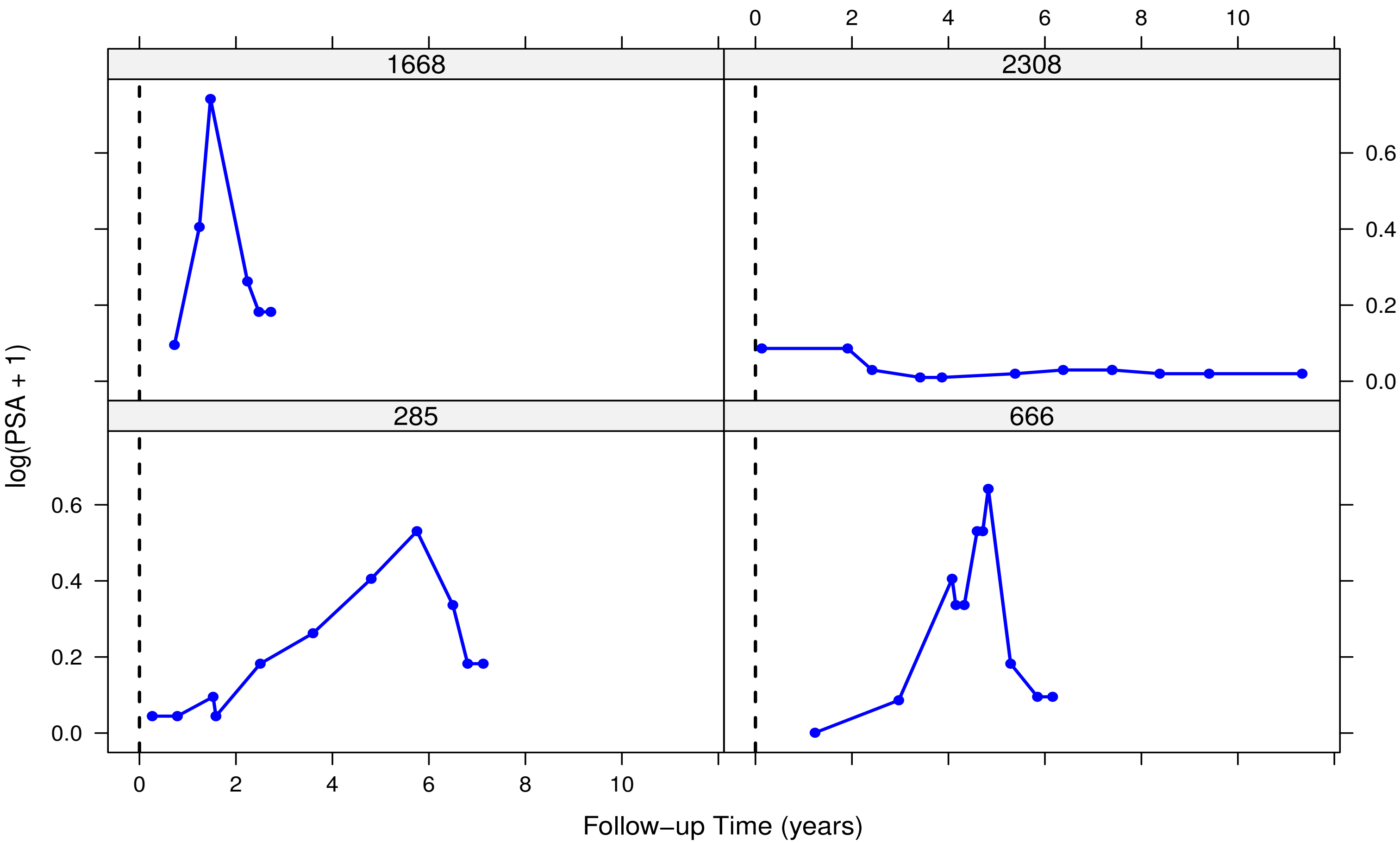
- University of Michigan Prostatectomy Data

  ▷ 3634 PCa patients followed-up in 1996–2013

  * aged 40 to 84 years with clinically localized cT1 to cT3 disease
  * received radical prostatectomy

  ▷ baseline variables: PSA, Gleason, T-stage, age, race, gland volume, perineural invasion, planned adjuvant therapy

1 Background & Aim (cont'd)



1 Background & Aim (cont'd)

- Challenges

  ▷ *Observational Data – no RCT*

  * selection bias
  * ascertainment bias

  ▷ *Time-Varying Salvage Therapy*

  * depends on previous PSA
  * PSA time-dependent confounder
  * endogeneity

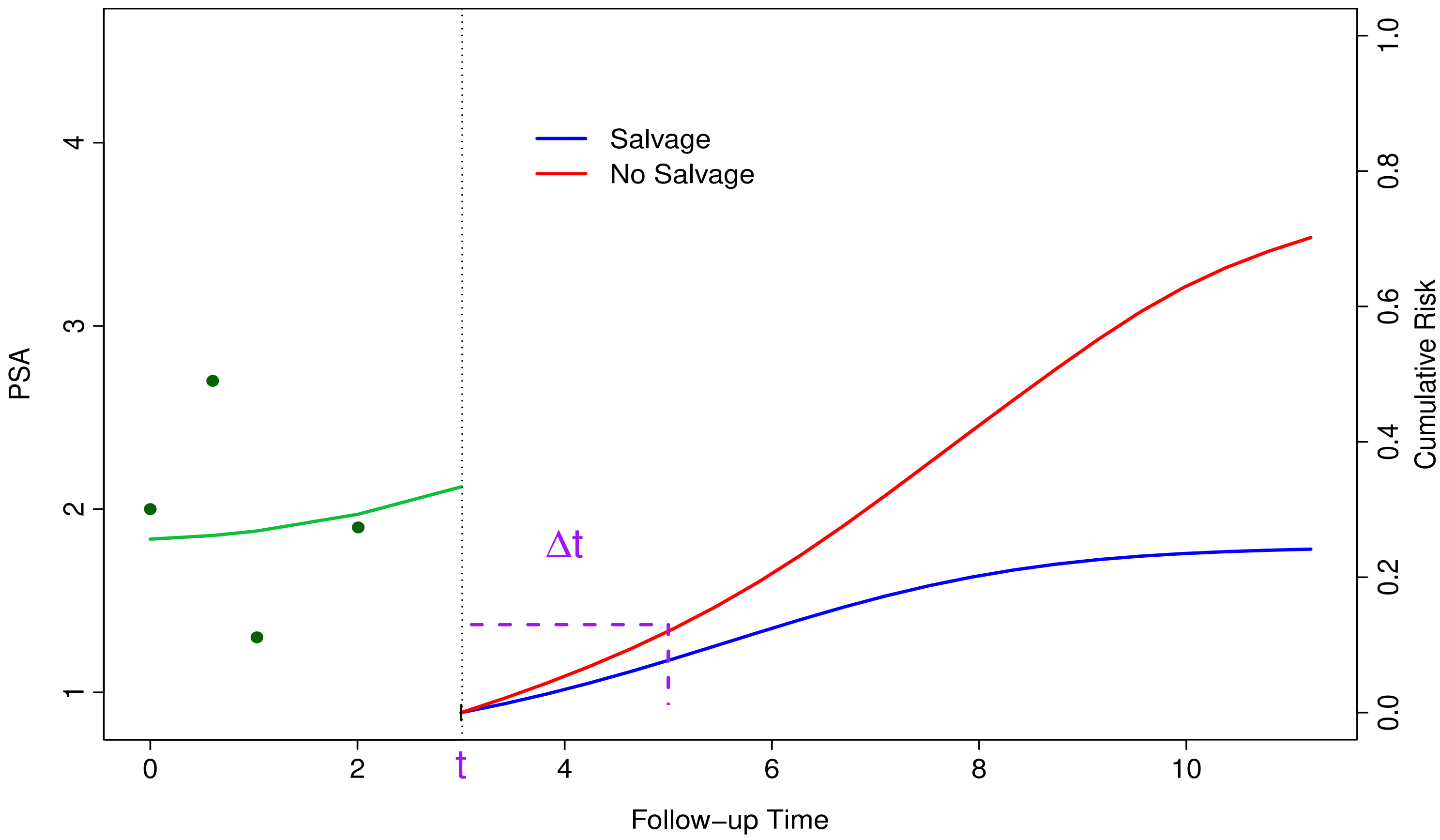
2 Causal ST Effects

- Standard assumptions for Causal Inference
  - ▷ *Consistency:* Observed outcomes equal the counterfactual outcomes for the actually assigned treatment
  - ▷ *Sequential Exchangeability:* The counterfactual outcomes are independent of the assigned treatment conditionally on the PSA history and baseline covariates
  - ▷ *Positivity:* Each patient has a nonzero probability of receiving ST at each time point t

2 Causal ST Effects (cont'd)

- Setting
  - ▷ PSA measurements up to t
  - ▷ no Salvage Therapy given up to t
  - ▷ we compare cumulative risk of metastasis in the medically-relevant interval $[t, t + \Delta t]$
  - ▷ under two regimes
    1. if Salvage Therapy is **not** given in the interval $[t, t + \Delta t]$
    2. if Salvage Therapy is given at t

2 Causal ST Effects (cont'd)



2 Causal ST Effects (cont'd)

Which is the target group?

- Notation
  - ▷ T_m: time to metastasis
  - ▷ T_d: time to death
  - ▷ $\mathcal{H}^*(t)$: a version of the PSA history up to t
  - ▷ $T_m^{(a)}$ and $T_d^{(a)}$ counterfactual outcomes
    - * $a = 1$, ST given at t
    - * $a = 0$, ST was not given in $[t, t + \Delta t]$

2 Causal ST Effects (cont'd)

- Marginal Salvage Therapy Effect

  ▷ we average over all PSA histories

$$\text{ST}^M(t + \Delta t, t) = \Pr\{T_m^{(1)} \leq t + \Delta t \mid T_m > t, T_d > t\} - \Pr\{T_m^{(0)} \leq t + \Delta t \mid T_m > t, T_d > t\}$$

- Notes:

  ▷ of lesser relevance to the urologists because they decide who gets ST based on PSA ⇒ **more bias**

  ▷ averages over a big group of patients ⇒ **less variance**

2 Causal ST Effects (cont'd)

- Conditional Salvage Therapy Effect

  ▷ we condition on the PSA history of a specific patient, i.e., $\mathcal{H}^*(t) = \mathcal{H}_i(t)$

$$\text{ST}^C(t + \Delta t, t) = \Pr\{T_m^{(1)} \leq t + \Delta t \mid T_m > t, T_d > t, \mathcal{H}_i(t)\}$$

$$- \Pr\{T_m^{(0)} \leq t + \Delta t \mid T_m > t, T_d > t, \mathcal{H}_i(t)\}$$

- Notes:

  ▷ much more relevant to the urologists ⇒ **less bias**

  ▷ averages over a narrow group of patients identified via modeling assumptions ⇒ **more variance**

2 Causal ST Effects (cont'd)

- Marginal-Conditional Salvage Therapy Effect

  ▷ consider ST for patients who had PSA levels above the threshold value c at their last visit, i.e., $\mathcal{H}^*(t) = \{Y(t) : Y(t) > c\}$

$$\text{ST}^{MC}(t + \Delta t, t) \;=\; \Pr\{T_m^{(1)} \leq t + \Delta t \mid T_m > t, T_d > t, \mathcal{H}^*(t)\}$$

$$- \Pr\{T_m^{(0)} \leq t + \Delta t \mid T_m > t, T_d > t, \mathcal{H}^*(t)\}$$

- Notes:

  ▷ relevant to the urologists ⇒ **compromised bias**

  ▷ averages over a bigger group of patients ⇒ **compromised variance**

3 Causal Effect Estimation

Standard Cox models not suitable

⇓

We need appropriate methods

3 Causal Effect Estimation (cont'd)

- Main approaches
  - ▷ Marginal Structural Models with IPW
  - ▷ G-Formula & G-Estimation
  - ▷ Model-based

- Structural Marginal Models & G-Formula

  ▷ *Advantage:* minimal/no assumptions for the outcome model

  ▷ *Disadvantage:*

  * it requires that the model for the weights is correct
  * requires correct models for other competing processes (e.g., censoring, visiting)

3 Causal Effect Estimation (cont'd)

- Model-based Estimation

  ▷ *Advantage:* it allows all competing processes to depend on the longitudinal history (in any complex manner)

  ▷ *Disadvantage:* it requires a correctly specified outcome model

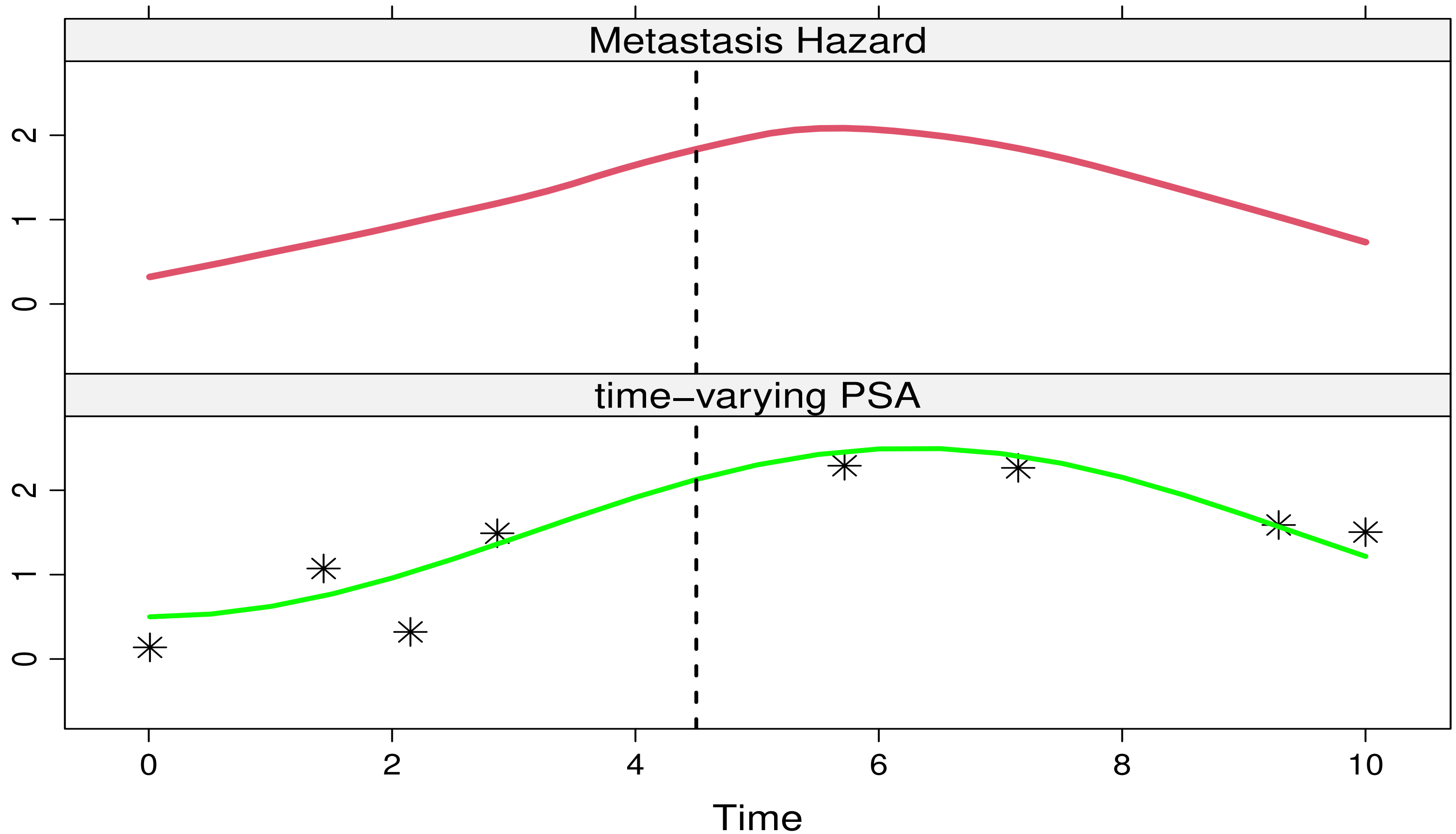
3 Causal Effect Estimation (cont'd)

Because salvage depends in a complex manner on the longitudinal history,
we opt for model-based estimation

$\Downarrow$

Joint Models for Longitudinal and Survival Data

3 Causal Effect Estimation (cont'd)



3 Causal Effect Estimation (cont'd)

Joint models completely specify the joint distribution of PSA, time-to-metastasis & time-to-death

- Under sequential ignorability,
  - ▷ they provide valid marginal distributions
  - ▷ *without requiring* to model the treatment assignment mechanism

4 PSA Sub-Model

- As PSA increases, patients may receive ST

- We let S_i denote the time a patient initiated ST
  - ▷ for patients who did not initiate ST, $S_i = \infty$

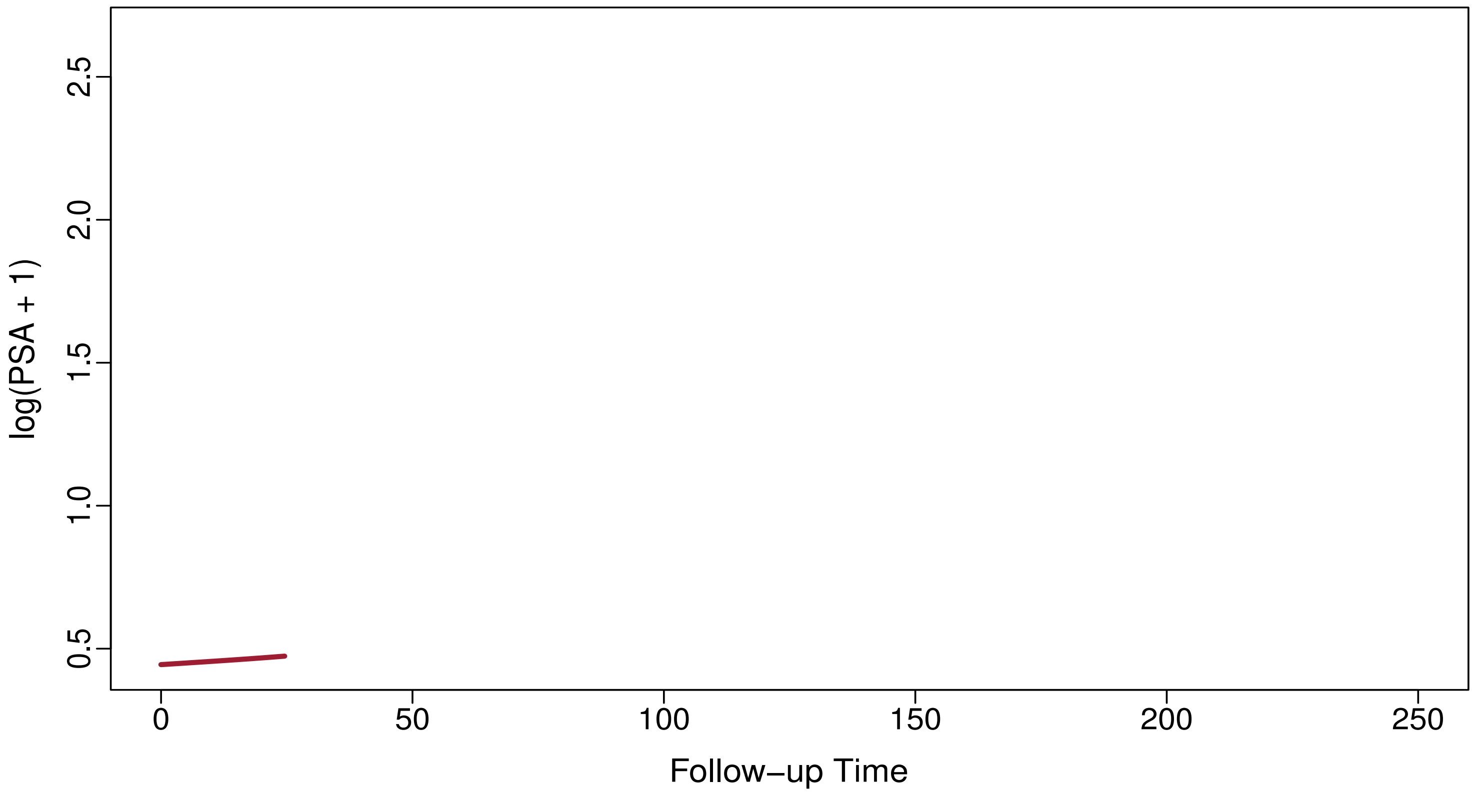
- After ST, PSA levels are expected to drop
  - ▷ but may rise again before metastasis

4 PSA Sub-Model (cont'd)

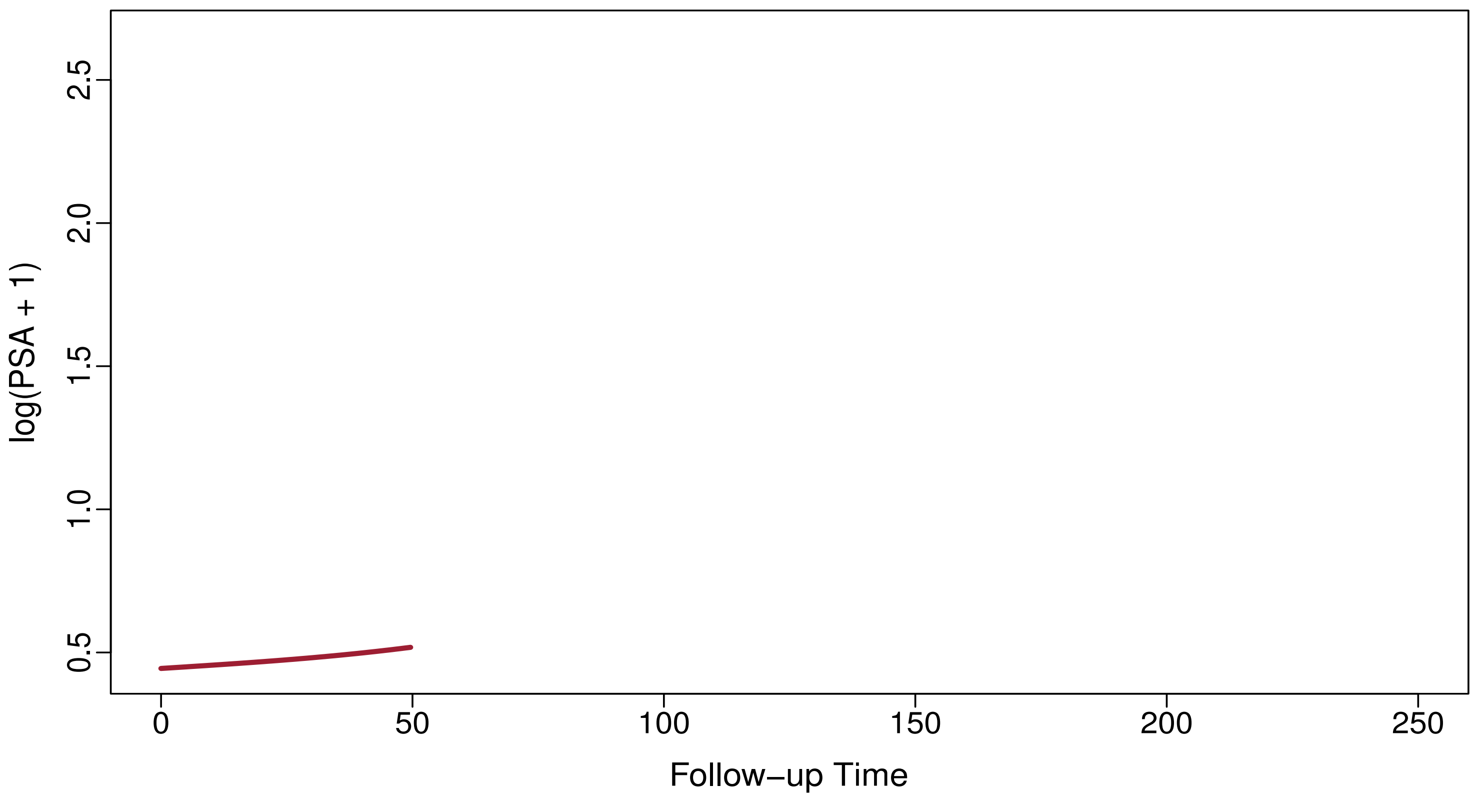
$$\log\{\mathsf{PSA}_i(t) + 1\} = \begin{cases} \eta_i(t) + \varepsilon_i(t) = \boldsymbol{x}_i(t)\boldsymbol{\beta} + \boldsymbol{z}_i(t)\boldsymbol{b}_i + \varepsilon_i(t), \; t < S_i \\ \\ \tilde{\eta}_i(t) + \varepsilon_i(t) = \\ \quad \eta_i(t) + \Big\{\tilde{\boldsymbol{x}}_i(\tilde{t})\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{z}}_i(t)\tilde{\boldsymbol{b}}_i\Big\} + \varepsilon_i(t), \quad t \geq S_i, \end{cases}$$

$$\boldsymbol{u}_i = (\boldsymbol{b}_i, \tilde{\boldsymbol{b}}_i) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$$

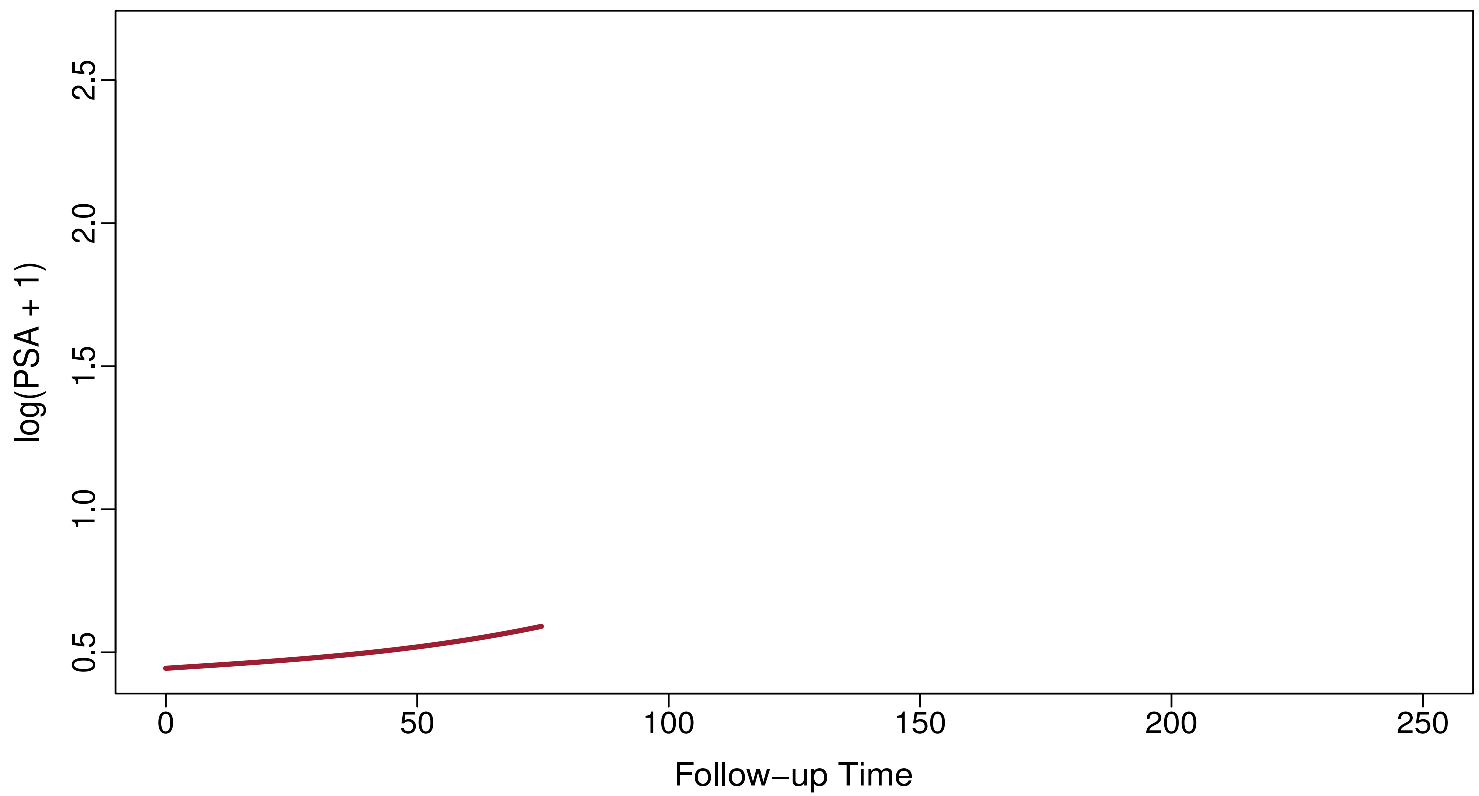
4 PSA Sub-Model (cont'd)



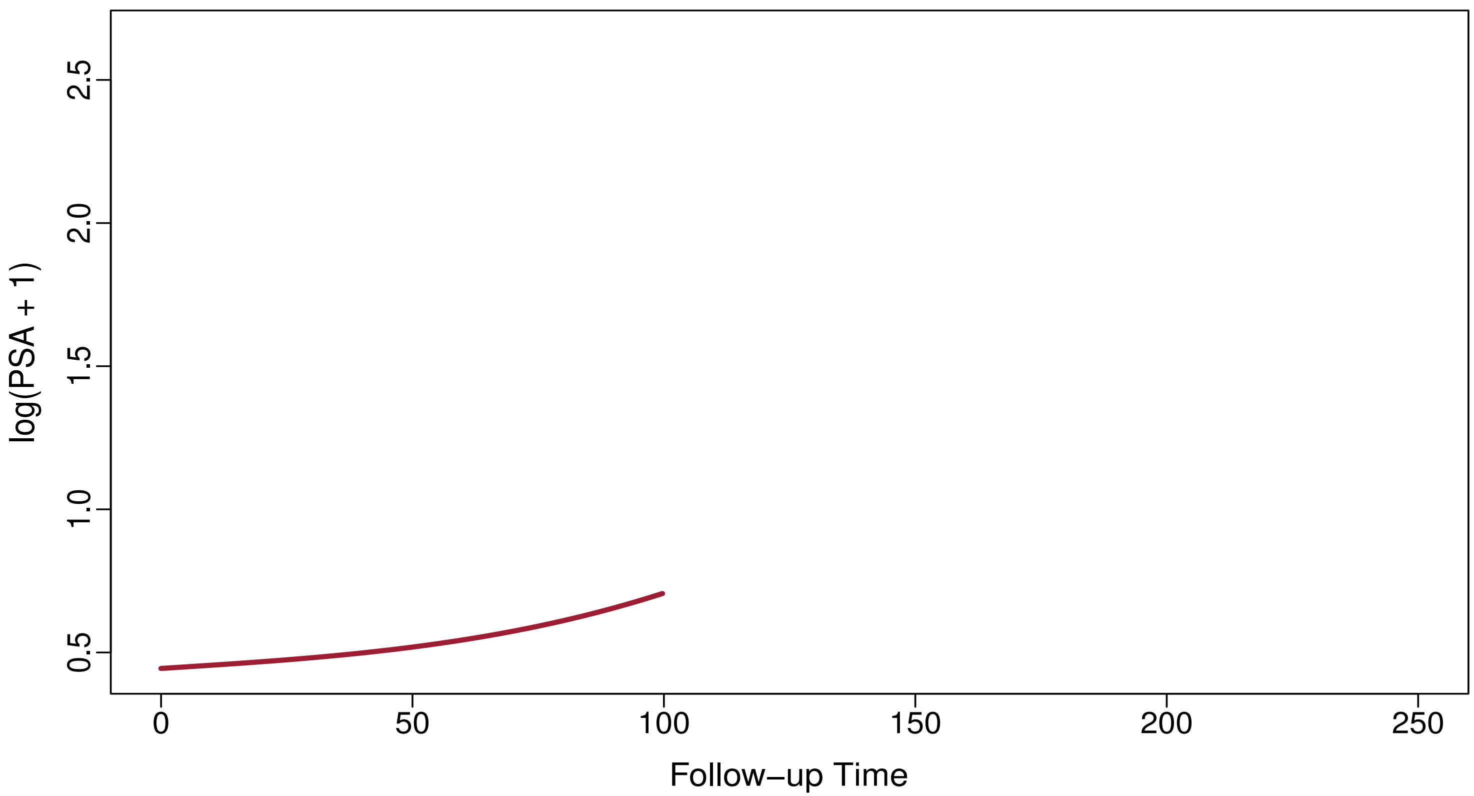
4 PSA Sub-Model (cont'd)



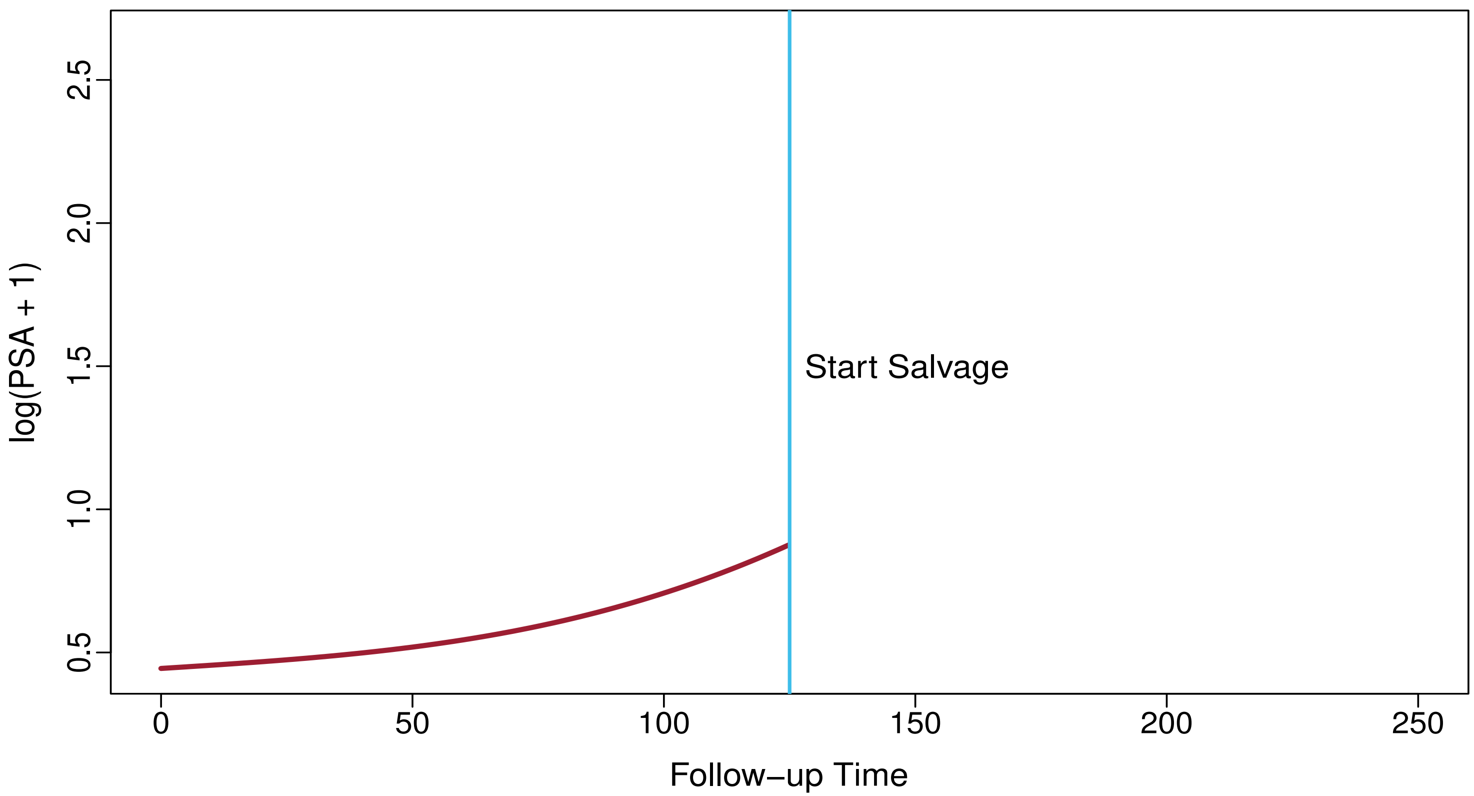
4 PSA Sub-Model (cont'd)



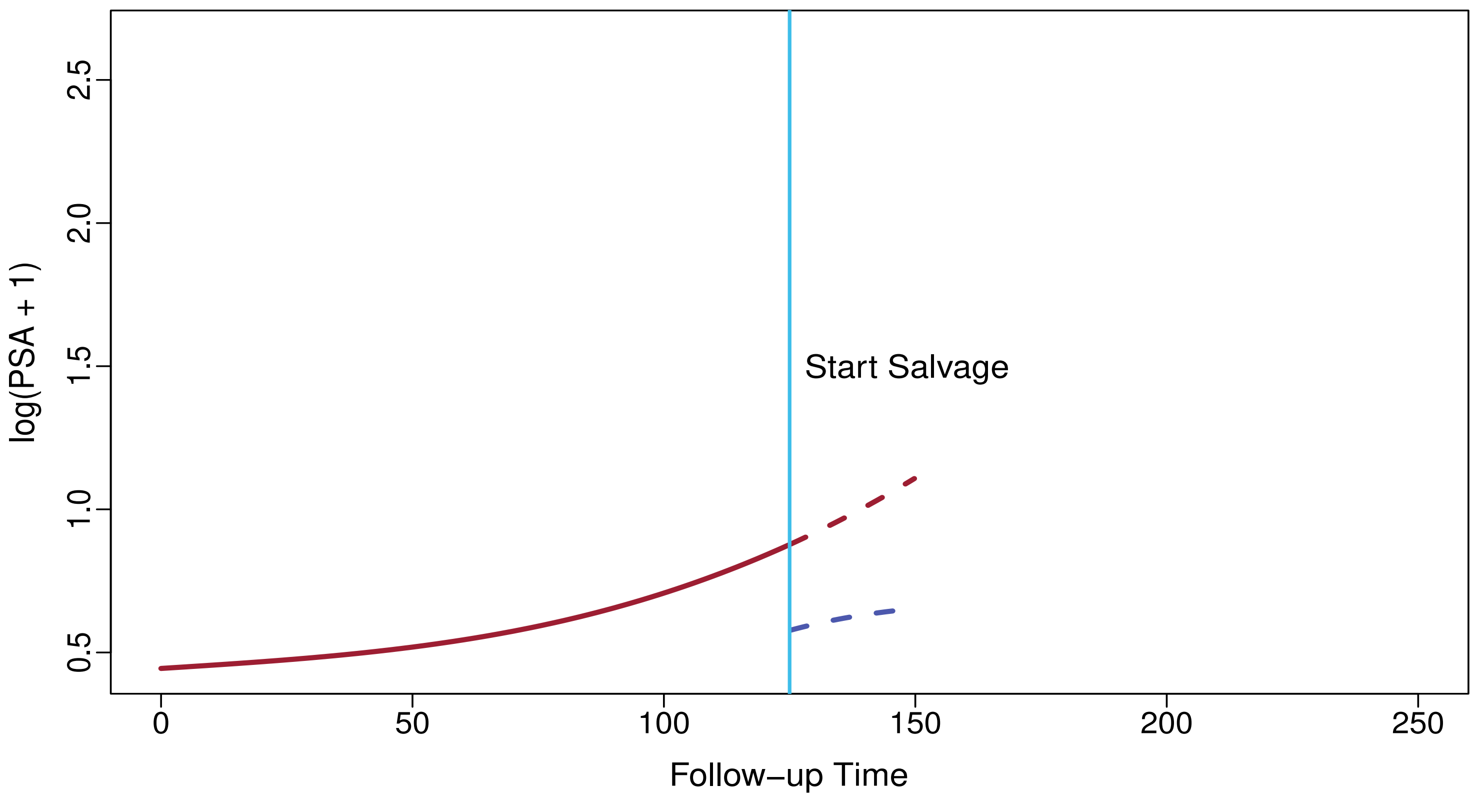
4 PSA Sub-Model (cont'd)



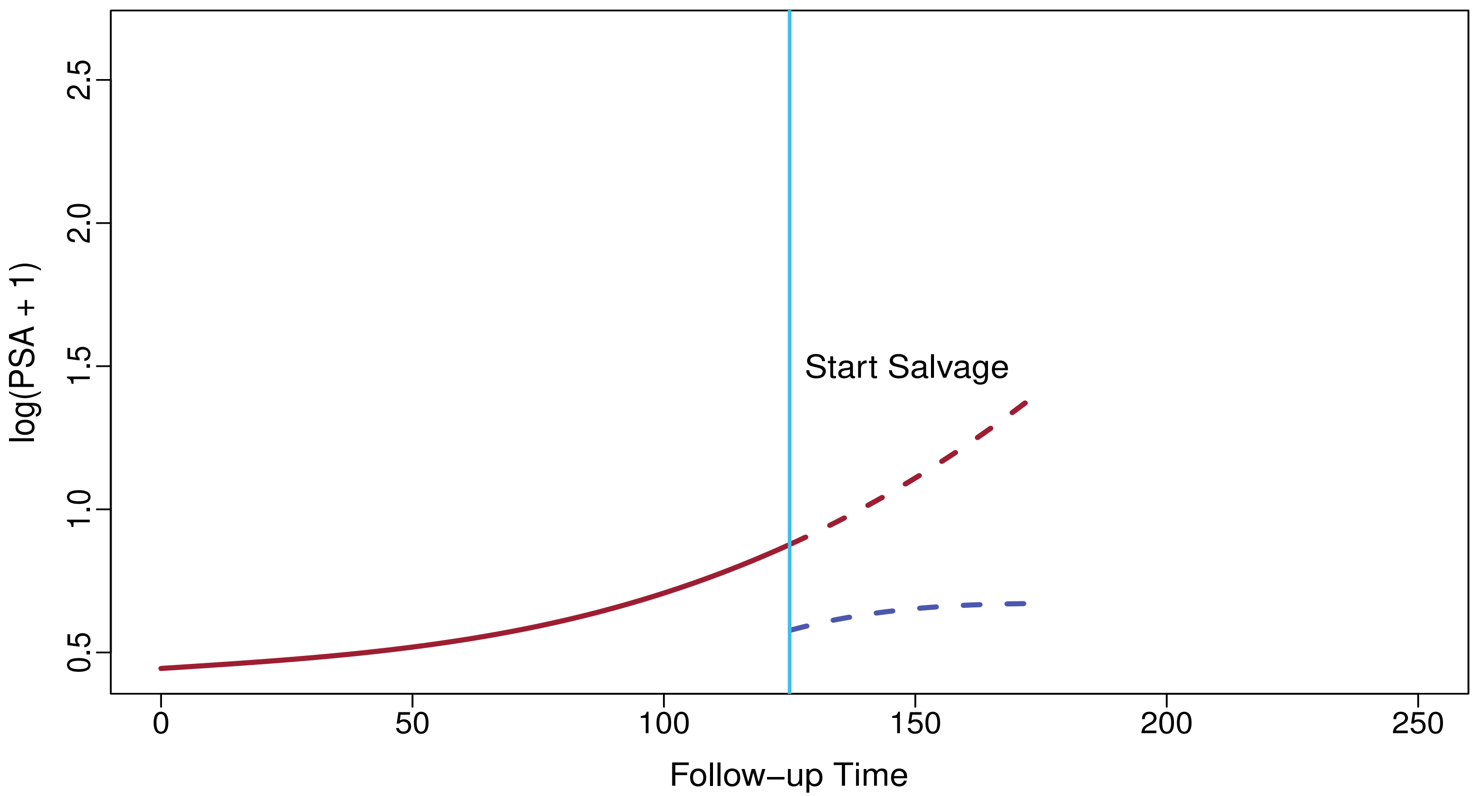
4 PSA Sub-Model (cont'd)



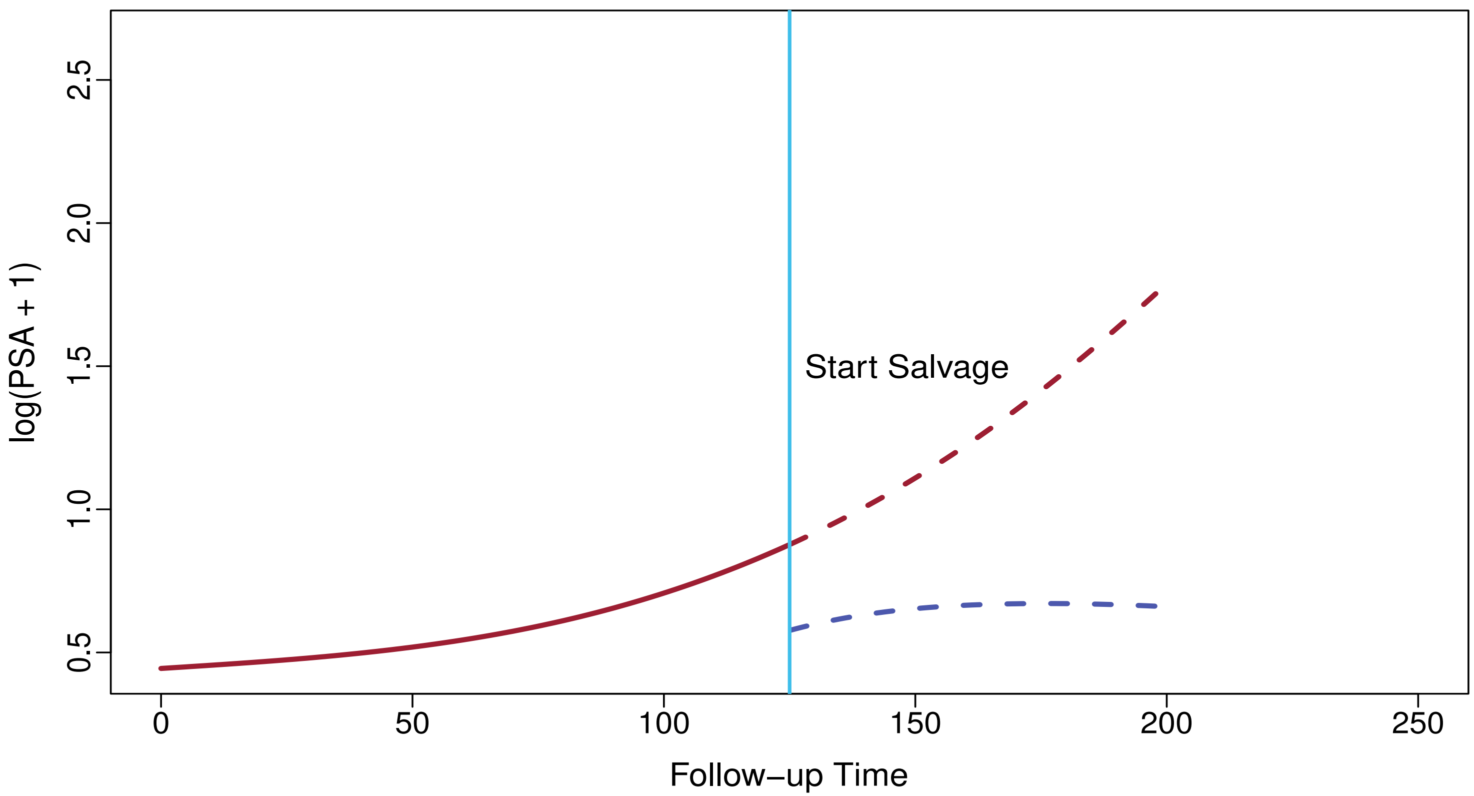
4 PSA Sub-Model (cont'd)



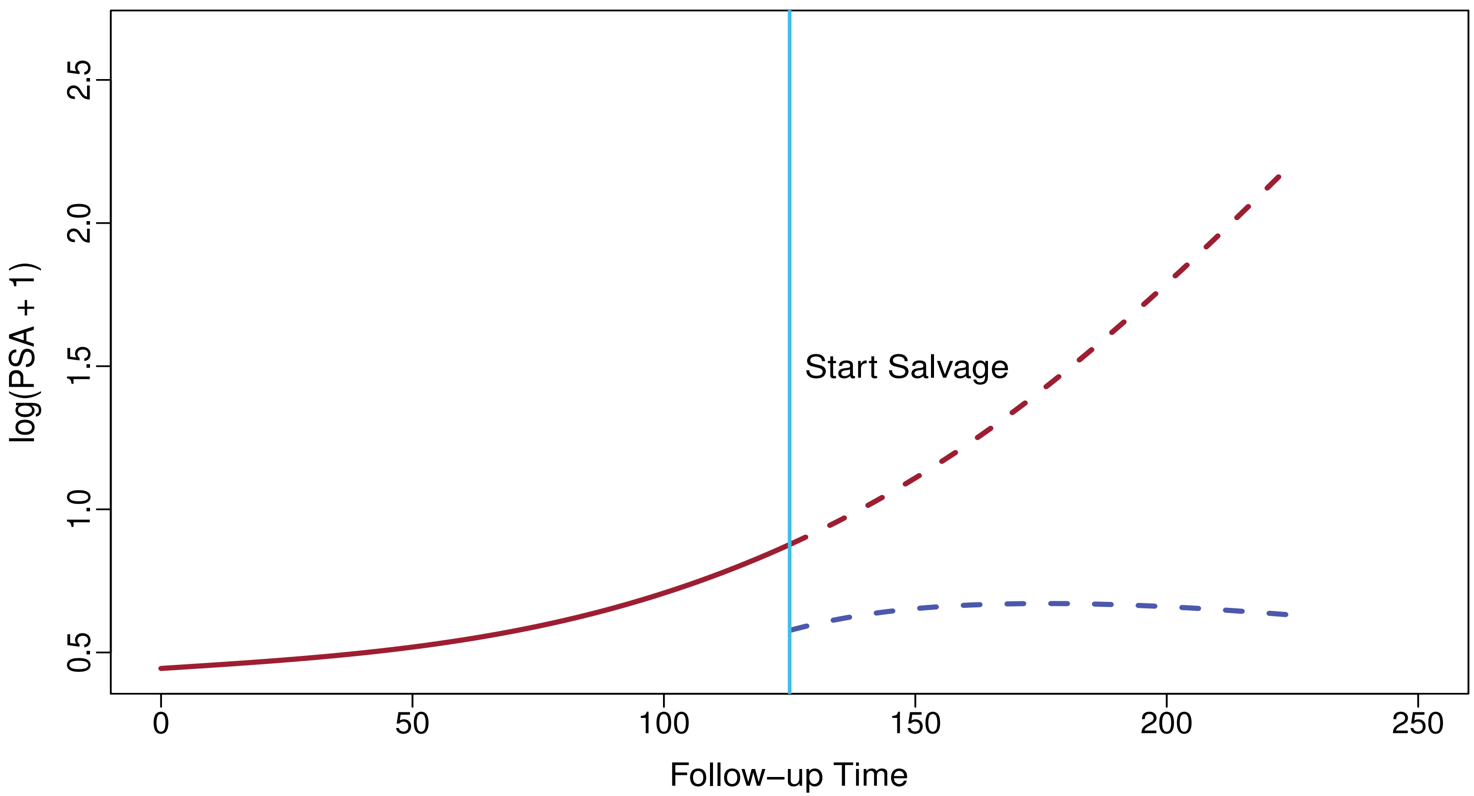
4 PSA Sub-Model (cont'd)



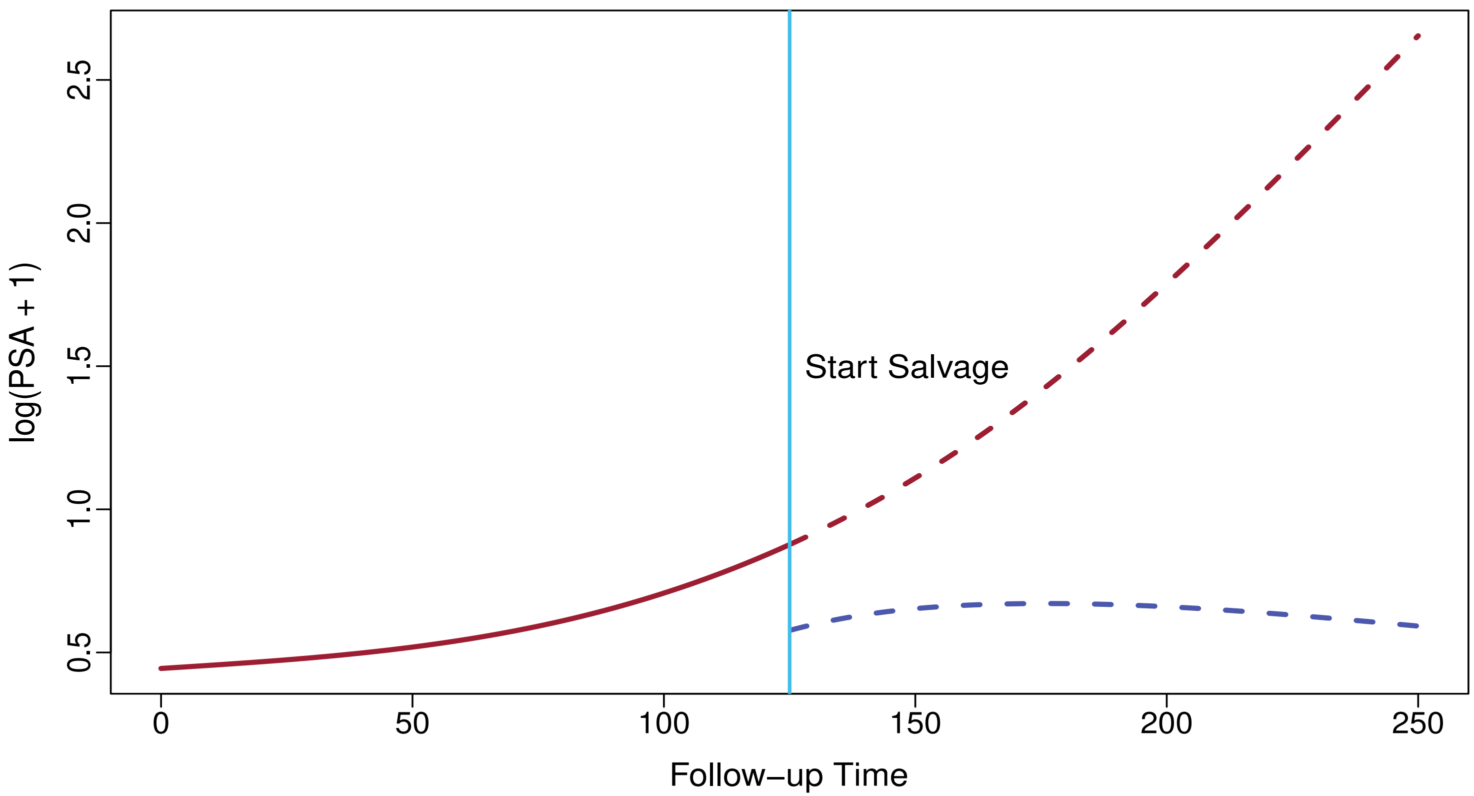
4 PSA Sub-Model (cont'd)



4 PSA Sub-Model (cont'd)



4 PSA Sub-Model (cont'd)



4 PSA Sub-Model (cont'd)

The model used in the UM data

- Fixed effects
  - ▷ *Before Salvage:* Nonlinear PSA evolution (B-spline with 6 internal knots)
  - ▷ *After Salvage:* pre-salvage evolution + drop in PSA, and change in linear evolution
  - ▷ baseline covariates: Age, baseline PSA, Gleason score, Charlson comorbidity index, perineural invasion

- Random effects
  - ▷ *the same time effect as in the fixed part*

5 Metastasis and Death Sub-Models

- Metastasis and Death treated as *Competing Risks*

- Separate hazard models for metastasis and death
  - ▷ linked with PSA and ST
  - ▷ baseline covariates

5 Metastasis and Death Sub-Models (cont'd)

- Metastasis Sub-Model linked to baseline covariates, Salvage and PSA

$$h_i^m(t) = \begin{cases} h_0^m(t) \exp\Big(\boldsymbol{\psi}_m^\top \boldsymbol{w}_i + \boldsymbol{\alpha}_m{}^\top f\{\eta_i(t)\}\Big), & t < S_i \\ h_0^m(t) \exp\Big(\boldsymbol{\psi}_m^\top \boldsymbol{w}_i + \gamma_m (t - S_i) + \boldsymbol{\xi}_m{}^\top g\{\tilde{\eta}_i(t)\}\Big), & t \geq S_i \end{cases}$$

5 Metastasis and Death (cont'd)

- Functions $f(\cdot)$ and $g(\cdot)$ specify the functional form
  - ▷ how PSA *before* and *after* Salvage is linked to metastasis

- Some options are...

5 Metastasis and Death (cont'd)

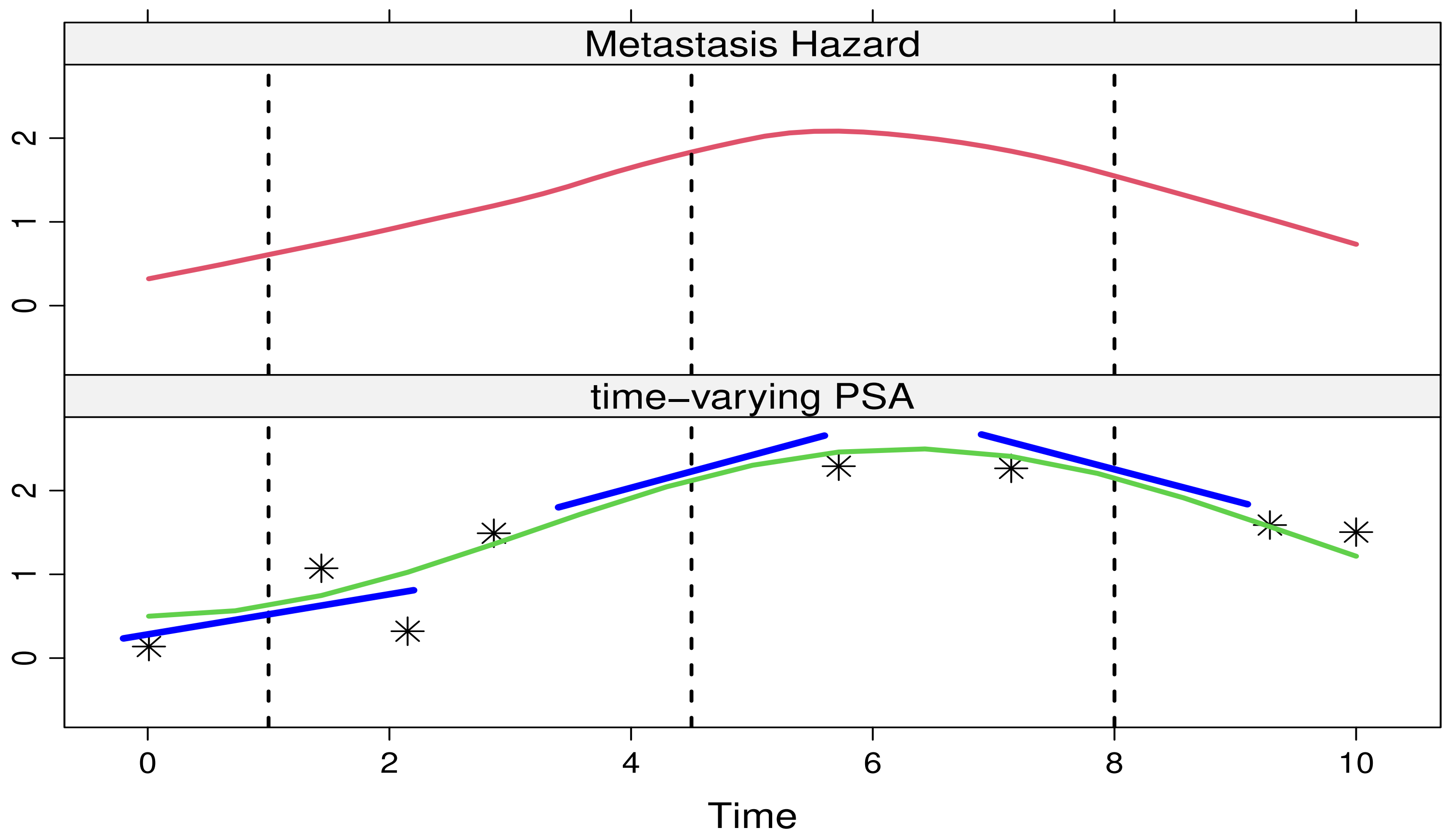
- *Time-dependent Slopes:* The hazard of metastasis at t is associated with both the current value and the slope of the PSA trajectory at t:

$$h_i^m(t \mid \mathcal{H}_i(t)) \; = \; h_0^m(t) \exp\{\boldsymbol{\psi}_m^\top \boldsymbol{w}_i + \alpha_{m1} \eta_i(t) + \alpha_{m2} \eta_i'(t)\},$$

where

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

5 Metastasis and Death (cont'd)



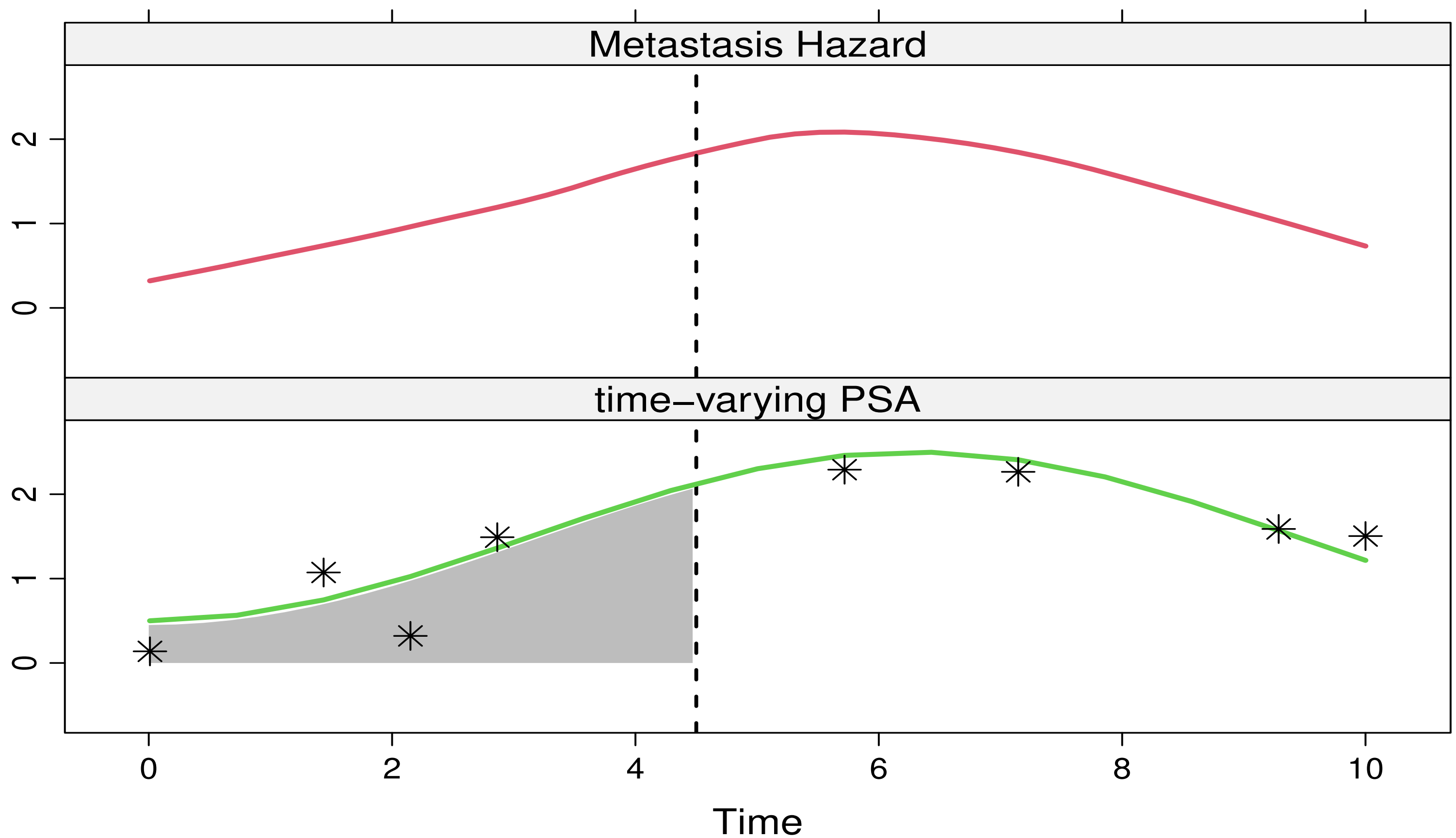
5 Metastasis and Death (cont'd)

- *Cumulative Effects:* The hazard of metastasis at t is associated with the area under the PSA trajectory up to t:

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

We account for the observation period

5 Metastasis and Death (cont'd)



5 Metastasis and Death (cont'd)

- Death Sub-Model linked to baseline covariates, Salvage *but not* PSA

$$h_i^d(t) = \begin{cases} h_0^d(t)\exp(\boldsymbol{\psi}_d^\top \boldsymbol{w}_i), & t < S_i \\ \\ h_0^d(t)\exp(\boldsymbol{\psi}_d^\top \boldsymbol{w}_i + \gamma_d), & t \geq S_i \end{cases}$$

6 Causal Effect Estimation

- From the joint model, we can obtain the conditional causal effect

$$\Pr\{T_{mi}^{(a)} \leq t + \Delta t \mid T_{mi} > t, T_{di} > t, \mathcal{H}_i(t), \mathcal{X}_i\} =$$

$$\int\int \Pr\{T_{mi}^{(a)} \leq t + \Delta t \mid T_{mi} > t, T_{di} > t, \boldsymbol{u}_i, \mathcal{X}_i, \boldsymbol{\theta}\}$$

$$\times\, p\{\boldsymbol{u}_i \mid T_{mi} > t, T_{di} > t, \mathcal{H}_i(t), \mathcal{X}_i, \boldsymbol{\theta}\}\; p(\boldsymbol{\theta} \mid \mathcal{D})\; \mathrm{d}\boldsymbol{u}_i \mathrm{d}\boldsymbol{\theta}$$

▷ $a = \{0, 1\}$

▷ $\mathcal{D} = \{T_i, \delta_i, Y_i; i = 1, \ldots, n\}$

▷ $p(\boldsymbol{\theta} \mid \mathcal{D})$ posterior

6 Causal Effect Estimation (cont'd)

- Monte Carlo scheme to estimate $\mathsf{ST}_i^C(t + \Delta t, t)$
  - ▷ sample $\breve{\boldsymbol{\theta}}^{(l)}$ from the posterior of the parameters $[\boldsymbol{\theta} \mid \mathcal{D}]$
  - ▷ sample $\breve{\boldsymbol{u}}_i^{(l)}$ from the posterior of the random effects $[\boldsymbol{u}_i \mid T_{mi} > t, T_{di} > t, \mathcal{H}_i(t), \mathcal{X}_i, \breve{\boldsymbol{\theta}}^{(l)}]$
  - ▷ calculate $\pi_i^{(l)}(t + \Delta t \mid t, a) = \Pr\{T_{mi}^{(a)} \leq t + \Delta t \mid T_{mi} > t, T_{di} > t, \breve{\boldsymbol{u}}_i^{(l)}, \mathcal{X}_i, \breve{\boldsymbol{\theta}}^{(l)}\}$

- We repeat L times and get

$$\widehat{\mathsf{ST}}_i^C(t + \Delta t, t) = \frac{1}{L} \sum_{l=1}^{L} \pi_i^{(l)}(t + \Delta t \mid t, a = 1) - \pi_i^{(l)}(t + \Delta t \mid t, a = 0)$$

6 Causal Effect Estimation (cont'd)

- Estimation of $\mathsf{ST}^M(t + \Delta t, t)$ and $\mathsf{ST}^{MC}(t + \Delta t, t)$ proceeds by averaging the conditional effects over the respective groups of patients

- For example, for $\mathsf{ST}^M(t + \Delta t, t)$
  - ▷ $\mathcal{R}(t)$ the subset of patients at risk at time t
  - ▷ for each patient in $\mathcal{R}(t)$, we calculate $\widehat{\mathsf{ST}}_i^C(t + \Delta t, t)$

$$\widehat{\mathsf{ST}}^M(t + \Delta t, t) = n_r^{-1} \sum_{i: i \in R(t)} \widehat{\mathsf{ST}}_i^C(t + \Delta t, t),$$

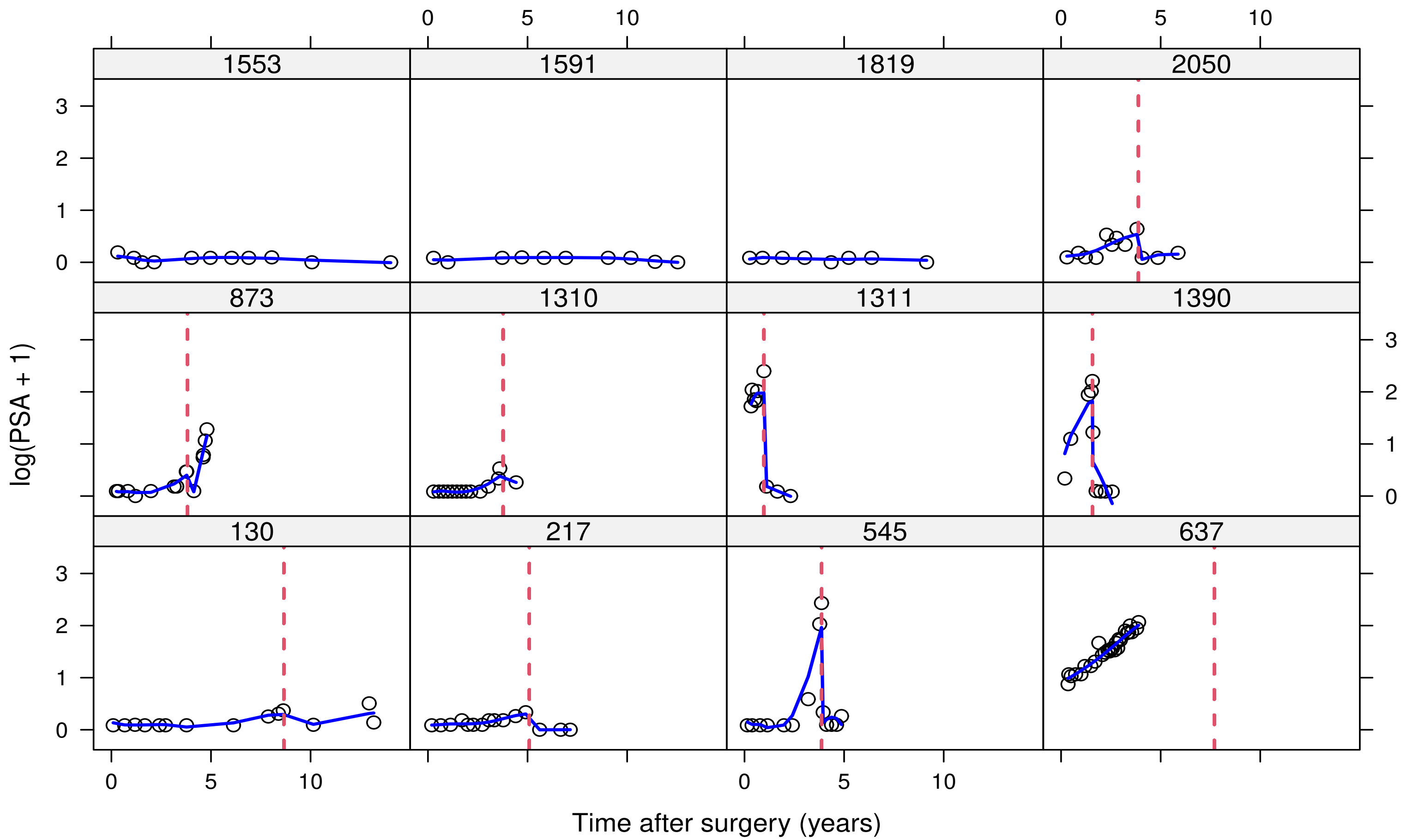
6 Causal Effect Estimation (cont'd)

- To estimate the variance of the causal effects, we need to take into account that they are a function of both the parameters $\boldsymbol{\theta}$ and the data $\mathcal{D}$

$$\text{Var}_{\mathcal{D}}\Big\{\widehat{\text{ST}}^M\big(t + \Delta t, t; \boldsymbol{\theta}, \mathcal{D}\big)\Big\} = \text{Var}_{\mathcal{D}}\Bigg[E_{\boldsymbol{\theta}|\mathcal{D}}\Big\{\text{ST}^M\big(t + \Delta t, t; \boldsymbol{\theta}, \mathcal{D}\big)\Big\}\Bigg]$$

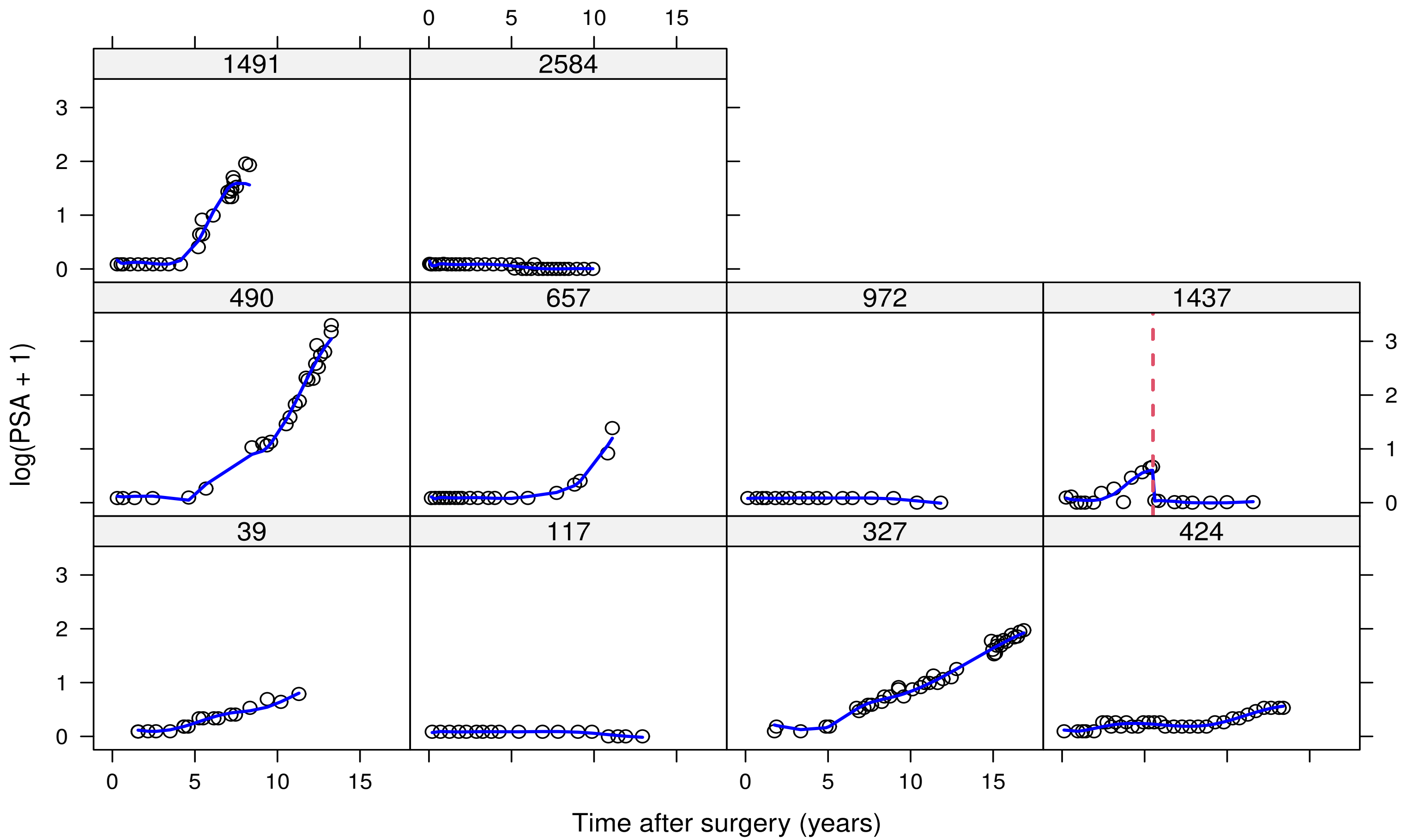
- We achieve this using an adaptation of the procedure of Antonelli et al. (2021)

7 Results

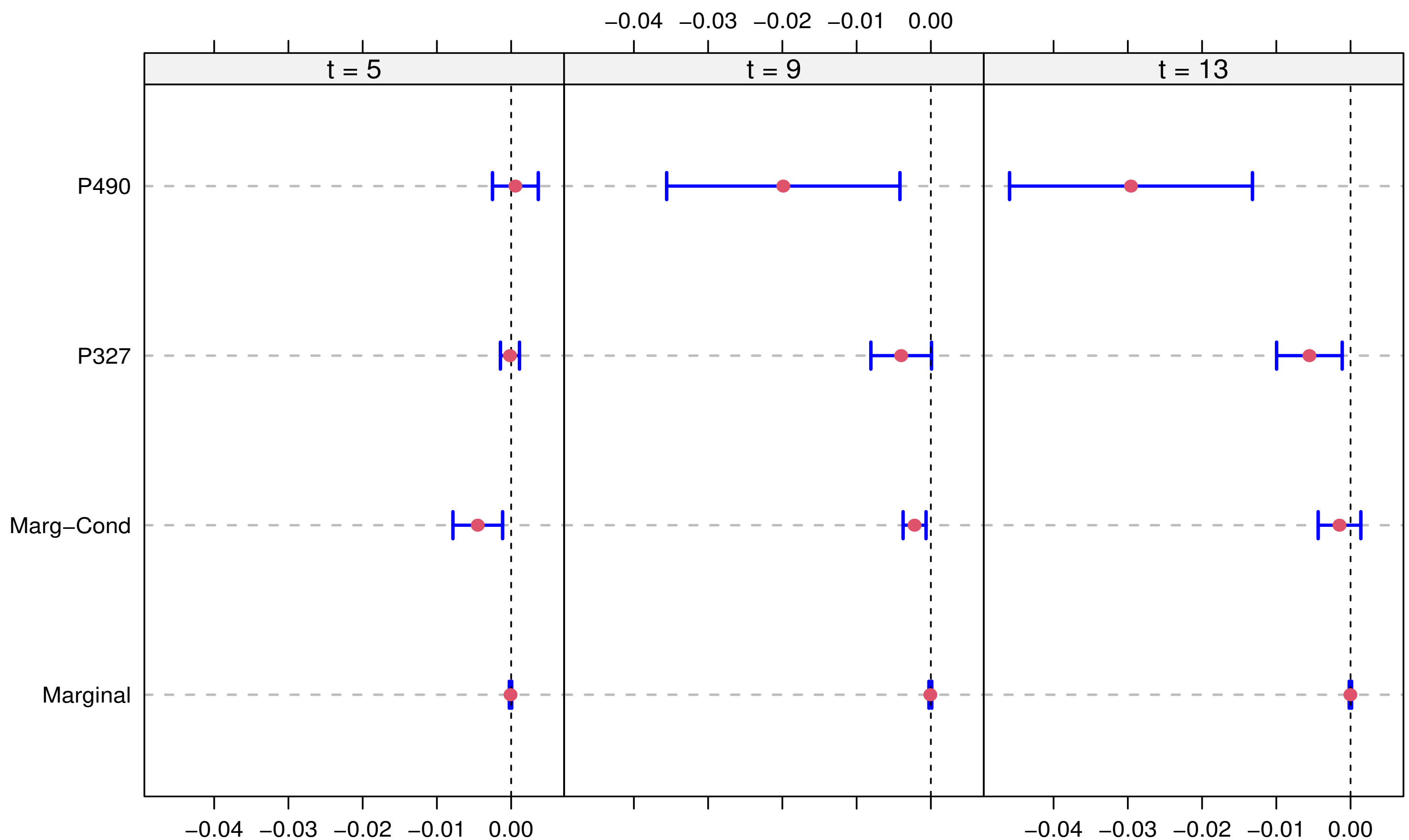


1553
1591
1819
2050
873
1310
1311
1390
130
217
545
637
log(PSA + 1)
Time after surgery (years)
0
5
10
0
1
2
3

7 Results (cont'd)



7 Results (cont'd)



8 Extensions & Discussion

- Implementation available in **JMbayes2**
  - ▷ `predict()` cumulative incidence risks
  - ▷ `causal_effects()` calculates the different causal effects (not yet in the package, but in GitHub)

- Shiny app. . .

Thank for your attention!

https://www.drizopoulos.com/

8 Causal Effect Estimation (cont'd)

- The first term is written as

$$\Pr\{T_{mi}^{(a)} \leq t + \Delta t, \mid T_{mi} > t, T_{di} > t, \boldsymbol{u}_i, \mathcal{X}_i, \boldsymbol{\theta}\} =$$

$$\frac{\displaystyle\int_t^{t+\Delta t} h_i^{m(a)}(v) \exp\Big(- \int_t^v \big\{h_i^{m(a)}(s) + h_i^{d(a)}(s)\big\} \, \mathsf{d}s - \int_0^t \big\{h_i^{m(0)}(s) + h_i^{d(0)}(s)\big\} \, \mathsf{d}s\Big) \, \mathsf{d}v}{\displaystyle\exp\Big(- \int_0^t \big\{h_i^{m(0)}(s) + h_i^{d(0)}(s)\big\} \, \mathsf{d}s\Big)}$$

- Using telescoping we get:

$$
\begin{aligned}
& p(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{\theta}_N \mid \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Y}, \boldsymbol{N}) \\
& \quad \propto \prod_{i=1}^{n} \prod_{j=1}^{n_i} p\{Y_i(t_{ij}), T_i, \delta_i \mid \mathcal{Y}_i(t_{i,j-1}), \mathcal{N}_i(t_{i,j-1}), \mathcal{X}_i, \boldsymbol{\theta}, \boldsymbol{u}_i\} \\
& \qquad\qquad \times \prod_{j=1}^{n_i} p\{N_i(t_{ij}) \mid \mathcal{Y}_i(t_{i,j-1}), \mathcal{N}_i(t_{i,j-1}), Y_i(t_{ij}), T_i, \delta_i, \mathcal{X}_i, \boldsymbol{\theta}_N, \boldsymbol{u}_i\} \\
& \qquad \times p(\boldsymbol{u}_i \mid \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \times p(\boldsymbol{\theta}_N)
\end{aligned}
$$

8 Causal Effect Estimation (cont'd)

- Under sequential exchangeability, we have that

$$p\{N_i(t_{ij}) \mid \mathcal{Y}_i(t_{ij}), \mathcal{N}_i(t_{i,j}), \mathcal{F}_i^{(a)}(v_{ij}), T_i^{(a)}, \delta_i^{(a)}, \mathcal{X}_i, \boldsymbol{\theta}_N, \boldsymbol{u}_i\} = p\{N_i(t_{ij}) \mid \mathcal{Y}_i(t_{ij}), \mathcal{N}_i(t_{i,j-1}), \mathcal{X}_i, \boldsymbol{\theta}_N\},$$

$\Rightarrow$ inference can be based on the first term (i.e., the observed data model) and ignore the second term

8 Computational Details (cont'd)

- Custom-made and tailored MCMC algorithm
  - ▷ Gibbs sampling (hierarchical centering for fixed effects)
  - ▷ adaptive Metropolis-Hastings
  - ▷ (Metropolis-adjusted Langevin algorithm for certain parameter)
  - ▷ centered design matrices

- Speed via parallel sampling of random effects

- Chains run in parallel

8 Results (cont'd)

https://emcbiostatistics.shinyapps.io/Plots_PSA/