Estimating survival from Gray's flexible model

Zdenek Valenta

Department of Medical Informatics
Institute of Computer Science
Academy of Sciences of the Czech Republic

Outline

I. Introduction

- Let Y be a random variable capturing time to occurrence of a certain event of interest. The **hazard function** $h(y)$ is at time y formally defined as follows:

$$h(y) = \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y | Y \geq y)}{\Delta y}, \qquad (1)$$

  where $P(.)$ denotes conditional probability, that an event of interest would occur immediately after time y, given it did not prior to this time.

- It follows from (1) that the hazard function may only take non–negative values.

I. Introduction

- Let $F(Y)$ denote a cumulative distribution function of the random variable Y, i.e. $F(Y) = P(Y \leq y)$. We assume that Y is absolutely continuous with density $f(y)$. The expression (1) may be then written as:

$$\begin{aligned} h(y) &= \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y | Y \geq y)}{\Delta y} \\ &= \lim_{\Delta y \to 0} \frac{P(y \leq Y < y + \Delta y)}{P(Y \geq y)\Delta y} \\ &= \frac{1}{P(Y \geq y)} \frac{d}{dy} F(y) = \frac{f(y)}{P(Y \geq y)} \\ &= \frac{f(y)}{S(y)}, \end{aligned} \qquad (2)$$

  where $S(y)$ denotes the value of the **survival function** at time y.

I. Introduction

- We define a **cumulative hazard function** $H(t)$ at time t as:

$$H(t) = \int_0^t h(y)dy \tag{3}$$

It follows from (2) that $S(.)$ and $H(.)$ capture equivalent information:

$$H(t) = -\ln(S(t)) \tag{4}$$

- Furthermore, it follows from (4) that we can determine the value of the **survival function** $S(t)$ at time t whenever we are able to evaluate the cumulative hazard function $H(t)$:

$$S(t) = \exp\{-H(t)\} \tag{5}$$

II. Semi–parametric survival models

- **Multiplicative models:**
  - **Cox PH model:**

$$h(y|Z) = h_0(y) \cdot \exp(\beta' Z) \tag{6}$$

  - **Gray's flexible model:**

$$h(y|Z) = h_0(y) \cdot \exp\{\beta(y)' Z\} \tag{7}$$

- **Additive models:**
  - **Aalen's linear model:**

$$h(y|Z) = h_0(y) + \beta(y)' Z \tag{8}$$

II. Semi–parametric survival models

Note: Aalen's linear model (8) may be embedded in the class of **multiplicative models**:

- **Aalen's model**:

$$h(y|Z) = h_0(y) + \beta(y)' Z$$

$$\exp(h(y|Z)) = \exp\{h_0(y) + \beta(y)' Z\} \tag{9}$$

$$h^1(y|Z) = h_0^1(y) \cdot \exp\{\beta(y)' Z\}$$

The class of multiplicative models represented by the Cox PH and Gray's flexible model **includes the whole class of models proposed by Aalen**.

III. Gray's model introduction

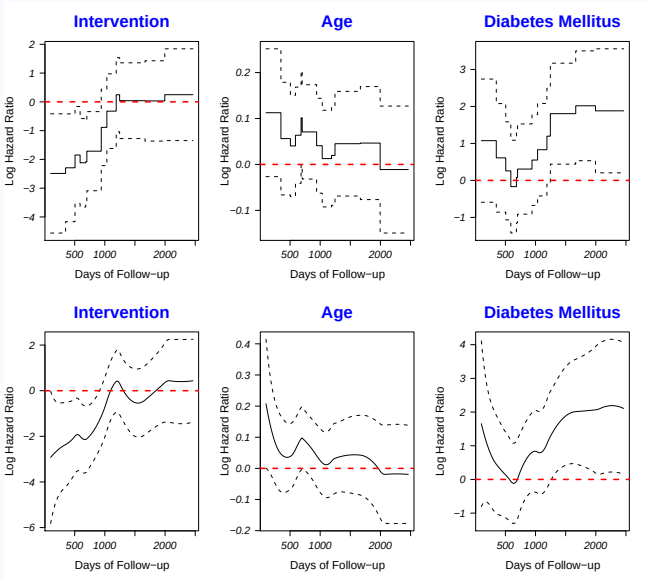
- Let us recall the definition of **Gray's flexible model**: (7):

$$h(y|Z) = h_0(y) \cdot \exp\{\beta(y)' Z\}$$

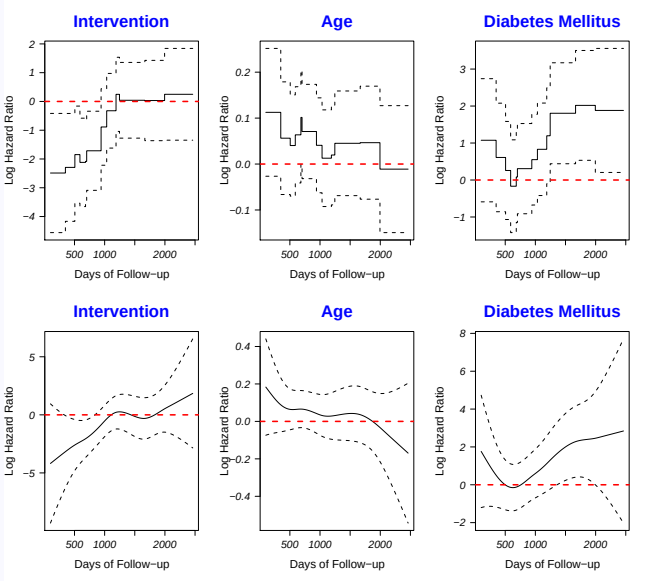
- Gray's model uses **penalized B-splines** for modelling time-varying effects $\beta(y)$. B-splines allow for flexible modelling of the covariate effects $\beta(y)$ and the hazard function over time.

- In the context of **Gray's model using piecewise-constant time-varying regression coefficients** the $\beta(y)$ remain constant for $y \in [\tau_{j-1}, \tau_j)$. We can thus write $\beta(y) = \beta_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp})$, where p denotes the number of model covariates and $j = 1, \ldots, M+1$ indexes the intervals on time axis. Here τ_j denote the knots that allow for a change of the regression coefficients β_j.

Piecewise-constant vs. quadratic penalized splines



Piecewise-constant vs. cubic penalized splines



IV. Survival estimates based on semi-parametric models

- **Cox PH model:**

$$\begin{aligned} S(t|Z) &= \exp\left\{-\int_0^t h_0(y)\cdot\exp\left(\beta' Z\right)dy\right\} \\ &= \exp\left\{-H_0(t)\cdot\exp\left(\beta' Z\right)\right\} \\ &= \left[S_0(t)\right]^{\exp(\beta' Z)}, \end{aligned} \tag{10}$$

where $S_0(t)$ represents **baseline survival function estimate** at time t.

IV. Survival estimates based on semi-parametric models

- **Aalen's linear model:**

$$\begin{aligned} h(y|Z) &= h_0(y) + \beta(y)'Z \\ h(y|Z) &= \widetilde{\beta}(y)'\widetilde{Z}, \end{aligned} \tag{11}$$

while $\widetilde{\beta}(y) = (h_0(y), \beta(y))$ a $\widetilde{Z} = (1, Z)$.

- Survival function estimates based on Aalen's model use cumulative regression coefficients $\widetilde{B}(t)$, where $\widetilde{B}_i(t) = \int_0^t \widetilde{\beta}_i(y)dy$. Estimating survival based on Aalen's model may then proceed as follows:

$$S\left(t|Z\right) = \exp\left\{-\widetilde{B}(t)'\widetilde{Z}\right\} \tag{12}$$

V. Estimating survival from Gray's model

- **Survival function estimate based on Gray's model**[a] using piecewise–constant penalized splines may be obtained as follows:

$$S\left(t|Z\right) = \exp\left\{-\sum_{j=1}^{M+1} H_{0j}\left(t\right) \cdot \exp\left(\beta_j' Z\right)\right\}, \qquad (13)$$

where Z denotes p-dimensional vector of patient's characteristics, and

$$H_{0j}(t) = \int\limits_{[\tau_{j-1},\tau_j)} I\left(u \leq t\right) dH_0(u) \qquad (14)$$

represents **a contribution to the cumulative baseline hazard function** $H_0(t)$ **on the interval** $[\tau_{j-1},\ \tau_j)$, $j = 1, \ldots, M+1$.

[a] Valenta Z et al, Statistics in Medicine 2002.

V. Estimating survival from Gray's model

- Derivation of **confidence limits** for the survival function estimate based on Gray's model uses the **Delta method**.
- Recall **the Taylor formulae** for a function $f(X)$ of a random variable X with expectation μ:

$$f(X) = \sum_{k=0}^{n} \frac{f^{(k)}(\mu)}{k!}(X-\mu)^k + R_n \qquad (15)$$

- **Delta method**:

$$\begin{aligned} \mathsf{Var}(f(X)) &\approx \mathsf{Var}\left[f(\mu) + f'(\mu)(X-\mu)\right] \\ &= \left[f'(\mu)\right]^2 \cdot \mathsf{Var}(X) \end{aligned} \qquad (16)$$

V. Estimating survival from Gray's model

- If X is a **random vector** the Delta method $G(X)$ takes the form:

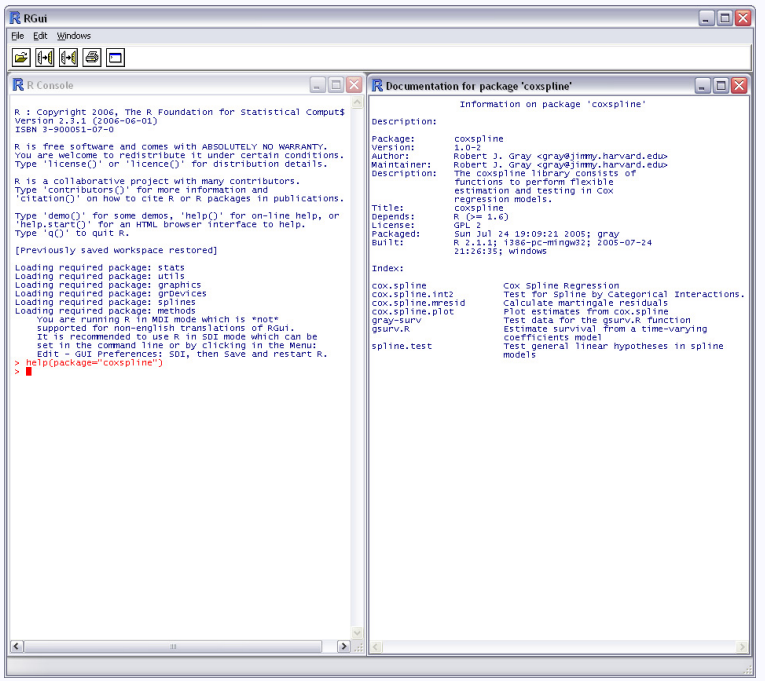
$$\mathsf{Var}(G(X)) \approx \nabla G'(\mu) \cdot \mathsf{Var}(X) \cdot \nabla G(\mu), \qquad (17)$$

where $\nabla G(\mu)$ is a column vector of first partial derivatives of G.

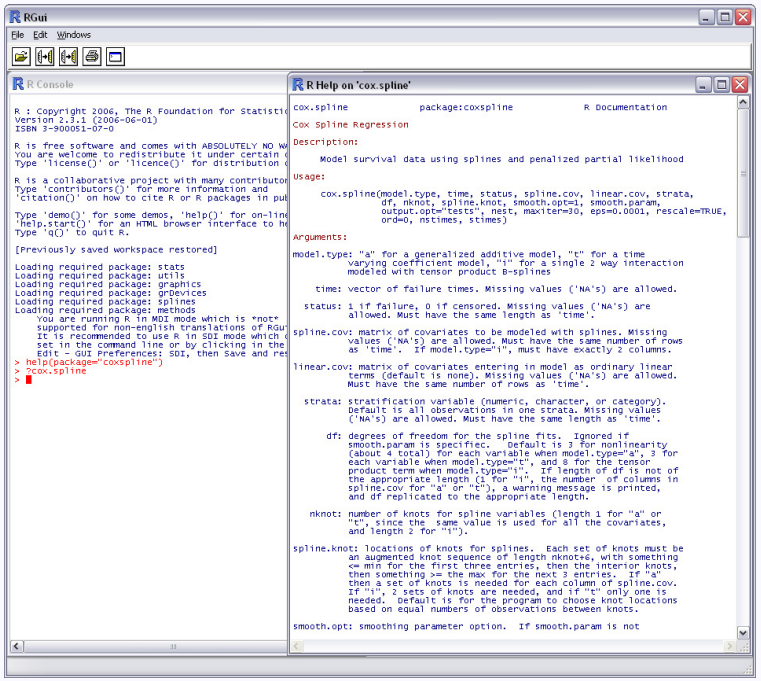
- **Confidence limits estimates**[a] were derived for a log– and log(- log)–transformed survival function $S(t)$ and are reported simultaneously in R.

[a] Valenta Z et al, Statistics in Medicine 2002.

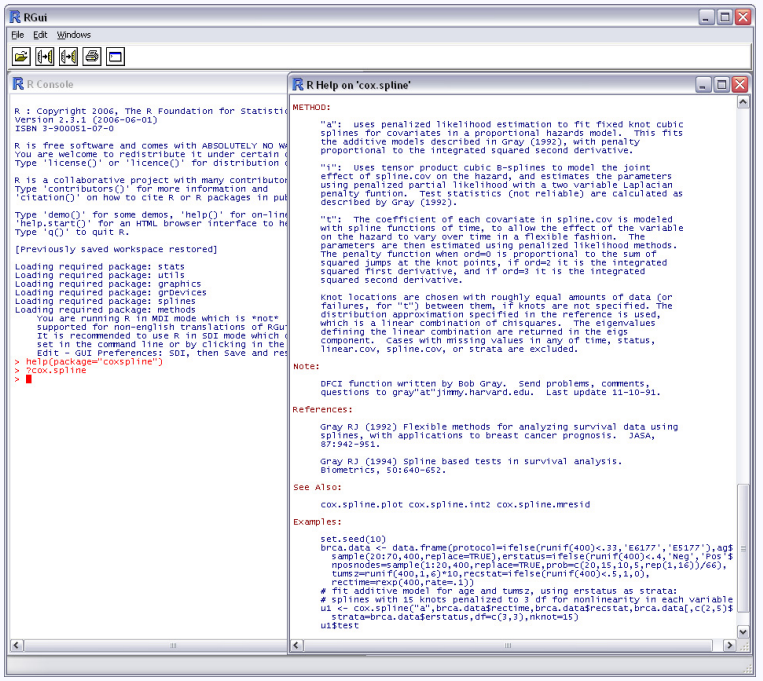
"coxspline" package in R



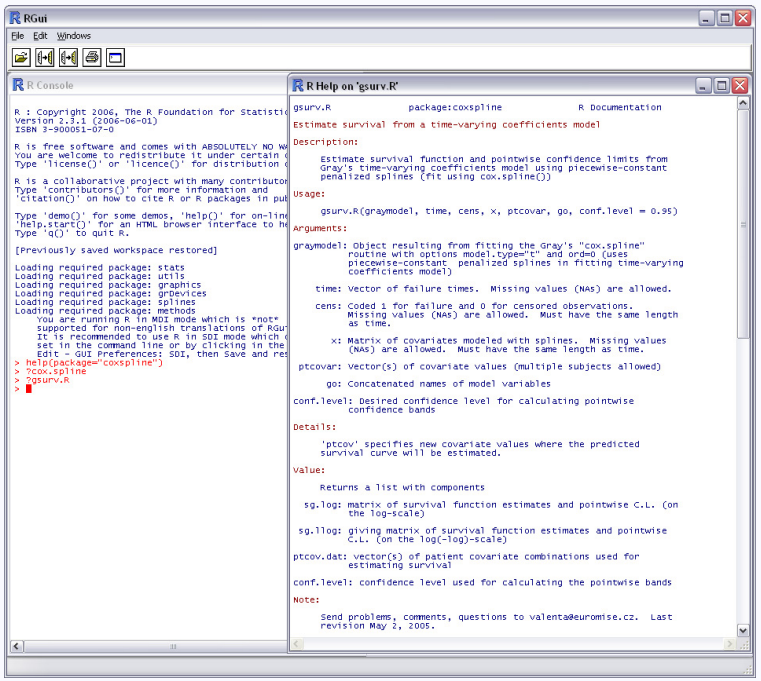
“cox.spline” R-routine



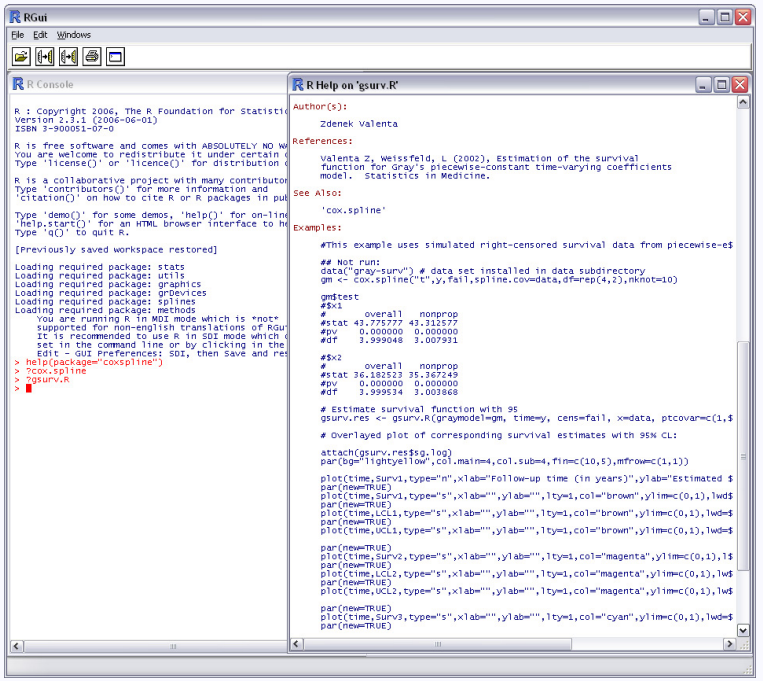
“cox.spline” R-routine (cont.)



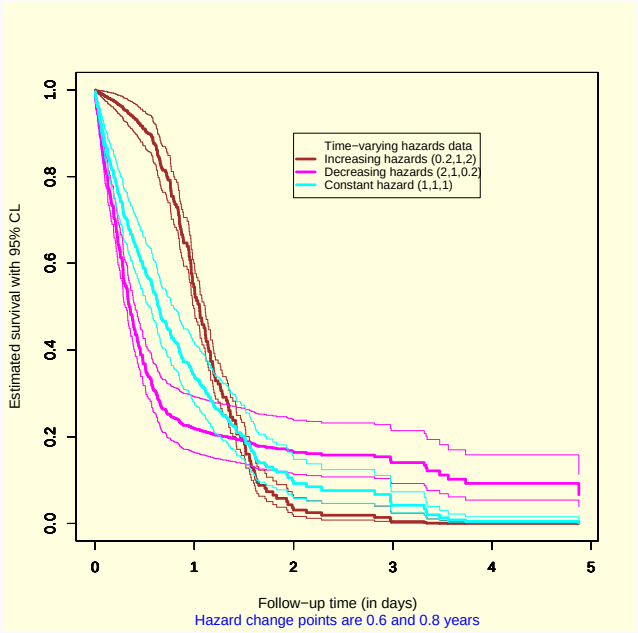
R-function “gsurv.R”



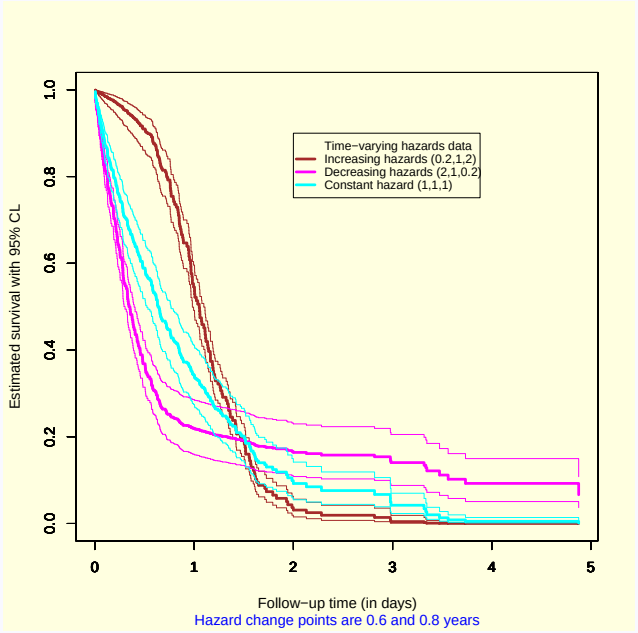
R-function “gsurv.R” (cont.)



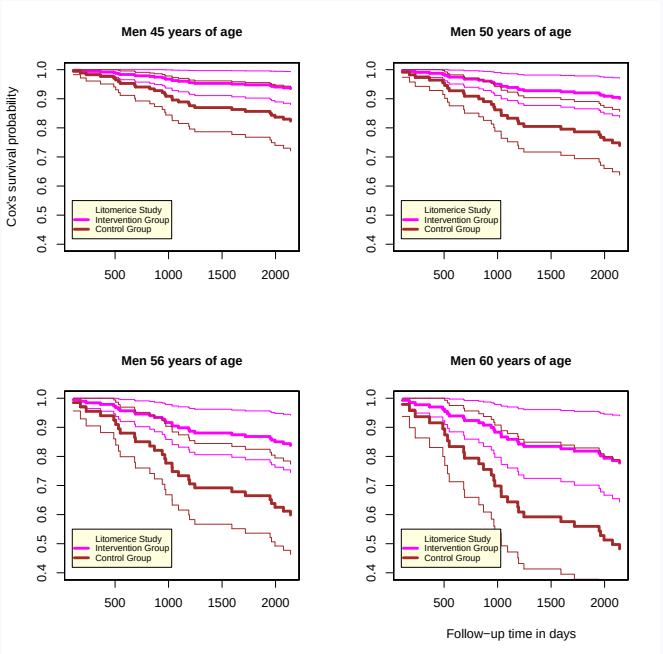
Example 1: survival estimates from Gray's model with 95% C.L. (log-transf.)



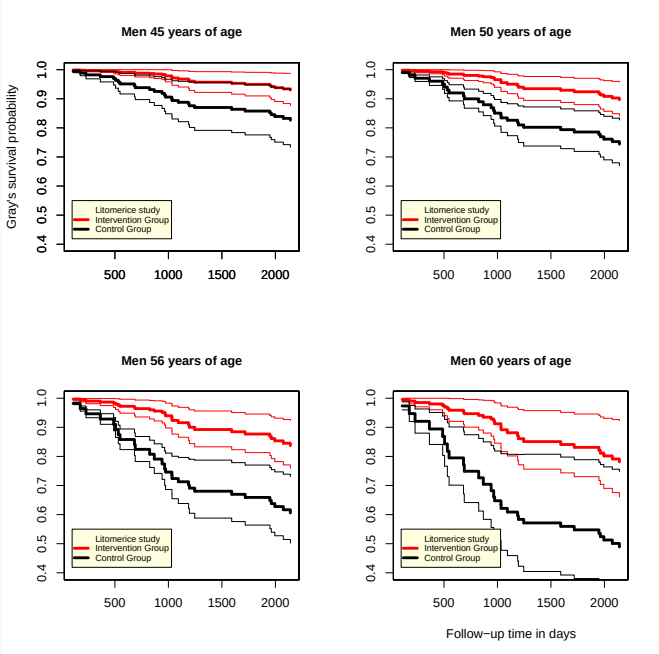
Example 1: survival estimates from Gray's model with 95% C.L. (log-log-transf.)



Example 2: Secondary prevention trial of CHD in Litomerice men after MI (Cox PH model results)



Example 2: Secondary prevention trial of CHD in Litomerice men after MI (Gray's model results)



V. Estimating survival from Gray's model

- Implementation of the **"coxspline" package for R statistical system** is available from the Dr. Gray's website *(Harvard University and Dana-Farber Cancer Institute, Boston, USA)*.
- **Web address**:

  http://biowww.dfci.harvard.edu/~gray/

- Package **"coxspline"**, version 1.0-2, implements Gray's model in R, including the survival function estimation using R-function **"gsurv.R"**.
- Current version of the **"coxspline"** package is compatible with the latest release of **R 2.3.1.** (2006-06-01):

  http://www.r-project.org/

VI. Impact of misspecifying the survival model[a]

- In three simulation studies we generated right-censored survival data that would satisfy exactly one of the semi-parametric survival models under consideration (i.e. **Aalen, Cox, Gray**).
- The data obtained were subsequently analyzed using each of the three models considered.
- The performance of each model was assessed using the **Bias** and **Mean Square Error** (MSE) of the estimated (conditional) survival distribution.

[a] Valenta Z et al, Model misspecification effect in univariable regression models for right-censored survival data, Proceedings of the 2002 Joint Statistical Meeting of the American Statistical Society.

VI. Impact of misspecifying the survival model

- **Bias** of the survival estimator $\hat{S}(t|Z)$:

$$\text{Bias}\left(\hat{S}(t|Z)\right) = \frac{1}{n_s}\sum_{i=1}^{n_s}\hat{S}^{(i)}(t|Z) - S(t|Z) \qquad (18)$$

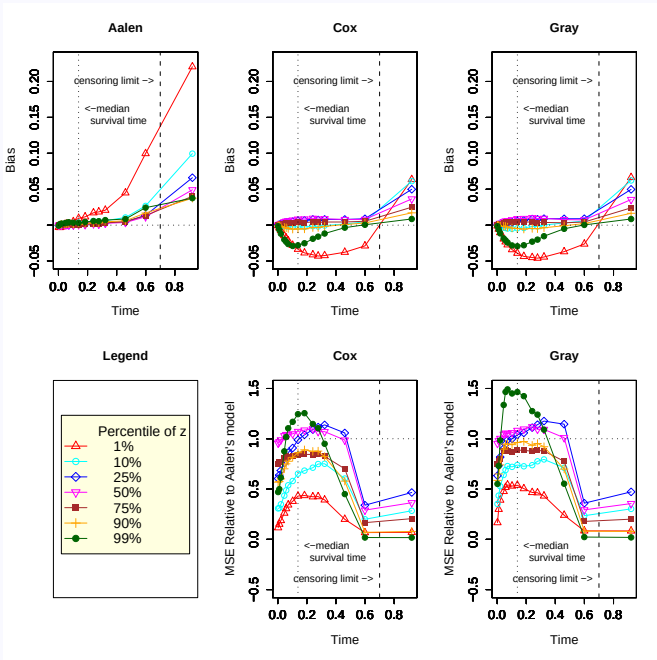
- **Mean Square Error** of the estimated survival $\hat{S}(t|Z)$:

$$\text{MSE}\left(\hat{S}(t|Z)\right) = \frac{1}{n_s}\sum_{i=1}^{n_s}\left(\hat{S}^{(i)}(t|Z) - S(t|Z)\right)^2 \qquad (19)$$

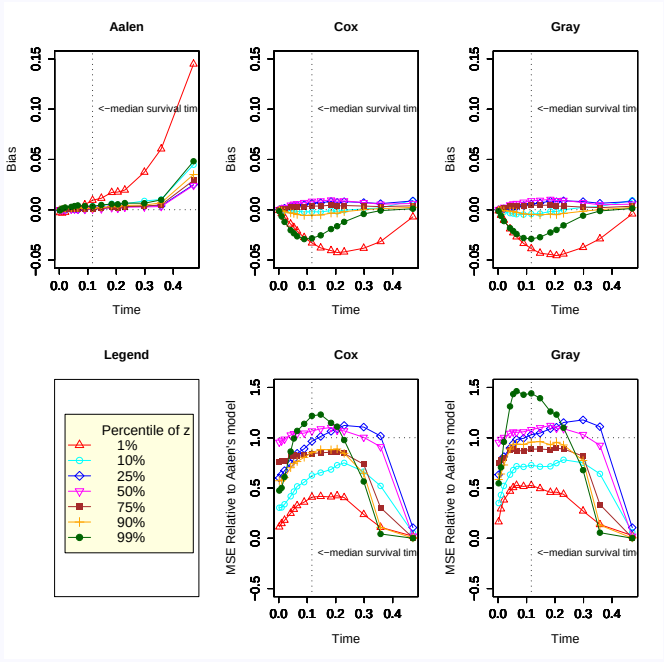
- **Bias-variance trade-off:**

$$\text{MSE}\left(\hat{S}(t|Z)\right) = \text{var}(\hat{S}) + \text{Bias}^2(\hat{S}) \qquad (20)$$

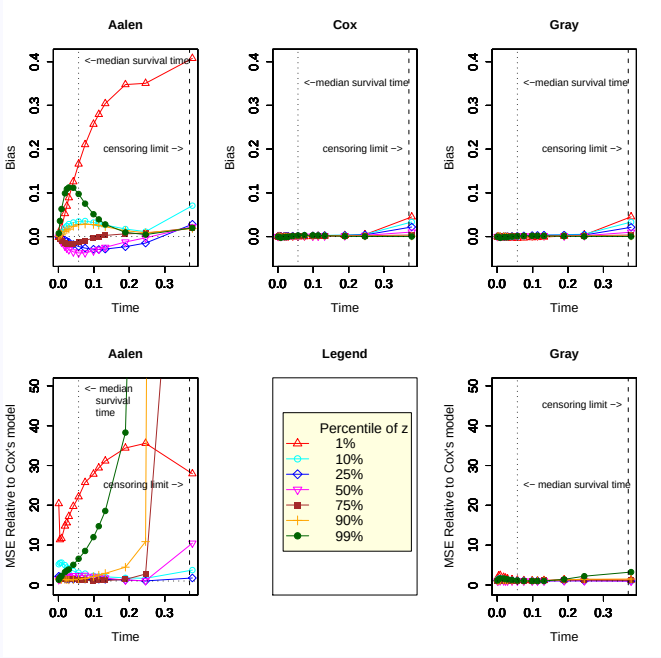
Aalen's model with constant β



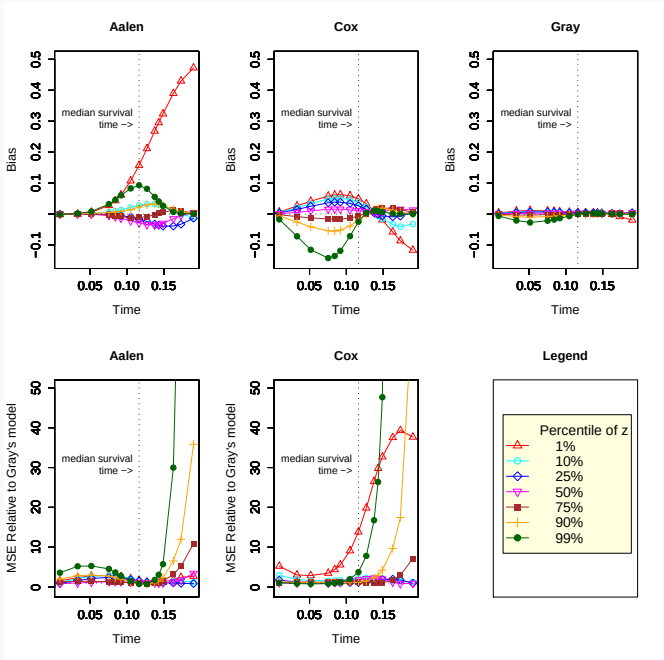
Aalen's model with time-varying $\beta(t)$



Cox PH model



Gray's model with time-varying $\beta(t)$



VII. Discussion

- When the data satisfied **Aalen's linear model**, both Cox's a Gray's model rendered biased survival estimates. They have, however, often shown a lower MSE than the native model for the data at hand.
- When analyzing **Cox PH model** data using the Gray's routine, we observed no dramatic increase in bias and MSE relative to native model, while using the same criteria the survival estimates based on Aalen's model appeared to be highly distorted.
- When the data followed **Gray's model** with time-varying covariate effects, both Cox's and Aalen's model rendered in terms of bias and MSE highly unreliable estimates of the conditional survival distribution. In other words, there was no alternative to using the native model in this instance.

References

[1] Cox DR: **Regression Models and Life Tables (with discussion)**, *Journal of the Royal Statistical Society*, 1972, Vol. 34, pp. 187–220.

[2] Aalen OO: **A linear regression model for the analysis of life times**, *Statistics in Medicine*, 1989, Vol. 8, pp. 907–925.

[3] Gray RJ: **Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis**, *Journal of the American Statistical Association*, 1992, Vol. 87, pp. 942–951.

[4] Gray RJ: **Spline-based tests in survival analysis**, *Biometrics*, 1994, Vol. 50., pp. 640–652.

[5] Valenta Z and Weissfeld LA: **Estimation of the Survival Function for Gray's Piecewise-Constant Time-Varying Coefficients Model**, *Statistics in Medicine*, 2002, Vol. 21(5), pp. 717-727.

Thank you for your attention!