Joint Models with Multiple Longitudinal Outcomes

Dimitris Rizopoulos

July 11, 2017

Outcomes in Follow-up Studies

- Often in follow-up studies different types of outcomes are collected
  - multiple longitudinal responses (e.g., markers, blood values)
  - time-to-event(s) of particular interest (e.g., death, relapse)

- Depending on the questions of interest, different types of statistical analysis are required

Outcomes in Follow-up Studies (cont'd)

- Focus *simultaneously* on multiple outcomes
  - association between longitudinal outcomes over time? *(evolution of the association)*
  - how longitudinal profiles interrelate with each other? *(association of the evolutions)*
  - which features of the longitudinal profiles are associated with the risk of death?

Goals of this talk

- *Our aims here are:*
    - brief review of joint models
    - two of their extensions
        - functional form
        - multiple longitudinal outcomes

Illustrative Case Study

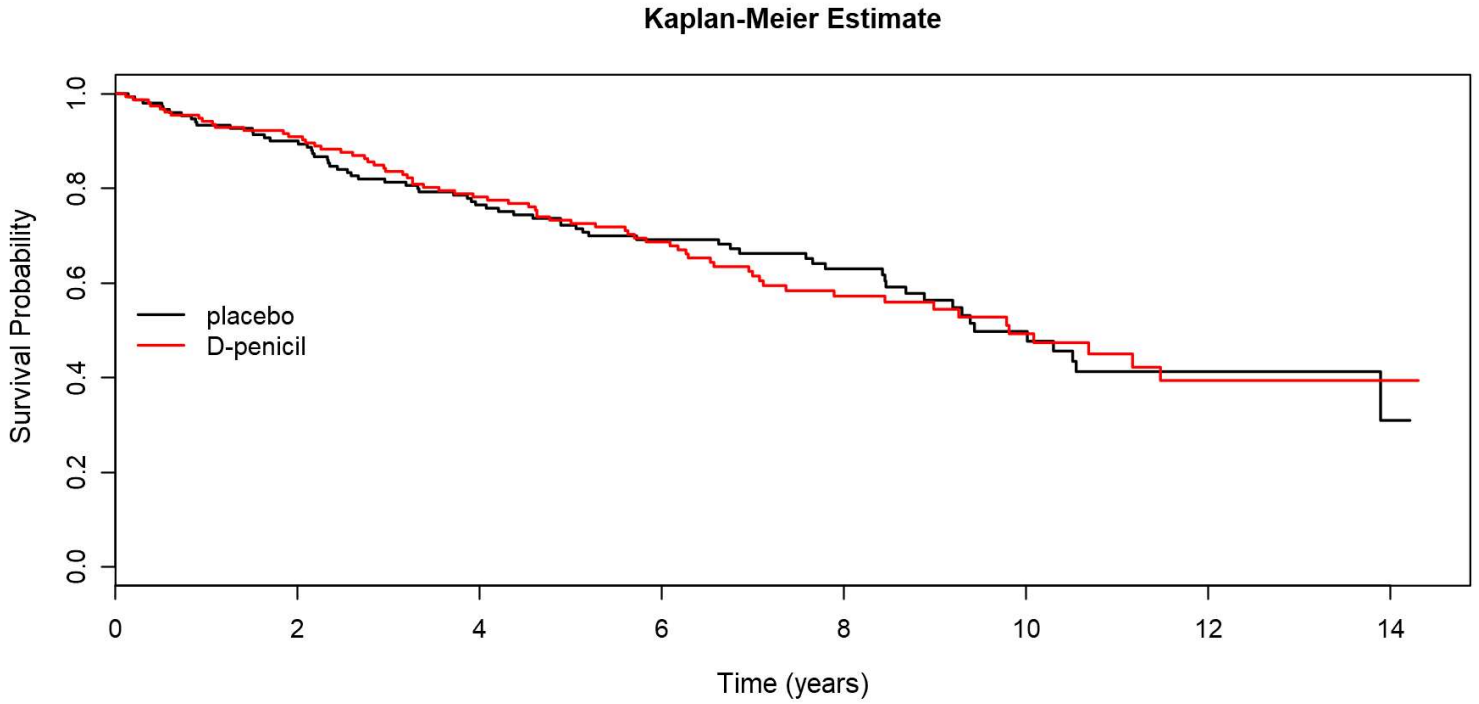
- Mayo Clinic PBC data: Primary Biliary Cirrhosis
  - a chronic, fatal but rare liver disease
  - characterized by inflammatory destruction of the small bile ducts within the liver

- Outcomes of interest:
  - time to death and/or liver transplantation
  - longitudinal
    - bilirubin, cholesterol, prothrombin time (continuous)
    - ascites, hepatomegaly, spiders (dichotomous)

Illustrative Case Study (cont'd)

Outcome:

survival



Illustrative Case Study (cont'd)

- Research Questions:
  - How strong is the association between the longitudinal biomarkers and the risk of death?
  - How the observed biomarker levels could be utilized to provide predictions of survival probabilities?

Time-varying Covariates

- To answer these questions we need to link
  - the survival outcome
  - the longitudinal biomarkers

- Biomarkers are *endogenous* time-varying covariates

Time-varying Covariates (cont'd)

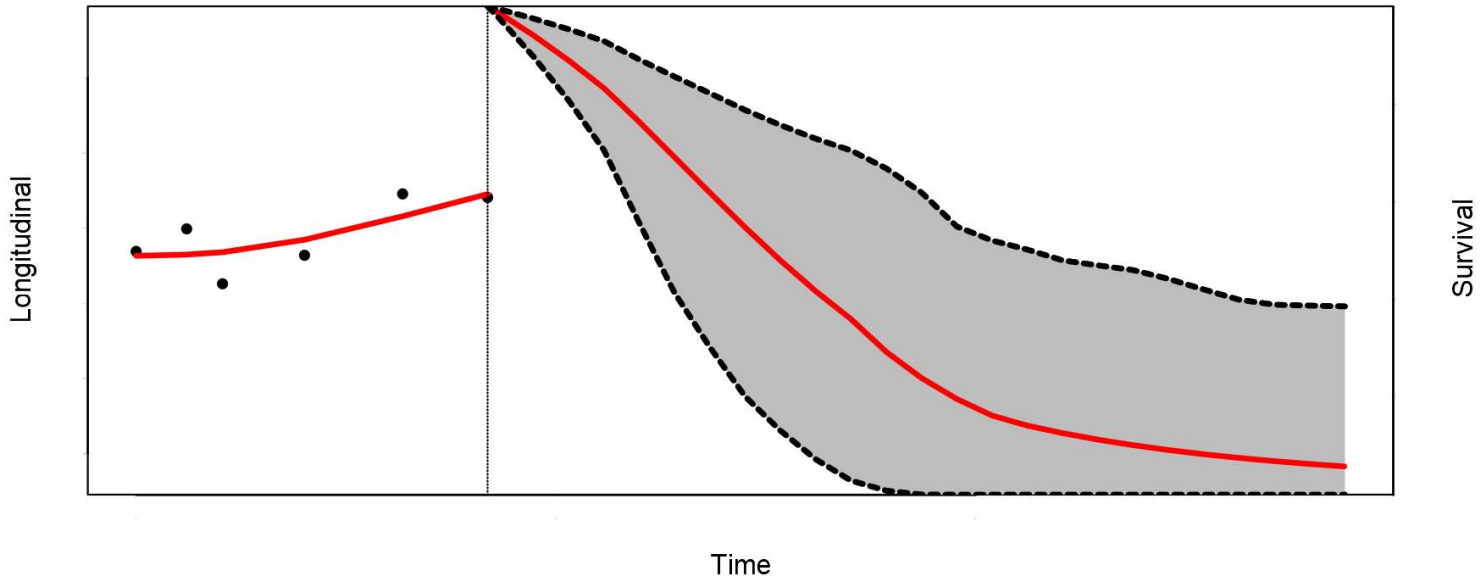
To account for endogeneity we use the framework of

Joint Models for Longitudinal & Survival Data

The Basic Joint Model

Measurements:

1 6 10



The Basic Joint Model (cont'd)

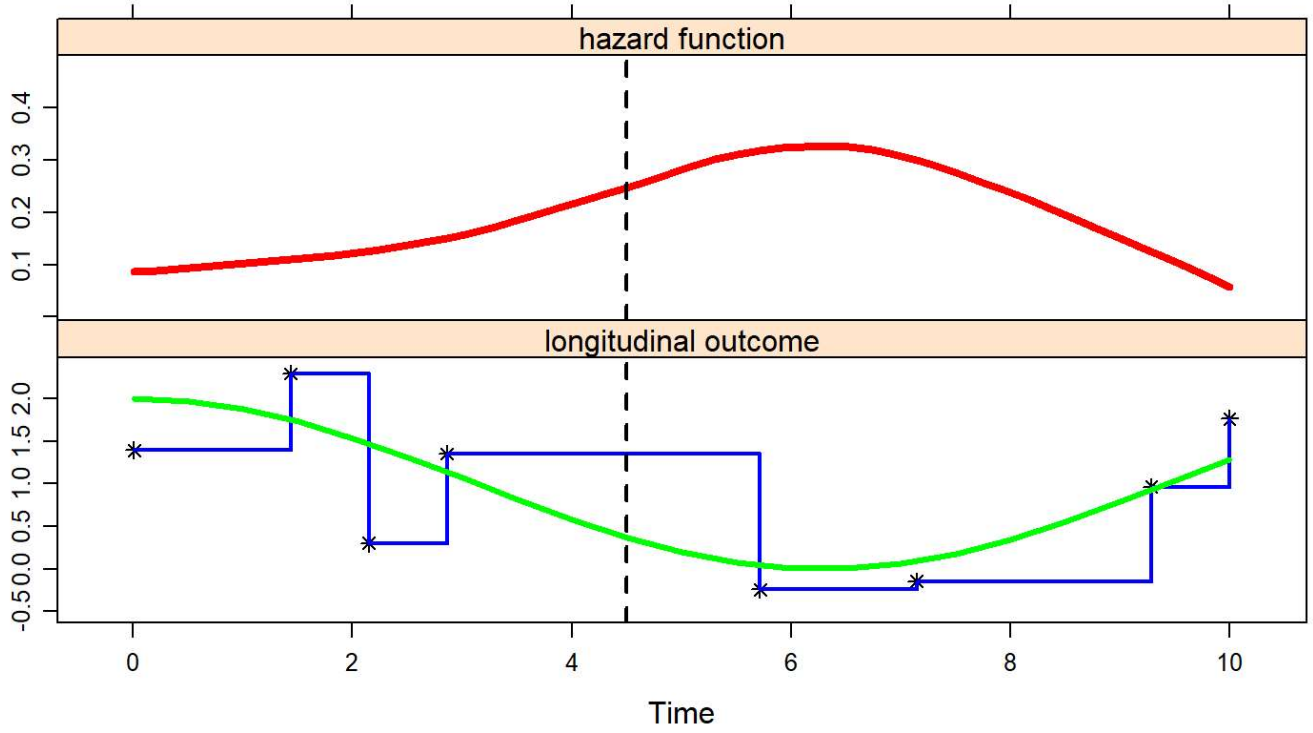
- We need some notation
  - T_i^* the true event times
  - T_i the observed event times
  - δ_i the event indicator
  - $\mathbf{y}_i$ the vector of longitudinal measurements
  - $\mathcal{Y}_i(t) = \{y_i(s), 0 \leq s < t\}$

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

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

  - 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…

The Basic Joint Model (cont'd)

- **Example:** A simple joint model for risk of death & serum bilirubin
  - Longitudinal outcome:

$$\begin{aligned} \log(\texttt{serBilir}_{ij}) &= \eta_i(t_{ij}) + \varepsilon_{ij} \\ & \beta_0 + \beta_1 N(t_{ij})_1 + \beta_2 N(t_{ij})_2 + \beta_3 \texttt{Female}_i + \\ & \beta_4 \texttt{Age}_i + b_{i0} + b_{i1} N(t_{ij})_1 + b_{i2} N(t_{ij})_2 + \varepsilon_{ij} \end{aligned}$$

    where
    - $N(t_{ij})_1$ and $N(t_{ij})_2$ denote the basis for a natural cubic spline with two degrees of freedom
    - $b_i \sim \mathcal{N}(0, D)$ and $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$

The Basic Joint Model (cont'd)

- **Example:** A simple joint model for risk of death & serum bilirubin
  - survival outcome:

$$h(t) = h_0(t) \exp\{\gamma_1 \texttt{Female}_i + \gamma_2 \texttt{Age}_i + \alpha \eta_i(t)\}$$

where

$$\log h_0(t) = \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, v)$$

with $B_q(t, v)$ denoting the q-th basis function of a B-spline with knots $v_1, \ldots, v_Q$

The Basic Joint Model (cont'd)

- Results: Survival submodel

	Post.Mean	2.5% CI	97.5% CI	P_tail
sex:Female	-0.016	-0.483	0.445	0.93
Age	0.066	0.047	0.084	0
α	**1.257**	**1.063**	**1.463**	**0**

- *Interpretation:* A unit increase of log(serBilir) at time t results in a 3.5-fold (95% CI: 2.9; 4.3) increase of the risk at t

Extensions

- Several extensions have been proposed in the literature - among others
  - competing risks & multistate models
  - frailty models
  - AFT models
  - latent classes
  - …

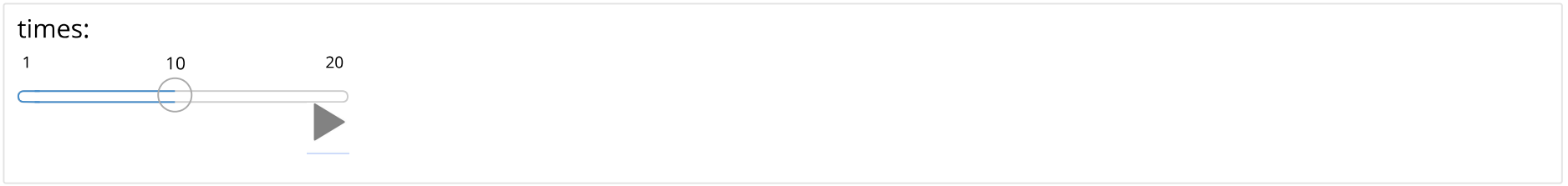
- We focus on two …

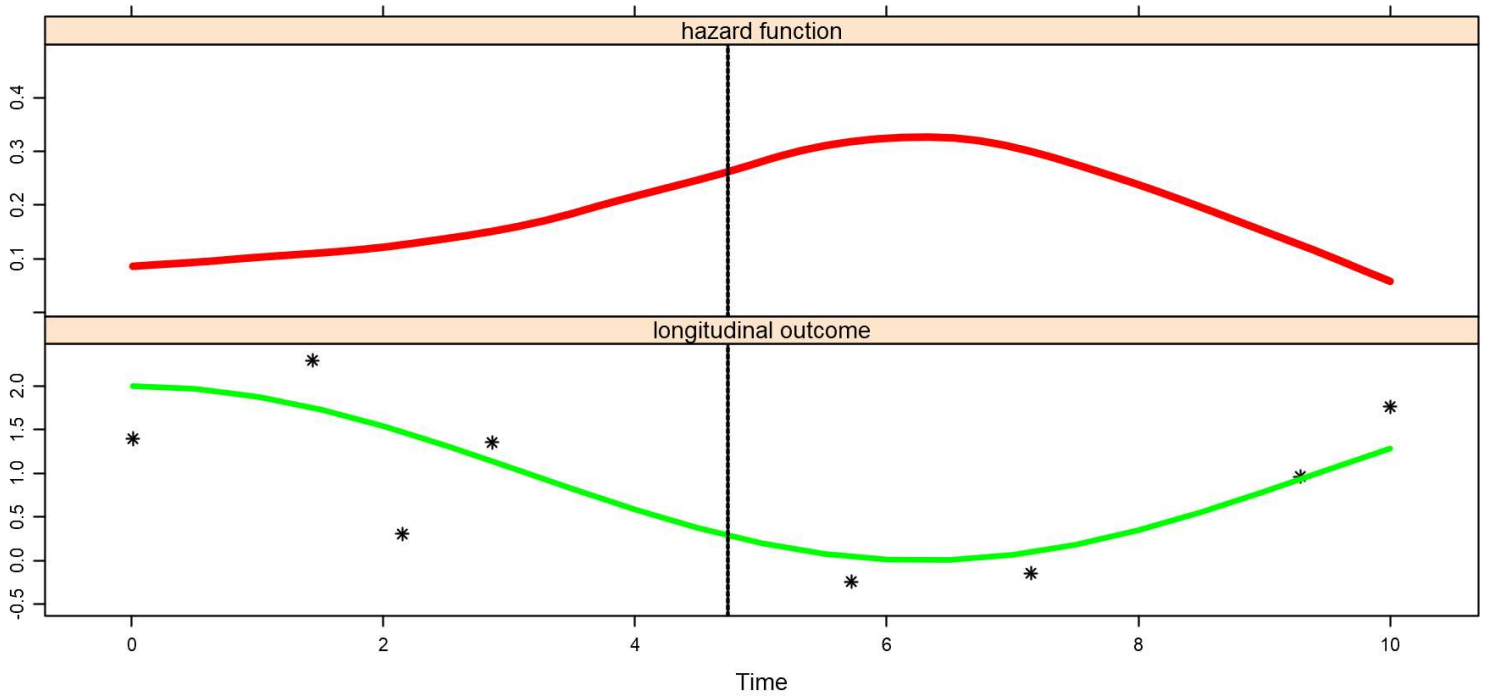
Functional Form

- 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)





Functional Form (cont'd)

Is this the only option?

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

- Let's see some possibilities…

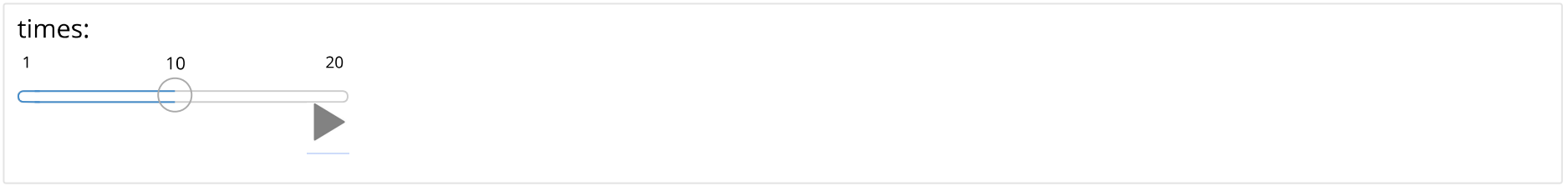
Functional Form (cont'd)

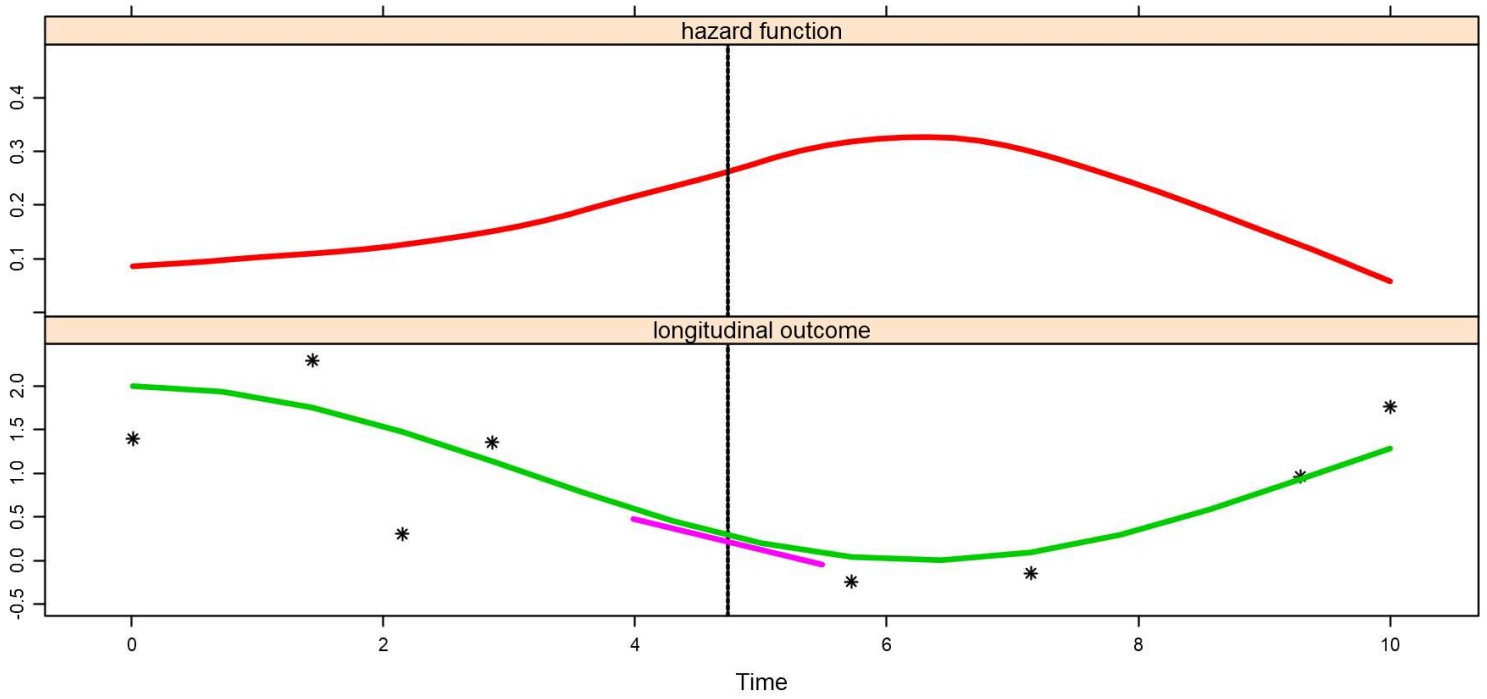
- Some options: Biomarker's rate of change
  - In prostate cancer, fast increasing PSA indicative of cancer

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

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

Functional Form (cont'd)





Functional Form (cont'd)

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

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

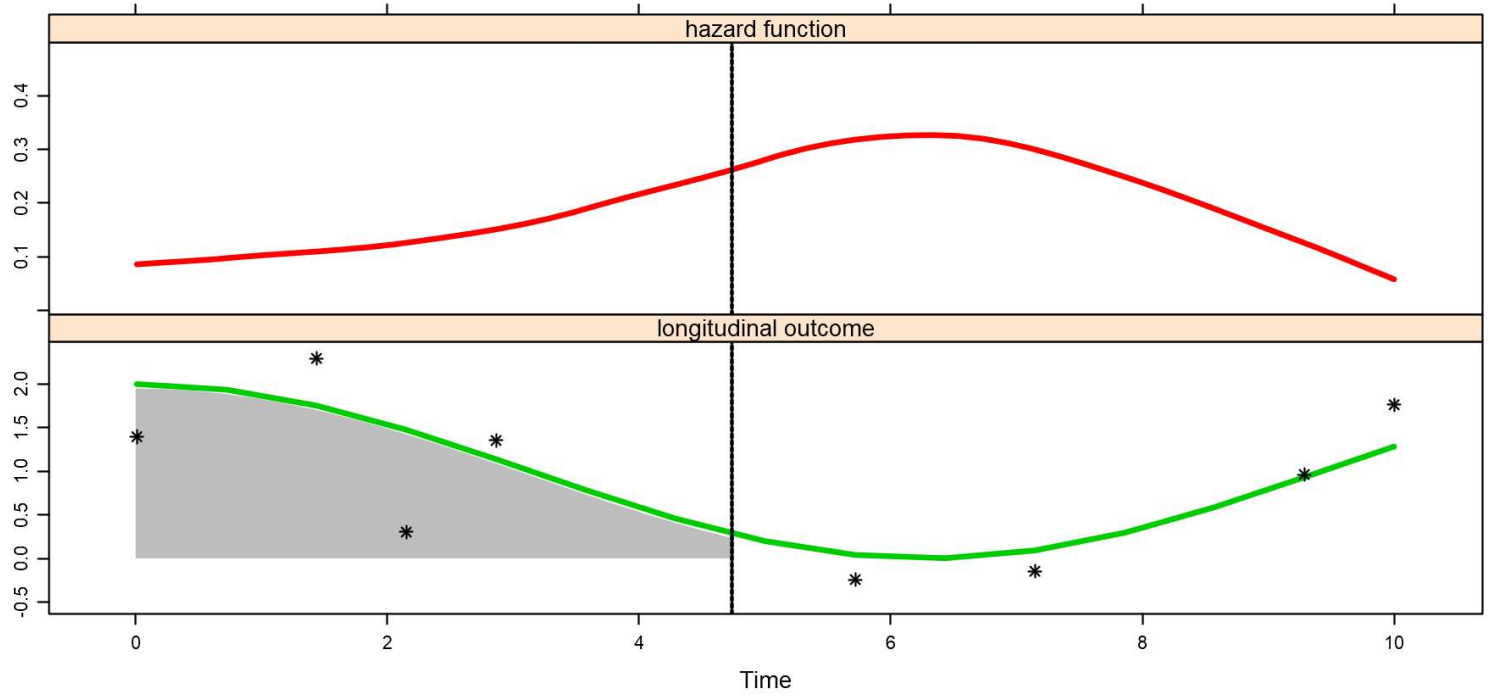
  - or even weighted cumulative effects

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

Functional Form (cont'd)

times:

1 10 20

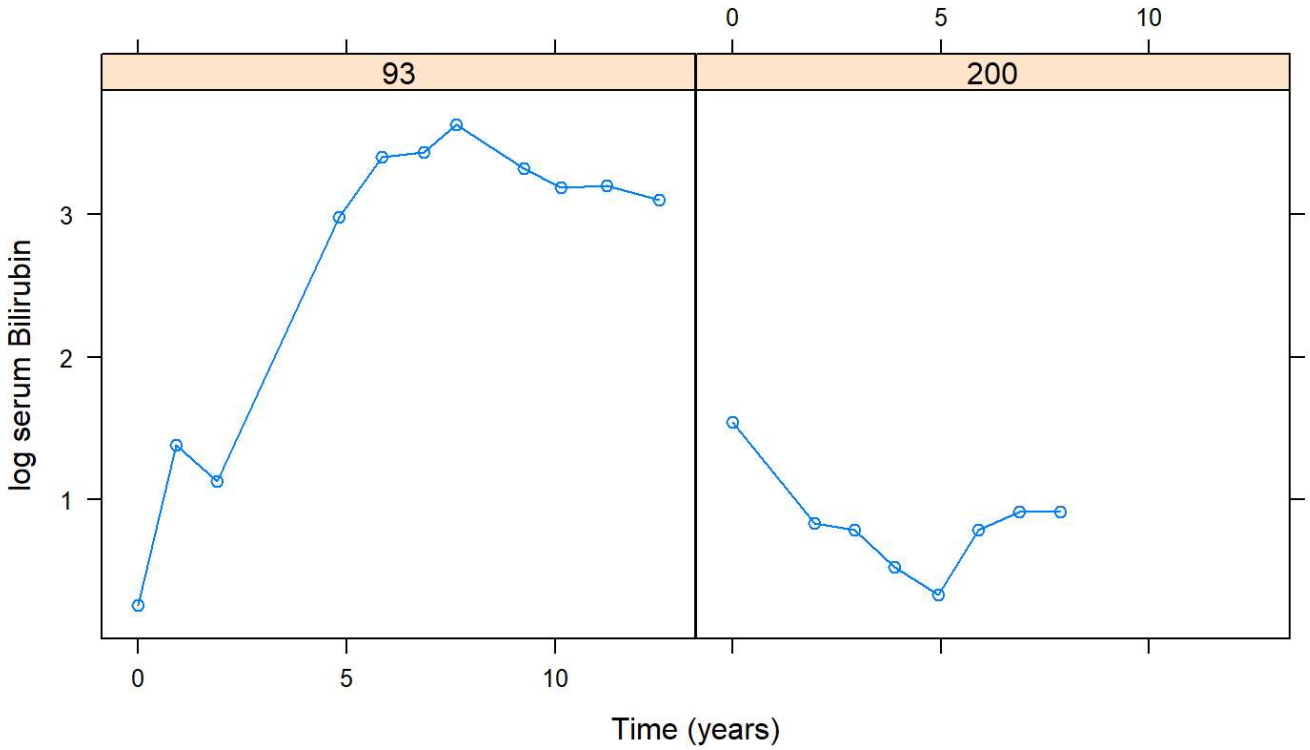


Functional Form (cont'd)

- **Example:** We extend the model we fitted for serum bilirubin
  - the same mixed model as before
  - Three functional forms for the relative risk model
    - current value (the one we have seen)
    - current value & current slope
    - cumulative effect

- We *dynamically* compare Patients 93 and 200

Functional Form (cont'd)



Functional Form (cont'd)

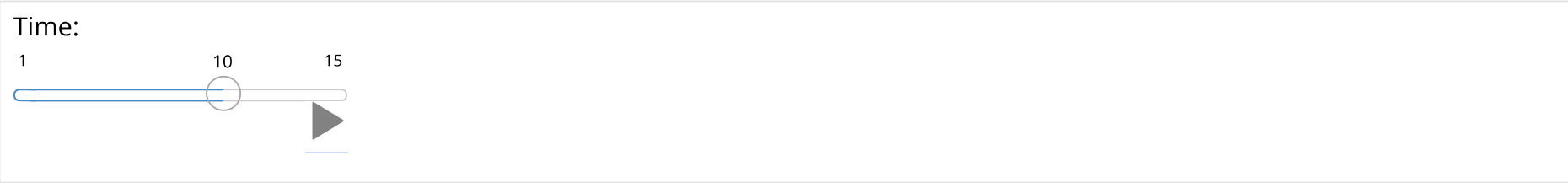
- We compute the dynamic 2-year Risk Ratio

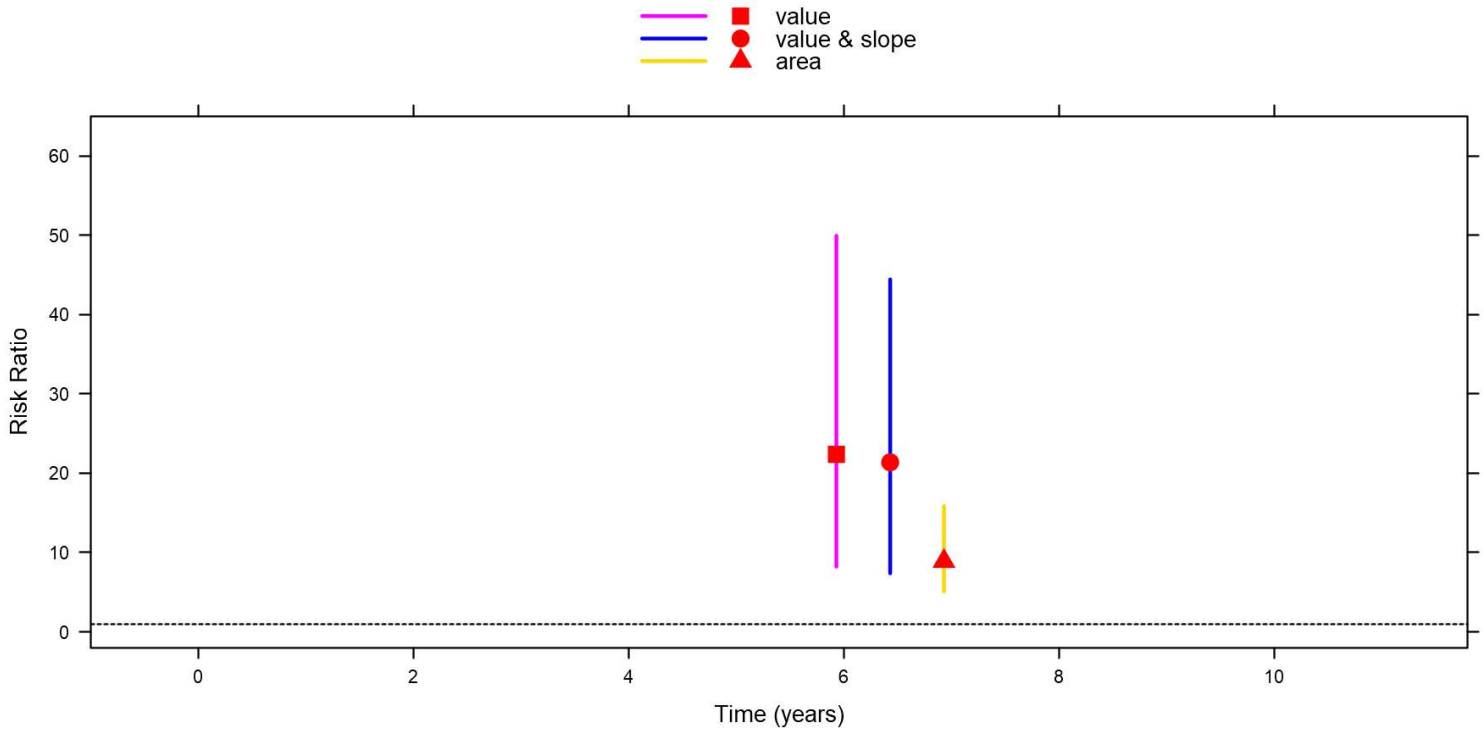
$$RR(t) = \frac{\Pr\{T_i^* \leq t+2 \mid T_i^* > t, \mathcal{Y}_i(t)\}}{\Pr\{T_j^* \leq t+2 \mid T_j^* > t, \mathcal{Y}_j(t)\}}$$

  where

  - i denotes Patient 93 and j Patient 200
  - $\mathcal{Y}_i(t)$, $\mathcal{Y}_j(t)$ denote their longitudinal measurements up to t

Functional Form (cont'd)





Multivariate Joint Models

- Up to now we have focused on a single longitudinal outcome

- However, very often, several biomarkers are relevant in predicting an event
  - e.g., in the PBC study
    - bilirubin, cholesterol, prothrombin time (continuous)
    - ascites, hepatomegaly, spiders (dichotomous)

Multivariate Joint Models (cont'd)

We need an extension of the basic joint model

Multivariate Joint Models (cont'd)

- Formally, we have
  - K possible longitudinal outcomes, i.e., $\mathbf{Y}_{1i}, \ldots, \mathbf{Y}_{Ki}$
  - multivariate generalized linear mixed model

$$\begin{cases} g_k\left[E\{y_{ki}(t) \mid \mathbf{b}_{ki}\}\right] = \eta_{ki}(t) = \mathbf{x}_{ki}^{\top}(t)\beta_k + \mathbf{z}_{ki}^{\top}(t)\mathbf{b}_{ki} \\ \\ h_i(t) = h_0(t)\exp\left\{\gamma^{\top}\mathbf{w}_i + \sum_{k=1}^{K}\alpha_k\eta_{ki}(t)\right\} \end{cases}$$

Multivariate Joint Models (cont'd)

- The association between the longitudinal outcomes is build via random effects

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_{1i} \\ \mathbf{b}_{2i} \\ \vdots \\ \mathbf{b}_{Ki} \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$$

- **(very) high-dimensional random effects**

Multivariate Joint Models (cont'd)

- Several papers on multivariate joint models
  - a couple under (pseudo) maximum likelihood
  - but mainly under the Bayesian approach or two-stage approaches
- Why?
  - high dimensional random effects
  - MCMC more robust than Gaussian quadrature

Multivariate Joint Models (cont'd)

- Bayesian approach - Practicalities
  - **advantages:**
    - it can be *generally* implemented in JAGS/WinBUGS
  - **disadvantages:**
    - zeros trick
    - painfully slow (2 hours even for just two longitudinal outcomes)

Multivariate Joint Models (cont'd)

- Even though in the majority of these papers the model is written for K longitudinal outcomes
- In practice it is only fitted for 2 or 3 outcomes ...

Multivariate Joint Models (cont'd)

Hence, a practical deadlock!

Multivariate Joint Models (cont'd)

- To overcome these difficulties some papers have proposed to work with **two-stage approaches**
  - fit the longitudinal outcomes in the first stage, and
  - then combine them with the survival one

- Computationally easier
  - it could be done with standard software
  - **however biased results!**

Multivariate Joint Models (cont'd)

It sounds like a lost cause!

Multivariate Joint Models (cont'd)

Our proposed solution

Corrected Two-Stage Approach

IS Two-Stage

- Why does the 2-stage approach give biased results?
  - because it **does not** work with the joint likelihood

- **Hence, to correct the two-stage approach we need the full likelihood**

- However, it is *not efficient* to work with the full joint likelihood due to the aforementioned computational problems

IS Two-Stage (cont'd)

- However, under a Bayesian approach there is a possible solution, namely

Importance Sampling (IS)

- IS allows to use a sample from a *wrong* distribution, and adjust it to look like a sample from the *correct* one

IS Two-Stage (cont'd)

- **Stage I:**
  - Fit a multivariate mixed effects model to the longitudinal outcomes alone
  - We obtain an MCMC sample from the distribution

$$\{\theta_y^{(m)}, \mathbf{b}^{(m)};\ m = 1, \ldots, M\} \ \sim \ [\theta_y, \mathbf{b} \mid \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}]$$

- **Stage II:**
  - For each MCMC realization from the first stage we obtain a value for the parameters of the survival model

$$\{\theta_t^{(m)};\ m = 1, \ldots, M\} \ \sim \ [\theta_t \mid T_i, \delta_i, \mathbf{b}^{(m)}, \theta_y^{(m)}]$$

IS Two-Stage (cont'd)

- The combined MCMC sample from the two-stage approach can be corrected with the weights

$$\widetilde{w}^{(m)} = \frac{p(\theta_t^{(m)}, \theta_y^{(m)}, \mathbf{b}^{(m)} \mid T_i, \delta_i, \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki})}{p(\theta_t^{(m)} \mid T_i, \delta_i, \theta_y^{(m)}, \mathbf{b}^{(m)})\; p(\theta_y^{(m)}, \mathbf{b}^{(m)} \mid \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki})}$$

$$w^{(m)} = \widetilde{w}^{(m)} \Big/ \sum_{m=1}^{M} \widetilde{w}^{(m)}$$

IS Two-Stage (cont'd)

- If you do the math …

$$\widetilde{w}^{(m)} = p(T_i, \delta_i \mid \mathbf{b}^{(m)}, \theta_y^{(m)})$$

$$= \int p(T_i, \delta_i \mid \theta_t, \mathbf{b}^{(m)}, \theta_y^{(m)}) p(\theta_t) \, d\theta_t$$

  - Hence, a marginal likelihood calculation

IS Two-Stage (cont'd)

- Approaches to estimate marginal likelihoods
  - Power posteriors
    - more accurate estimate of marginal likelihood
    - but computationally intensive

  - **Laplace approximation**

IS Two-Stage (cont'd)

- Notes:
  - **Stage I** can be easily performed in STAN or JAGS/WinBUGS
    - quite fast; no requirement for the *zeros trick*
  - **Stage II** separate sampling for each realization from Stage I
    - Embarrassingly parallel problem
    - parallel computing utilizing CPU cores

IS Two-Stage (cont'd)

- *OK, how to do it in practice?*
- A suit of functions has been added in the JMbayes (https://cran.r-project.org/package=JMbayes) package
  - `mvglmer()` fits multivariate mixed models using JAGS or STAN using parallel computing for the multiple chains
  - lme4 (https://cran.r-project.org/package=lme4)-like syntax
  - for example, two longitudinal outcomes, one continuous & one binary using JAGS

```
multMixed <- mvglmer(list(y1 ~ group * time + (time | id),
                          y2 ~ group * time + (1 | id)),
                     data = dat, n.processors = 2,
                     families = list(gaussian, binomial))
```

IS Two-Stage (cont'd)

- To fit the same model with STAN, we simply set the `engine` argument

```
multMixed <- mvglmer(list(y1 ~ group * time + (time | id),
                          y2 ~ group * time + (1 | id)),
                     data = dat, n.processors = 2,
                     families = list(gaussian, binomial),
                    engine = "STAN")
```

IS Two-Stage (cont'd)

- The MCMC sample from `multMixed` is then used in `mvJointModelBayes()`
  - MCMC sampling of θ_t written in C++ based on Rcpp (https://cran.r-project.org/package=Rcpp) and RcppArmadillo (https://cran.r-project.org/package=RcppArmadillo)
  - parallel computing using package foreach (https://cran.r-project.org/package=foreach) with back-end package parallel (https://cran.r-project.org/)

```
CoxFit <- coxph(Surv(Time, event) ~ group, dat.id, model = TRUE)

multJM <- mvJointModelBayes(multMixed, CoxFit, timeVar = "time", update_RE = FALSE)
```

IS Two-Stage (cont'd)

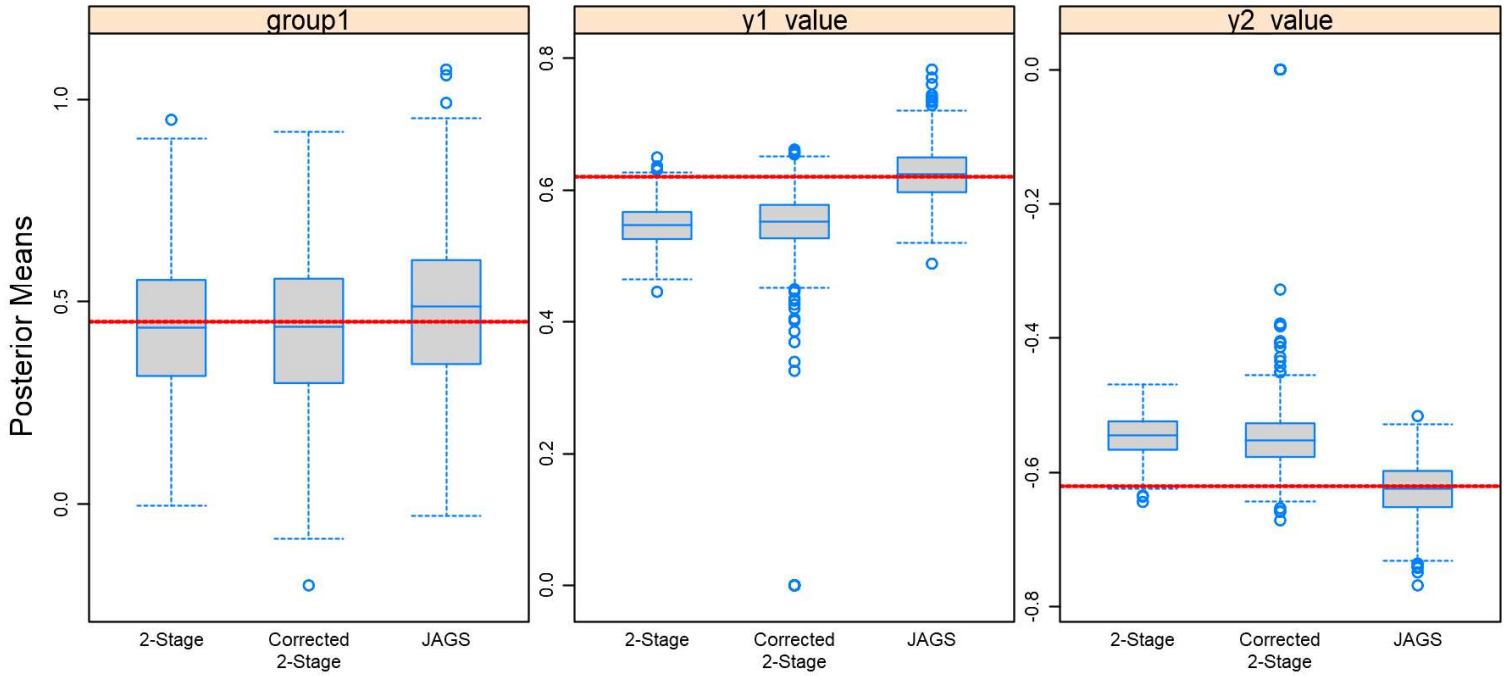
- *OK, how does it perform?*

- Simulation study
  - 2 longitudinal outcomes (both normal)
  - compare corrected two-stage approach with full Bayesian
  - Stage I: JAGS 2 chains run in parallel
  - Stage II: run in parallel using 4 cores

IS Two-Stage (cont'd)

Performance:

- Time ○
- Bias ◉

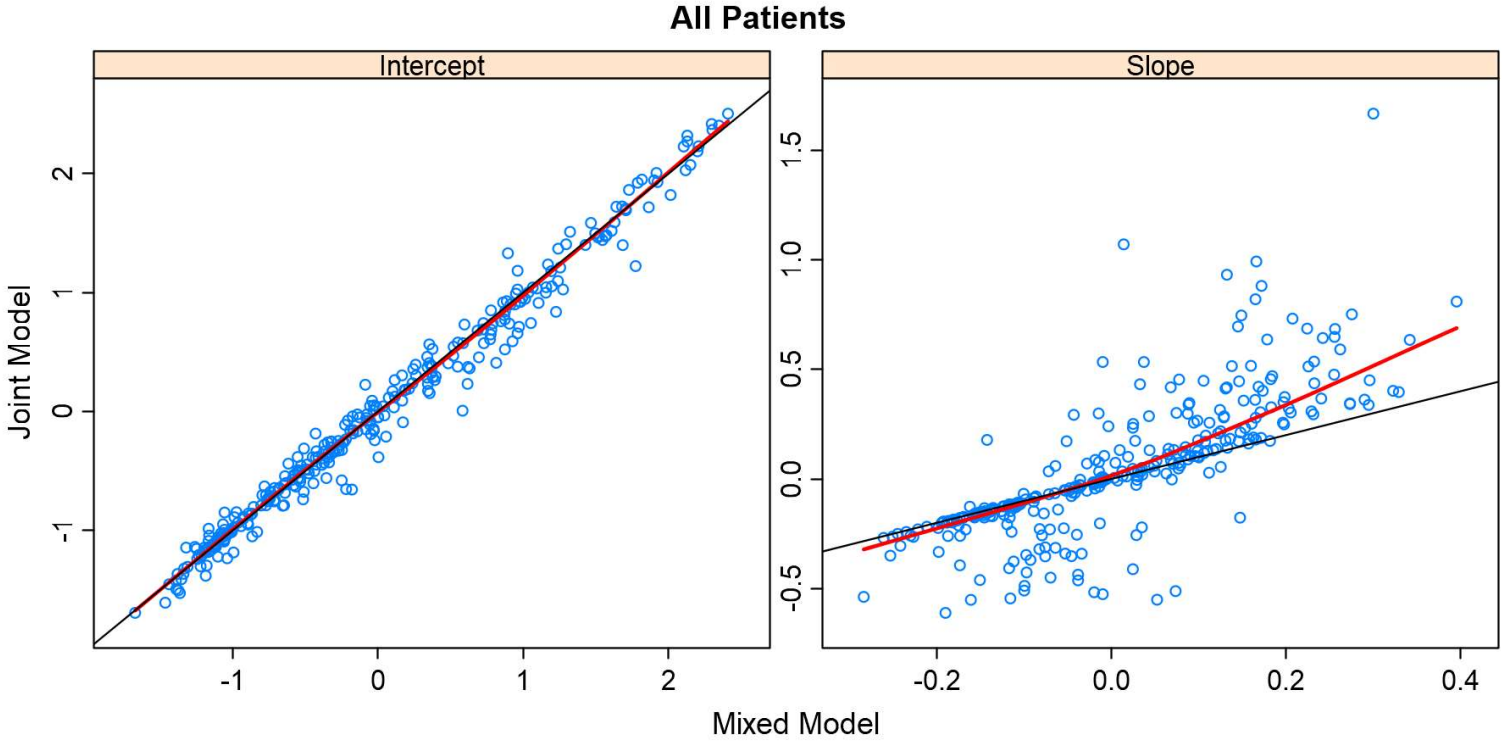


IS Two-Stage (cont'd)

- The correction does not seem to help much!!
- Why is that?
  - detective work …

IS Two-Stage (cont'd)

Sub-group:
- All
- Event-free
- Event



IS Two-Stage (cont'd)

- **Stage I:**
  - Fit a multivariate mixed effects model to the longitudinal outcomes alone
  - We obtain an MCMC sample from the distribution

$$\{\theta_y^{(m)}, \mathbf{b}^{(m)};\ m = 1, \ldots, M\} \ \sim \ [\theta_y, \mathbf{b} \mid \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}]$$

- **Stage II:**
  - For each MCMC realization from the first stage we obtain a value for the parameters of the survival model **and** the random effects

$$\{\theta_t^{(m)}, \mathbf{b}^{(m)};\ m = 1, \ldots, M\} \ \sim \ [\theta_t, \mathbf{b} \mid T_i, \delta_i, \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}, \theta_y^{(m)}]$$

IS Two-Stage (cont'd)

- Now Stage II is more challenging
  - Stage II-a: $\mathbf{b}^* \sim [\mathbf{b} \mid T_i, \delta_i, \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}, \theta_y^{(m)}, \theta_t^*]$
  - Stage II-b: $\theta_t^* \sim [\theta_t \mid T_i, \delta_i, \theta_y^{(m)}, \mathbf{b}^*]$

- Stage II-a: entails calculating the multivariate density of *all* longitudinal outcomes

IS Two-Stage (cont'd)

- The combined MCMC sample from the two-stage approach can be corrected with the weights

$$\widetilde{w}^{(m)} = \frac{p(\theta_t^{(m)}, \theta_y^{(m)}, \mathbf{b}^{(m)} \mid T_i, \delta_i, \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki})}{p(\theta_t^{(m)}, \mathbf{b}^{(m)} \mid T_i, \delta_i, \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}, \theta_y^{(m)})\; p(\theta_y^{(m)}, \mathbf{b}^{(m)} \mid \mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki})}$$

$$w^{(m)} = \widetilde{w}^{(m)} \Big/ \sum_{m=1}^{M} \widetilde{w}^{(m)}$$

IS Two-Stage (cont'd)

- Again we obtain a marginal likelihood computation

$$\widetilde{w}^{(m)} = \frac{p(\mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}, T_i, \delta_i \mid \theta_y^{(m)})}{p(\mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki} \mid \mathbf{b}_i^{(m)}, \theta_y^{(m)})\, p(\mathbf{b}_i^{(m)} \mid \theta_y^{(m)})}$$

where

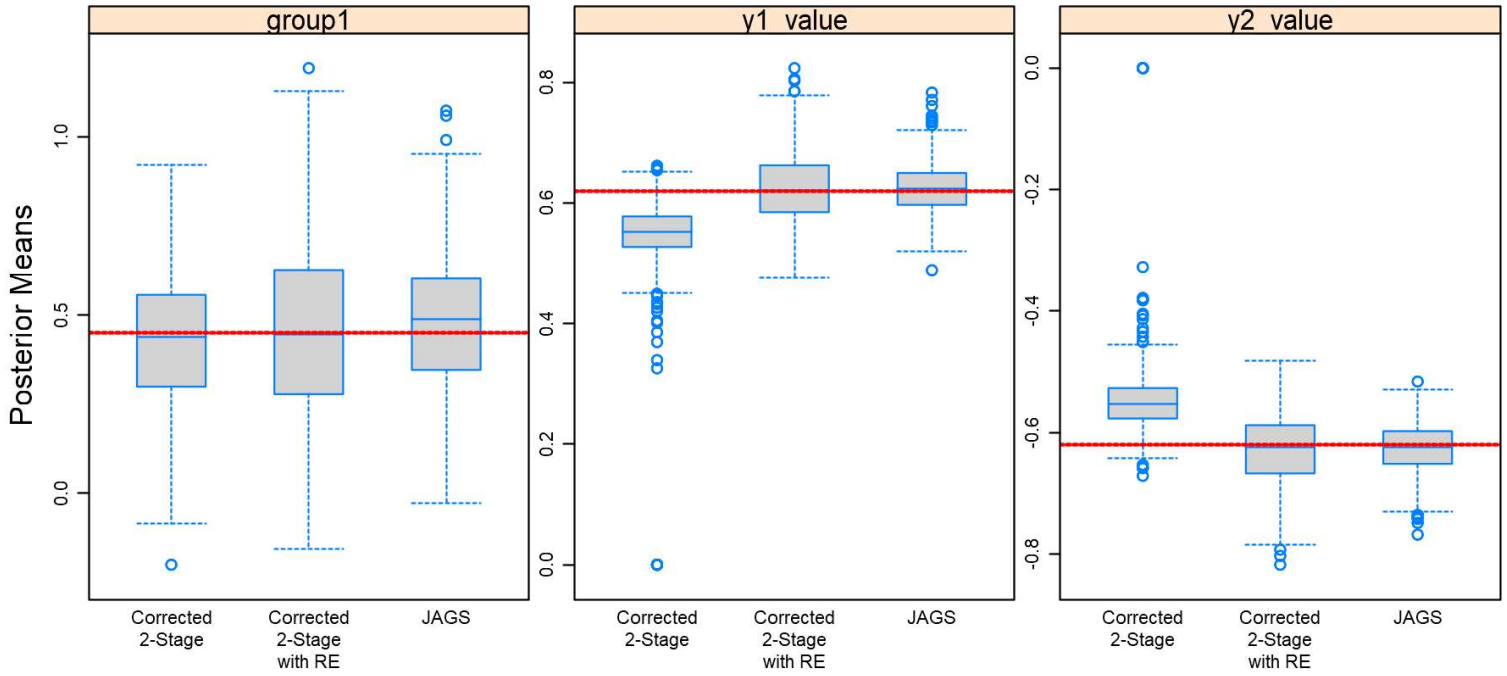
$$p(\mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki}, T_i, \delta_i \mid \theta_y^{(m)}) =$$

$$\int\int p(\mathbf{y}_{1i}, \ldots, \mathbf{y}_{Ki} \mid \mathbf{b}_i, \theta_y^{(m)}) p(T_i, \delta_i \mid \mathbf{b}_i, \theta_t, \theta_y^{(m)}) p(\mathbf{b}_i \mid \theta_y^{(m)}) p(\theta_t)\; d\mathbf{b}_i d\theta_t$$

IS Two-Stage (cont'd)

Performance:

- Time ○
- Bias ◉



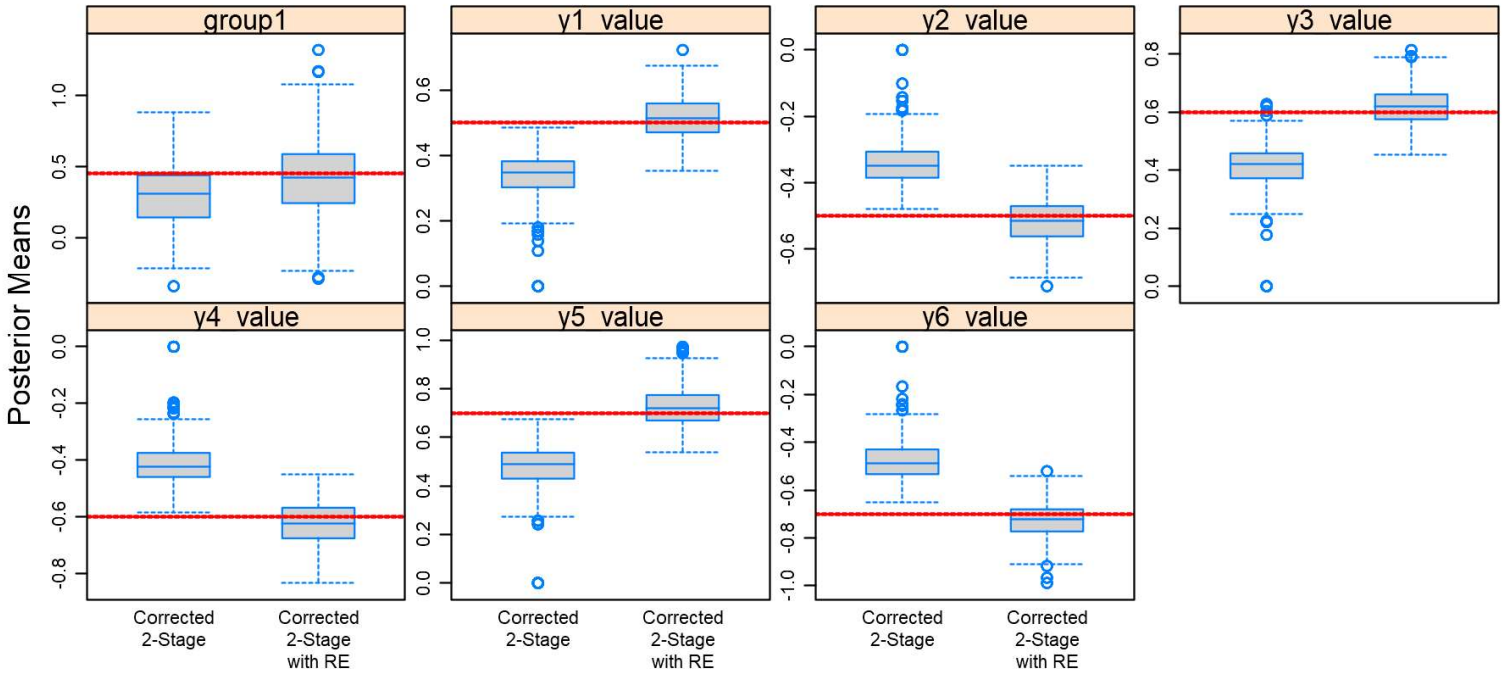
IS Two-Stage (cont'd)

- Extra simulation study
  - 6 longitudinal outcomes (all normally distributed)

IS Two-Stage (cont'd)

Performance:

- Time ○
- Bias ◉



IS Two-Stage (cont'd)

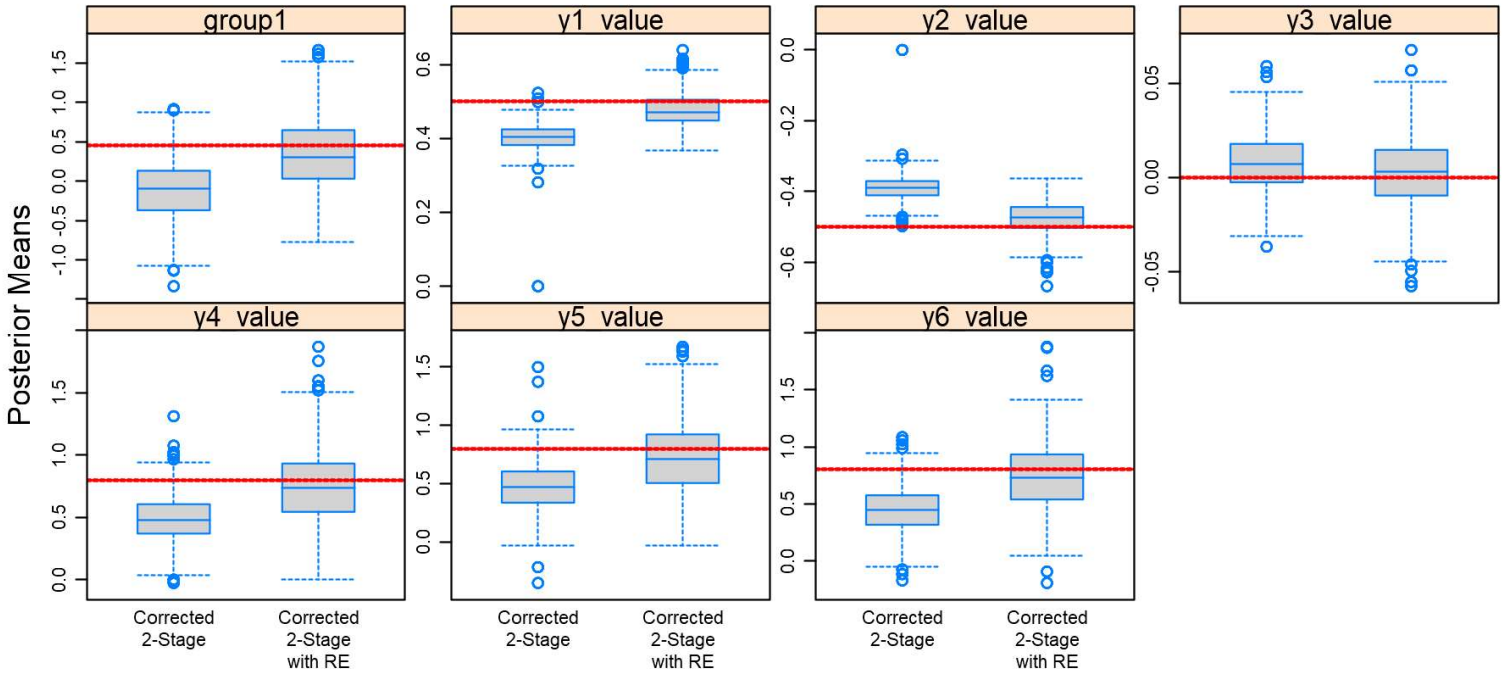
- Extra simulation study
  - 6 longitudinal outcomes
  - 3 continuous
  - 2 binary
  - 1 Poisson

- Run with STAN
  - much better than JAGS for mixed-type multivariate mixed models

IS Two-Stage (cont'd)

Performance:

- Time ○
- Bias ◉



IS Two-Stage (cont'd)

- Also implemented within `mvJointModelBayes()` by setting argument `update_RE` to `TRUE` (which is actually the default)

```
multJM <- mvJointModelBayes(multMixed, CoxFit, timeVar = "time", update_RE = TRUE)
```

IS Two-Stage (cont'd)

- **Example:** We fit a multivariate joint model for the PBC with the longitudinal outcomes
  - serum bilirubin (continuous)
  - serum cholesterol (continuous)
  - prothrombin time (continuous)
  - ascites (dichotomous)
  - hepatomegaly (dichotomous)
  - spiders (dichotomous)

IS Two-Stage (cont'd)

	Post.Mean	2.5% CI	97.5% CI	P_tail
log_serBilir	0.176	-0.207	0.572	0.336
sqrt_serChol	-0.03	-0.09	0.09	0.984
prothro_time	1.235	0.55	1.659	0
ascites	0.336	0.003	0.5	0.046
hepatomegaly	-0.016	-0.084	0.251	0.288
spiders	-0.072	-0.168	0.062	0.368

Combo of Extensions

- So far we have considered the standard functional form, i.e.,

$$\begin{cases} g_k\left[E\{y_{ki}(t) \mid \mathbf{b}_{ki}\}\right] = \eta_{ki}(t) = \mathbf{x}_{ki}^{\top}(t)\beta_k + \mathbf{z}_{ki}^{\top}(t)\mathbf{b}_{ki} \\ \\ h_i(t) = h_0(t)\exp\left\{\gamma^{\top}\mathbf{w}_i + \sum_{k=1}^{K}\alpha_k\eta_{ki}(t)\right\} \end{cases}$$

Combo of Extensions (cont'd)

- However, for each of the K outcomes we may consider several functional forms simultaneously

$$\begin{cases} g_k\left[E\{y_{ki}(t) \mid \mathbf{b}_{ki}\}\right] = \eta_{ki}(t) = \mathbf{x}_{ki}^{\top}(t)\beta_k + \mathbf{z}_{ki}^{\top}(t)\mathbf{b}_{ki} \\ \\ h_i(t) = h_0(t)\exp\left\{\gamma^{\top}\mathbf{w}_i + \sum_{k=1}^{K}\sum_{l=1}^{L} f_{kl}(\mathcal{H}_{ki}(t), \alpha_{kl})\right\} \end{cases}$$

$\mathcal{H}_{ki}(t) = \{\eta_{ki}(s), 0 \leq s < t\}$ history k-th longitudinal outcome up to t

Combo of Extensions (cont'd)

- Functions $\{f_{kl}(\cdot); l = 1, \ldots, L\}$ define which components of the history of outcome k are associated with the hazard

- Choice of the optimal functional form(s) per longitudinal outcome using suitable priors for α_{kl} (Andrinopoulou and Rizopoulos, 2016, SiM, 4813–4823)
  - Bayesian lasso
  - elastic net
  - Horseshoe prior
  - ridge
  - …

Combo of Extensions (cont'd)

- `mvJointModelBayes()` offers the option for a global-local ridge-type shrinkage prior, i.e.,

$$\begin{cases} \alpha_{kl} \sim \mathcal{N}(0, \tau\psi_{kl}) \\ \tau^{-1} \sim Gamma(0.1, 0.1) \\ \psi_{kl}^{-1} \sim Gamma(1, 0.01) \end{cases}$$

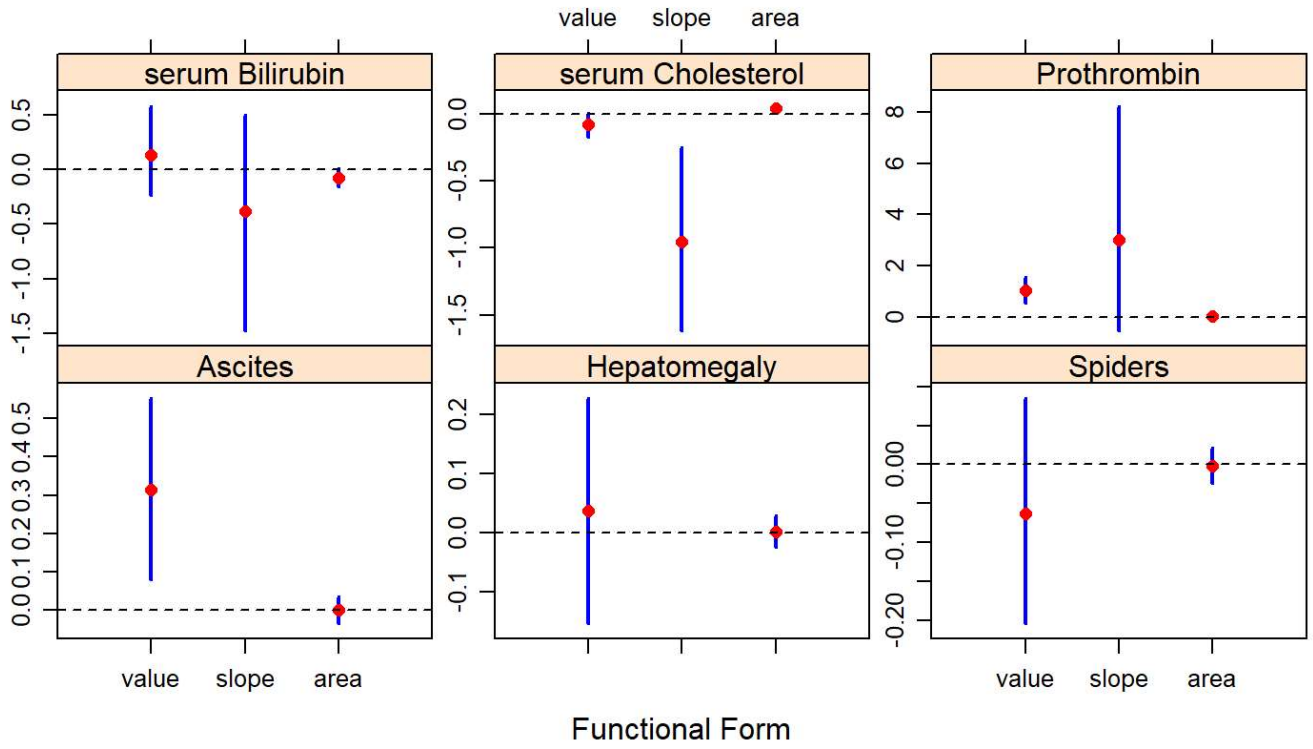
  using

```
mvJointModelBayes(..., priors = list(shrink_alphas = TRUE))
```

Combo of Extensions (cont'd)

- **Example:** We extend the multivariate joint model fitted to the PBC dataset
  - bilirubin, cholesterol, prothrombin time (continuous)
    - current value, current slope & cumulative effect
  - ascites, hepatomegaly, spiders (dichotomous)
    - current value & cumulative effect

Combo of Extensions (cont'd)



Thank you for your attention!

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