Joint Modeling of Efficacy and Laboratory Data in Clinical Trials

Kao-Tai Tsai, Ph.D.
JPHCOPH GSU & Celgene Corporation

Presentation at the 2016 BASS at Washington, D.C.

October 24, 2016

Outline of Presentation

Objective of this research

Motivation:

- Clinical trials collect lots of lab test data, however, they are rarely analyzed in full extend.
- Many institutions (e.g., NIH) had published various laboratory test guide regarding to various diseases.

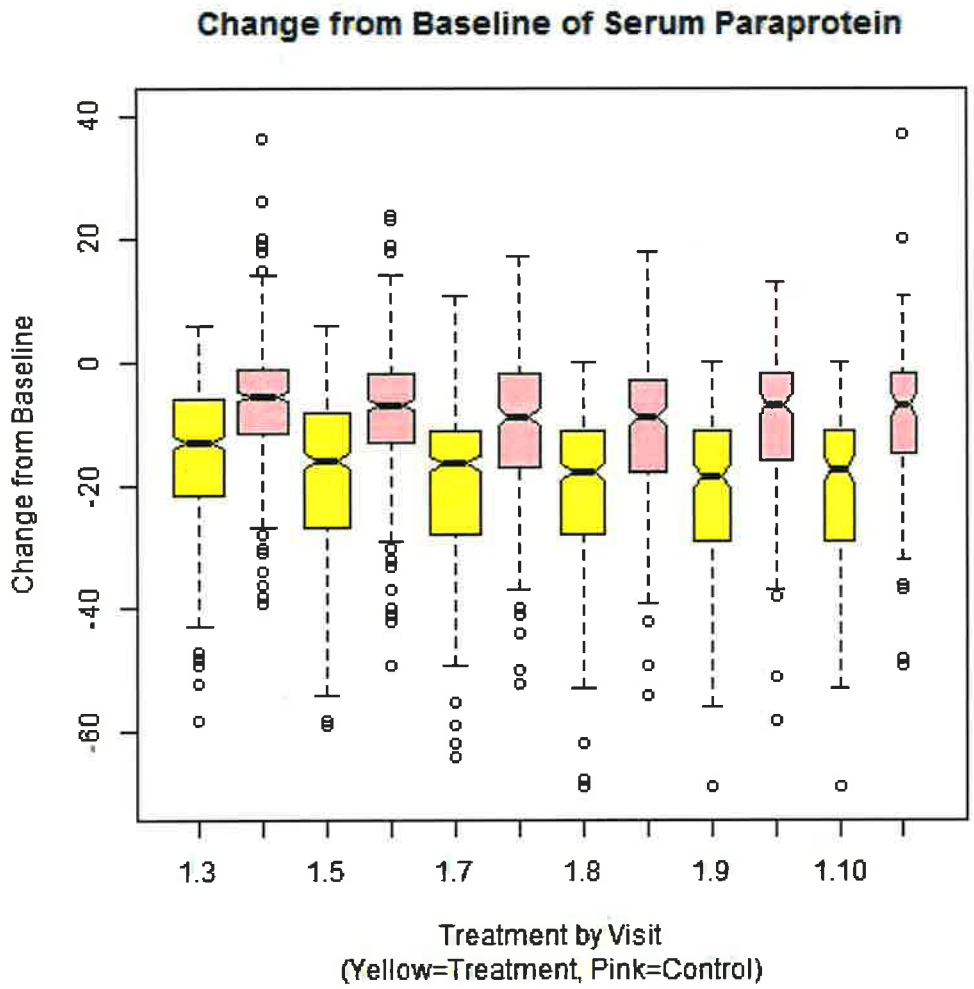
Objective:

- To propose a joint modeling of efficacy endpoint and multiple longitudinal processes to estimate treatment effect.

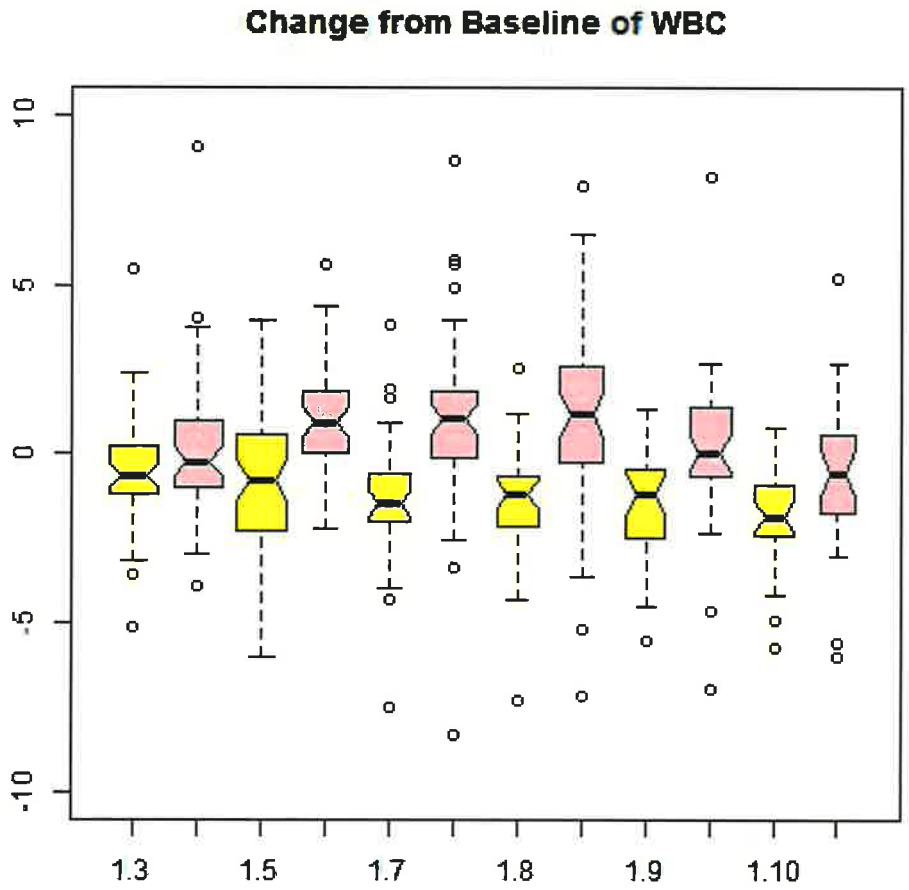
Data source:

- Clinical trial data of sample size about 650 for multiple myeloma.
- Efficacy data: time to disease progression (PFS).
- Various lab test longitudinal data series.

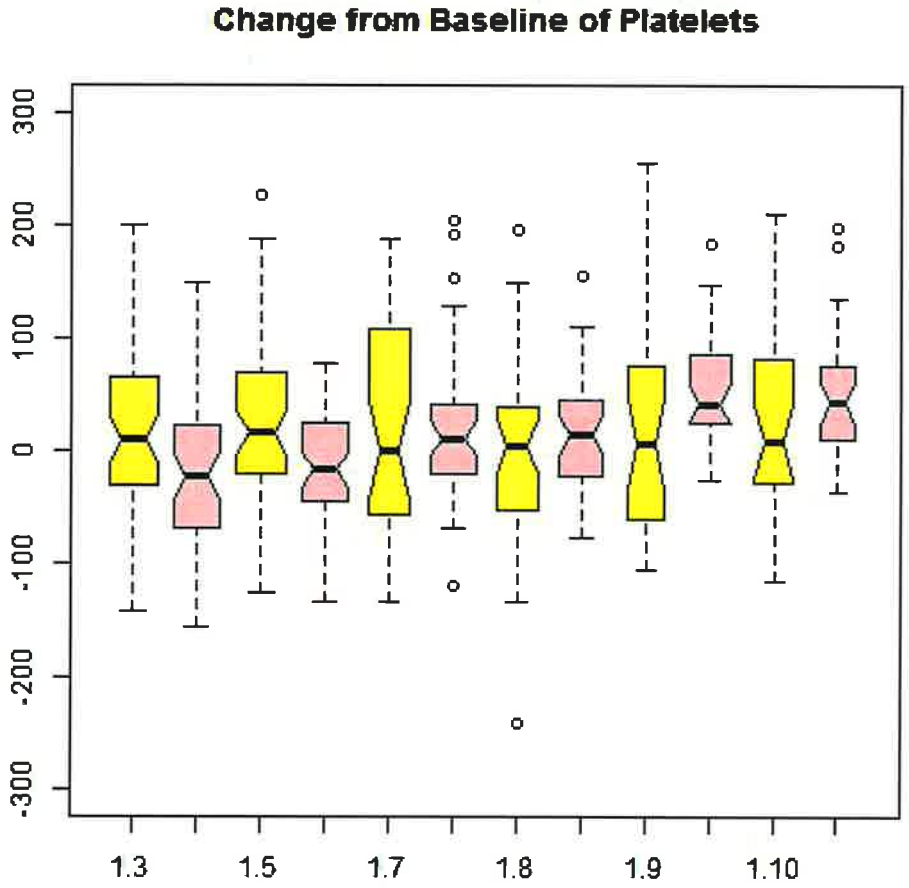
Longitudinal lab test values



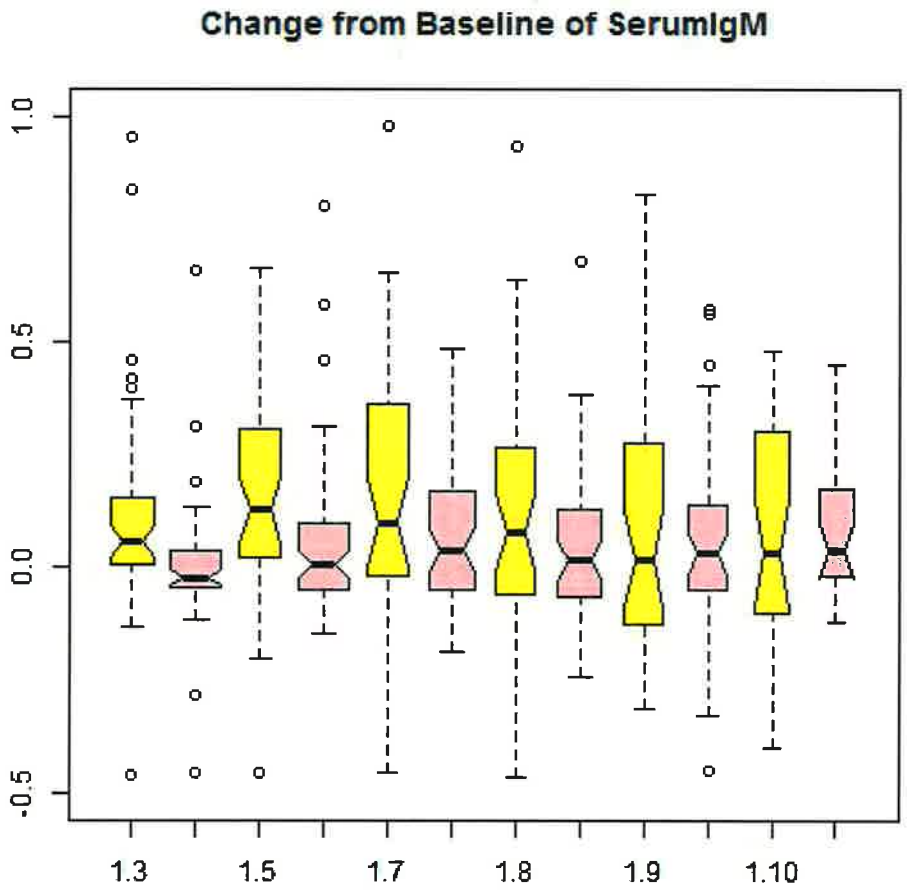
Longitudinal lab test values



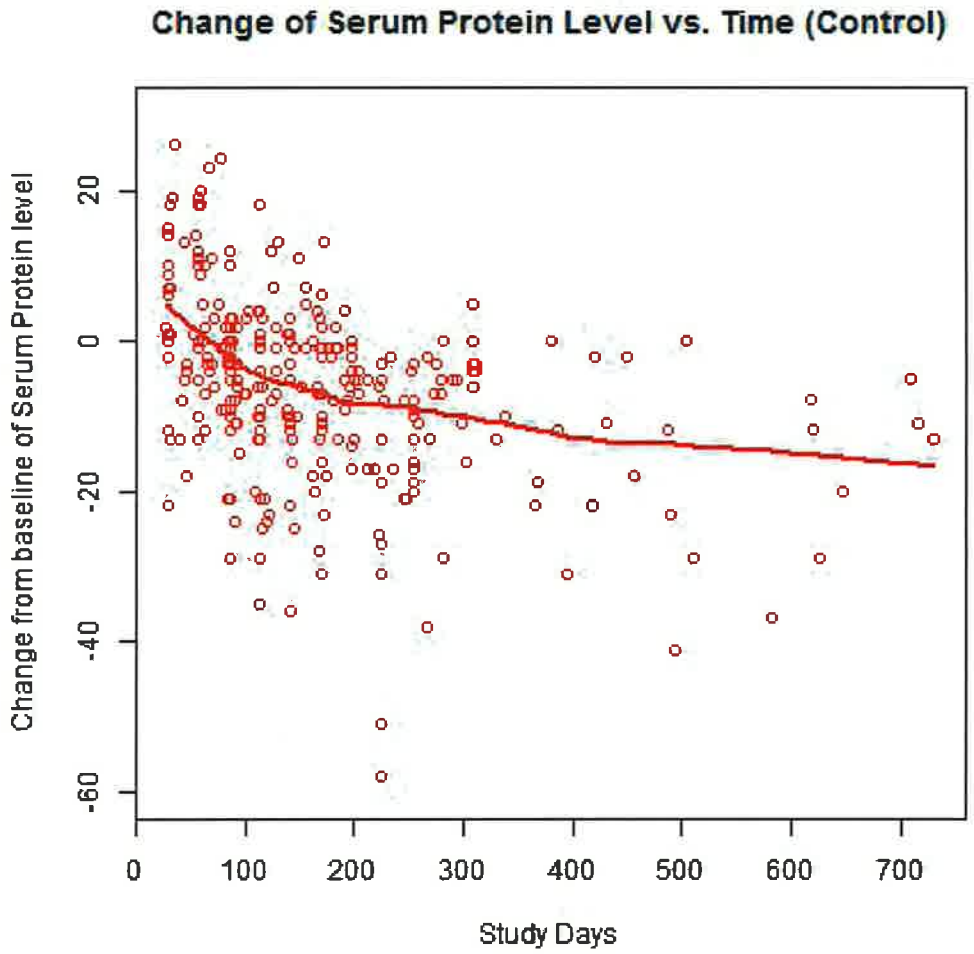
Longitudinal lab test values



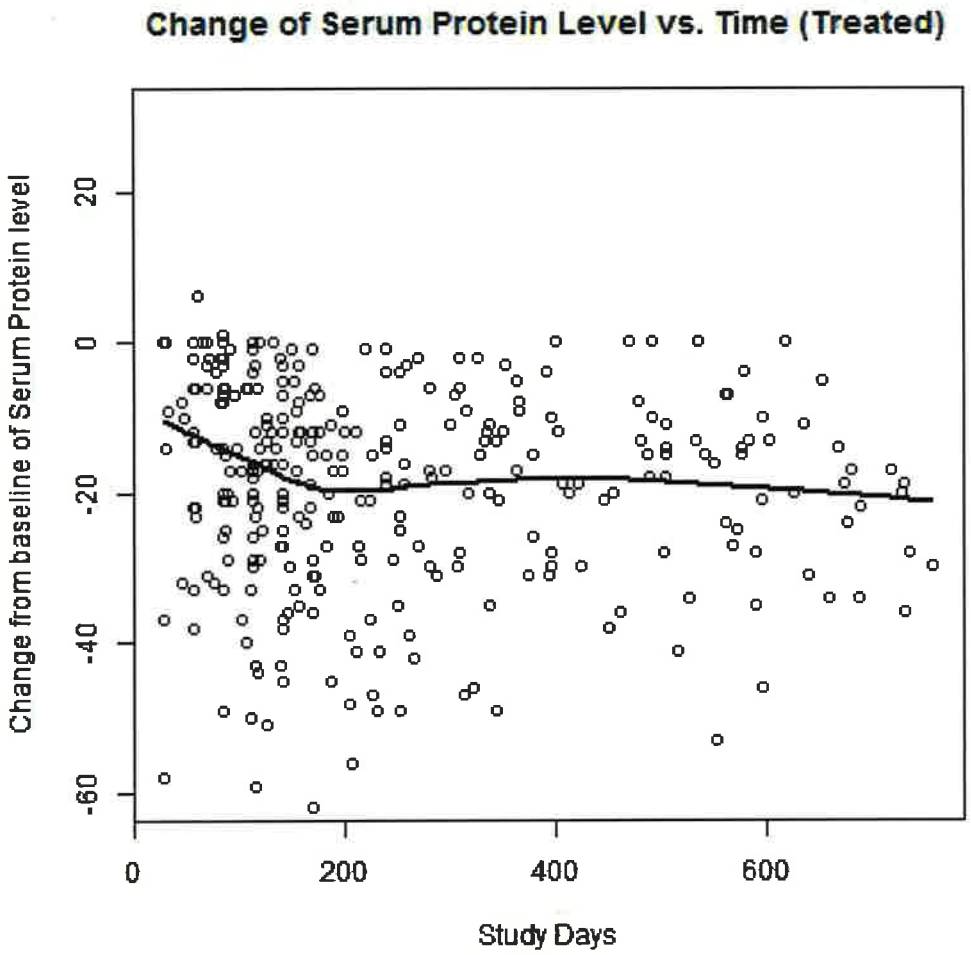
Longitudinal lab test values



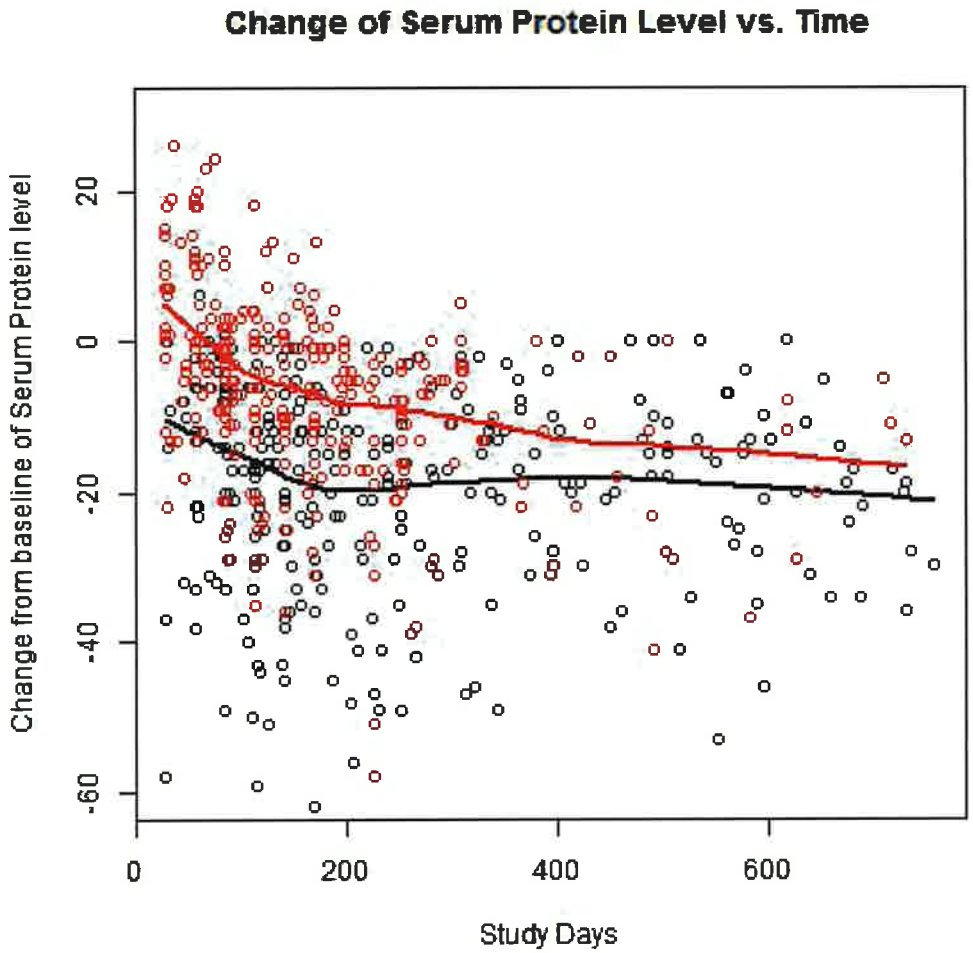
Serum Protein Change vs Time (Ctrl)



Serum Protein Change vs Time (Rx)



Serum Protein Change vs Time (Ctrl & Rx)



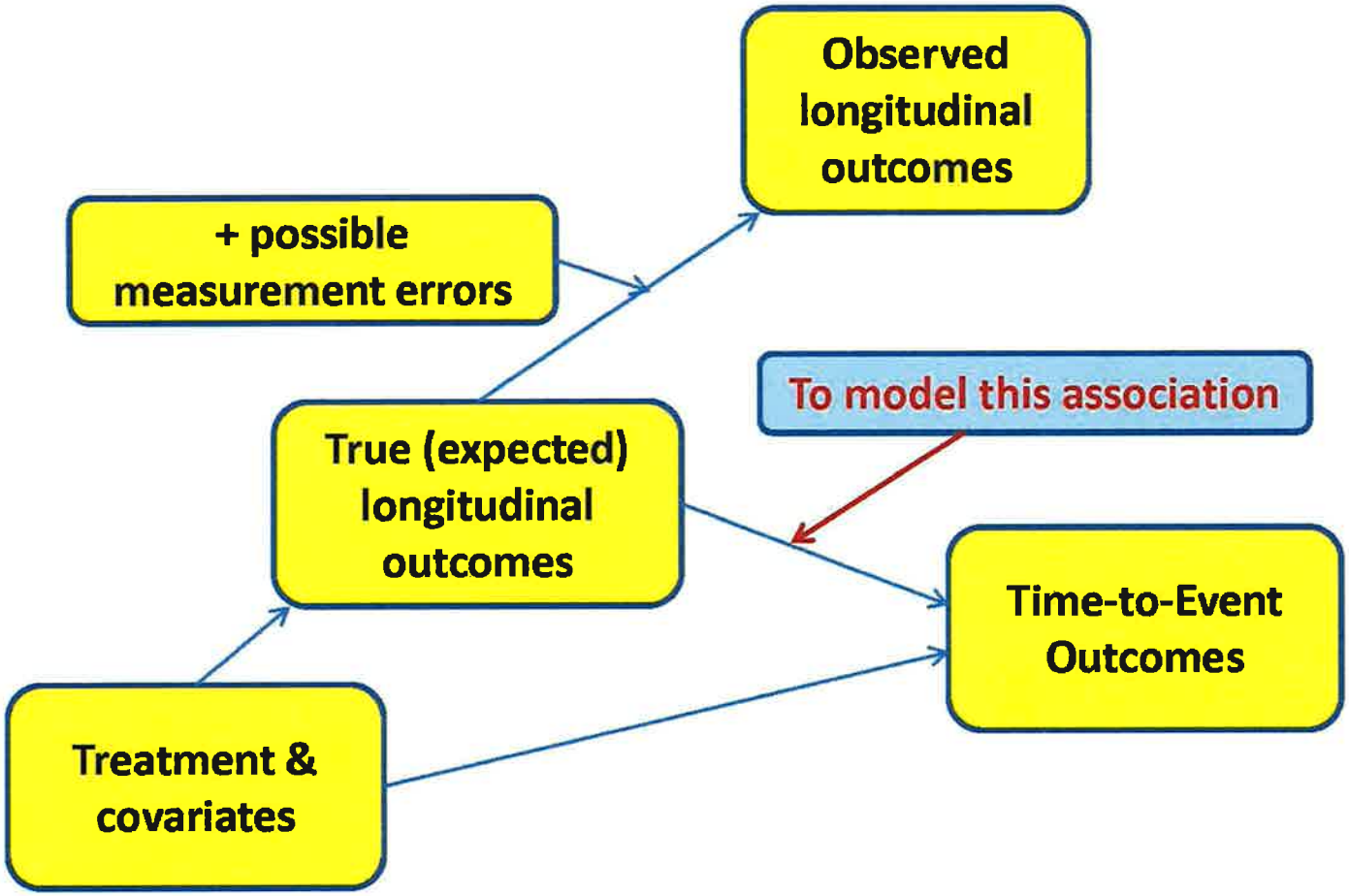
Rationale for Joint Modeling - 1

- Cancer studies very often collect time-to-event data and various repeated measurements of longitudinal data for each subject simultaneously.
- The longitudinal data, such as lab tests, genetic biomarker, or a health outcome, can be important predictors or surrogates of an event of interest, such as progression-free survival, relapse-free survival, or overall survival.
- Joint models for longitudinal data and time-to-event data are commonly used that bring these two data types together (simultaneously) into a single model so that one can infer the association between them, and to better assess the effect of a treatment.

Rationale for Joint Modeling - 2

- Joint models are increasingly used in clinical trial data analysis because they
  - provide more efficient estimate of the treatment effects on the event of interest,
  - provide more efficient estimate of the treatment effects on the longitudinal data series,
  - provide more detailed relationship on how hazard of events are affected by longitudinal process in dimension of time,
  - can potentially reduce the bias in the estimates of the overall treatment effect (however, which is not always equivalent to a more favorable result).

Rationale for Joint Modeling - 3



Notations and Models - 1

For subject i, $(i = 1, ..., N)$,

- For the time-to-event process, let
  - T_i^* denote the true event time,
  - C_i be the censoring time,
  - the distributions of T_i^* and C_i are independent,
  - the observed $T_i = \min(T_i^*, C_i)$,
  - $\delta_i = I(T_i^* \leq C_i)$ be the event indicator,
  - with hazard function $\lambda_i(t)$.
- For the longitudinal process, let $y_{ij}(t)$ denote the value of the longitudinal outcome at time point t_{ij} with $j = 1, \cdots, n_i$.

Notations and Models - 2

The basic idea of joint model assumes

- a longitudinal process

$$y_i(t) = \mathcal{F}_{1i}(t) + \mathcal{R}_{1i}(t) + \epsilon_i(t), \tag{1}$$

where $\mathcal{F}_{1i}(t)$ is a fixed effect, $\mathcal{R}_{1i}(t)$ is an unobserved random effect, and $\epsilon_i(t)$ is random measurement error.

- an event process, such as survival, with hazard function

$$\lambda_i(t) = \lambda_0(t) \exp\{\mathcal{F}_{2i}(t) + \mathcal{R}_{2i}(t)\}, \tag{2}$$

where $\mathcal{F}_{2i}(t)$ is a fixed effect, $\mathcal{R}_{2i}(t)$ is an unobserved random effect.

- the random effects $(\mathcal{R}_{1i}, \mathcal{R}_{2i}) \sim N(0, \Sigma)$.

Notations and Models - 3

- Specifically, for each $t \in \{t_{ij} \mid j = 1, \cdots, n_i\}$,

$$\begin{aligned} y_i(t) &= \mathcal{F}_{1i}(t) + \mathcal{R}_{1i}(t) + \epsilon_i(t) \\ &= x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t), \end{aligned} \tag{3}$$

where

- $x_i(t)$ is the design matrix for the fixed effect,
- β is the vector of the unknown fixed effect parameters,
- $z_i(t)$ is the design matrix for the random effect, and
- $b_i \sim N(0, \Sigma)$ is a vector of random effect parameters,
- $\epsilon_i(t) \sim N(0, \sigma^2)$ is the measurement error independent of b_i.

Model 1: Therneau and Grambsch - 1

- To quantify the effect of $\mathcal{F}_{1i}(t) + \mathcal{R}_{1i}(t)$ on the risk of an event, a common option is to use a relative risk model of the form (Therneau and Grambsch 2000):

$$h_i(t|\mathcal{M}_i(t), \omega_i) = \lim_{dt \to 0} Pr\{t \leq T_i^* < t + dt \mid T_i^* \geq t, \mathcal{M}_i(t), \omega_i\}/dt$$
$$= h_0(t) \exp\{\gamma^T \omega_i + \alpha(\mathcal{F}_{1i}(t) + \mathcal{R}_{1i}(t))\} \qquad (4)$$

where

  - $\mathcal{M}_i(t) = \{\mathcal{F}_{1i}(u) + \mathcal{R}_{1i}(u), 0 \leq u < t\}$ denotes the history of the true unobserved longitudinal process up to time t.
  - $h_0(t)$ denotes the baseline risk function at time t, and
  - $\omega_i = \mathcal{F}_{2i}(t)$ is a vector of baseline covariates with a corresponding vector of regression coefficients γ .
  - parameter α quantifies the effect of the underlying longitudinal outcome to the risk of an event.

Model 1: Therneau and Grambsch - 2

Remark:

- The baseline hazard function $h(\cdot)$ can be estimate at each time point t, namely $h(t)$.
- One can also estimate $h(\cdot)$ based on the cumulative information of hazard up to time t, namely $h(\mathcal{C}_i(t))$, where

$$\mathcal{C}_i(t) = \int_0^t \exp\{\gamma^T \omega_i + \alpha(\mathcal{F}_{1i}(s) + \mathcal{R}_{1i}(s))\} ds.$$

  Hence,

$$h_i(t|\mathcal{M}_i(t), \omega_i) = h_0(\mathcal{C}_i(t)) \exp\{\gamma^T \omega_i + \alpha(\mathcal{F}_{1i}(t) + \mathcal{R}_{1i}(t))\} \quad (5)$$

- If prior knowledge about $h(\cdot)$ is available, extra specification of h can increase the efficiency of estimation.

Model 2: Wulfsohn & Tsiatis and Henderson, et al. - 1

Wulfsohn & Tsiatis (1997) used the 2-stage method proposed by Laird & Ware (1982), and Henderson, et al. (2000) extended Wulfsohn & Tsiatis' approach and proposed the following approach.

With the latent bivariate Gaussian process $\mathcal{R}_i(t) = (\mathcal{R}_{1i}(t), \mathcal{R}_{2i}(t))$, such that

- the longitudinal process

$$Y_{ij} = x_{1i}(t)'\beta_1 + \mathcal{R}_{1i}(t_{ij}) + e_{ij} \tag{6}$$

with

$$\mathcal{R}_{1i}(t) = V_{1i}(t) + d_{1i}(t)'V_{2i}(t)$$

and

$$V_{1i} \sim N(0, \Sigma_{v1}), \quad V_{2i} \sim N(0, \Sigma_{v2}).$$

Model 2: Wulfsohn & Tsiatis and Henderson, et al. - 2

- the event process

$$\lambda_i(t) = H_i(t)\alpha_0 \exp\{x_{2i}(t)'\beta_2 + \mathcal{R}_{2i}(t)\}. \tag{7}$$

- In Henderson et al., they assume

$$V_{1i}(t) = U_1,\ V_{2i}(t) = U_2,\ d_{1i}(t) = t,$$

and

$$\mathcal{R}_{2i}(t) = \gamma_1 U_1 + \gamma_2 U_2 + \gamma_3(U_1 + U_2 t) + U_3$$

with U_3 being another error term.

Model 2: Wulfsohn & Tsiatis and Henderson, et al. - 3

Remarks:

- One of the differences between these two approaches is the latter approach allows extra random effects in the time-to-event process in addition to that from the longitudinal process to increase the flexibility of individual effect.
- The formal approach uses the MLE and the latter approach uses EM algorithm to estimate the parameters.
- Both methods only analyzed one longitudinal process.

Parameter estimation - 1

The joint likelihood function contribution from the i-th subject can be formulated as

$$\mathcal{L} = \int p(T_i; \delta_i | \mathcal{R}_i; \theta_t, \beta) \times \prod_j p(y_i(t_{ij}) | \mathcal{R}_i, \theta_y) \times p(\mathcal{R}_i, \theta_{\mathcal{R}}) d\mathcal{R}_i, \quad (8)$$

namely,

$$\begin{aligned} \mathcal{L} = & P(\text{event process}) \\ & \times P(\text{longitudinal process}) \\ & \times P(\text{latent random processes}) \end{aligned}$$

Parameter estimation - 2

Question: *How to incorporate more than one longitudinal process into the model?*

It is not as easy to simply extend Eq (8) to

$$\begin{aligned}\mathcal{L} = \int p(T_i; \delta_i | \mathcal{R}_i; \theta_t, \beta) &\times \prod_j p(y_i(t_{ij}) | \mathcal{R}_i, \theta_y) \\ &\times \prod_j p(z_i(t_{ij}) | \mathcal{R}_i, \theta_z) \\ &\times \cdots \\ &\times p(\mathcal{R}_i, \theta_{\mathcal{R}}) d\mathcal{R}_i ? \end{aligned} \tag{9}$$

Proposal: estimate the association of each longitudinal process with event process one-by-one using Eq (8) and combine them to obtain the joint effect.

Modeling Multivariate Longitudinal Data - 1

The general linear mixed model for repeated measures can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \tag{10}$$

where $\boldsymbol{u} \sim \boldsymbol{N}(0, \boldsymbol{G})$, $\boldsymbol{e} \sim \boldsymbol{N}(0, \boldsymbol{R})$, and $\boldsymbol{Cov}(\boldsymbol{u}, \boldsymbol{e}) = 0$.

- Eq (10) includes parameters in the fixed effects vector β and all unknowns in the covariance matrices $\boldsymbol{G}$ and $\boldsymbol{R}$.
- The number of parameters to be estimated increases almost exponentially when the number of sequences and the number of repeats increased that can cause substantial computational challenges in convergence.

Modeling Multivariate Longitudinal Data - 2

The GLMM can be expressed as:

$$\begin{bmatrix} y \\ u \end{bmatrix} \sim N\left(\begin{bmatrix} X\beta \\ 0 \end{bmatrix}, \begin{bmatrix} (Z\psi_\theta Z' + \sigma I) & \Sigma_{yu} \\ \Sigma_{yu} & \psi_\theta \end{bmatrix}\right) \tag{11}$$

with

$$\Sigma_{yu} = Z\psi_\theta, \quad E(u|y) = \psi_\theta Z'\Sigma_\theta^{-1}(y - X\beta)/\sigma^2. \tag{12}$$

By plugging the values of the estimated parameters, the random effects can be predicted as

$$\hat{u} = \hat{\psi}_\theta Z'\Sigma_\theta^{-1}(y - X\hat{\beta})/\hat{\sigma}^2 \quad \text{with} \quad \Sigma_{u|y} = \psi_\theta - \psi_\theta Z'\Sigma_\theta^{-1} Z\psi_\theta/\sigma^2. \tag{13}$$

The fitted values can be predicted as

$$\hat{y} = X\hat{\beta} + Z\hat{u}. \tag{14}$$

Modeling Multivariate Longitudinal Data - 3

Remark:

A few potential computational difficulties:

- The number of parameters need to be estimated can increase exponentially with increase of the number of series, the number of measures, and the complexities of the covariance matrix.
- With the data from these two studies, $N \approx 650$ and two data series, SAS could converge sometimes with only a few repeated measures (e.g., 2 or 3) and simple covariance matrix (e.g., compound symmetry). R also suffers the same challenges.

Modeling Multivariate Longitudinal Data - 4

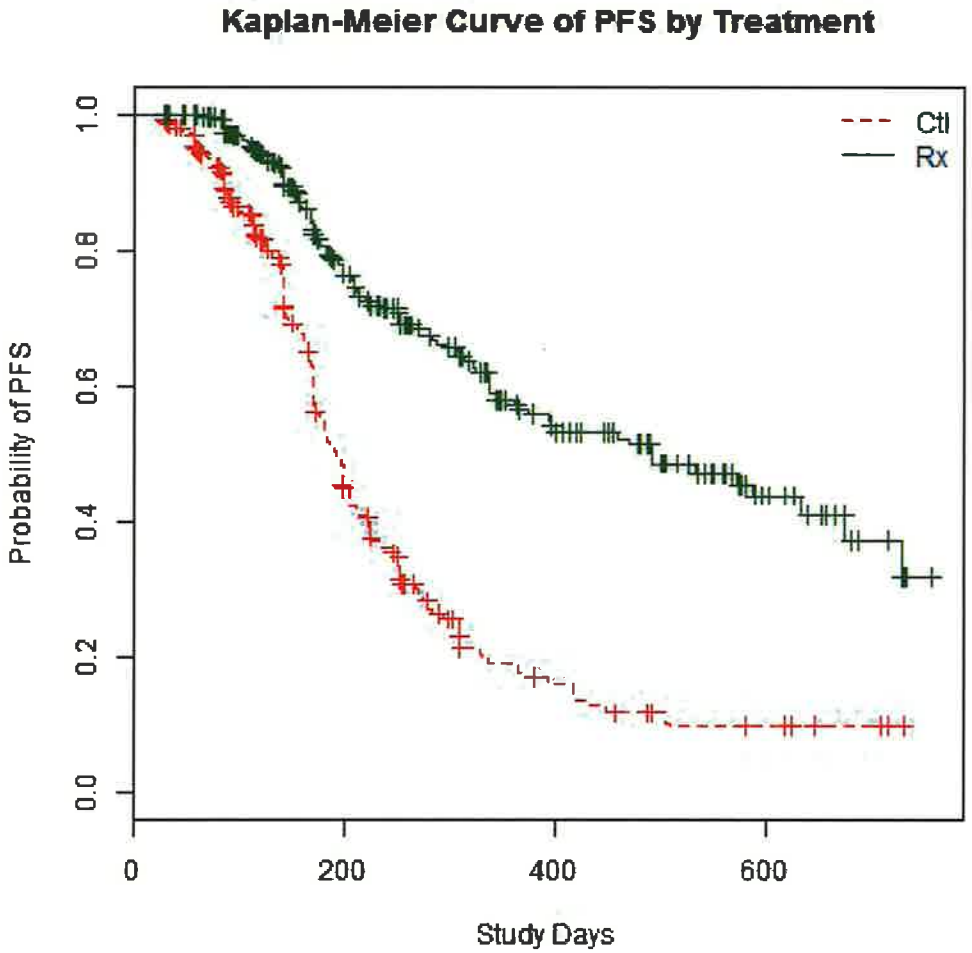
Remark 2:

Some approaches in modeling multiple longitudinal sequences:

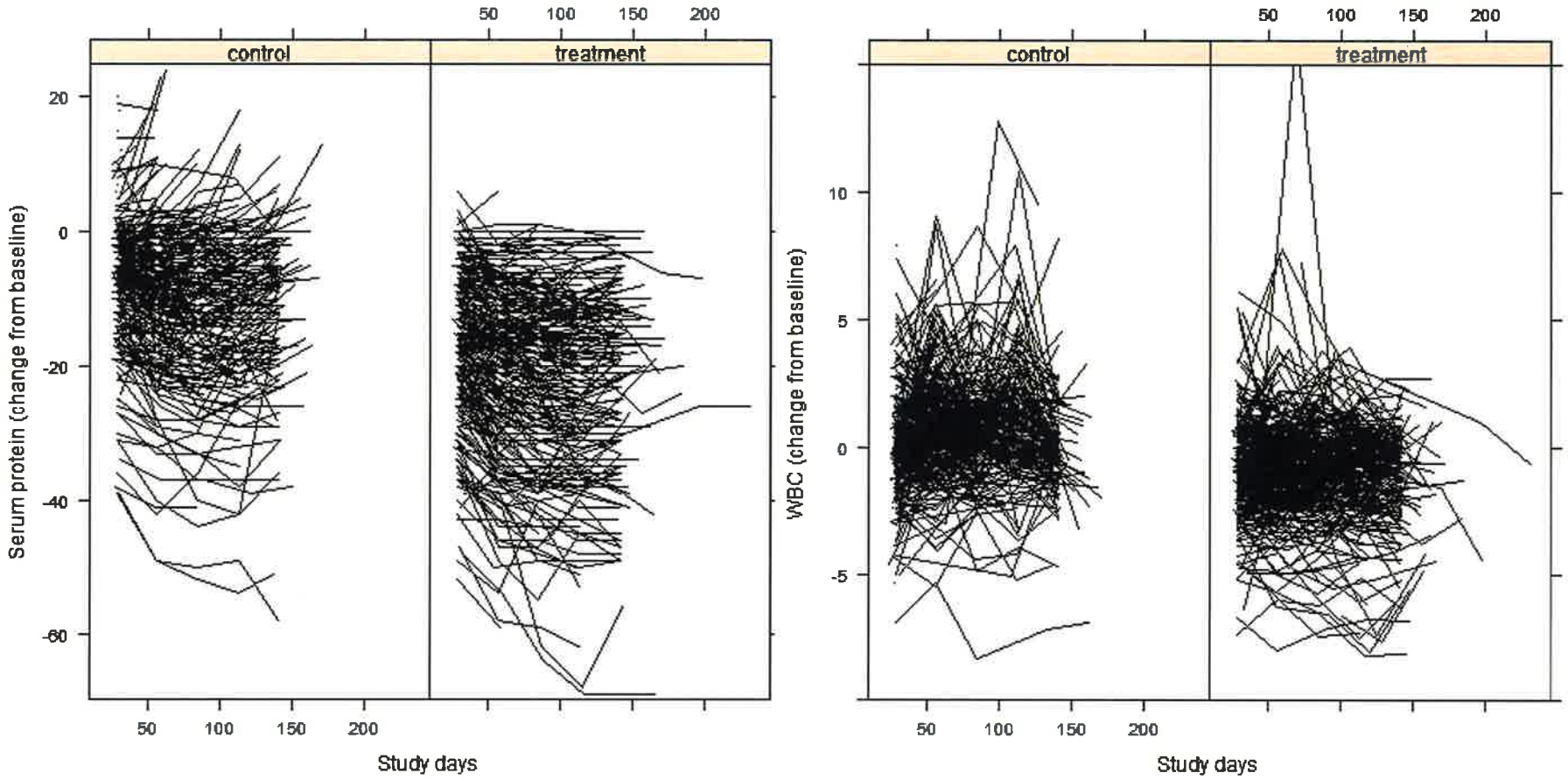
- Assume one sequence is the response and other sequences as covariates. (Note: this can run into time-dependent issues. Transitional Markov model is a better alternative.)
- Sum up the values of various sequences and treat it as one sequence. (Note: highly problematic.)
- Let $\boldsymbol{Y} = (\boldsymbol{Y_1}, \boldsymbol{Y_2}, \cdots, \boldsymbol{Y_r})$ be r different longitudinal repeated series, T and b be the event times and random effects one can assume conditional independence of different series and time given random effects:

$$\boldsymbol{f(Y_1, Y_2, \cdots, Y_r, T|b)} = \Big(\prod_{\boldsymbol{i=1}}^{\boldsymbol{r}} \boldsymbol{f(Y_i|b)}\Big) \cdot \boldsymbol{f(T|b)}.$$

PFS Curve by Treatment



Lab Test Values by Treatment



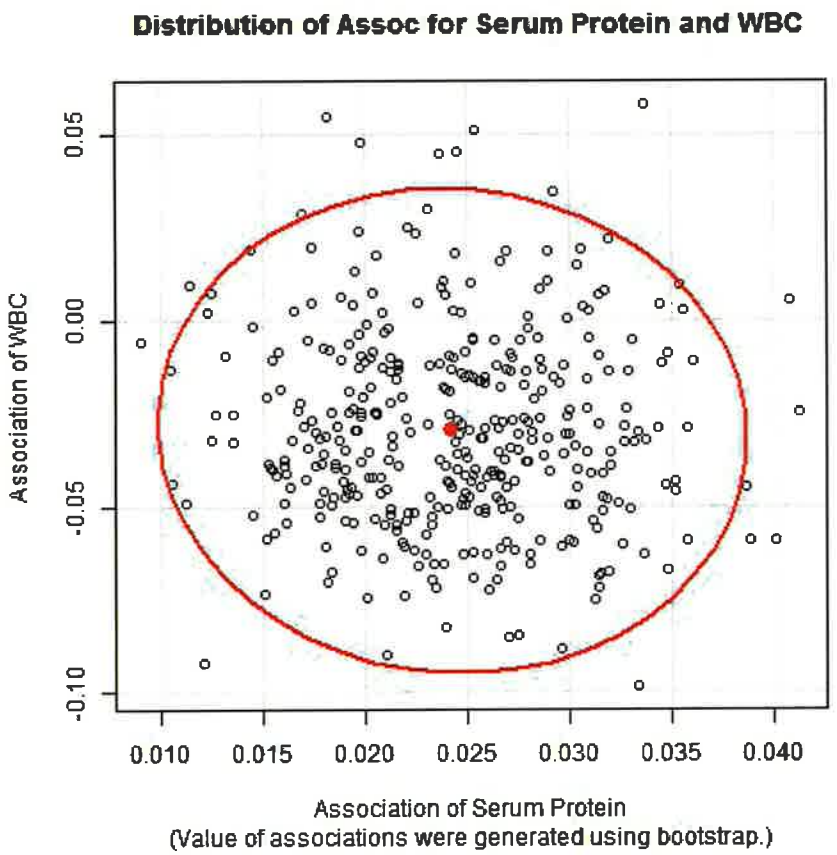
Longitudinal and Event Processes (Serum)

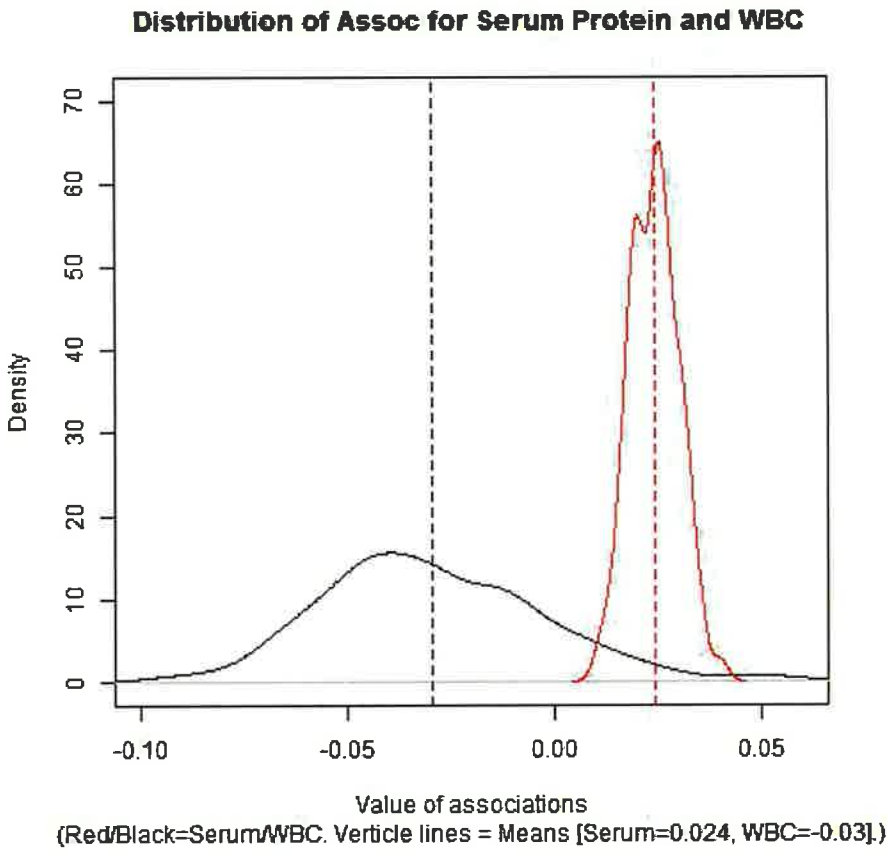
```
Joint Model Summary:

Longitudinal Process: Linear mixed-effects model
Event Process: Relative risk model with piecewise-constant
                baseline risk function

Longitudinal Process
               Value Std.Err  z-value p-value
(Intercept) -10.0584  0.3024 -33.2590 <0.0001
day           0.0038  0.0031   1.2074  0.2273
day:trtgrp   -0.0346  0.0031 -11.0981 <0.0001

Event Process
            Value Std.Err  z-value p-value
trtgrp    -0.8966  0.1263  -7.0986 <0.0001
Assoct     0.0236  0.0044   5.3997 <0.0001
```

Longitudinal and Event Processes (WBC)

```
Joint Model Summary:

Longitudinal Process: Linear mixed-effects model
Event Process: Relative risk model with piecewise-constant
                baseline risk function

Longitudinal Process
              Value Std.Err z-value p-value
(Intercept) -0.2173  0.0986 -2.2039  0.0275
day          0.0053  0.0012  4.4575 <0.0001
day:trtgrp  -0.0098  0.0012 -8.4021 <0.0001

Event Process
            Value Std.Err  z-value p-value
trtgrp    -1.2105  0.1336  -9.0624 <0.0001
Assoct    -0.0285  0.0233  -1.2227  0.2214
```

Associations of Serum Protein and WBC





Combining the Associations - 1

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be the associations of serum protein and WBC, respectively, from their models, the BLUE of the combined association can be estimated by:

$$\mathcal{A} = (1 - \beta)\mathcal{A}_1 + \beta\mathcal{A}_2$$

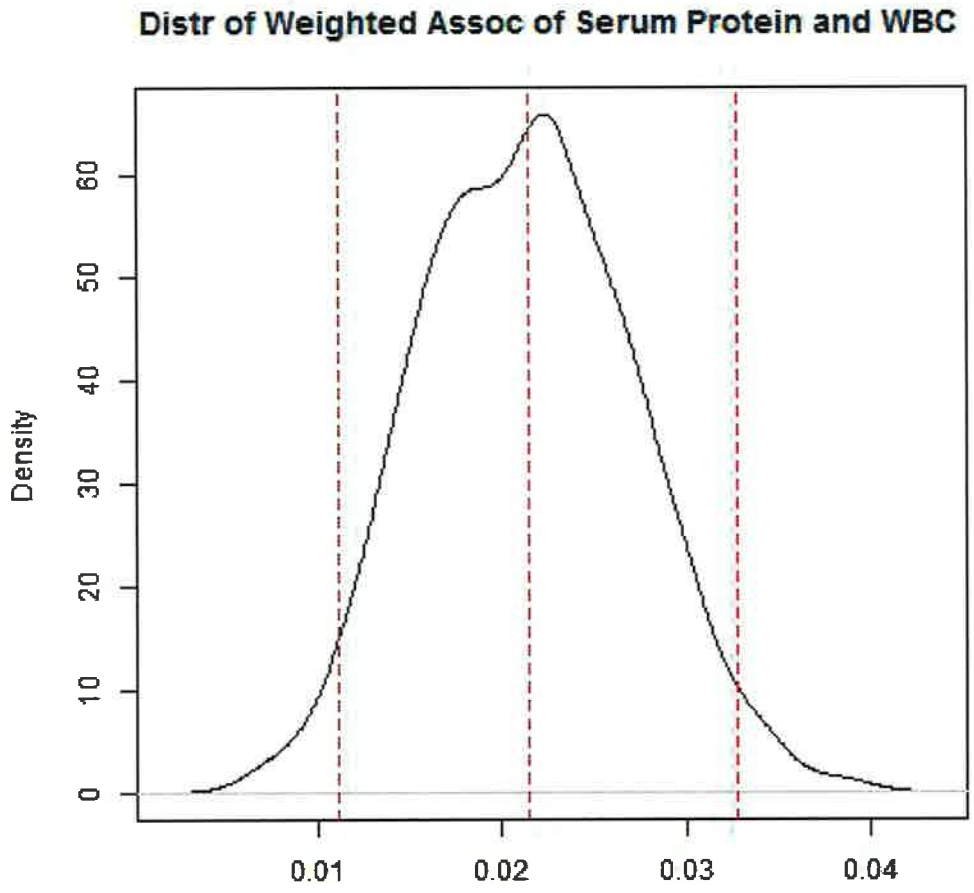
with the estimated weight given by

$$\beta = \frac{1 - \rho(\sigma_1/\sigma_2)}{1 - 2\rho(\sigma_1/\sigma_2) + (\sigma_1/\sigma_2)^2}.$$

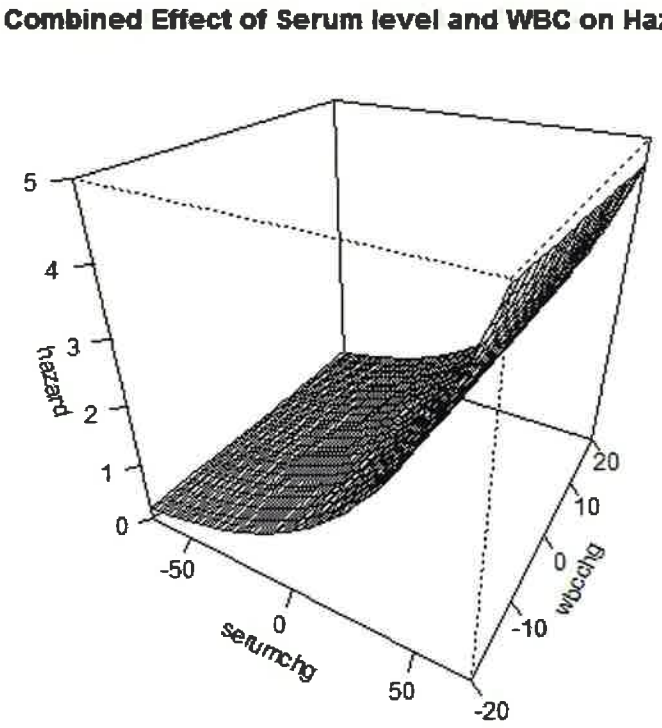
Combining the Associations - 2

```
Based on bootstrap sampling estimates:

Covariance matrix of associations from serum and WBC:
               [,1]          [,2]
[1,]  3.436944e-05 -5.768486e-06
[2,] -5.768486e-06  6.990862e-04

Correlation matrix of associations from serum and WBC:
             [,1]        [,2]
[1,]  1.00000000 -0.03721433
[2,] -0.03721433  1.00000000

The estimated weight for WBC:
[1] 0.05387694
```

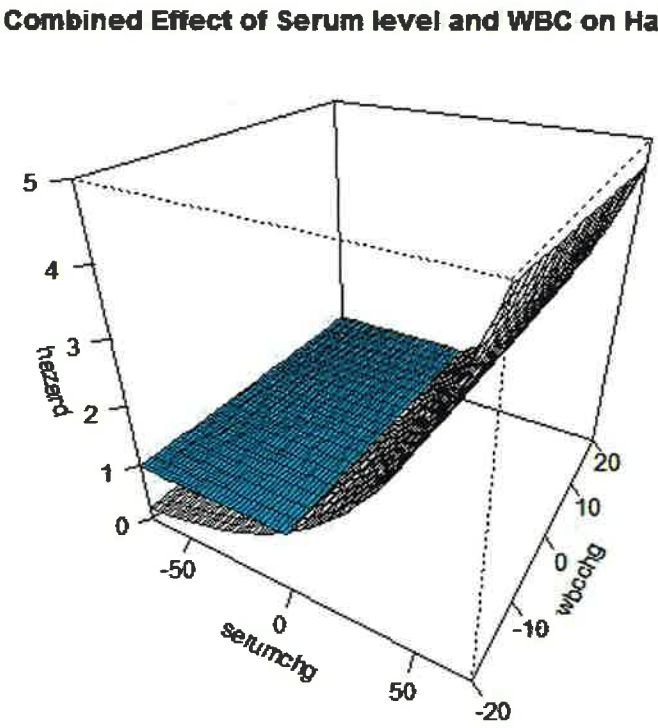
Weighted Associations



Combined Effect of Serum Protein and WBC on Hazard

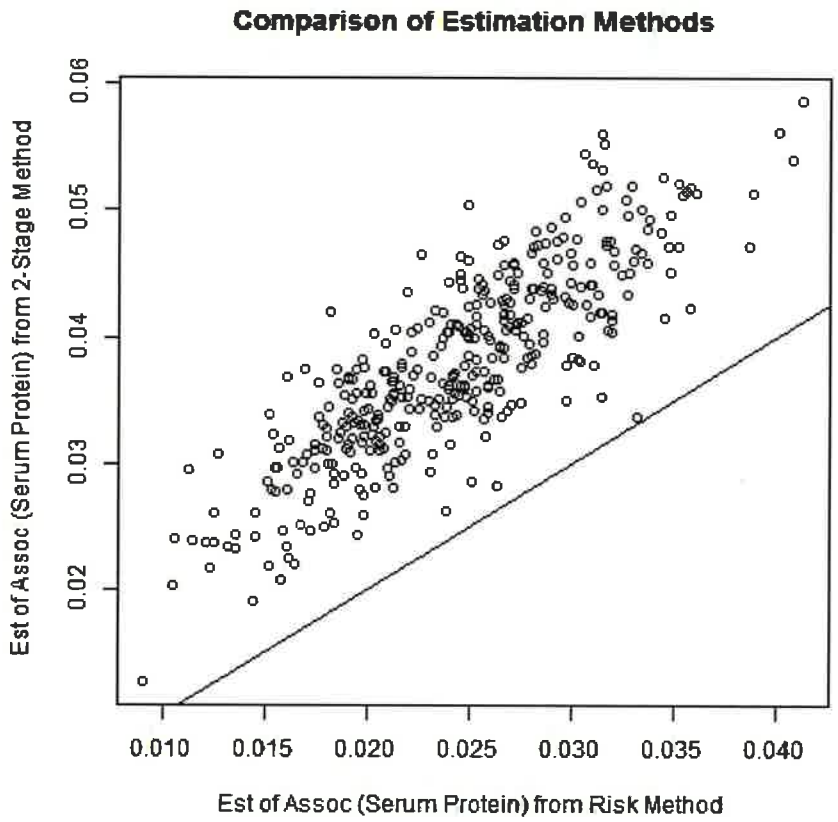


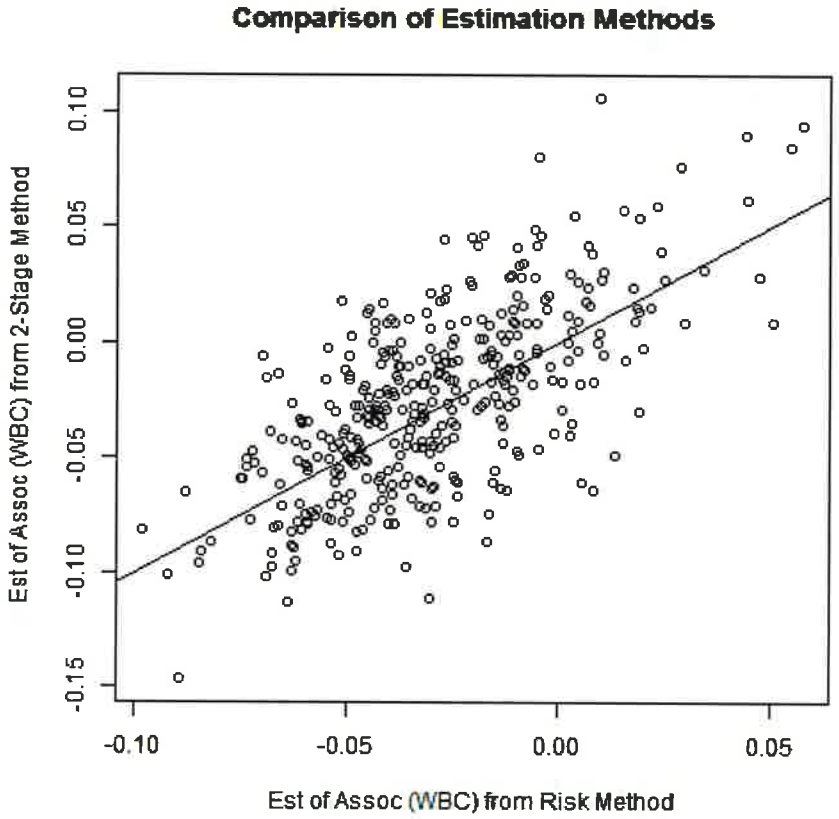
Hazard = exp(0.946*serumchg*0.0236+0.054*wbcchg*-0.0285)



Cyan plate: exp(0.946*serumchg*0.0236+0.054*wbcchg*-0.0285)=1

Comparison of Estimation Methods





Estimate Hazard Function of PFS

- Hazard estimators can then be obtained by smoothing the increments of the Nelson-Aalen estimator

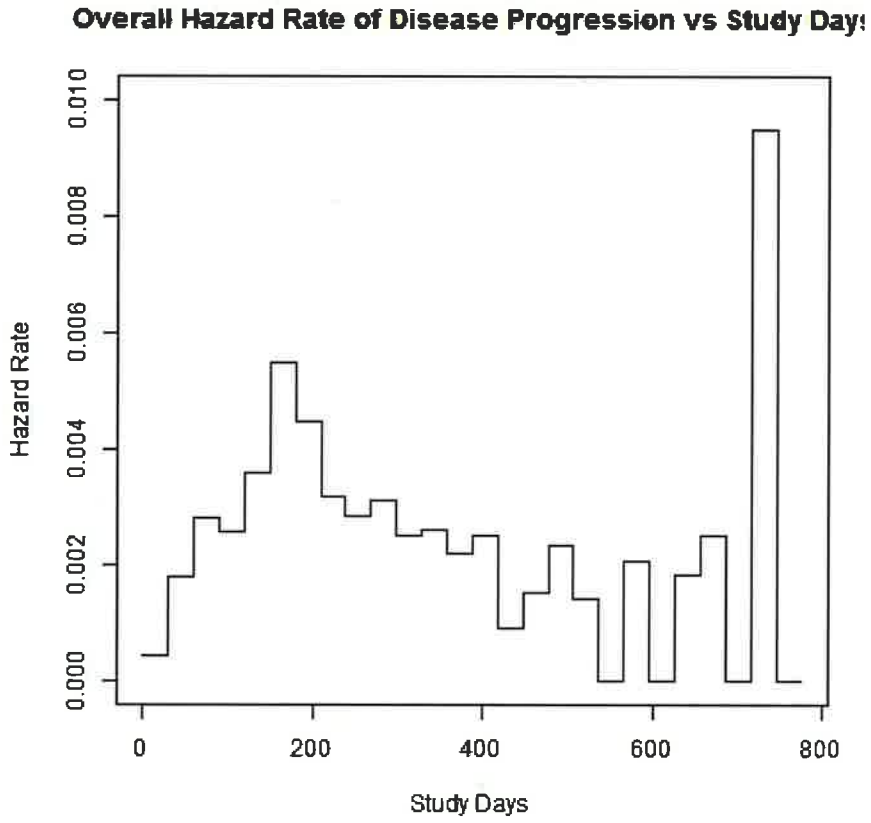
$$\Lambda_n(t) = \sum_{i=1}^{n} \delta_{[i]} I_{(T_{(i)} \leq t)} / (n - i + 1). \tag{15}$$

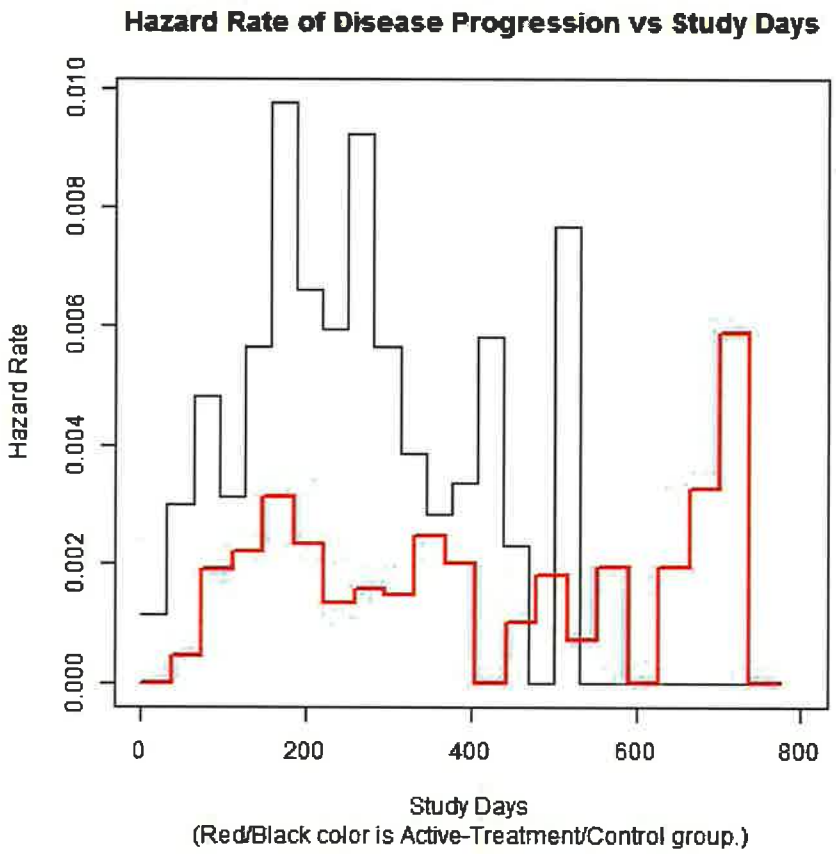
of the cumulative hazard function $\Lambda(t)$, where $\delta_{[i]}$ is the censoring indicator of $T_{(i)}$.

- Using the kernel method, one can estimate the kernel hazard rate function by

$$\Lambda_n(t) = \sum_{i=1}^{n} \{\delta_{[i]} / (n - i + 1)\} (1/h) K((t - T_{(i)})/h). \tag{16}$$

Hazard Function Estimate





Summary

- Clinical trials collect huge amount of data and lots of them were left unanalyzed.
- To better understand the overall treatment effects, one needs to analyze efficacy and safety data, pre and post-treatment data together, as they quite often interact with each other.
- Joint model of various data types collected in clinical studies has been well-established in both theory and practices. We propose a general approach to combine the effect of multiple longitudinal processes on the estimation of treatment effect.