On the Analysis of Longitudinal Clinical Lab Data with Latent Mixture Models

Jonathan S. Schildcrout, Cathy A. Jenkins, Jack Ostroff, Frank E. Harrell, Donald C. Trost

Department of Biostatistics, Vanderbilt University School of Medicine
Pfizer Global Research and Development

BASS XIV November 5, 2007

Outline

Pre-marketing analysis of safety data

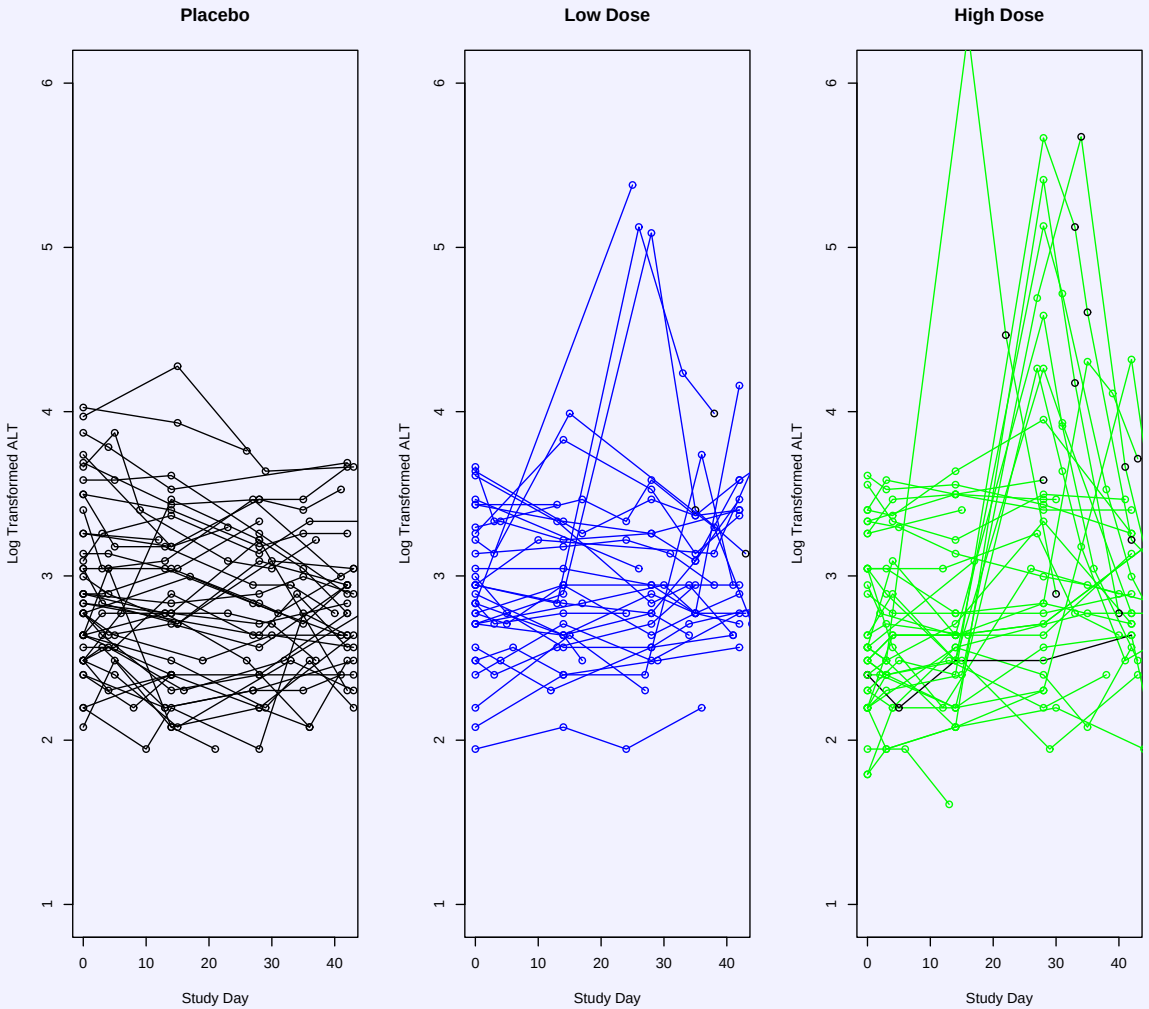
- Towards reporting to governing agencies and on label inserts, rates of adverse events (AE) over the course of followup during pre-marketing trials are used
- Often, the binary AEs are derived from continuous longitudinal clinical laboratory data
  - Liver function: alanine aminotransferase (ALT)) elevations are used to characeterize liver damage.
  - An hepatotoxic event may be defined by ALT concentrations crossing some threshold value (e.g. 3X ULN) at some time during follow-up.

- Considerations
  - Continuous data are dichotomized
    - Information loss
    - The dichotomization point (3X ULN) has no intrinsic meaning
    - Reference ranges: uncertainty / multiple labs and transformations
  - Longitudinal data are ignored.
    - Ignoring the added information by taking repeated measures on individuals.
    - Dynamic treatment effects cannot be captured (does ALT elevations at 28 days mean the same thing as ALT elevations at 7 days?)
  - Lack of power: combine multiple studies on the same product or on the same class of products
  - There is little effort to characterize the real biological impact of the products on typical subjects.

Motivating Example: Impact of drug X on liver function

	Placebo	Low Dose	High Dose
N	135	66	76
Female (%)	48	21	57
Ethnicity (%)			
White	81	85	84
Black	10	6	8
Other	9	9	8
Age (years)	50 (27, 65)	36 (24, 64)	53 (36, 66)
Weight (kilograms)	82 (60, 106)	83 (64, 105)	80 (56, 98)
Log-transformed, baseline ALT	2.8 (2.2, 3.5)	2.9 (2.4, 3.5)	2.6 (1.9, 3.4)

Motivating Example: Impact of drug X on liver function, *cont.*



A longitudinal data analysis approach

$$Y_i(t) = \mathbf{X}_i(t) \cdot \beta(t) + \epsilon_i(t)$$

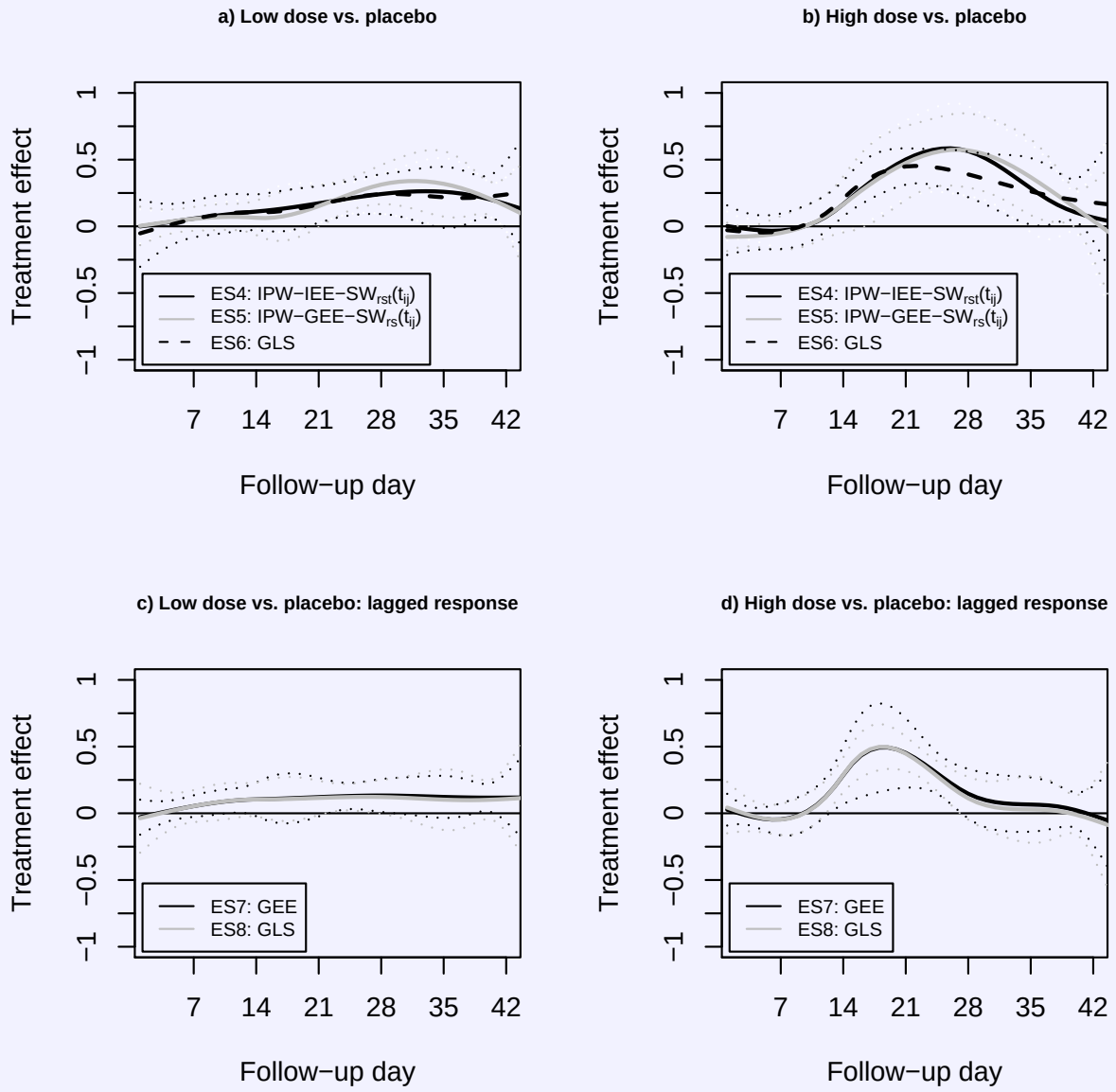
- $Y_i(t)$: response value for subject i at time t
- $\mathbf{X}_i(t)$: $p-$vector of covariates subject i at time t
- $\epsilon_i(t)$: mean-0 error process
  - Assuming a distributional form $\longrightarrow$ parametric estimation
  - Not assuming a distributional form $\longrightarrow$ semi-parametric estimation
- $\beta(t)$ parameter vector that is the target of inference
  - Flexibility in the functional form of these parameter is critical (parametrically, semi-parametrically, or non-parametrically)

Selection mechanisms

- Schildcrout JS et.al, *Statistics in Medicine (DOI: 10.1002/SIM.3071)*
- Potentially, a large number of selection mechanisms... many of which may be response history dependent

Type	Approach
Dropout	IPW-GEE
	ML
Visit rate	IPW-GEE
	ML
Tx discontinuation	IPTW-GEE
	Drop subsequent follow-up + GEE
	Drop subsequent follow-up + ML

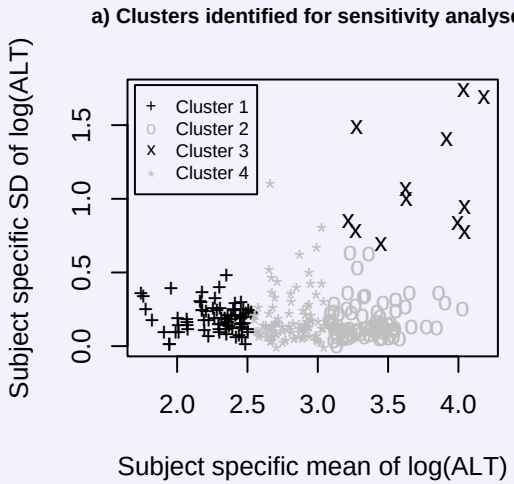
Dynamic treatment effects

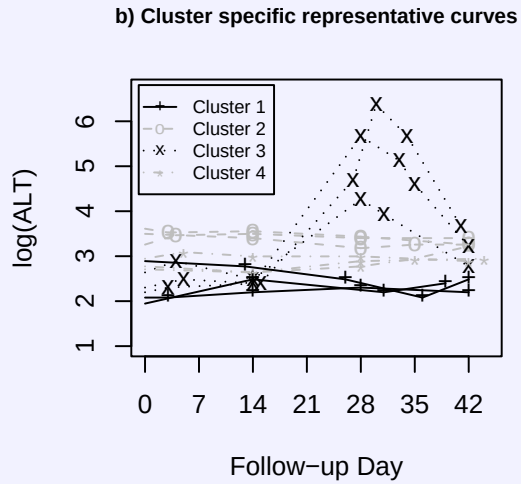


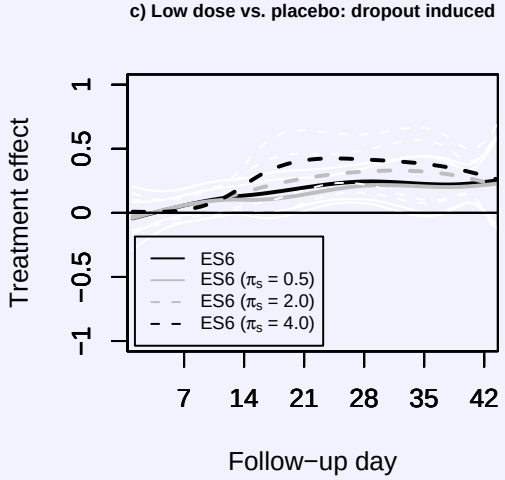
Selection mechanisms *(cont.)*

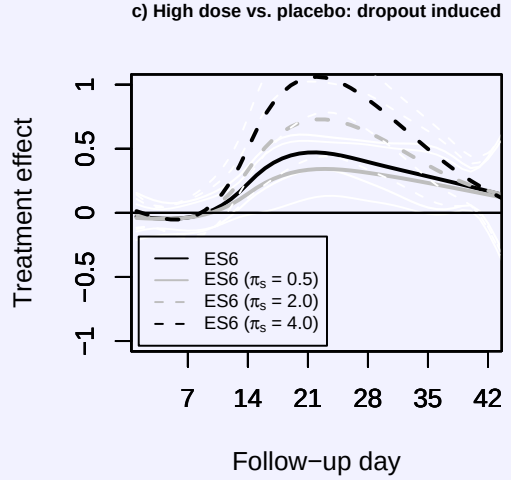
- Non-ignorable
  - Patients in clinical trials may be substantially different from those to whom we would like to generalize results
  - Exclusion criteria: counter-indicated medications, lifestyle choices
  - Post-marketing, medications are prescribed to unstudied patient populations
  - Evaluation strategies
    - Posit selection mechanisms and perform sensitivity analyses
    - Weighted estimating equations
    - Weighted resampling based approaches (e.g., tilted bootstrap)

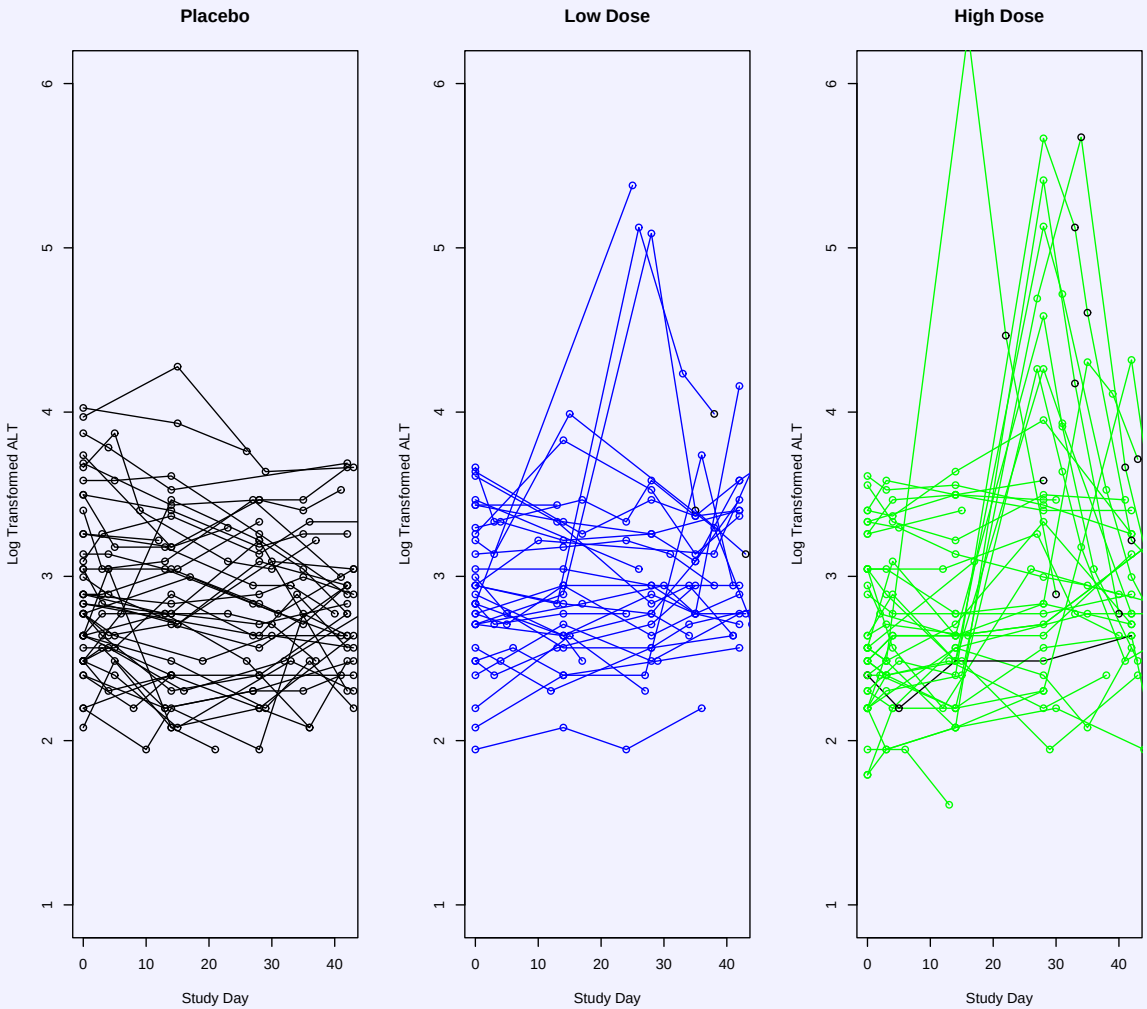
Sensitivity Analysis










Placebo
Low Dose
High Dose
Log Transformed ALT
Study Day
1
2
3
4
5
6
0
10
20
30
40

The underlying model assuming two levels of susceptibility

- Let R_i be an indicator that subject i is someone who is susceptible to the effects of treatment if given and r_i is the realized value if we knew it.

$$\begin{aligned} Y_i(t) =& (\beta_0 + \beta_1 tx_i + \beta_2 r_i + \beta_3 tx_i r_i) \\ &+ (\beta_4 + \beta_5 tx_i + \beta_6 r_i + \beta_7 tx_i r_i) \cdot f(t) + \epsilon_i(t) \\ \mu^r_{i,tx}(t) =& E(Y_i(t) \mid tx_i, r_i) \end{aligned}$$

- Treatment effect for non-susceptible subjects

$$\mu^0_{i,1}(t) - \mu^0_{i,0}(t) = \beta_1 + \beta_5 f(t)$$

- Treatment effect for susceptible subjects

$$\mu^1_{i,1}(t) - \mu^1_{i,0}(t) = (\beta_1 + \beta_3) + (\beta_5 + \beta_7) \cdot f(t)$$

The underlying model assuming two levels of susceptibility

- By randomization,

$$\pi_i \equiv p(R_i = 1) = p(R_i = 1 \mid tx_i = 1) = p(R_i = 1 \mid tx_i = 0)$$

- Marginal mean

$$\mu_{i,tx}(t) = \mu^1_{i,tx}(t)\pi_i + \mu^0_{i,tx}(t)(1 - \pi_i)$$

- Marginal treatment effect

$$\mu_{i,1}(t) - \mu_{i,0}(t) = (\mu^1_{i,1}(t) - \mu^1_{i,0}(t))\pi_1 + (\mu^0_{i,1}(t) - \mu^0_{i,0}(t))(1 - \pi_1)$$

- If it is reasonable to assume $\mu^0_{i,1}(t) = \mu^0_{i,0}(t)$,

$$\mu^1_{i,1}(t) - \mu^1_{i,0}(t) = \frac{\mu_{i,1}(t) - \mu_{i,0}(t)}{\pi_i}$$

- We may then be interested in π_i and $\mu^1_{i,1}(t) - \mu^1_{i,0}(t)$
- The challenge: R_i is unknown

Mixture models and (ML) estimation

- π_k = overall probability of membership in class k
- The mixture model density is given by,

$$g(y_i(t) \mid \mathbf{x}_i(t), \theta) = \sum_{k=1}^{K} \pi_k f(y_i(t) \mid \mathbf{x}_i(t), \theta_k)$$

where $\theta = \{\theta_1, \theta_2, \ldots, \theta_k, \pi_1, \pi_2, \ldots, \pi_k\}$

- With N subjects and n_i repeated measures for subject i, we may consider maximum likelihood estimation with the log-likelihood

$$l = \sum_{i=1}^{N} \sum_{t} log[g(y_i(t) \mid \mathbf{x}_i(t), \theta)],$$

Mixture models and (ML) estimation

- This is very similar to a missing data problem where r_i is missing for each subject, and $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{r}_i)$ would be the complete data
- The EM algorithm is generally used for the maximization.
- Expectation step
  - Given the parameter estimates at the $m-1^{th}$ iteration, $\theta^{(\mathbf{m-1})}$ calculate the (*posterior*) probabilities of group membership

$$p_{ik}^{(m)} = \frac{\pi_k^{(m-1)} f(\mathbf{y}_i \mid \mathbf{x}_i, \theta_k^{(m-1)})}{\sum_{w=1}^{K} \pi_w^{(m-1)} f(\mathbf{y}_i \mid \mathbf{x}_i, \theta_w^{(m-1)})}$$

  and calculate overall class membership probabilities (prior probability at iteration $m+1$),

$$\pi_k^{(m)} = \frac{1}{N} \sum_{i=1}^{N} p_{ik}^{(m)}$$

Mixture models and (ML) estimation

- Maximization step
  - Given p_{ik}^m for all i, maximize the weighted likelihood,

$$l = \sum_{i=1}^{N} \sum_{t_{ij}} p_{ik}^{(m)} f(y_i(t) \mid \mathbf{x}_i(t), \theta_k)$$

    w.r.t. θ_k for each of the k classes.
- Notice the number of estimated parameters estimated with large K.

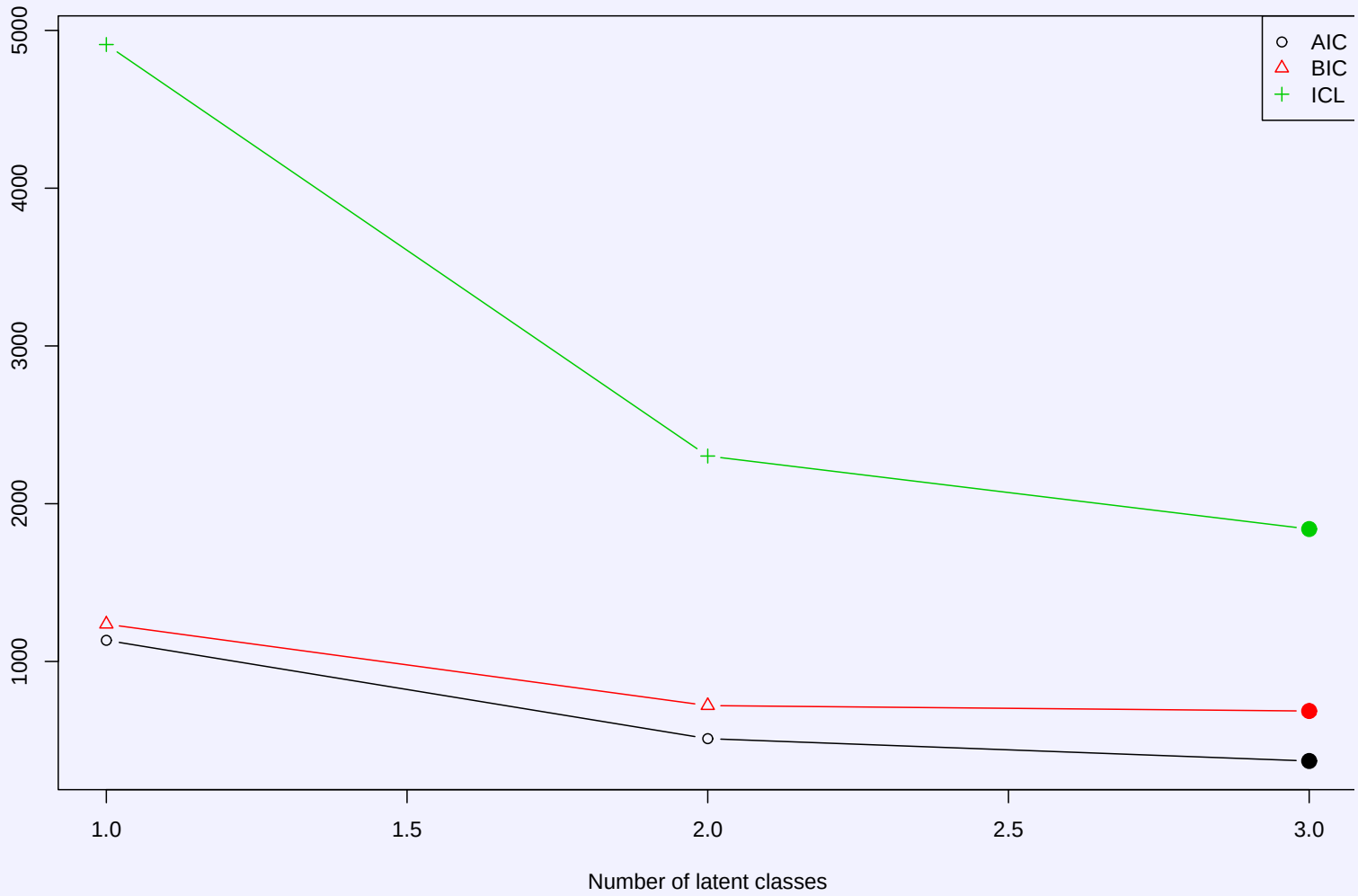
The model being used

- r_i is categorical (and not observed)

$$E(Y_i(t)) = (\beta_0 + \beta_1 tx_i + \beta_2 r_i + \beta_3 tx_i r_i) + (\beta_4 + \beta_5 tx_i + \beta_6 r_i + \beta_7 tx_i r_i) \cdot ns(t)$$

- ns : natural spline matrix with knots at days 14,28,35, and 42.
- Two factors to describe the three levels of Tx_i: $I(Tx_i = low)$ and $I(Tx_i = high)$
- The number of classes: use objective measures to determine an appropriate number of classes while acknowledging the limited sample size.

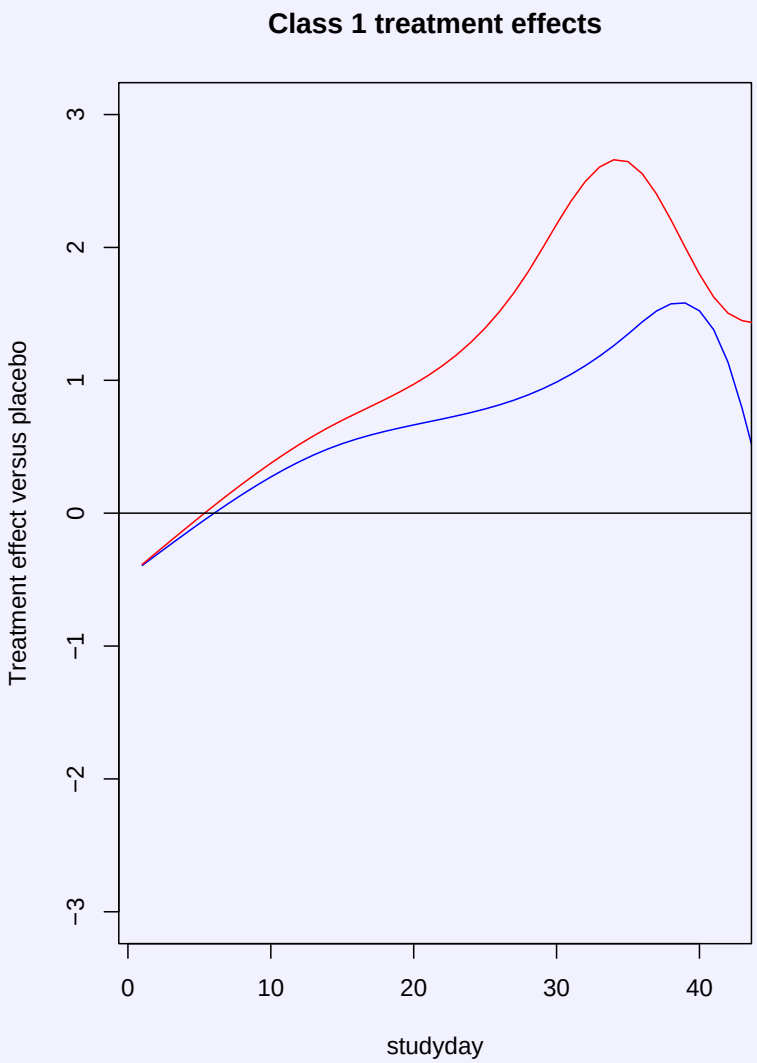
Number of Classes

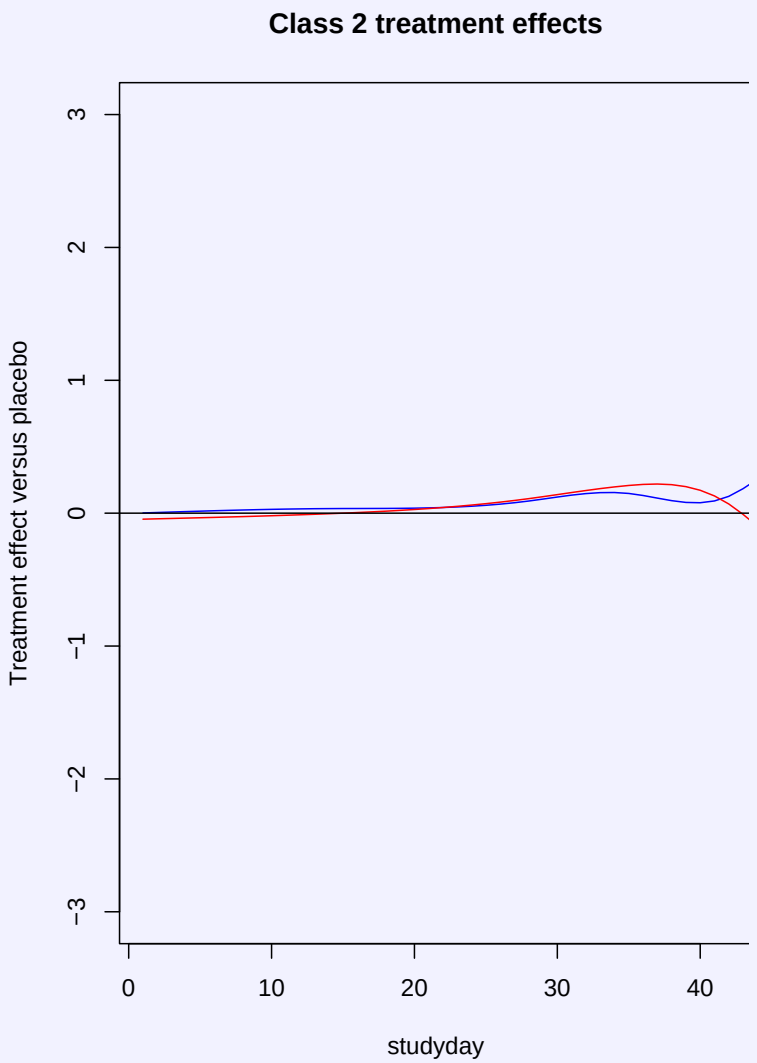


Two latent class model

- Prior class membership (π_k):
  1. 0.14
  2. 0.86

Two latent class model

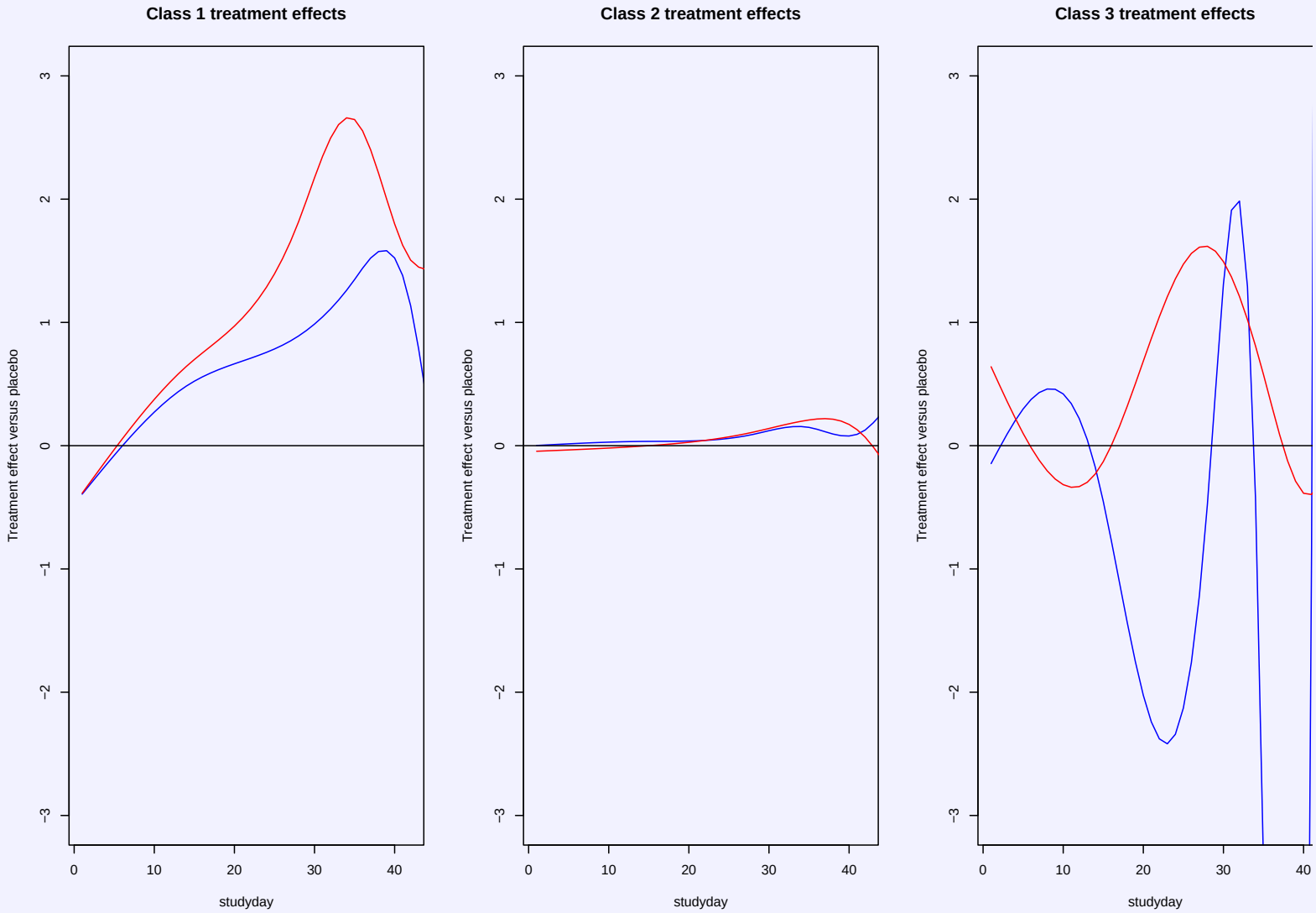




Three latent class model

- Prior class membership (π_k):
  1. 0.098
  2. 0.852
  3. 0.051

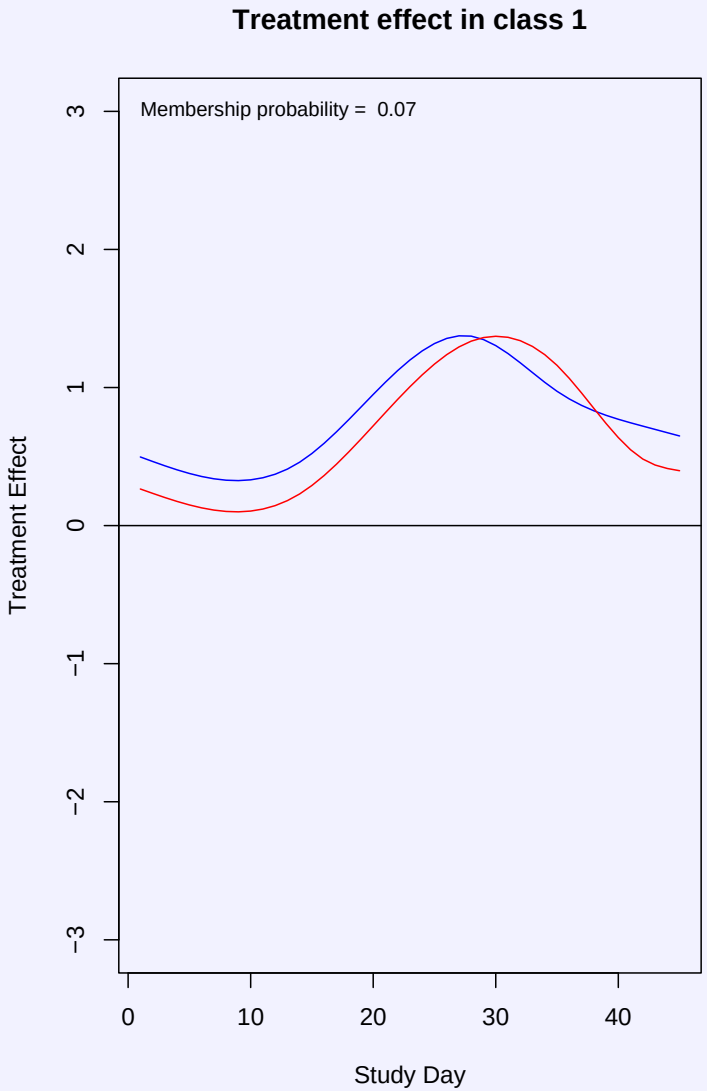
Three latent class model

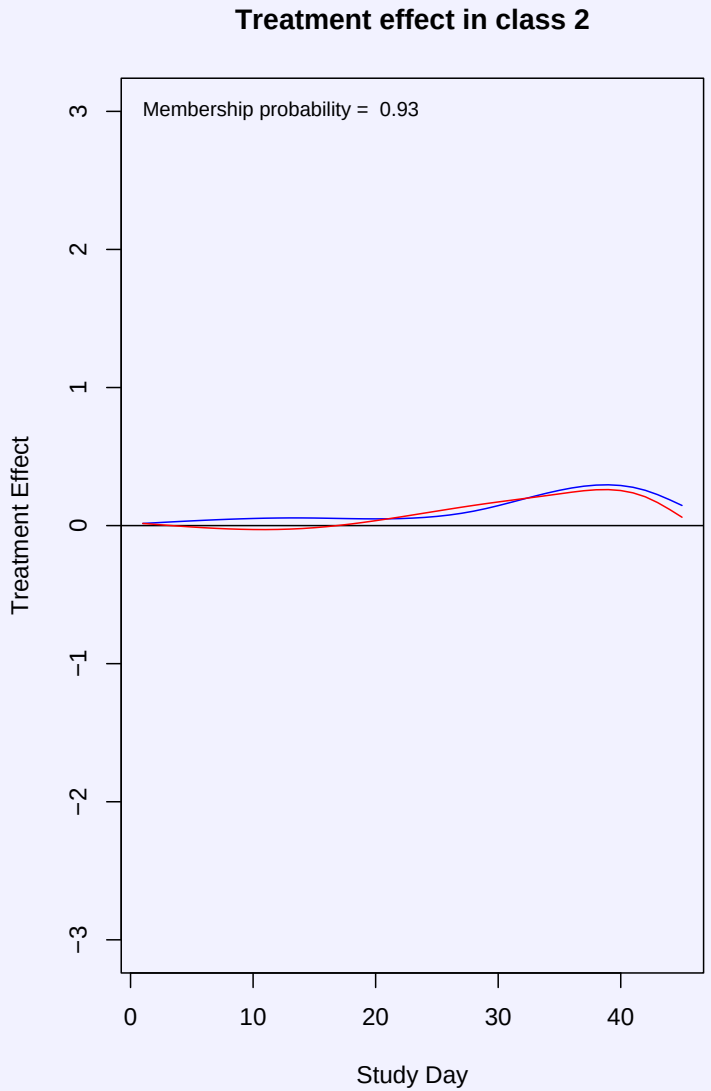


Bayesian implementation

$$
\begin{aligned}
Y_i(t) &\sim N(\mu_i(t), \tau) \\
\mu_i(t) =&(\beta_0 + \beta_1 tx_i + \beta_2 r_i + \beta_3 tx_i r_i) + \\
&(\beta_4 + \beta_5 tx_i + \beta_6 r_i + \beta_7 tx_i r_i) \cdot ns(t) \\
R_i \mid p &\sim bernoulli(p_i) \\
p_i &\sim beta(\alpha, \beta) \\
\sigma^2 &\sim U(0, 100) \\
\tau &= 1/\sigma^2 \\
\beta_j \mid \mu_j, \tau_j &\sim N(\mu_j, \tau_j) \\
\mu_j &\sim N(0, 100) \\
\sigma_j^2 &\sim U(0, 100) \\
\tau_j &= 1/\sigma_j^2
\end{aligned}
$$

Treatment Effects based on means of posteriors

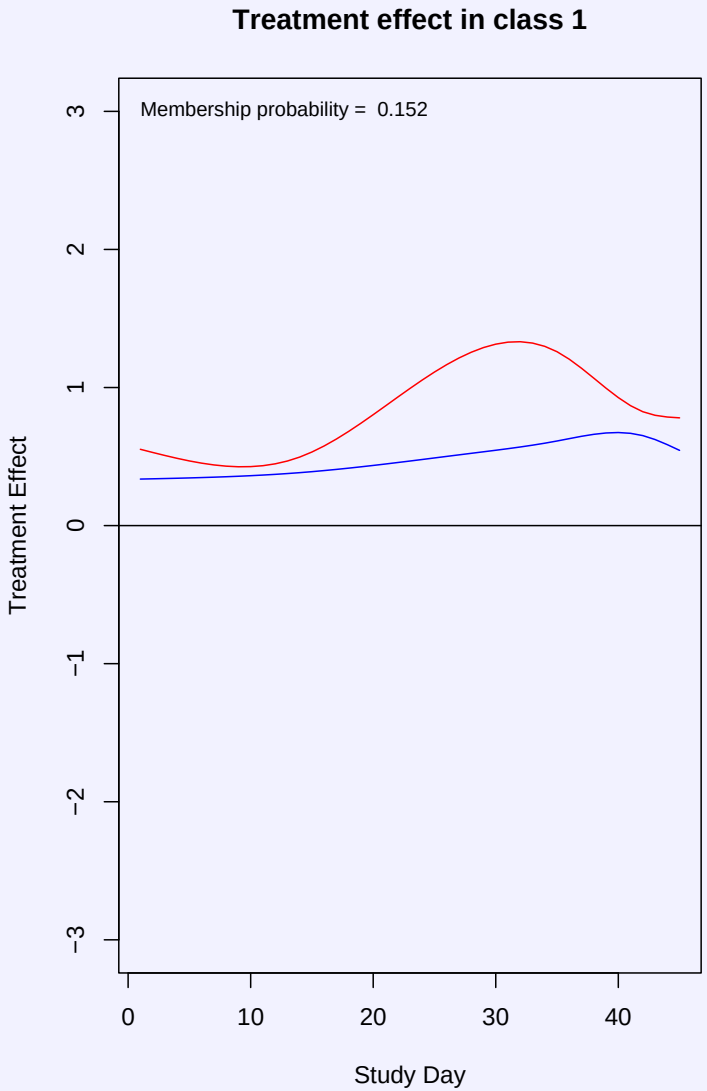


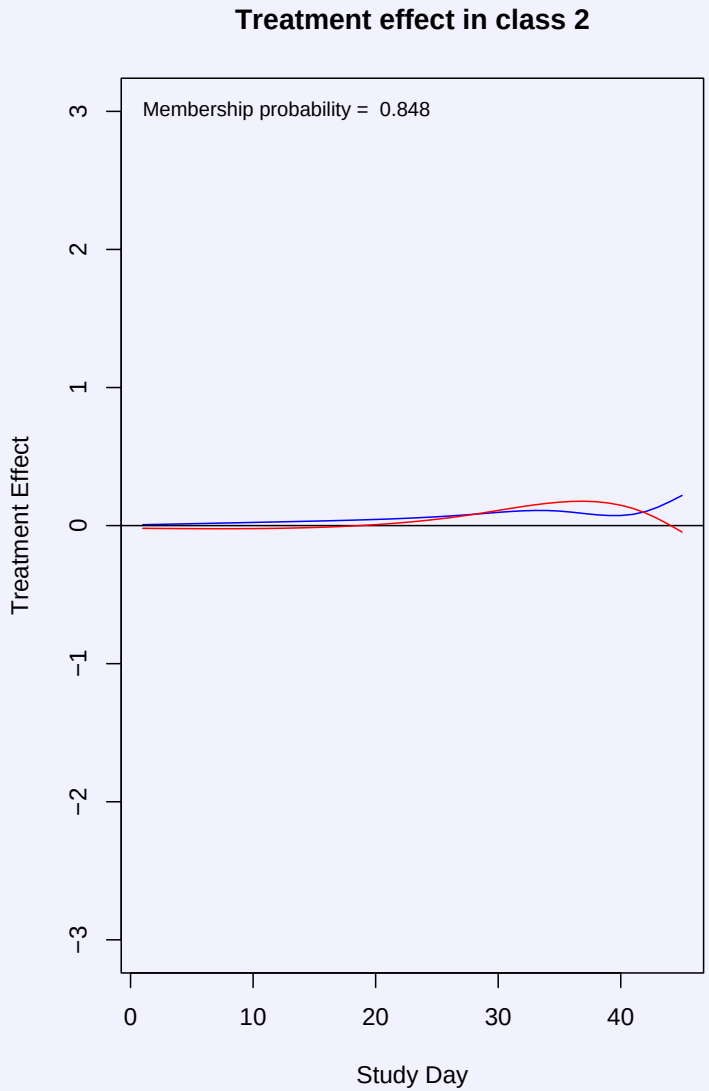


Bayesian implementation (unequal variance)

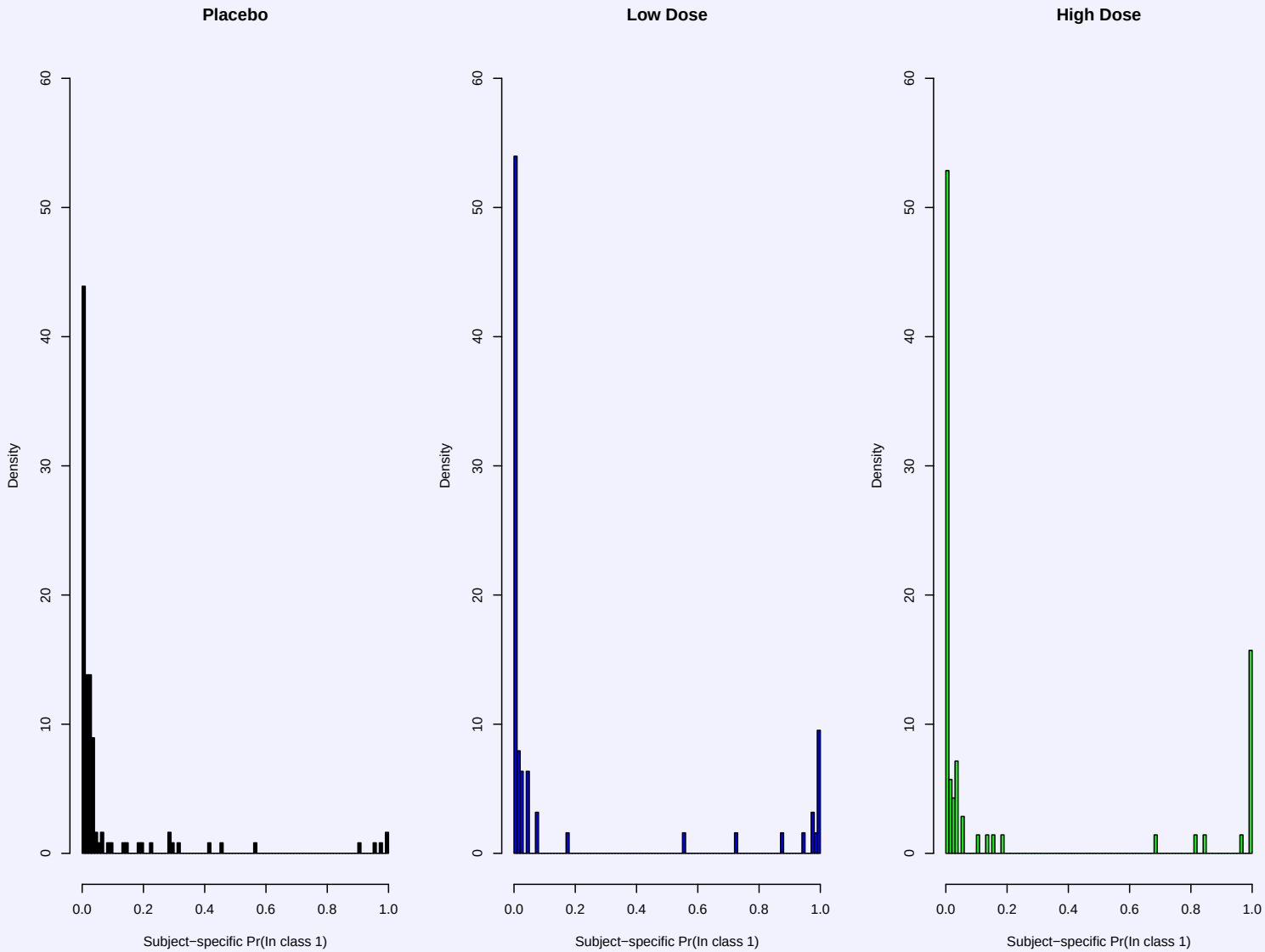
$$
\begin{aligned}
Y_{it} &\sim N(\mu_i(t), \tau_r) \\
\mu_i(t) =&(\beta_0 + \beta_1 tx_i + \beta_2 r_i + \beta_3 tx_i r_i) + \\
&(\beta_4 + \beta_5 tx_i + \beta_6 r_i + \beta_7 tx_i r_i) \cdot ns(t) \\
\sigma_r^2 \mid r &\sim U(0, 100) \\
\tau_r \mid r &= 1/\sigma_r^2 \\
R_i \mid p &\sim bernoulli(p_i) \\
p_i &\sim beta(1, 1) \\
\beta_j \mid \mu_j, \tau_j &\sim N(\mu_j, \tau_j) \\
\mu_j &\sim N(0, 100) \\
\sigma_j^2 &\sim U(0, 100) \\
\tau_j &= 1/\sigma_j^2
\end{aligned}
$$

Treatment Effects based on means of posteriors

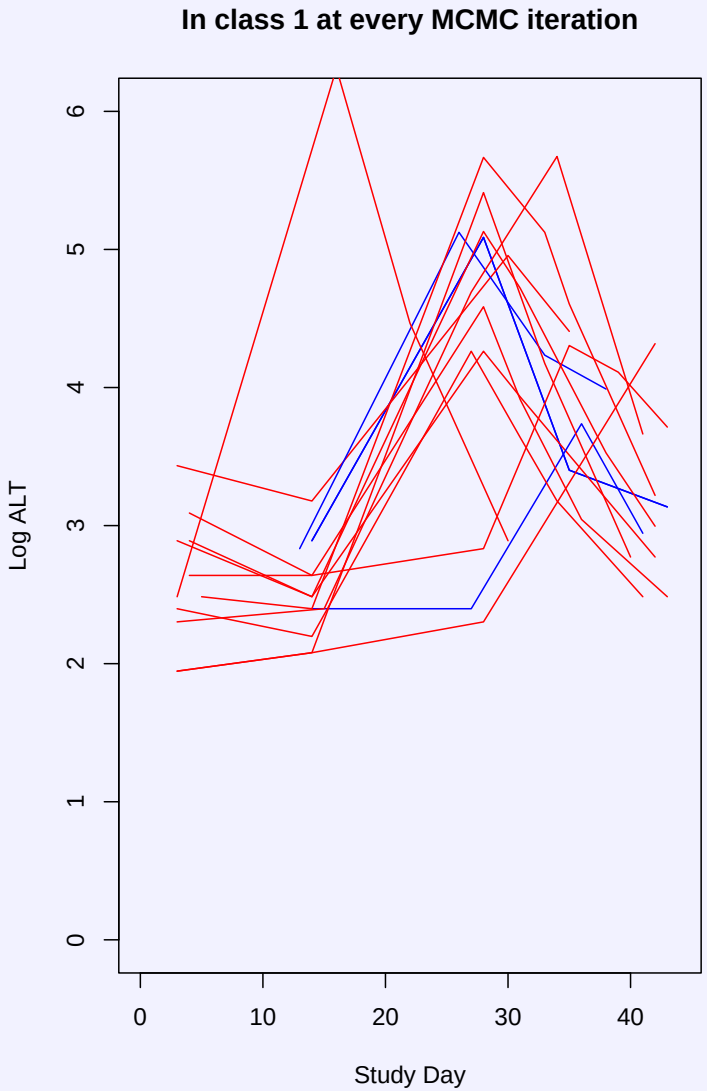


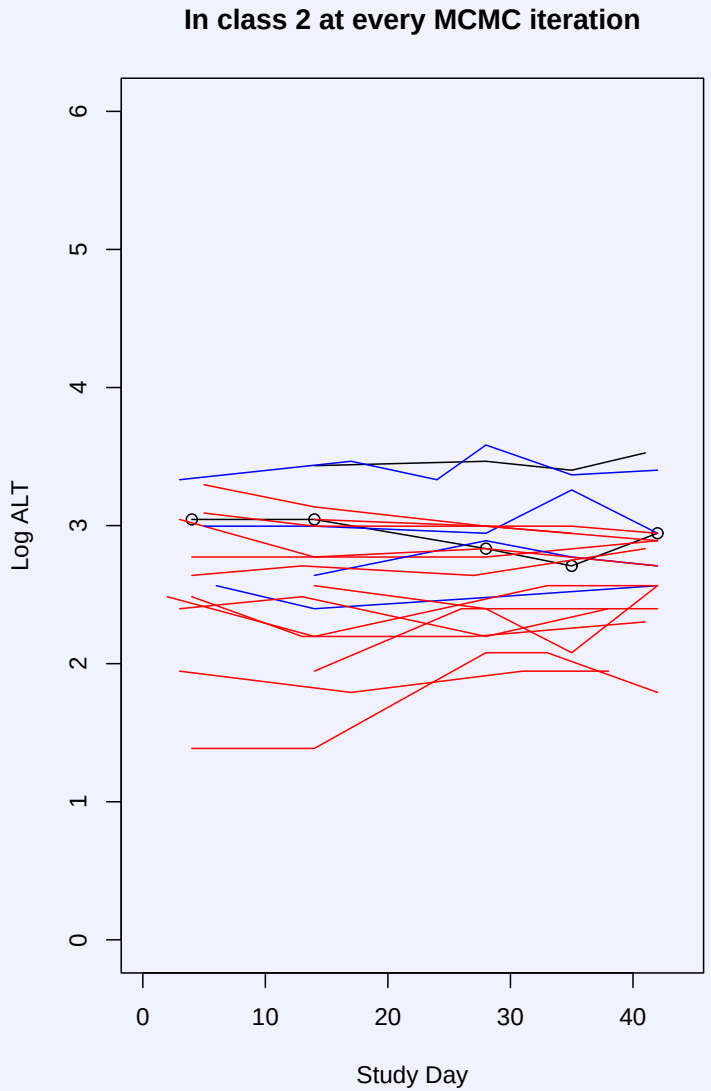


Classification Probabilities



Profiles for subjects classified with certainty





Identifiability

- Maximum Likelihood
  - Overfitting: Number of classes
    - Can be dealt with by requiring $\pi_k > 0$ and $\theta_k \neq \theta_{k'}$
  - Component relabelling
    - Compenent labels are interchangeable e.g., there are $K!$ label combinations that lead to a common maximized likelihood
    - Affects interpretation and not fitting of the EM algorithm... or we can impose restrictions
- Bayesian approach
  - Label reshuffling
    - No real way to discriminate between components of the mixture belongng to the same parametric family
    - During MCMC, labels can be permuted which can be very problematic (e.g., bimoddal posterior distributions)
    - Post-processing may be recommended

Summary

- Routinely collected laboratory data are not utilized effectively
- Longitudinal data analysis (acknowledging mechanisms of selection) improves our ability to characterize pharmaceutical safety
- Latent class regression seems to have the potential to improve this characterization in a very meaningful way.
- There are a number of challenges with it practically: primarily associated with the fact that class membership is not observed