A New High Dimensional Surrogacy Measure Based on Bayesian Variable Selection Approach

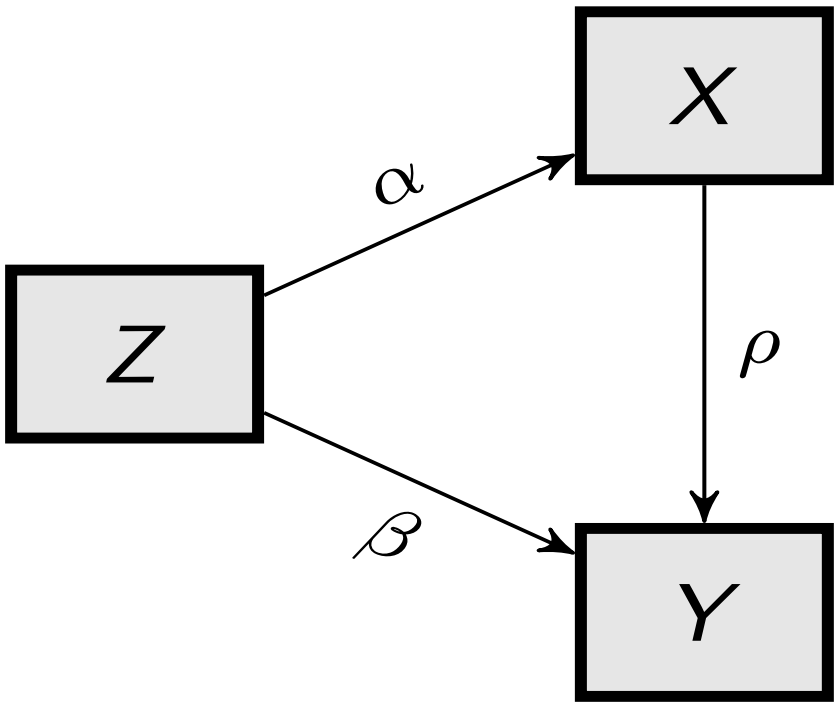
Olajumoke Evangelina Owokotomo[1] Ziv Shkedy[1] Adetayo Kasim[2]

[1]Interversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat) Center for Statistics (CenStat), Hasselt University, Diepenbeek, Belgium.

[2]Wolfson Research Institute, Durham University, Durham, United Kingdom.

40th Annual Conference of the International Society for Clinical Biostatistics on July, 17th 2019

Surrogacy measure and evaluation



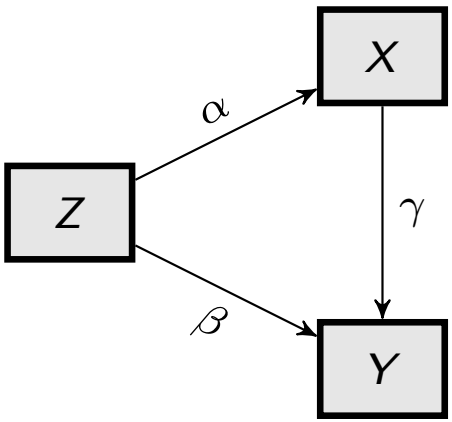
- Different evaluation approach in the literature
- Goal : A new measure of surrogacy

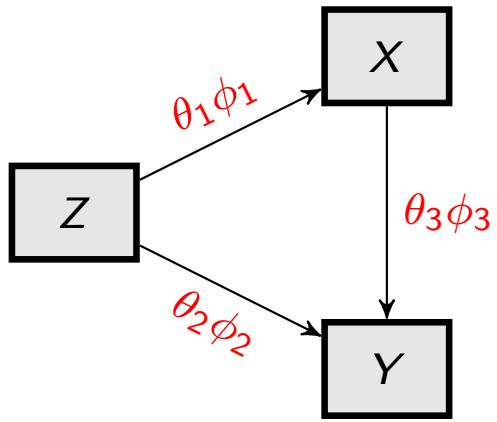
What is new?

- An easy and simple measure.
- Applied to any type of endpoints combination.
- Model uncertainty is taken into account.
- Based on a probability measure which gives the importance of an endpoint as a biomarker.
- It is related to other measures for surrogacy
  Buyse and Molenberghs, 1998, Alonso and Molenberghs, 2001.

Biomarker/Surrogacy setting

- A Bayesian Variable Selection approach.





$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} \mu_X + \alpha Z \\ \mu_Y + \beta Z + \gamma X \end{matrix} , \quad \Sigma \right) \qquad \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} \mu_X + \theta_1\phi_1 Z \\ \mu_Y + \theta_2\phi_2 Z + \theta_3\phi_3 X \end{matrix} , \quad \Sigma \right)$$

$$\alpha = \phi_1{}^*\theta_1, \qquad \beta = \phi_2{}^*\theta_2, \qquad \gamma = \phi_3{}^*\theta_3.$$

Bayesian Variable Selection (prior specification)

$$\phi_i \sim B(\pi_i) \quad \text{and} \quad \pi_i \sim U(0,1)$$

$$\phi_i = \begin{cases} 1, & \theta_i \text{ is included in the model,} \\ 0, & \theta_i \text{ is not included in the model.} \end{cases}$$

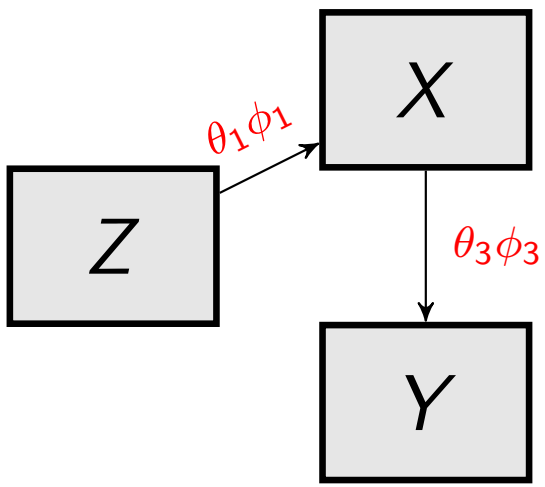
$$\theta_1, \theta_2, \theta_3 \sim \mathcal{N}(0, 0.00001)$$
$$\Sigma^{-1} \sim dwish(W, df)$$

$$W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad df = 2$$

Configuration of the inclusion parameters

$$\phi = \{\phi_1, \phi_2, \phi_3\}$$

- The configuration of ϕ define uniquely all possible models;
- Example: If $\phi = \{1, 0, 1\}$ then;

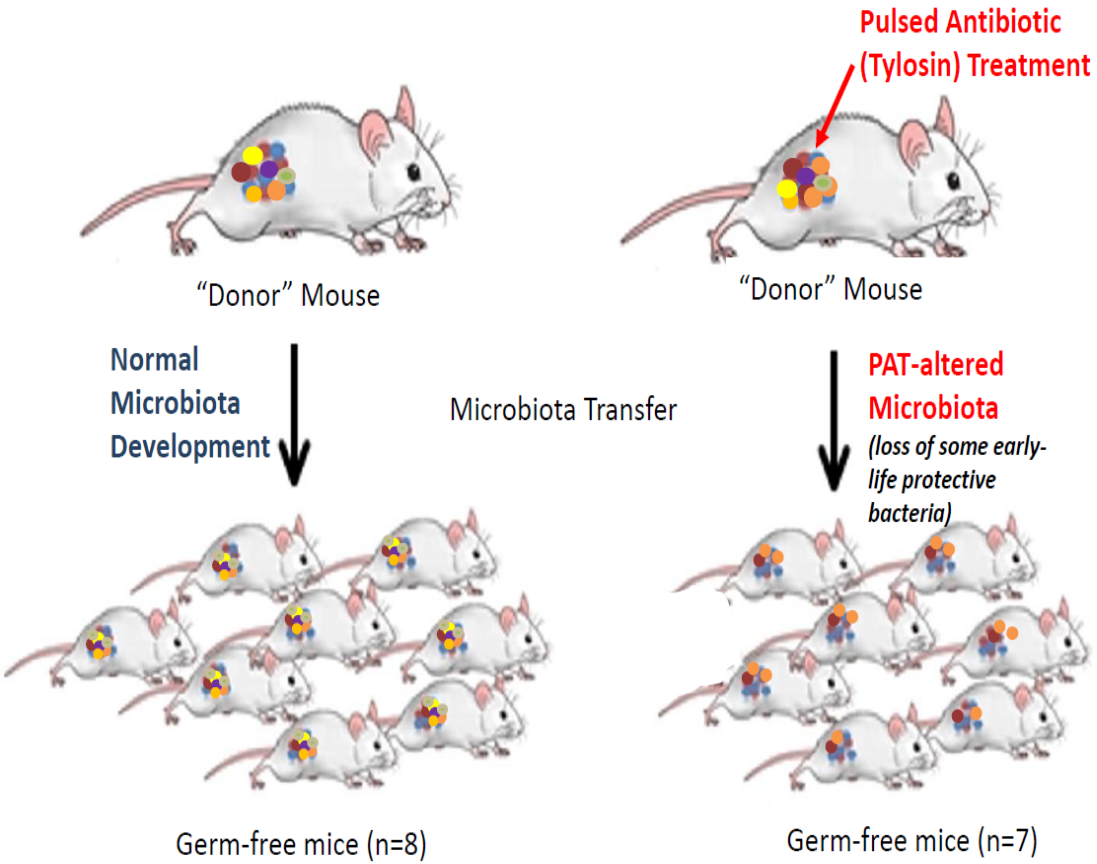


- Z influences Y indirectly via X.

Inclusion probability

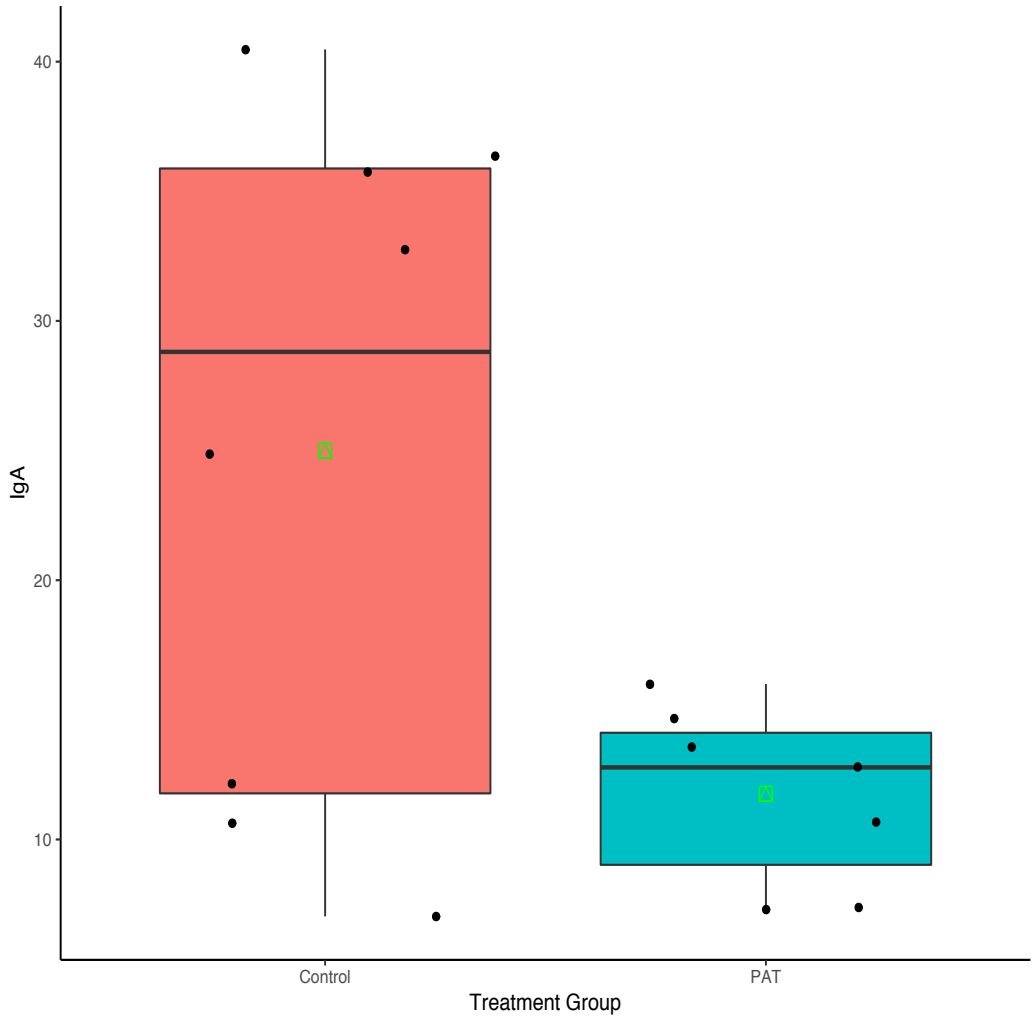
- Inclusion probability as measure of surrogacy.
- $\pi_i = P(\phi_i = 1 | \text{data}, \text{model paramters})$.
- Example:
  $P(\phi_3 = 1 | \text{data}, \text{model paramters}) =$
  $P(\text{X is included in the model})$.

The transPAT study

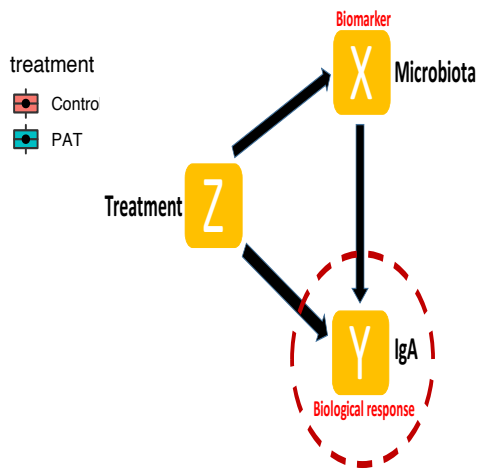


- Measurements:
  1. Microbiome Data
  2. Immunological Data
- Research question: Is the PAT altered Microbiota sufficient to alter Intestinal Immunity?

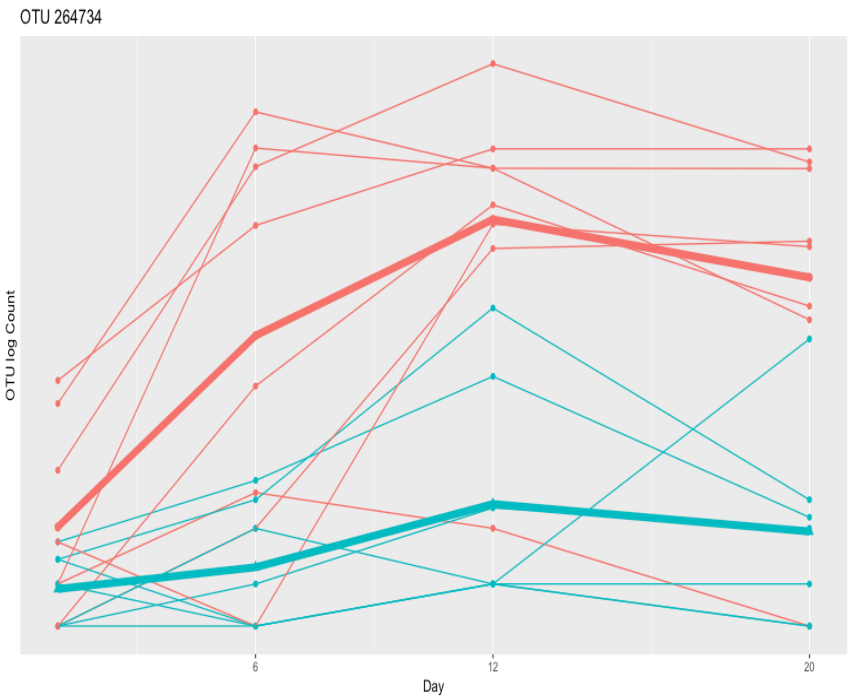
Immunological data

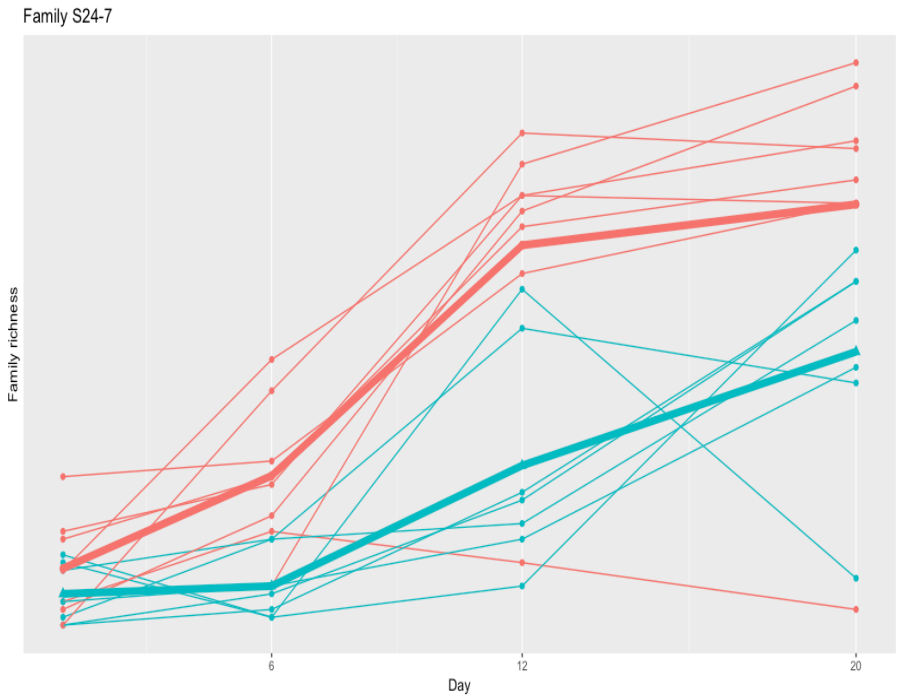


- Immunity: Measured by IgA level
- Analysis: Truncated at Day20



Microbiome data(1)





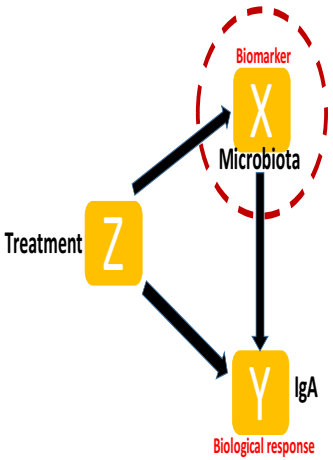
OTU Level (OTU count)

FAMILY Level (Richness)

Microbiome data(2)

$$
\underbrace{\left(\begin{array}{cccc|cccc}
x_{1,1} & x_{1,2} & \cdots & x_{1,8} & x_{1,9} & x_{1,10} & \cdots & x_{1,15} \\
x_{2,1} & x_{2,2} & \cdots & x_{2,8} & x_{2,9} & x_{2,10} & \cdots & x_{2,15} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{j,1} & x_{j,2} & \cdots & x_{j,8} & x_{j,9} & x_{j,10} & \cdots & x_{j,15} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
x_{355,1} & x_{355,2} & \cdots & x_{355,8} & x_{355,9} & x_{355,10} & \cdots & x_{355,10}
\end{array}\right)}_{\text{Control Group} \qquad \text{PAT altered Group}}
\quad m = 355 \text{ OTUs}
$$

- Similar in structure to other omics data: gene expression data, metabolic data...



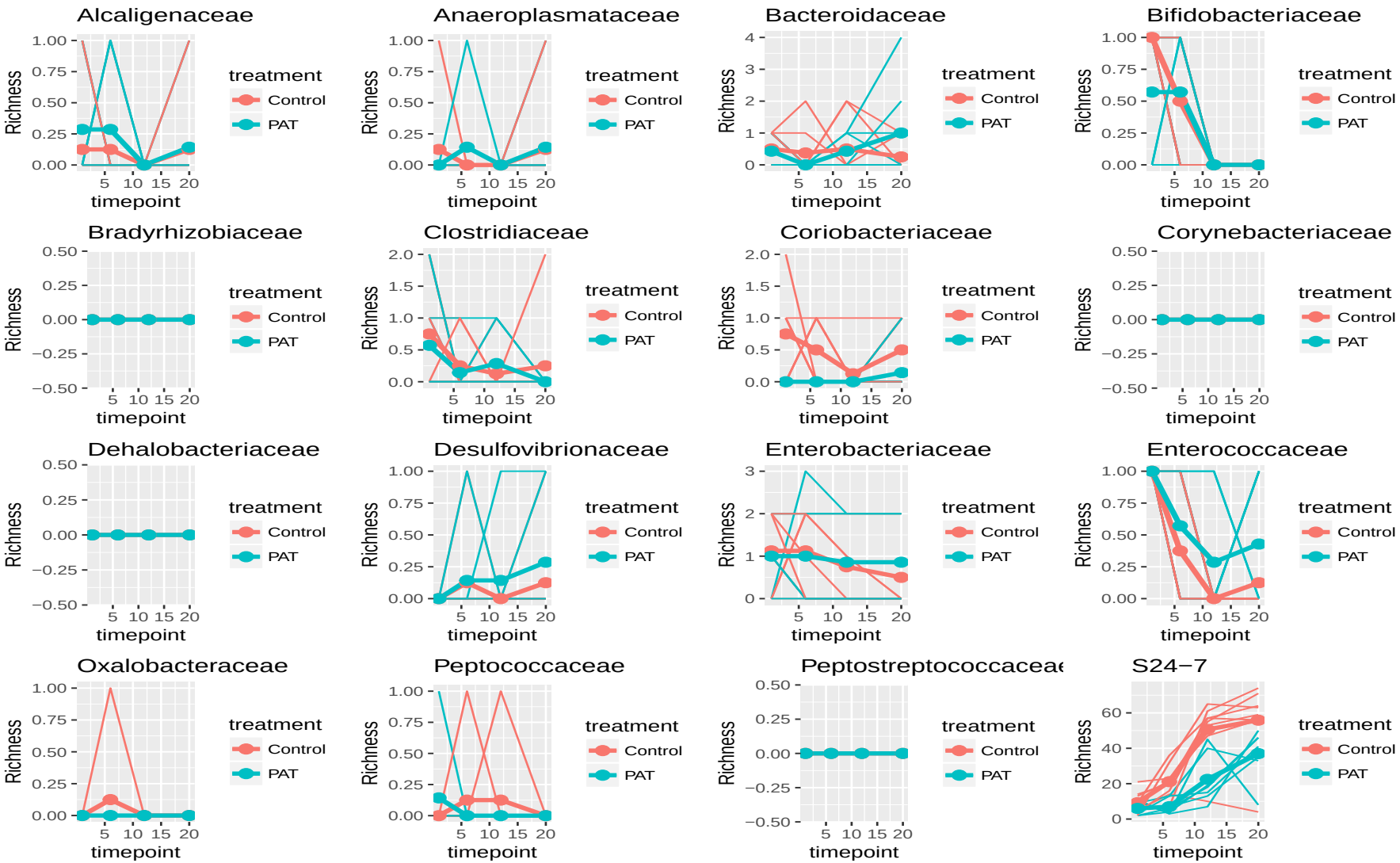
Data structure

- Repeated measurements at 4 time points
- 355 OTU's = 30 families
- A subject : Mouse
- Observation unit = $\{Z_i, Y_i, X_{ij}\}$

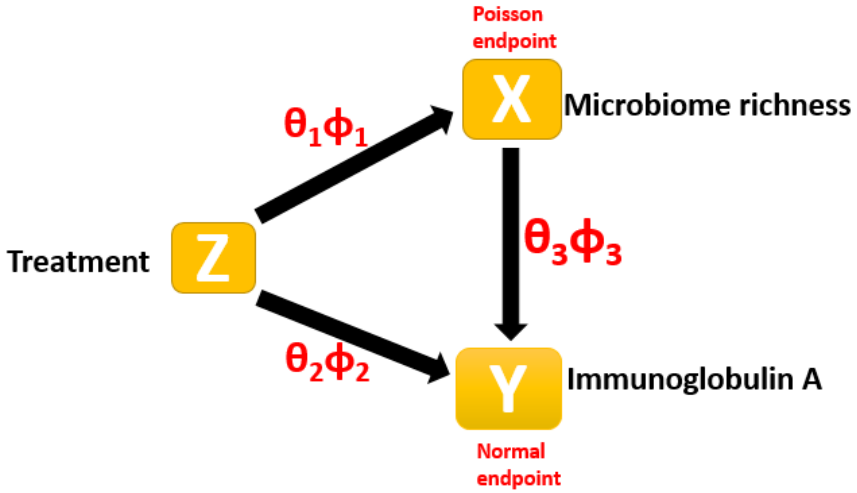
$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{15} \end{bmatrix}, X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,15} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,15} \\ \vdots & \vdots & \vdots & \vdots \\ X_{30,1} & X_{30,2} & \cdots & X_{30,15} \end{bmatrix}, Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{15} \end{bmatrix}$$

- Richness: Number of nonzero OTUs for a subject
- Family Level richness: Richness belonging to a particular family

Family richness over time



Bayesian Variable Selection formulation



$X_i \sim Pois(\lambda_i)$
$log(\lambda_i) = \mu_x + \phi_1{}^*\theta_1 Z_i$

$Y_i \sim N(\mu_i, \tau)$
$\mu_i = \mu_y + \phi_2{}^*\theta_2 Z_i + \phi_3{}^*\theta_3 X_i$

$\alpha = \phi_1{}^*\theta_1$; ϕ_1=1 then $\alpha = \theta_1$
$\beta = \phi_2{}^*\theta_2$; ϕ_2=1 then $\beta = \theta_2$
$\gamma = \phi_3{}^*\theta_3$; ϕ_3=1 then $\gamma = \theta_3$

Priors specification

$\tau \sim Gamma(0.00001, 0.00001)$ $\quad\quad$ $\mu_x, \mu_y, , \alpha, \beta, \gamma \sim N(0, 0.00001)$
$\phi_i \sim B(\pi_i)$ and $\pi_i \sim U(0, 1)$

Design matrix for the indicators

- Matrix of indicator for the 8 models:

$$\Phi = \left\{ \begin{array}{ccc} \phi_1 & \phi_2 & \phi_3 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right\}$$

- Unique identification of a model

$$C = \{1, 2, 4\}$$
$$T_r = 1 + \Phi C^T$$

Models' posterior probability

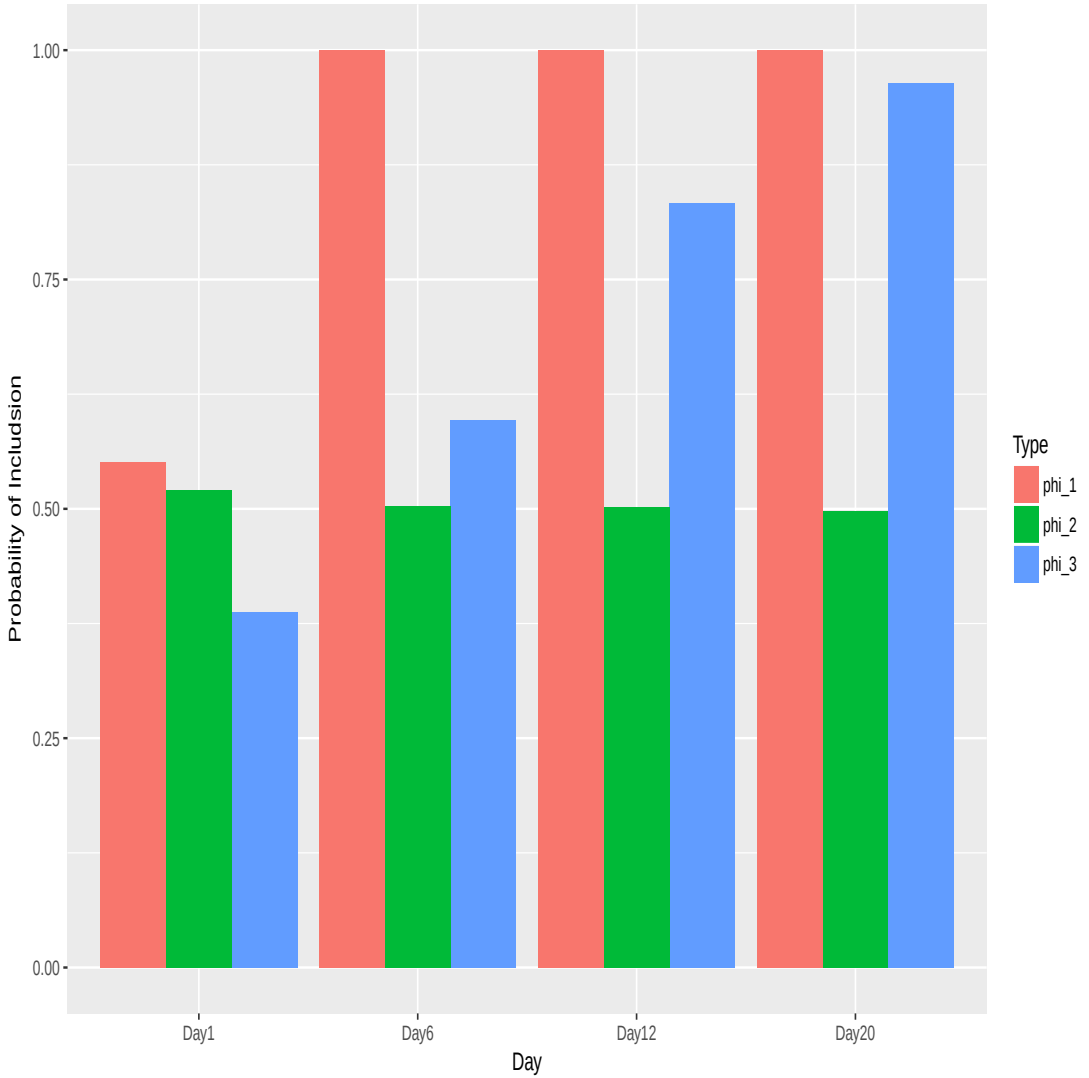
- Transformation rule for each model

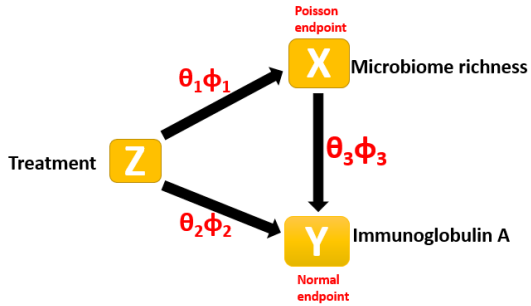
$$T_r = \begin{cases} 1, & \text{for } \phi = (\phi_1 = 0, \ \phi_2 = 0, \ \phi_3 = 0), & \text{model } m_1 \\ 2, & \text{for } \phi = (\phi_1 = 1, \ \phi_2 = 0, \ \phi_3 = 0), & \text{model } m_2 \\ 3, & \text{for } \phi = (\phi_1 = 0, \ \phi_2 = 1, \ \phi_3 = 0), & \text{model } m_3 \\ 5, & \text{for } \phi = (\phi_1 = 0, \ \phi_2 = 0, \ \phi_3 = 1), & \text{model } m_4 \\ 6, & \text{for } \phi = (\phi_1 = 1, \ \phi_2 = 0, \ \phi_3 = 1), & \text{model } m_5 \\ 4, & \text{for } \phi = (\phi_1 = 1, \ \phi_2 = 1, \ \phi_3 = 0), & \text{model } m_6 \\ 7, & \text{for } \phi = (\phi_1 = 0, \ \phi_2 = 1, \ \phi_3 = 1), & \text{model } m_7 \\ 8, & \text{for } \phi = (\phi_1 = 1, \ \phi_2 = 1, \ \phi_3 = 1), & \text{model } m_8 \end{cases}$$

- Posterior probability of transformation

$$p(m_4|\phi, \text{data}) = p(T_r = 5|\phi, \text{data})$$

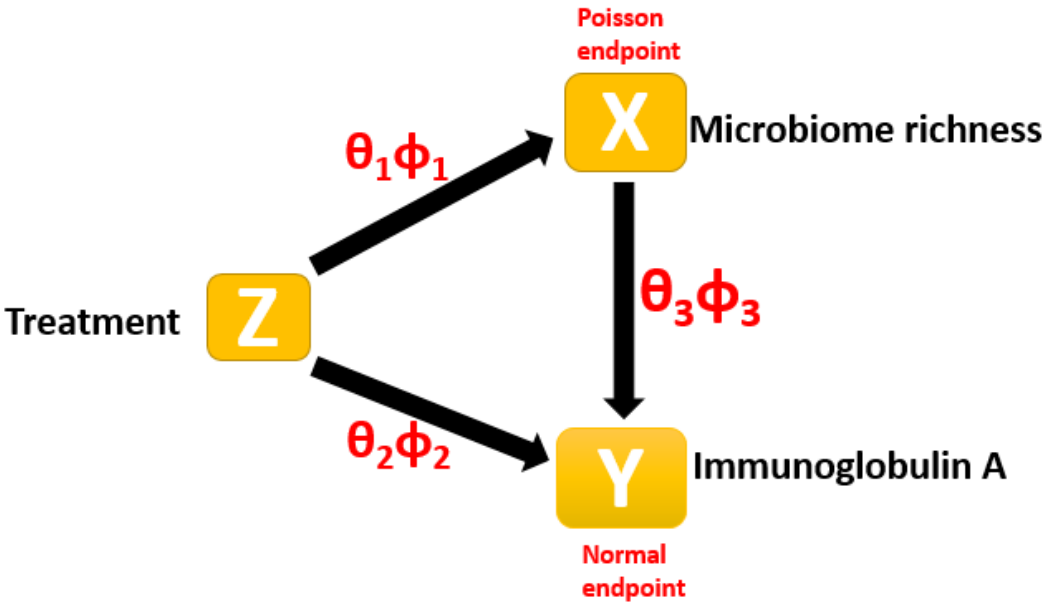
Inclusion probability (per day)





- $P(\phi_3 = 1|data)$ develops over time

What is the probability for X to be a biomarker at day 20?



$$P(\phi_1 = 1|data) = P(m_2) + P(m_5) + P(m_6) + P(m_8) = 0.9996$$
$$P(\phi_2 = 1|data) = P(m_3) + P(m_6) + P(m_7) + P(m_8) = 0.5007$$

$$\underline{\text{Probability that X is a biomarker}}$$
$$P(\phi_3 = 1|data) = P(m_4) + P(m_5) + P(m_7) + P(m_8) = 0.96045$$

Conclusion and Further Research

Conclusion

- Bayesian extension of the JM for all type of distribution
- The probability of inclusion as a measure of surrogacy
- Richness as a biomarker was seen to be highly related to the IgA over time

Further research

- To take into account the longitudinal information
- Incorporate other features
- Testing the method on a better dataset (small sample and few active features)
- Establish a connection between our method and the information criteria approach through simulation.