A new (high dimensional) surrogacy measure based on Bayesian Variable Selection Approach: An application to Microbiome experiment

Olajumoke Evangelina Owokotomo
Rudradev Sengupta
Thi Huyen Nguyen
Ziv Shkedy
Adetayo Kasim

Non-Clinical Statistics Conference, October 20th, 2022

UHASSELT
I-BioStat
KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Background

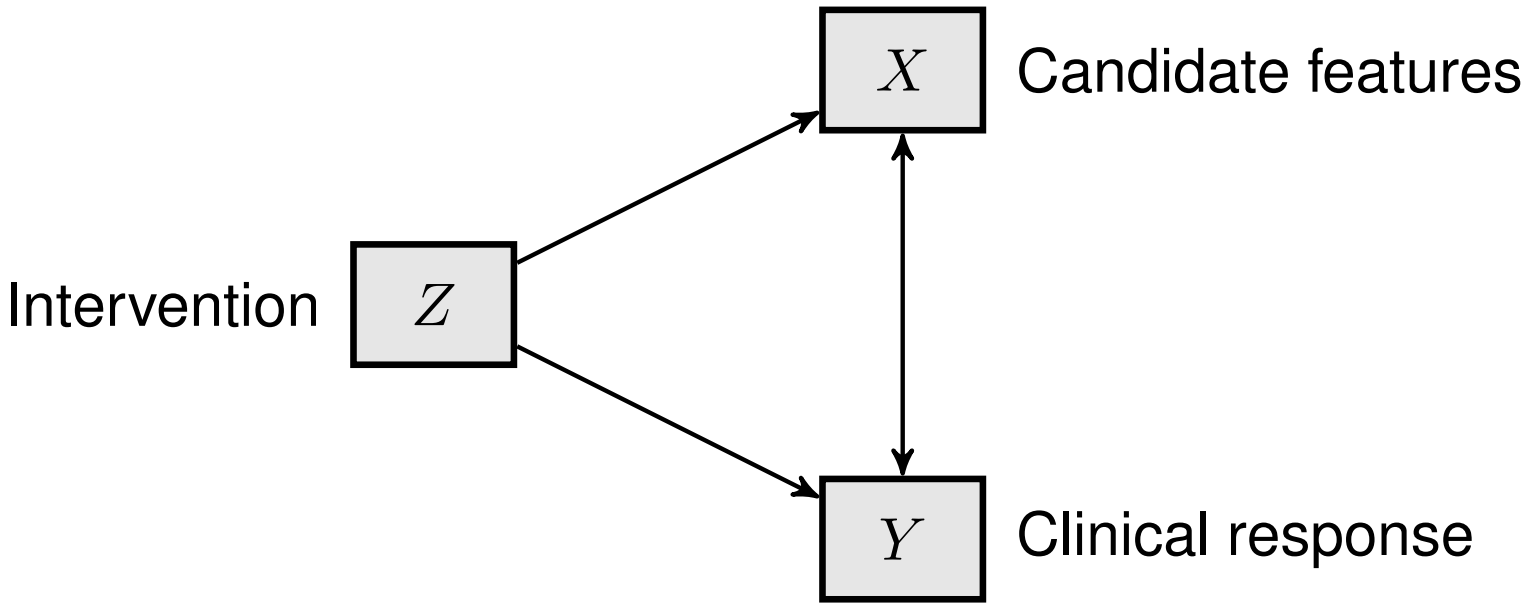
- A biomarker as an attribute that can be objectively measured as an indicator of a healthy or pathological biological processes or pharmacological responses (BWG).
- In the search for surrogates, biomarkers are initially developed and are then further validated to be recognized as surrogates.
- Blood sugar: to identify patients with Type 2 diabetes mellitus.

High Dimensional Biomarkers

- High-throughput biological and medical experiments often result in high dimensional data.
- Discovering biomarker(s) from high dimensional experimental settings for clinical response of interest.
- Approach: Apply and extend methodology develop for clinical trails for ”omics” experiments.

The Biomarker Setting I

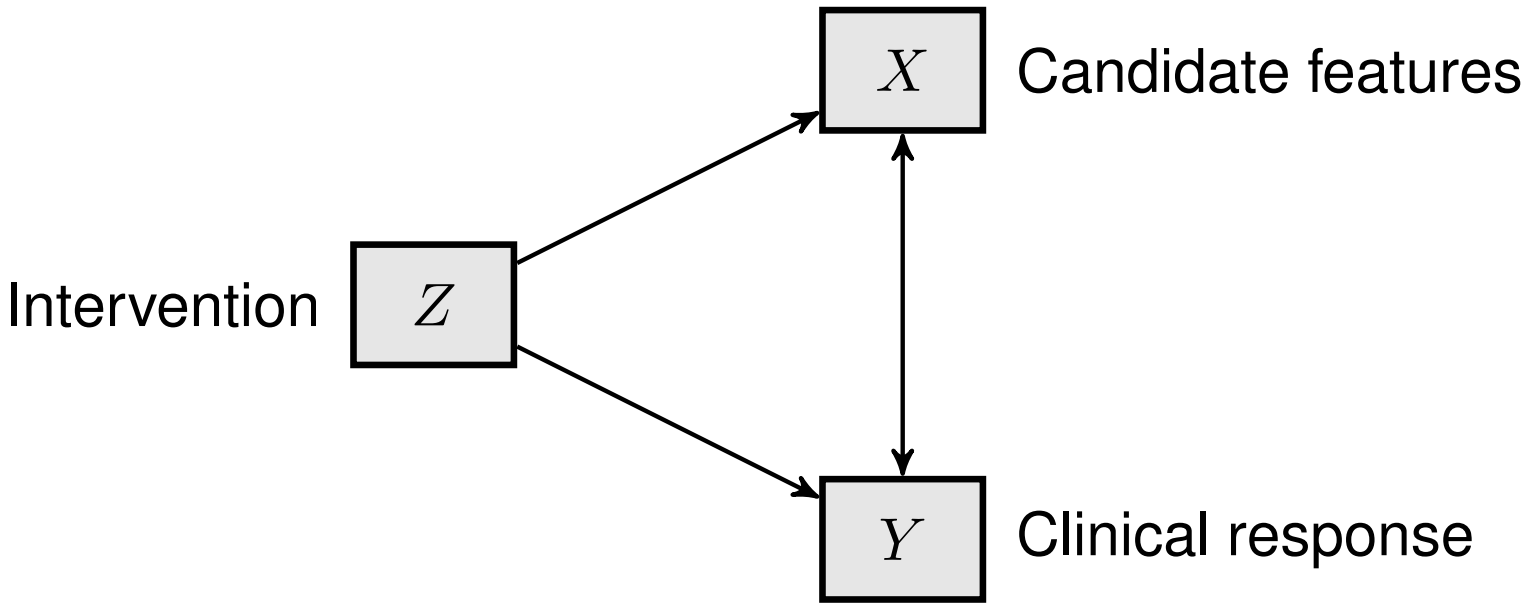
- Response of interest : Y.
- Biomarker matrix: X.
- Intervention: in our case, treatment (Z).



Owokotomo et. al (2022)

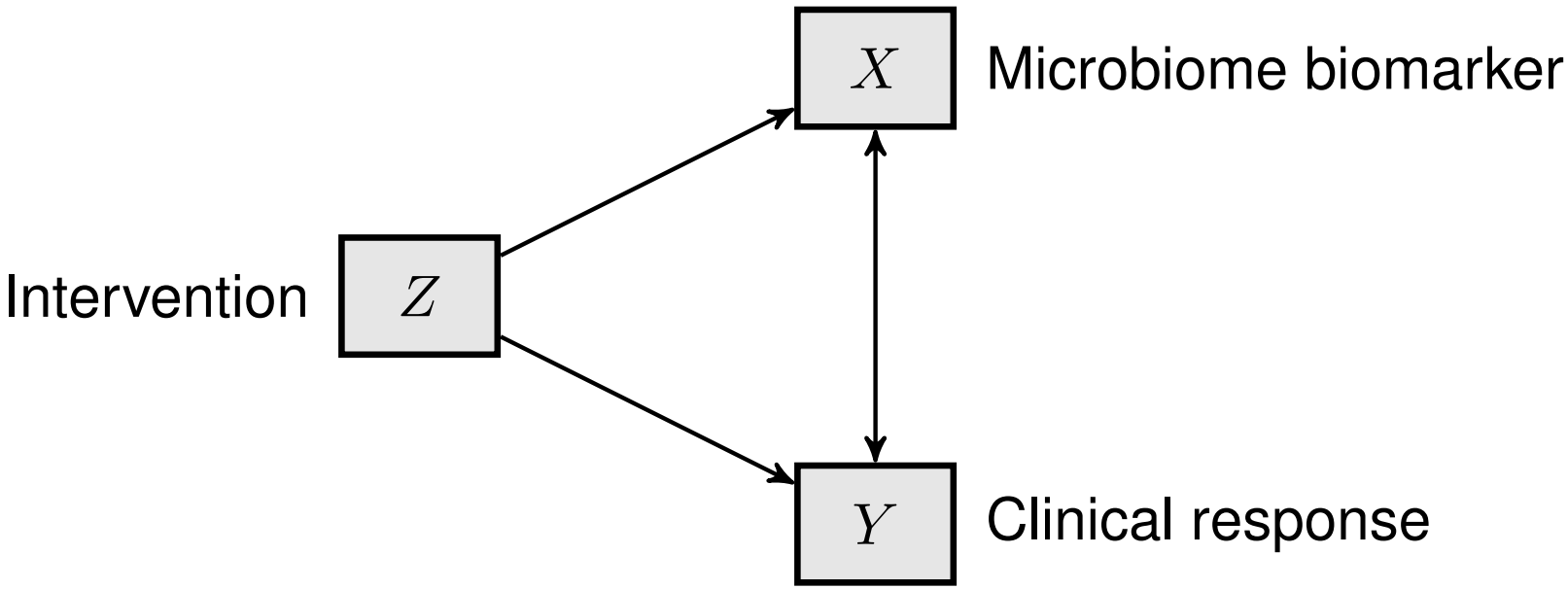
The Biomarker Setting II

- Explore how intervention factors (treatment) influence the biological and clinical response.
- The relationship between the features and clinical response given the intervention.
- The intervention can influence both biological activity and clinical response.

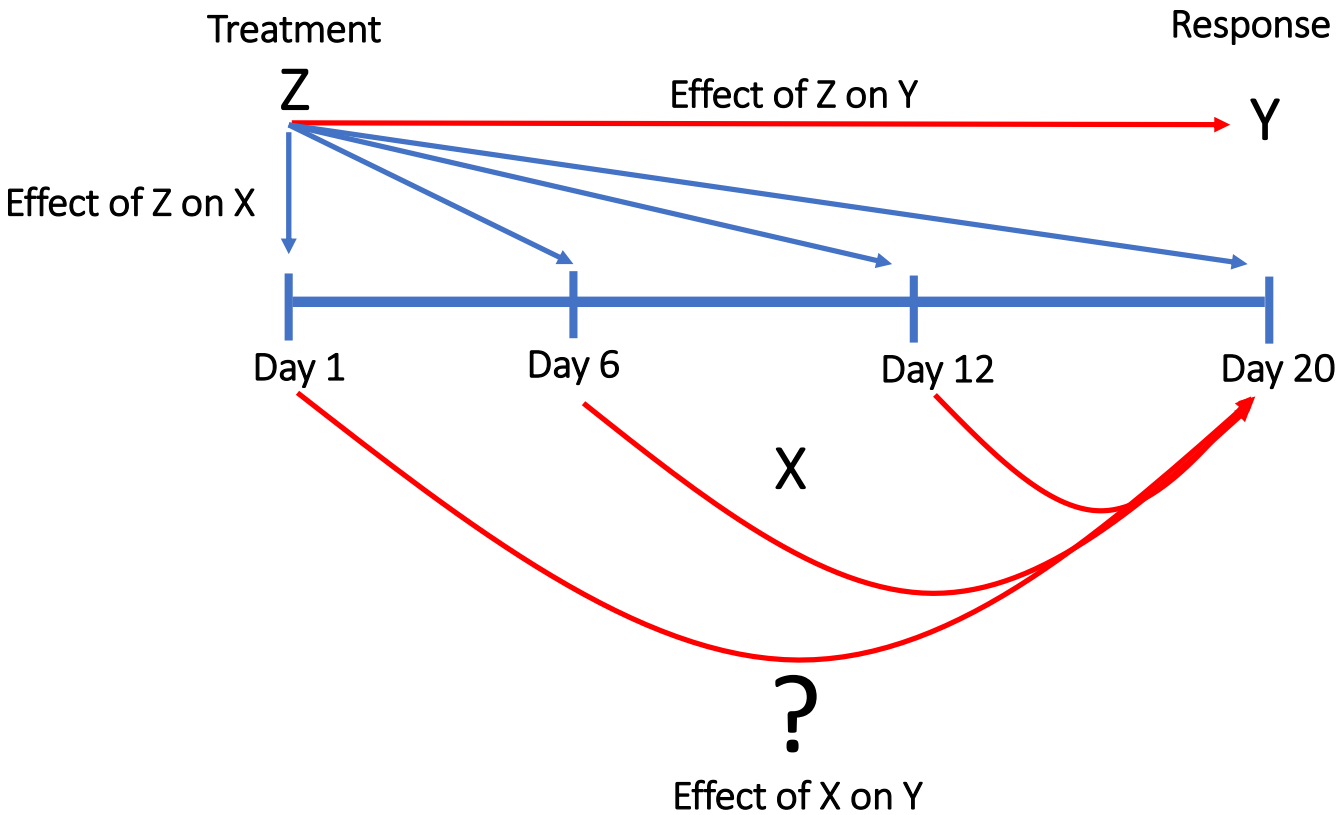


Biomarker Setting for a Microbiome Intervention Study

- Analysis : different level of the microbiome phylogenetic tree.
- Clinical response: binary, continuous, survival.
- Goal of the study.



The Microbiome Setting

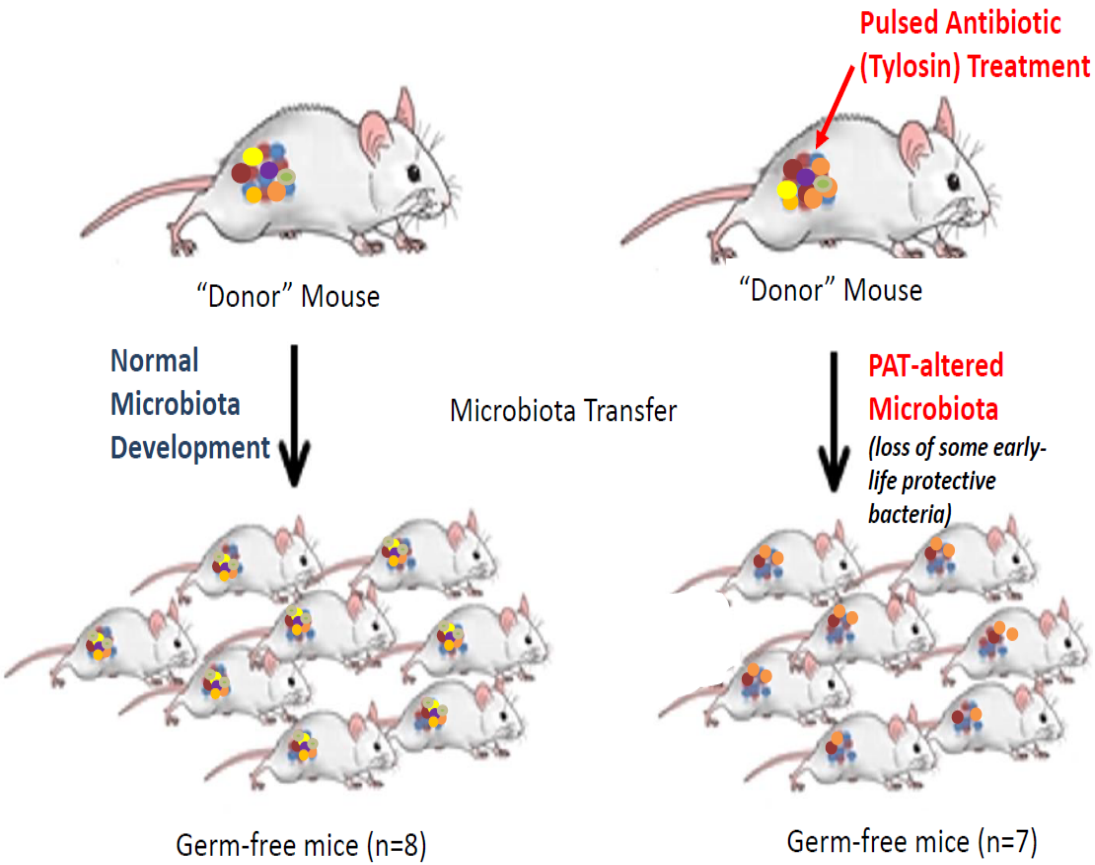


- Can we use microbiome information at a particular day to predict the clinical response of interest?

The Pulse Antibiotics Treatment Study: the TransPat Study

- Experiment conducted at the New York University Langone Medical center human microbiome program laboratory.
- Hypothesized that a single pulse of macrolide antibiotics (tylosin), administered early in life, could perturbed the intestinal microbiota.
- The study focused on the effects of tylosin on the subject's immune system and the microbiome. *Ruiz et. al (2017)*

The TransPAT Study



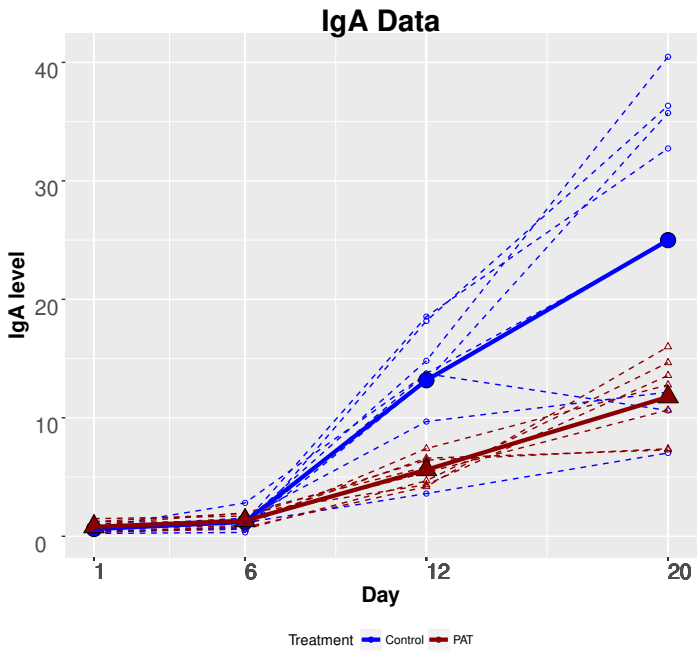
- Measurements:
  1. Microbiome Data.
  2. Immunological Data.
  3. Treatment information.

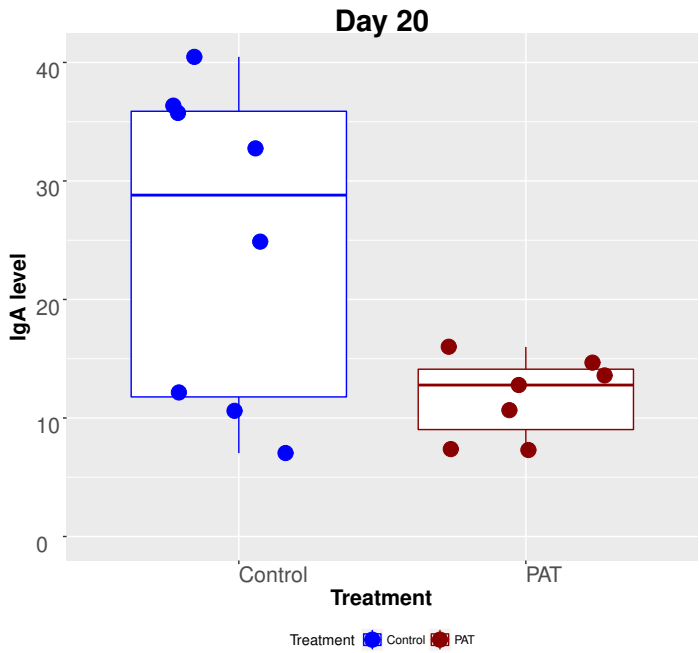
- Research question:
  Is the PAT altered Microbiota sufficient to alter Intestinal Immunity?

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

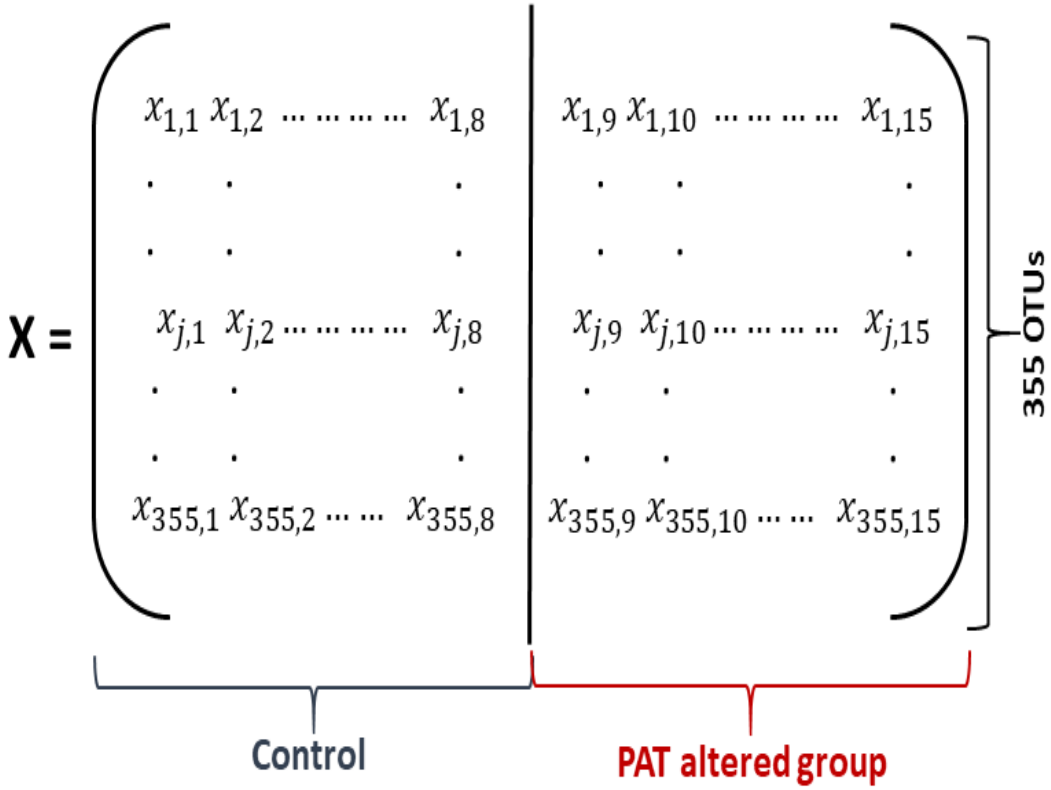
Immunological Data

- Immunity: measured by IgA level.
- Response: IgA level Day20.





Microbiome Data Structure



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

Data Structure

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

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

- Richness: number of nonzero OTUs for a subject.
- Family Level richness: richness of a particular family.

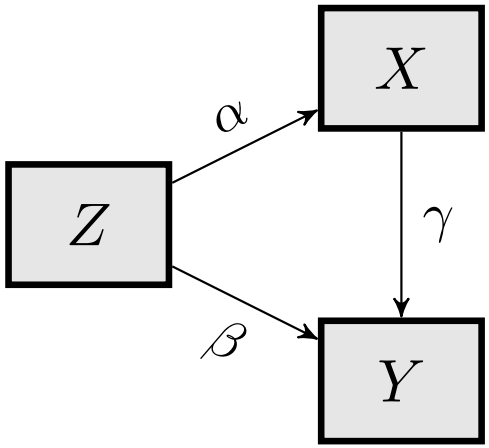
Bayesian Variable Selection Approach (BVS)

What Is New?

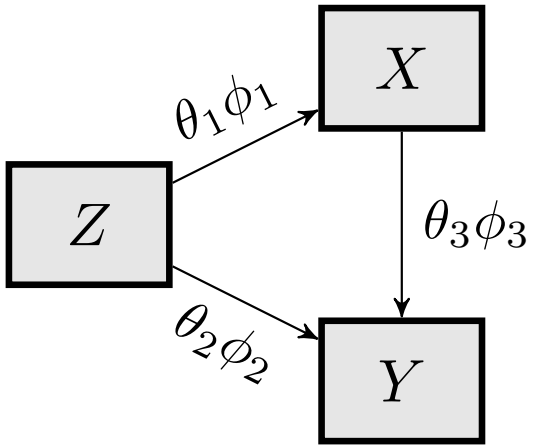
- Applied to any type of outcome combination.
- Model uncertainty is taken into account.
- Based on a probability measure which gives the importance of an endpoint as a biomarker.
- It is related to other measures for surrogacy:
  - The adjusted association measure. *Buyse and Molenberghs (1998)*
  - The information theory approach. *Alonso and Molenberghs (2007)*

Biomarker Surrogacy Setting to BVS

- A path analysis model



- A BVS path analysis model



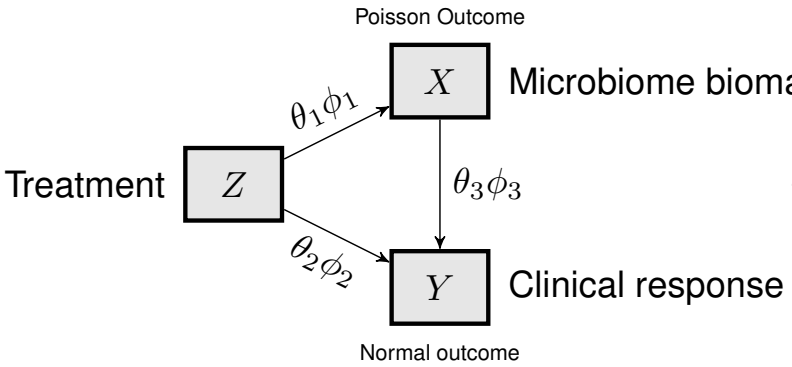
$$\phi_s = \begin{cases} 1, & \theta_s \text{ is included in the model,} \\ 0, & \theta_s \text{ is not included in the model.} \end{cases}$$

- $\boldsymbol{\phi} = \{\phi_1, \phi_2, \phi_3\}$ and $\boldsymbol{\Theta} = \{\theta_1, \theta_2, \theta_3\}$. *(Kuo and Mallick (1998), O'Hara and Sillanpää (2009)*

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Bayesian Variable Selection Formulation

- For the TransPat data (Y: continuous and X: Count variable).



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

- Likelihoods and linear predictors:

$$X_i \sim Pois(\lambda_i),$$
$$log(\lambda_i) = \mu_x + \phi_1\theta_1 Z_i.$$

$$Y_i \sim N(\mu_i, \tau),$$
$$\mu_i = \mu_y + \phi_2\theta_2 Z_i + \phi_3\theta_3 X_i.$$

$\alpha = \phi_1\theta_1$; 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 $\gamma = \theta_3$.

Prior Specifications

- Non-informative sets of priors.

$$\tau \sim Gamma(0.00001, 0.00001),$$
$$\mu_x, \mu_{y,}, \alpha, \beta, \gamma \sim N(0, 0.00001).$$

- Joint prior:

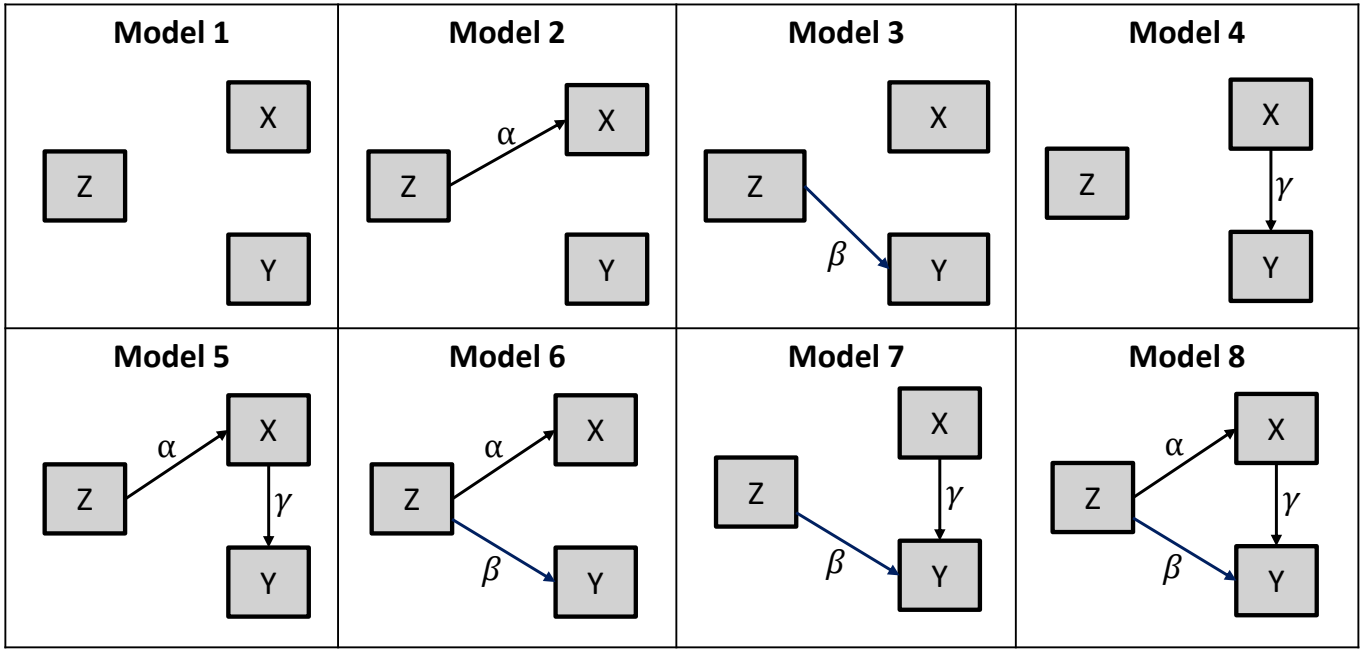
$$P(\theta_s, \phi_s) = P(\theta_s) \times P(\phi_s),$$
$$\theta_s \sim N(0, 0.00001).$$

- Inclusion parameters:

$$\phi_i \sim B(\pi_i),$$
$$\pi_i \sim U(0, 1).$$

Graphical Representation of All Possible Path Models

- Eight path analysis models represent the possible association patterns.



Models' Posterior Probability

- The 8 possible models:

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

- Posterior probability of transformation:

$$p(m_4|\phi, \text{data}) = p(\phi = (0,0,1)|\text{data}) = p(T_r = 5|\phi, \text{data}).$$

The Bayesian Variable Selection as Surrogacy Measure

- Posterior probability that X is a biomarker:

$$p(\phi_3 = 1|\boldsymbol{\phi}, \text{data}) = p(m_4|\boldsymbol{\phi}, \text{data}) + p(m_5|\boldsymbol{\phi}, \text{data}) + p(m_7|\boldsymbol{\phi}, \text{data}) + p(m_8|\boldsymbol{\phi}, \text{data}).$$

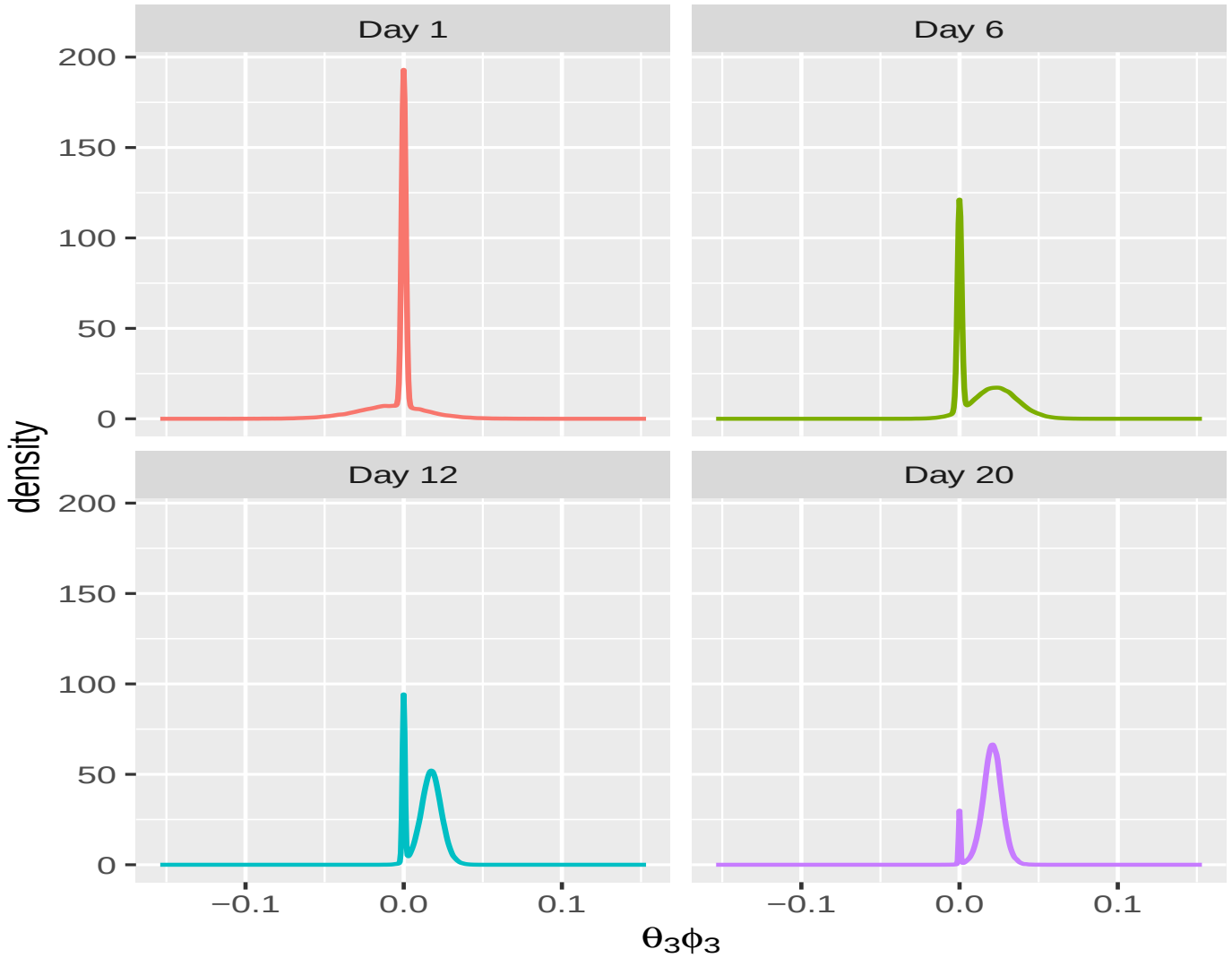
- The inclusion probability of γ, the path from X to Y.

Application to the Data

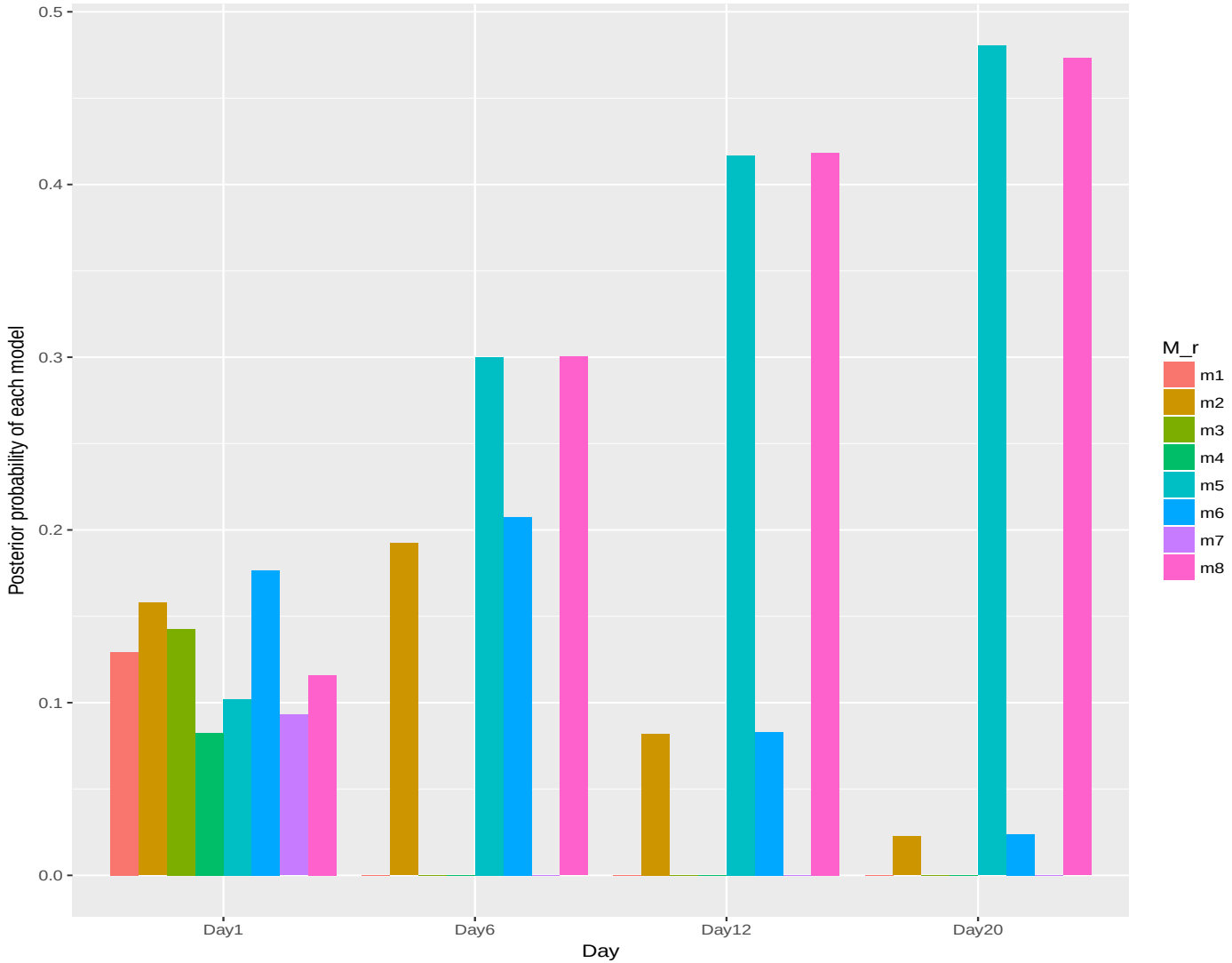
Application to the Data (S24-7 family)

- Response: log IgA at day 20.
- Biomarker: observed richness.
- Analysis at each time point.
- JAGS, convergence satisfied.

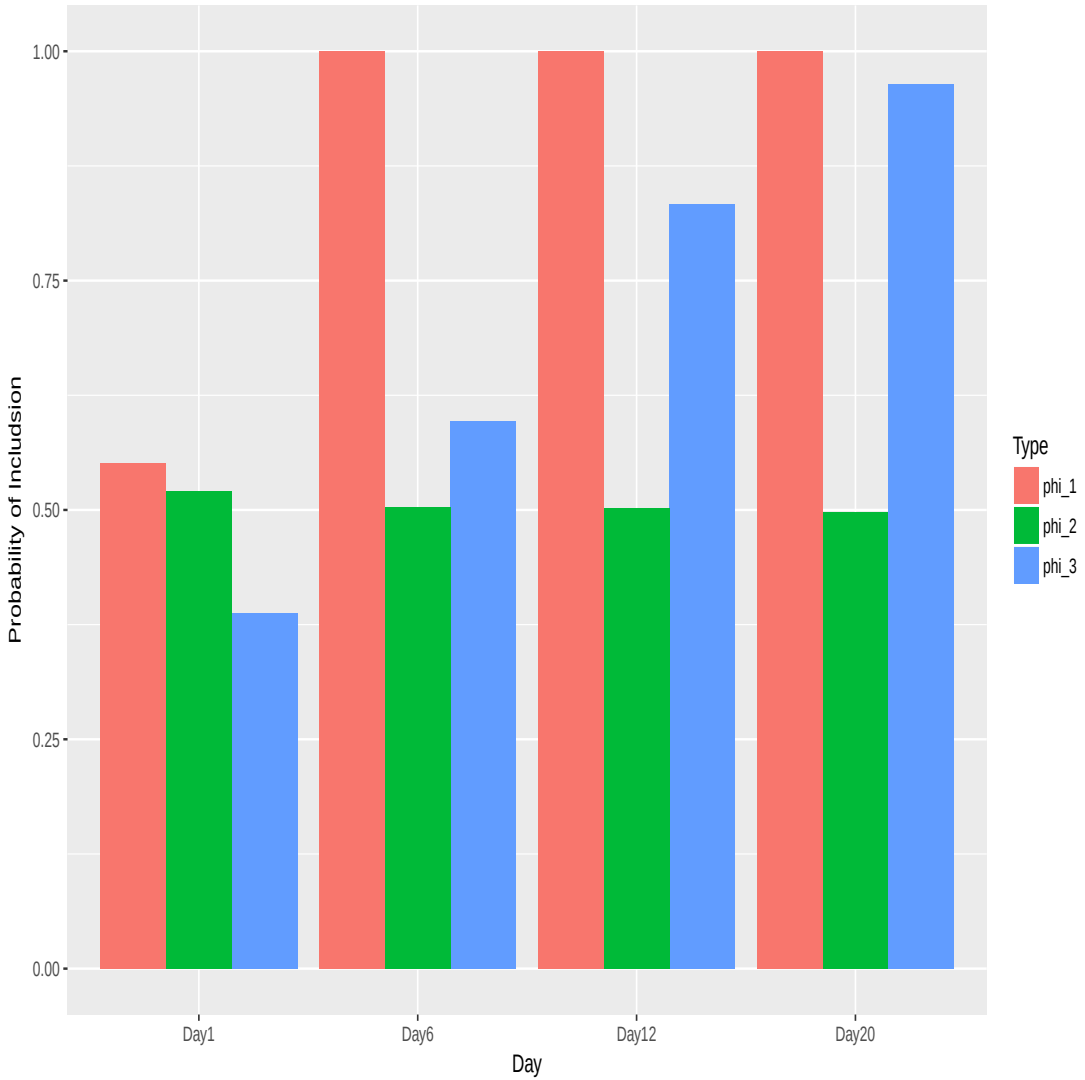
Density plot of $\gamma = \theta_3 * \phi_3$

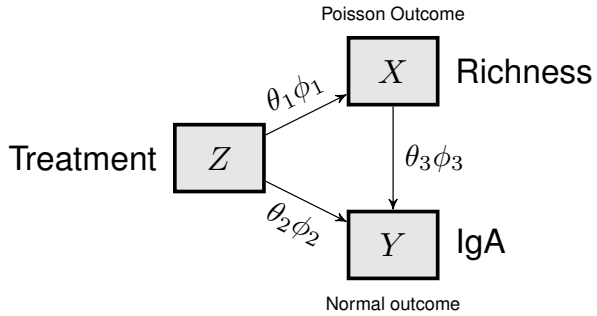


Posterior Model probabilities



Inclusion probabilities





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

Summary

- A Bayesian method to identify biomarkers using the BVS.
- The probability of inclusion for identifying biomarkers.
- This approach takes model uncertainty into account with ease of computation when the outcomes are both not of the normal type.
- Applied to a microbiome dataset where intestinal microbiota was assessed as a biomarker for the Immunological response.

Ongoing research

- The impact of the prior distributions.

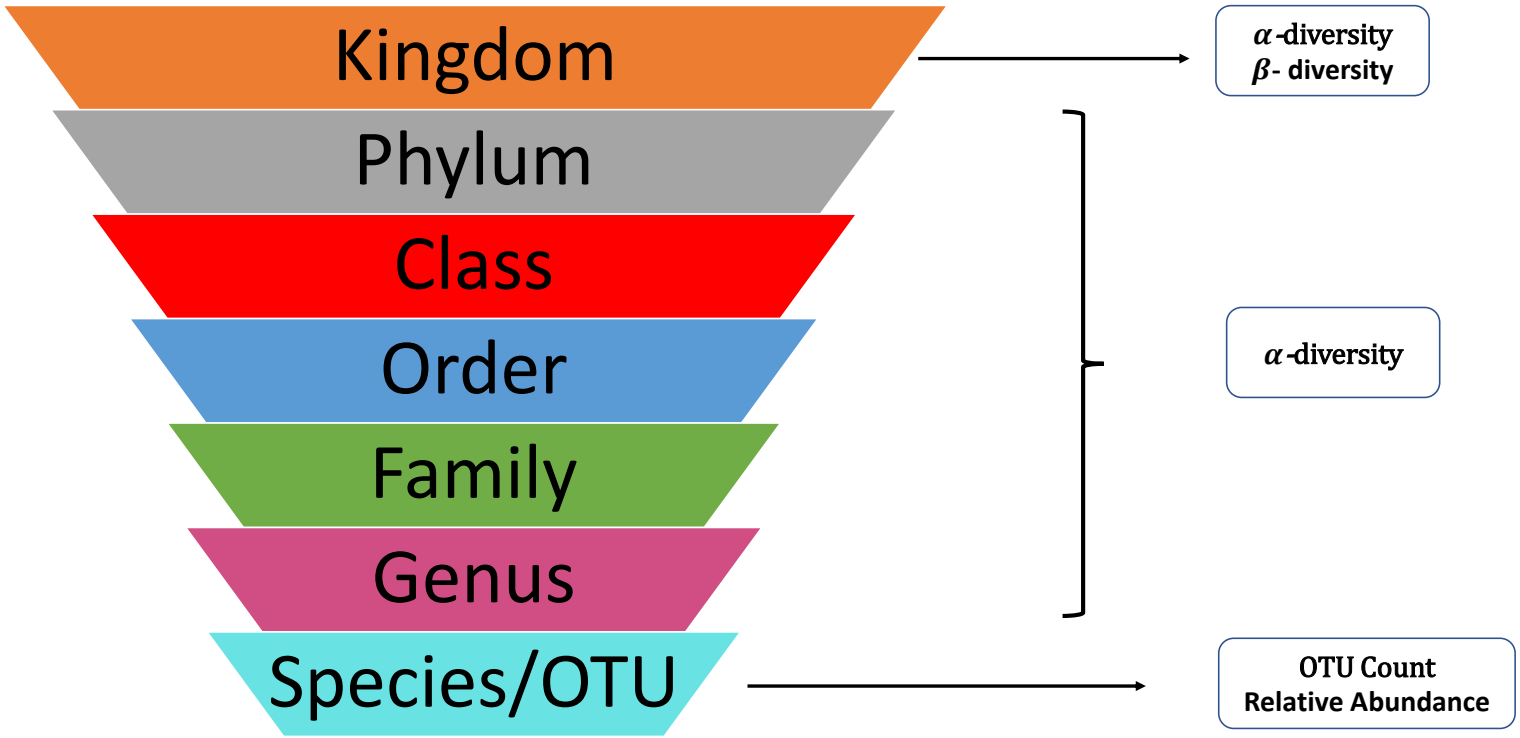
Thank you for your Attention!!!

Bedank!!!

Merci Beacoup!!!

Ese Pupo!!!

Features of Microbiome Data



- Compositional in nature.
- Sparce.
- Over dispersion.
- Vary library size.

Configuration of the Inclusion Parameters

- 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

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