Biomarkers detection in Microbiome experiments using Structural equation modelling

Thi Huyen Nguyen, Rudradev Sengupta, Olajumoke Evangelina Owokotomo, and Ziv Shkedy

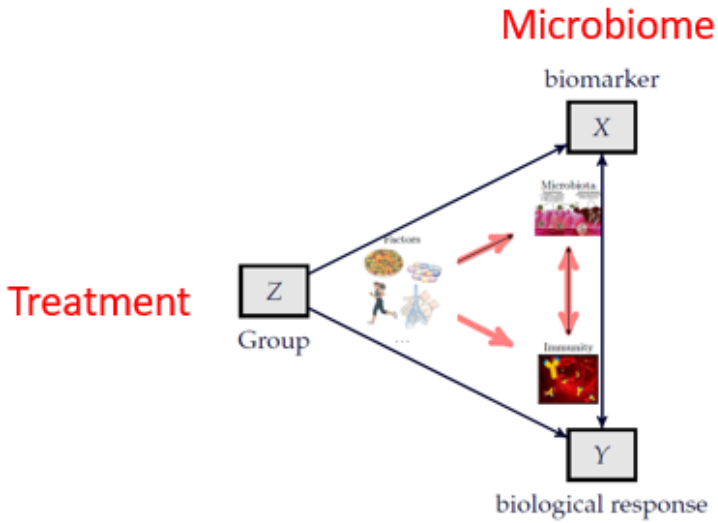
Hasselt University, J&J, GSK

NCS conference 2022 - Belgium

Table of Contents

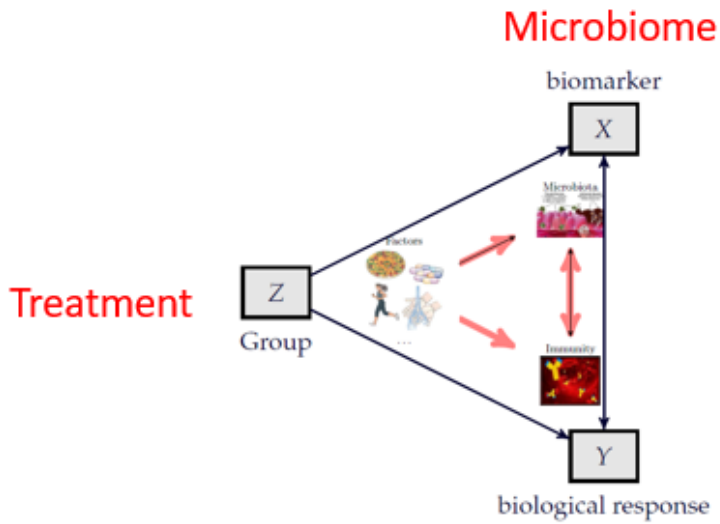
Table of Contents

The basic data structure



- Observation unit: (Z_i, Y_i, X_{ij})
- For each subject in the study:
  - Treatment (intervention).
  - Response/clinical outcome.
  - Microbiome variables.

Type of response



- Y – continuous response

  ↓

  Normal/normal model

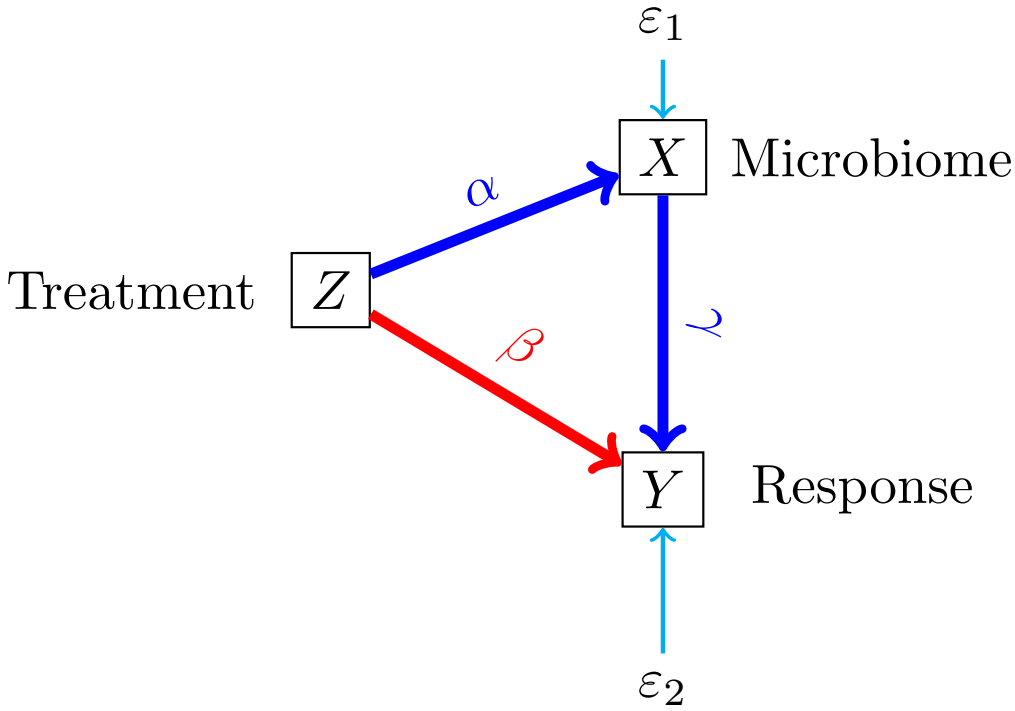
- Y – binary response

  ↓

  Binary/normal model

Path analysis models

- For a continuous response variable:



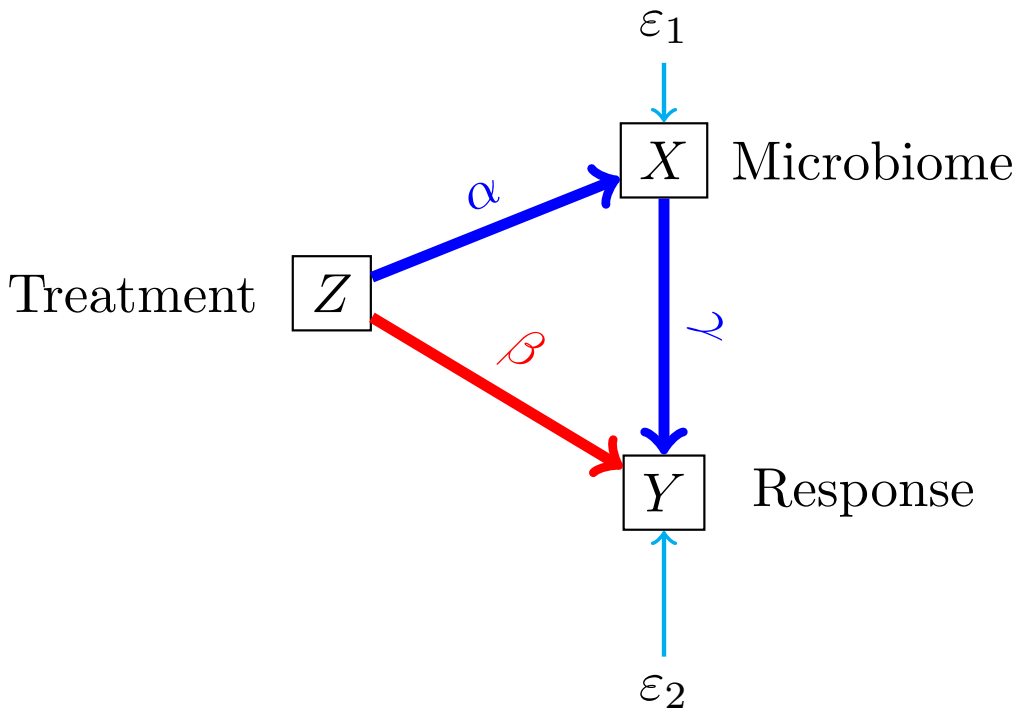
- A feature-specific structural equation model:

$$X_i = \mu_X + \alpha Z_i + \varepsilon_{1i},$$
$$Y_i = \mu_Y + \beta Z_i + \gamma X_i + \varepsilon_{2i}.$$

assuming

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right].$$

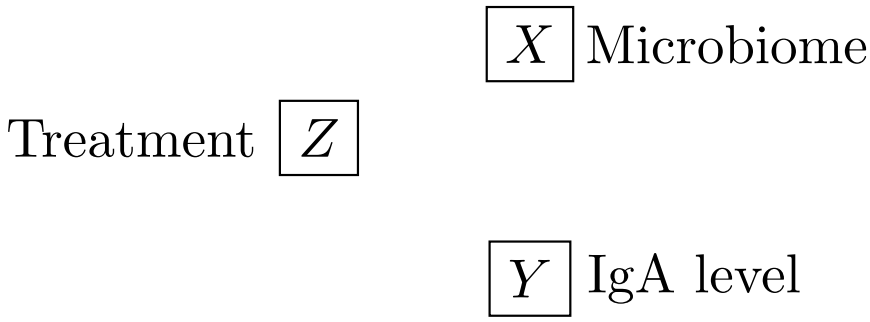
Path analysis models



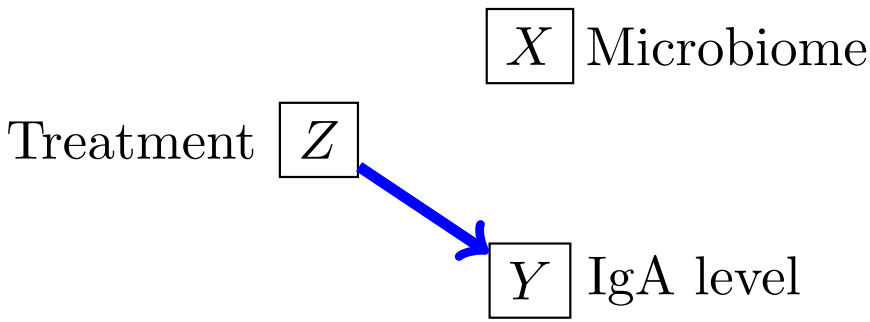
- Direct effect: influence of treatment on the IgA Level that is unmediated with by the microbiome: β.
- Indirect effect: the effect of the Treatment is mediated by the microbiome: $\alpha * \gamma$.
- Total effect: sum of the direct and indirect effects: $\beta + \alpha * \gamma$.

Potential path models

- Eight possible path models

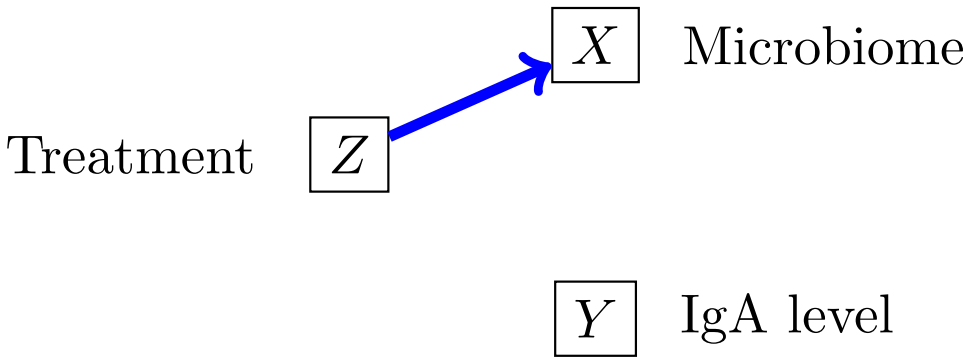


- Independent model (model 1): No relation at all.

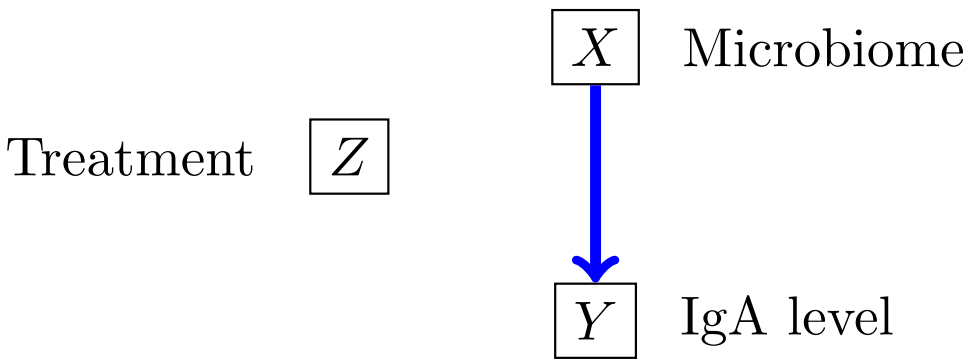


- Single effect model (model 2): Direct effect between Z and Y.

Potential path models

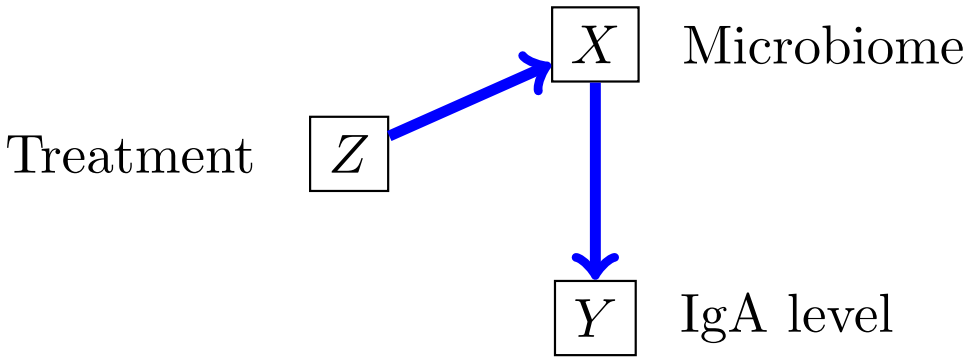


- Single effect model (model 3): Direct effect between Z and X.

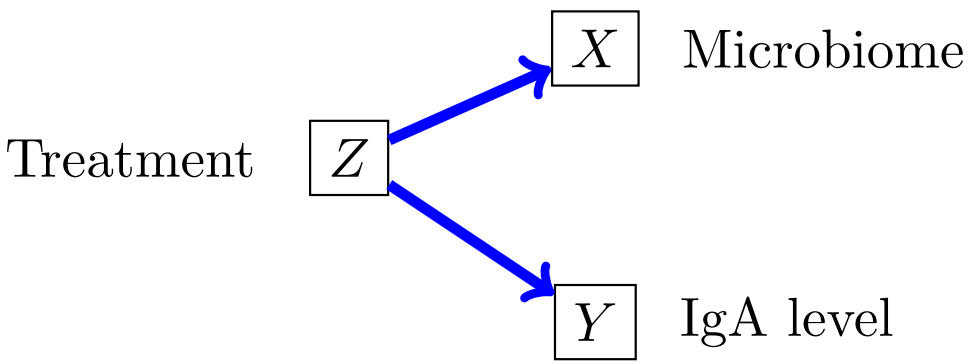


- Single effect model (model 4): Direct effect between X and Y.

Potential path models

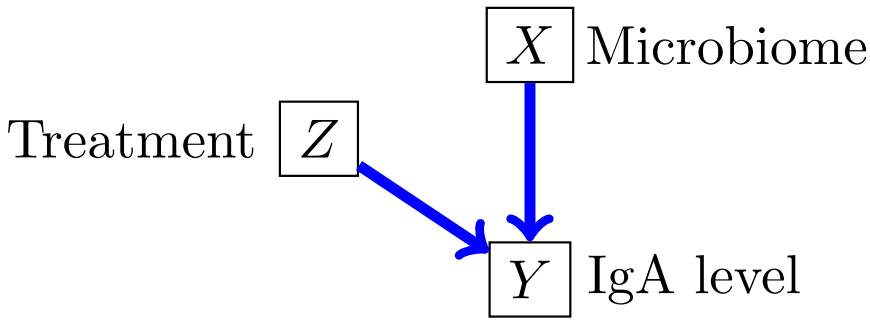


- Indirect effect model (model 5): Indirect effect between Z and Y (complete mediation).

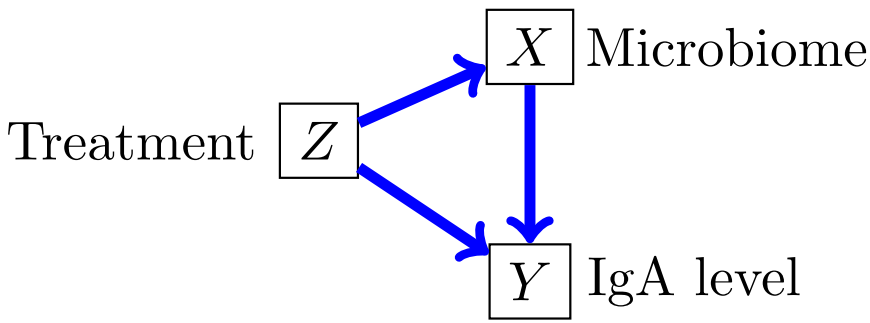


- Common effect model (model 6): Z affects X as well as Y. Conditional independence: given Z, X and Y are independent.

Potential path models



- Common effect model (model 7): Z as well as X influence Y.



- Partial mediation model (model 8): Z affects X which in turn affects Y. In addition Z affects Y.

The posterior probability

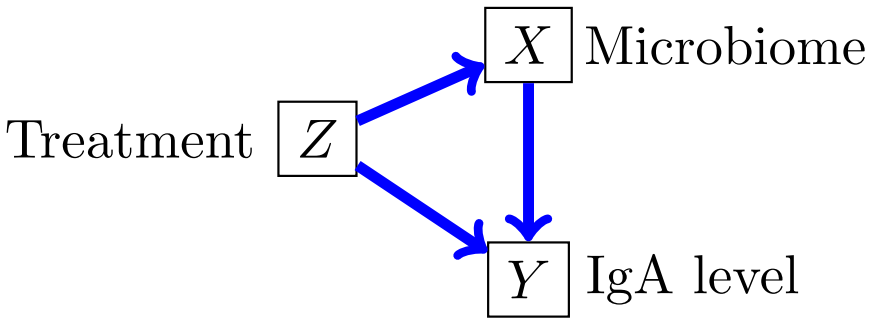
- Let $M_1, \cdots, M_R$ be a set of R candidate models fitted to the data.

$$w_r = P_{IC}(M_r|data) = \frac{\exp\left(-\frac{1}{2}\Delta IC_r\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta IC_r\right)}, \quad r = 1, \cdots, R.$$

- Non informative prior $P(M_r) = 1/R$ for all models.
- The model likelihood is approximated by $P_{IC}(data|M_r) = \exp\left(-\frac{1}{2}\Delta IC_r\right)$, where $\Delta IC_r = IC_r - IC_{min}$ with $IC_{min} = min(IC_1, \cdots, IC_R)$.
- Information criterion (IC) can be AIC, BIC, DIC, WAIC, ...

Table of Contents

The TransPAT study



Y – continuous response

Normal/normal model

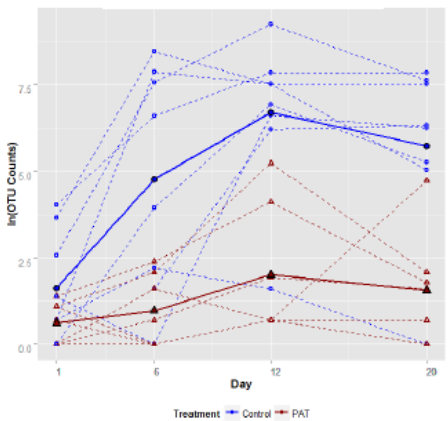
- Pulsed antibiotic treatment (PAT) model of pediatric exposures.
- Hypothesis: A single pulse of macrolide antibiotics (tylosin) administered early in life will change the intestinal microbiota and lead to long lasting alterations in immune profiles.
- Intervention: one dose of antibiotic.

Different resolutions

The analysis will be focused on:

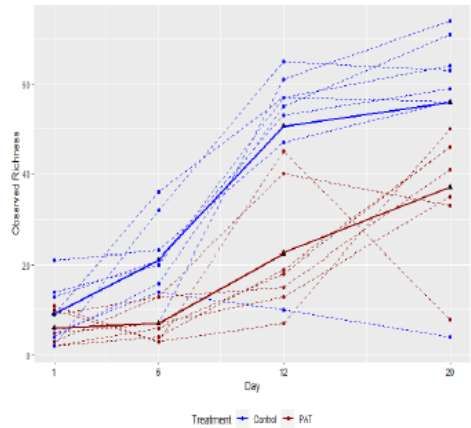
- Family level (a subset of bacteria which belong to the same family).
- The same analysis can be done on OTU and Kingdom level.

OTU 264734



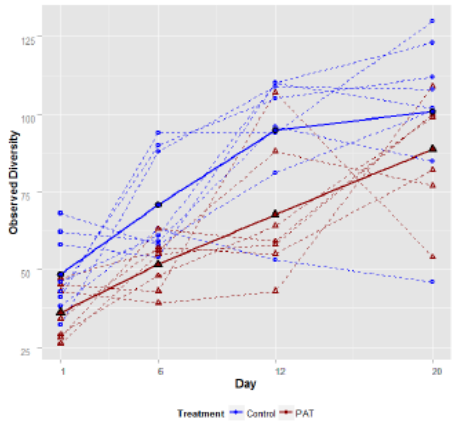
The OTU is a part of the S24-7 family

Family S24-7



Number of OTUs with non-zero count in the family

Alpha Diversity



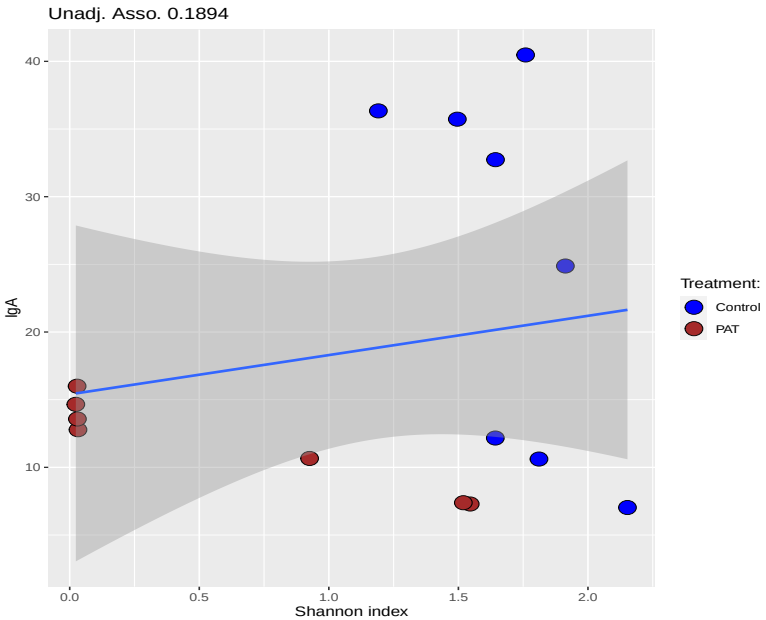
Number of OTUs with non-zero count in the data.

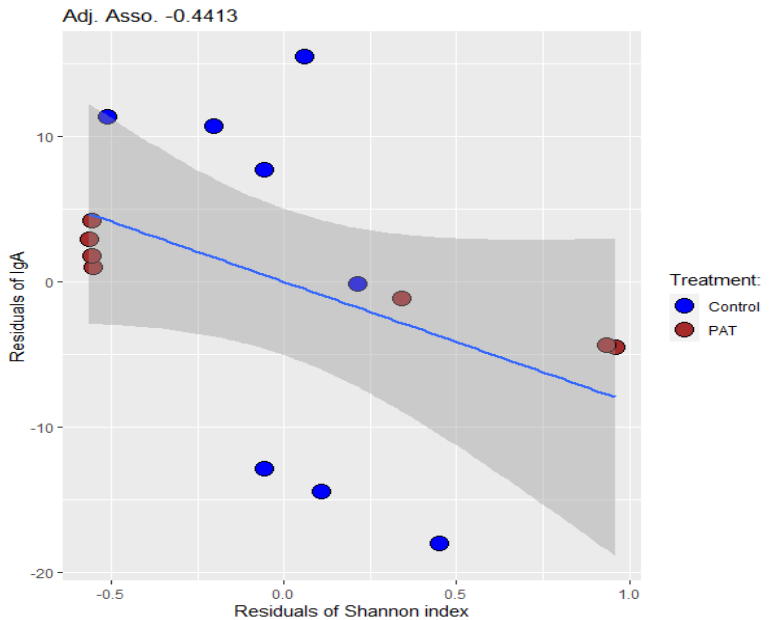
Results of eight SEMs for family S24-7 at day 12

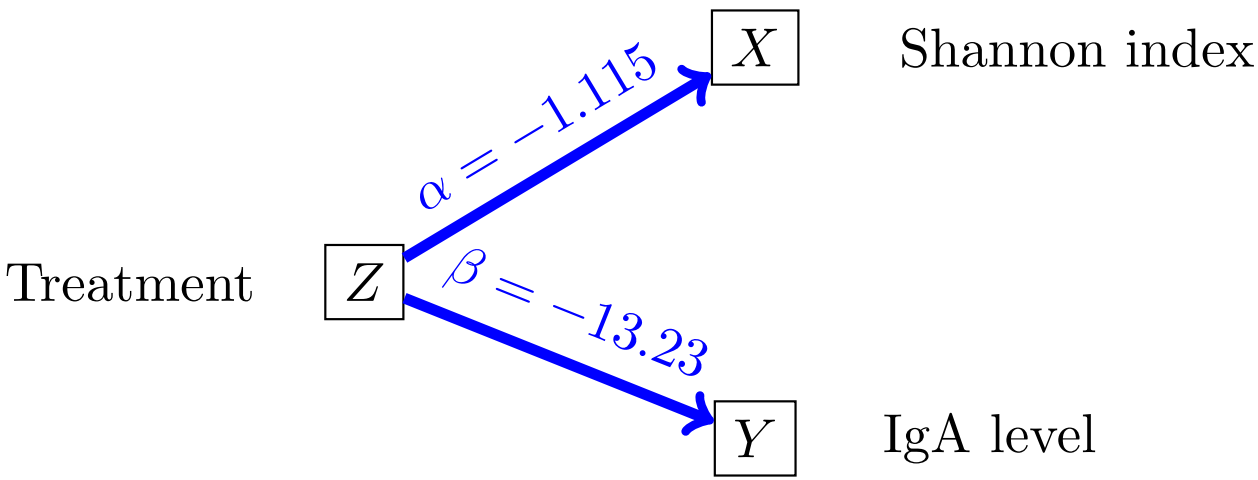
Model	AIC	BIC	w_{AIC}	w_{BIC}	α(p-value)	γ(p-value)	β(p-value)
1	157.4780	160.3102	0.0002	0.0006			
2	153.4003	156.9406	0.0016	0.0031			-13.23(0.0062)
3	147.3415	150.8818	0.0330	0.0641	-1.115(<0.0001)		
4	158.9300	162.4703	0.0001	0.0002		2.901(0.455)	
5	148.7935	153.0418	0.0160	0.0218	-1.115(<0.0001)	2.901(0.455)	
6	142.0141	146.9705	0.4731	0.4531	-1.115(<0.0001)		-13.23(0.0062)
7	152.1506	156.3989	0.0030	0.0041		-8.273(0.0043)	-22.45(<0.0001)
8	142.0141	146.9705	0.4731	0.4531	-1.115(<0.0001)	-8.273(0.0568)	-22.45(0.0006)

- Shannon index as potential biomarker.
- Possible models: 6 and 8.
- Shannon index of S24-7 at day 12 can be a biomarker for IgA.
- Similar results for days 1, 6, and 20.

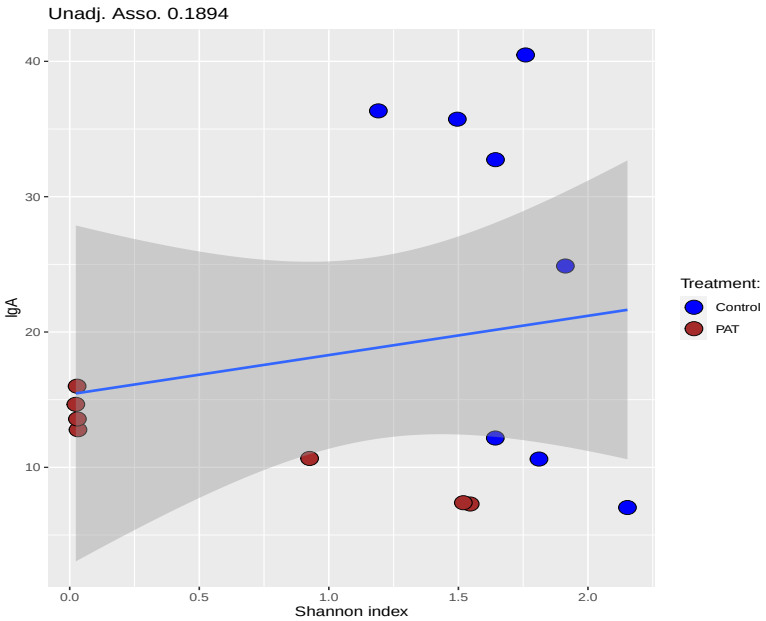
Result of model 6 day 12

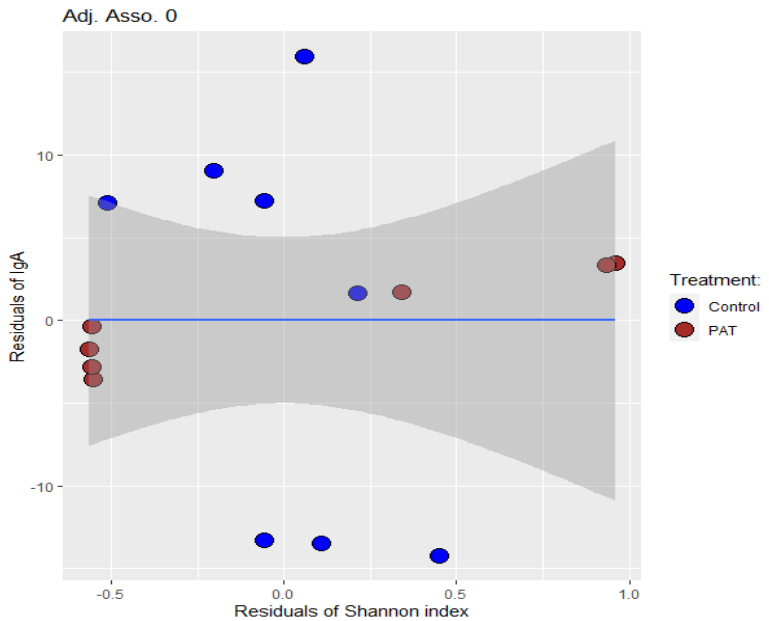






Result of model 8 day 12





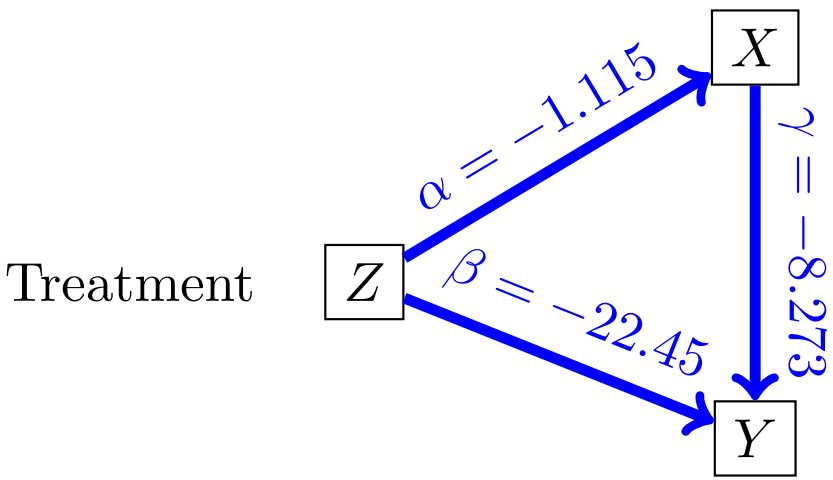
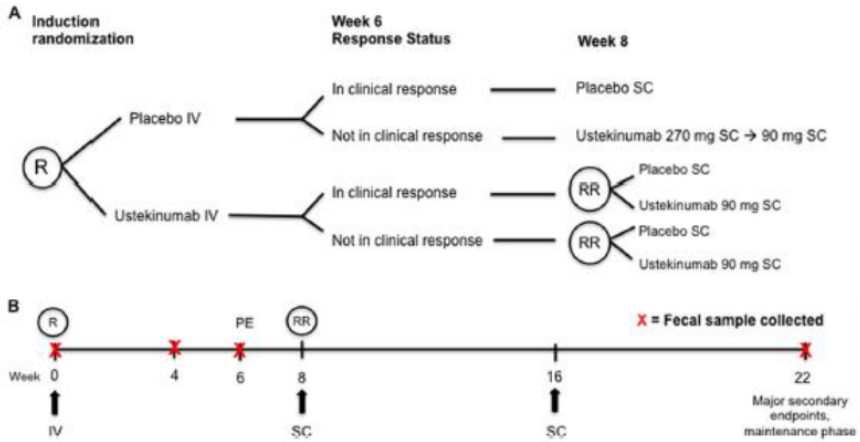


Table of Contents

The Doherty study



- 130 participants were divided into two treatment groups:
  - Placebo (n = 29).
  - UST intravenously for therapy (n = 101).
- Longitudinal microbiome measurements at week 0, 4, and 6.
- Clinical endpoint: remission at week 6.

The endpoint and research question

- Endpoint of interest: remission at week 6.
- Binary outcome:

$$Y_i = \begin{cases} 1 & \text{remission,} \\ 0 & \text{non remission.} \end{cases}$$

- Association between the binary clinical outcome and the microbiome at week 0, 4, and 6 taking into account the treatment effect.
- Analysis at each time point separately.
- Analysis at a kingdom level.
- X_i: Inverse Simpson index at week 0, 4, and 6.

Path analysis models for a binary/normal variables

- The corresponding probit model:

$$\begin{cases} X_i \sim N(\mu_X + \alpha Z_i, \sigma^2), \\ Y_i \sim B(p_i) \\ \Phi^{-1}(p_i) = \mu_Y + \beta Z_i + \gamma X_i. \end{cases}$$

- $B(p_i)$: Bernoulli distribution with the success probability

$$p_i = P(Y_i = 1).$$

- Φ: the standard normal cumulative distribution function.

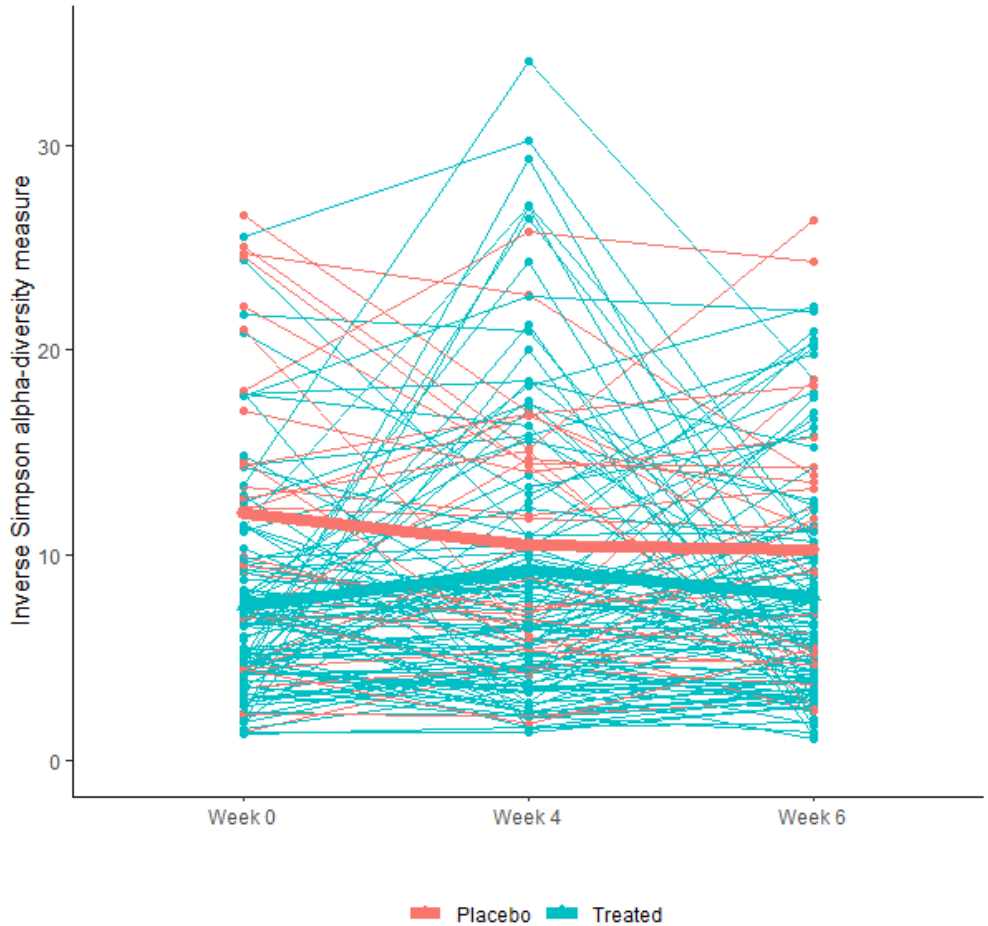
Bayesian approach

- Parameters of the two models in the system are estimated simultaneously by using a single variance-covariance matrix.

- Assume independence and normality of errors.

- Non-informative priors:
  - $\mu_X, \mu_Y, \alpha, \gamma, \beta \sim N(0, 10^6)$.
  - $\sigma \sim \text{Cauchy}(0, 10)$.

Profile plot of Inverse Simpson at kingdom level



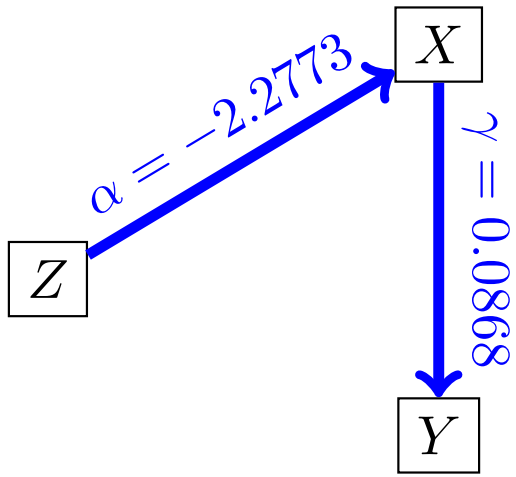
Results on kingdom level at week 6

Model	WAIC	w_{WAIC}	α(sd)	γ(sd)	β(sd)
1	1264.0725	0.1701			
2	1266.1939	0.0589			-0.1646 (0.3127)
3	1263.1064	0.2758	-2.3019 (1.1763)		
4	1267.3996	0.0322		0.0865 (0.0245)	
5	1263.8000	0.1950	-2.2773 (1.1995)	0.0868 (0.0238)	
6	1264.2623	0.1547	-2.2825 (1.1924)		-0.1794 (0.3178)
7	1266.9722	0.0399		0.0871 (0.0237)	0.0438 (0.3588)
8	1265.7560	0.0733	-2.2860 (1.2100)	0.0879 (0.0244)	0.0582 (0.3554)

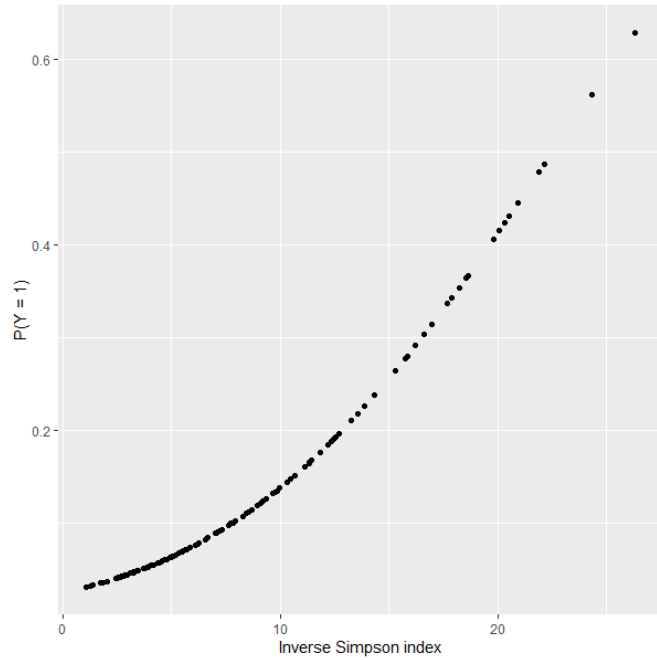
- Inverse Simpson index as potential biomarker.
- Possible model: 3 and 5.
- Inverse Simpson on kingdom level at week 6 can be a biomarker for remission.

Model 5 at week 6

- Model 5:



- Predicted probability of remission using model 5:



- Only indirect effect of the treatment on the response via the microbiome.

- Positive relationship between the probability to have the response and value of inverse Simpson.

Discussion

- Path analysis model for detection of biomarkers.
- Different types of responses.
- Different microbiome indices (Shannon, Inverse Simpson, etc.) can lead to different results.
- Still to do:
  - How to take longitudinal process into account.
  - Modelling time to event responses.

THANK YOU
FOR
YOUR ATTENTION

Modelling approach

- Variables: X_i, Y_i, and Z_i are microbiome, remission, and treatment of subject i.

- Let Y_i^* be the level of a latent continuous variable underlying the response Y_i. Assuming Y_i^* is normally distributed.

$$Y_i = \begin{cases} 1 & Y_i^* > 0, \\ 0 & Y_i^* \leq 0. \end{cases}$$