Microbiome Project:

Structural Equation Model

Rudradev Sengupta, Nolen Joy Perualila and Ziv Shkedy

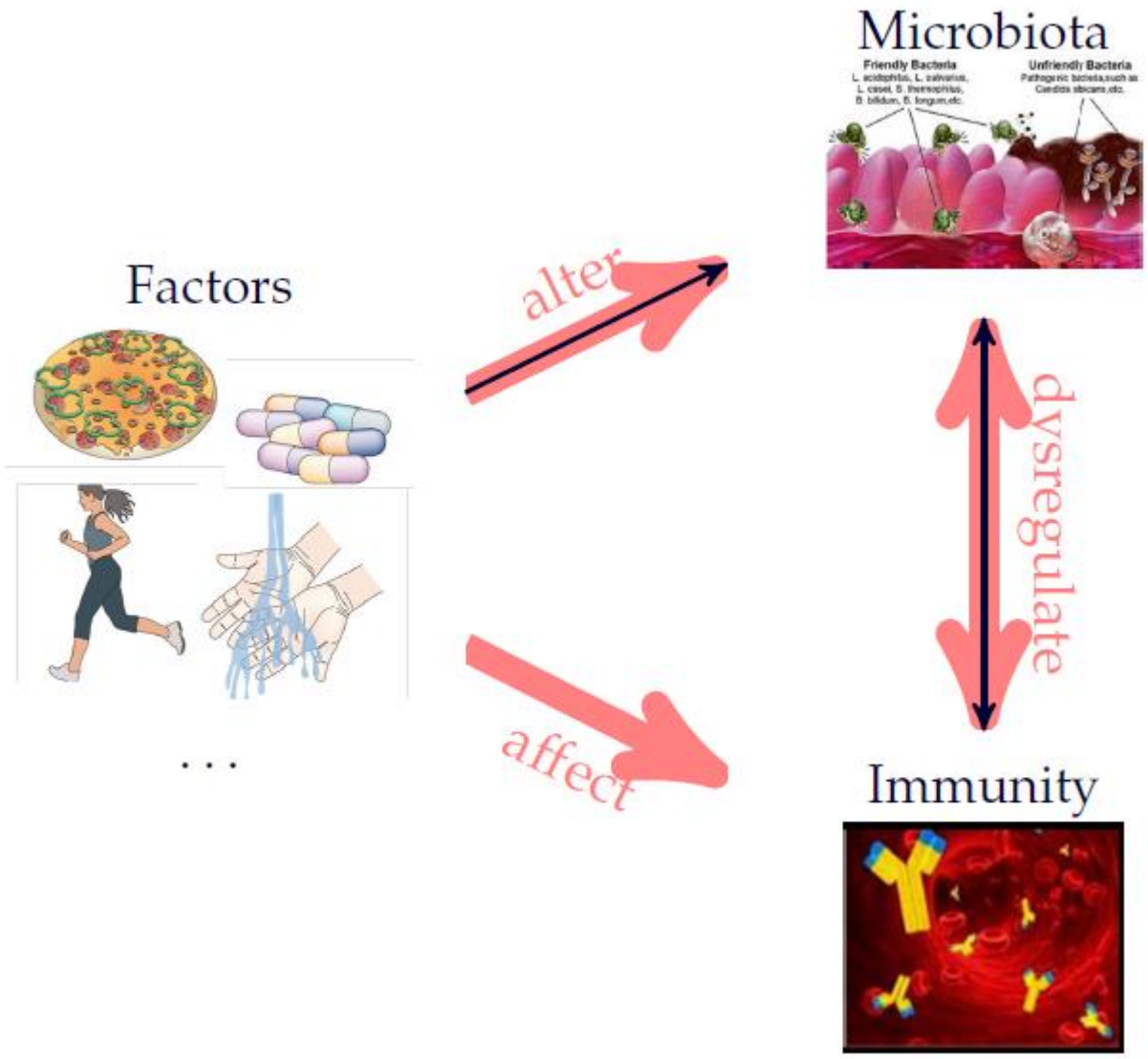
Microbiome Statistics Meeting

16.01.2017

universiteit hasselt
KU LEUVEN
I-BioStat

Part 1: The Setting

Interdependence in the Gut



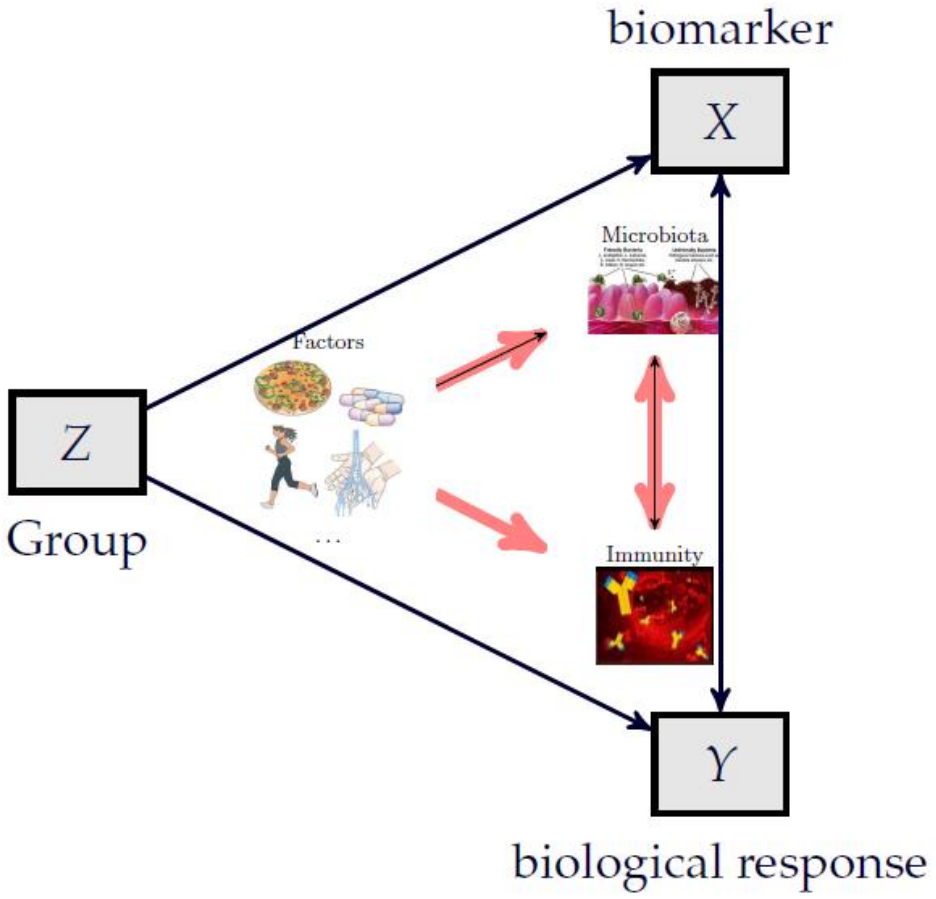
The Biomarker Setting

Typical Data:

Y = Biological Response (IgA)

X = Candidate biomarker for Y (Microbiome data)

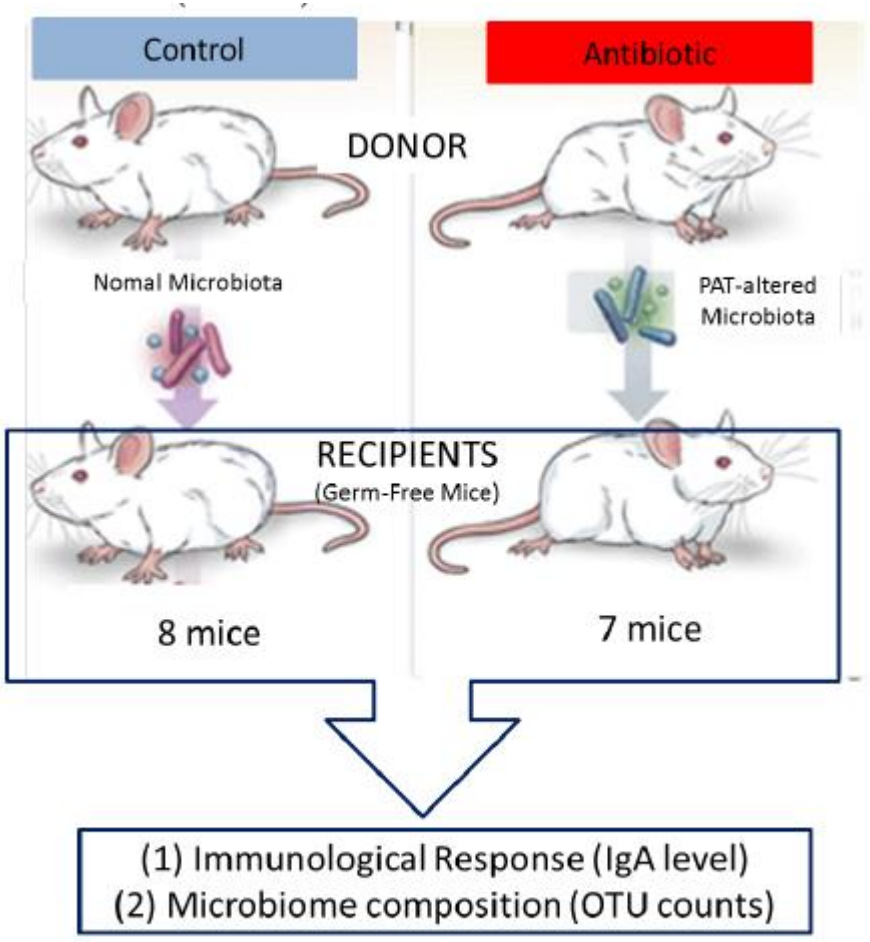
Z = Factor (Treatment)

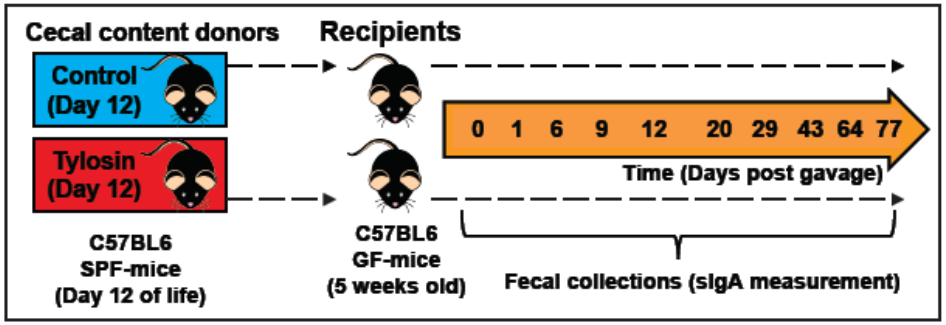


PAT Studies

- PAT - Pulsed Antibiotic Treatment model of pediatric exposures.
- Hypothesis: A series of short, therapeutic-dose pulses of antibiotic administered early in life will perturb the intestinal microbiota and lead to long-lasting alterations in metabolic and immune profiles.
- exPAT with 2 antibiotics – amoxicillin & tylosin - tylosin was found to be effective and hence continued.
- pulsePAT - tylosin at 2 different dose levels over time - PAT1 & PAT3.
- transPAT – tylosin only at one dose level.

transPAT Study



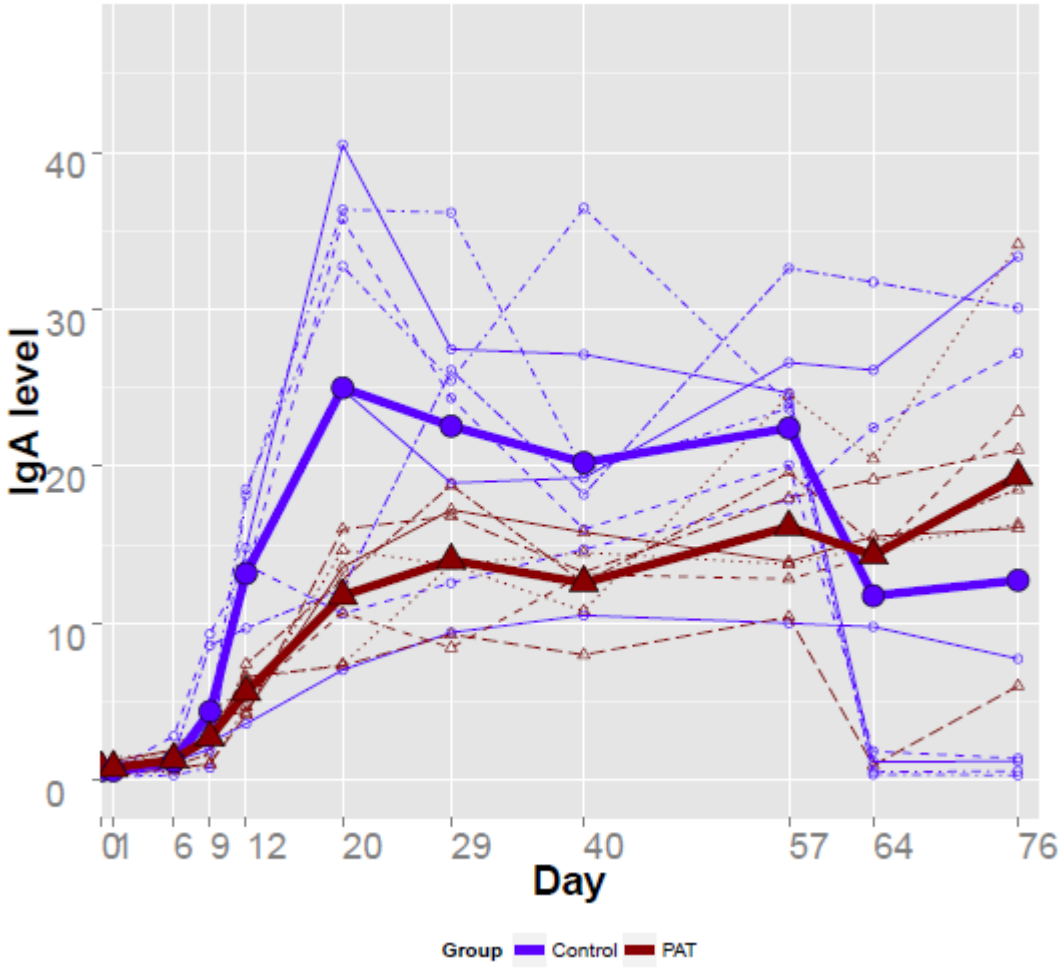


(2 days post-PAT)

Is the (PAT)-altered microbiota sufficient to alter intestinal immunity?

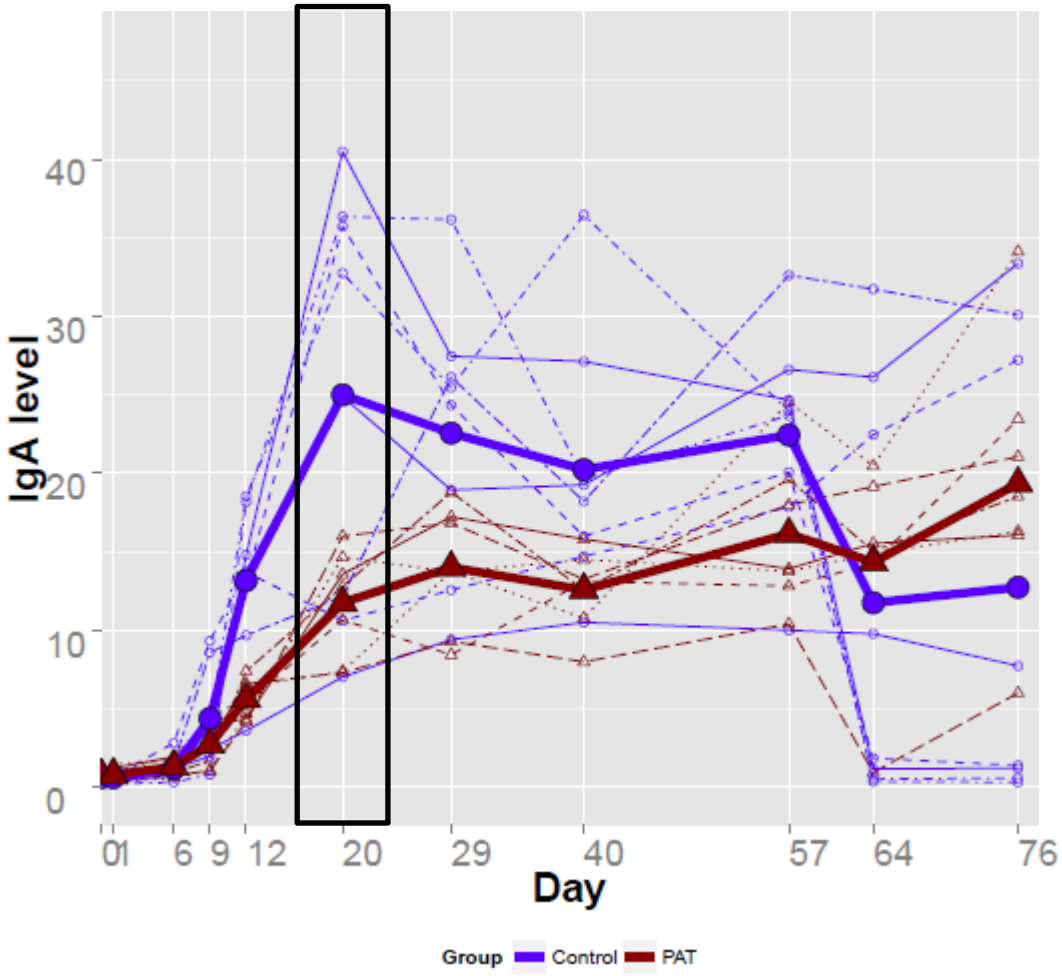
Biological Response: IgA

- *Post-transplant:* Development of the immune system as microbiome colonizes the gut.
- Low IgA Level could mean the immune system is under stress or compromised.



Biological Response: IgA

- *Post-transplant:* Development of the immune system as microbiome colonizes the gut.
- Low IgA Level could mean the immune system is under stress or compromised.
- Day 20: largest observed difference (weakened host immunity for PAT-purturbed microbiota).



Aim: Discover OTUs associated with IgA at day 20.

Data

- transPAT study.
- Data available on 355 OTUs and IgA level for 15 subjects.
- 2 treatment groups – Control (8), PAT (7).
- 6 available common timepoints between IgA and OTU:
  Day 1, Day 6, Day 12, Day 20, Day 64 and Day 76(14 subjects).
- Analysis based on initial 4 timepoints.

$$
\underbrace{\left(\begin{array}{cccc|cccc}
x_{11} & x_{12} & \cdots & x_{18} & x_{19} & x_{1,10} & \cdots & x_{1,15} \\
x_{21} & x_{22} & \cdots & x_{28} & x_{29} & x_{2,10} & \cdots & x_{2,15} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
x_{j1} & x_{j2} & \cdots & x_{j8} & x_{j9} & x_{j,10} & \cdots & x_{j,15} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
x_{355,1} & x_{355,2} & \cdots & x_{355,8} & x_{355,9} & x_{355,10} & \cdots & x_{355,15}
\end{array}\right)}_{\text{CONTROL}} \quad \underbrace{\qquad\qquad}_{\text{PAT}}
\qquad m = 355\,\text{OTUs}
$$

Biomarker Setting: Data Structure

Per timepoint:

X = Microbiome Data

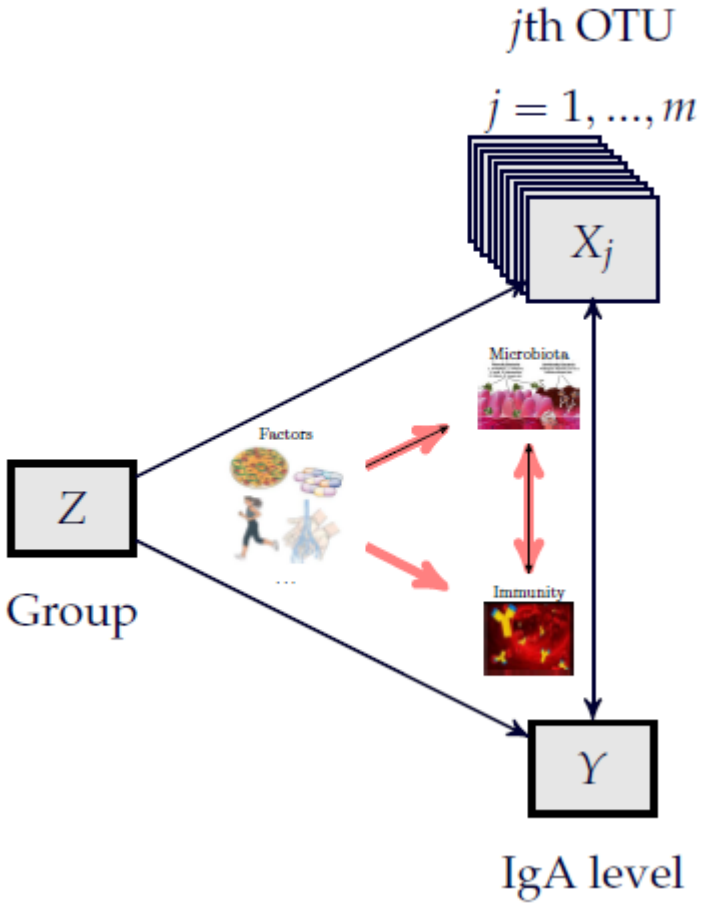
$$\text{m OTUs} \left\{ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{j1} & x_{j2} & \cdots & x_{jn} \\ \cdot & \cdot & \cdot & \cdot \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \right.$$

Y = IgA Level

$y_1, y_2, \ldots, y_n$

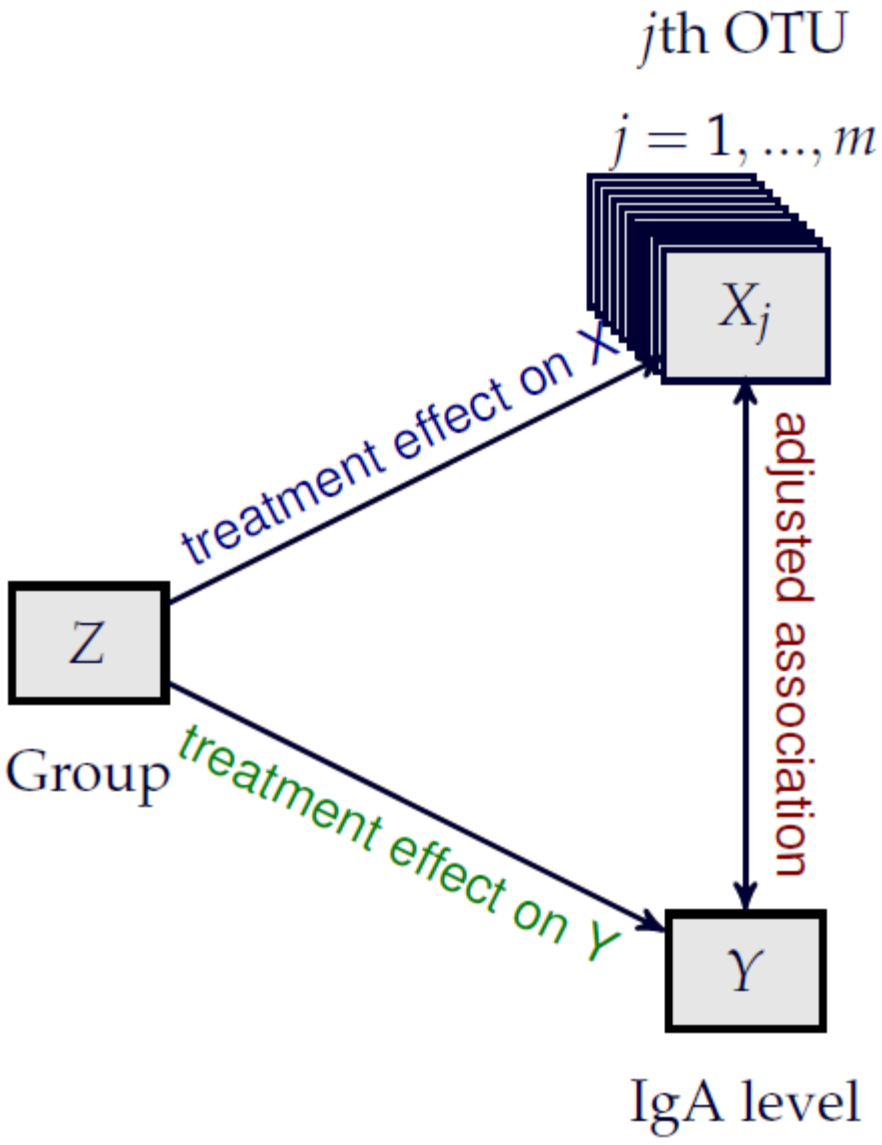
Z = Treatment Vector (Binary)

$z_1, z_2, \ldots, z_n$



OTU-level Biomarker Setting

- IgA level at day 20 is significantly different between the treatment groups.
- Which of the OTUs are differentially abundant ?
- Which of the OTUs are significantly associated with the IgA level after taking treatment effect into account ?
- What is the source of the correlation ?



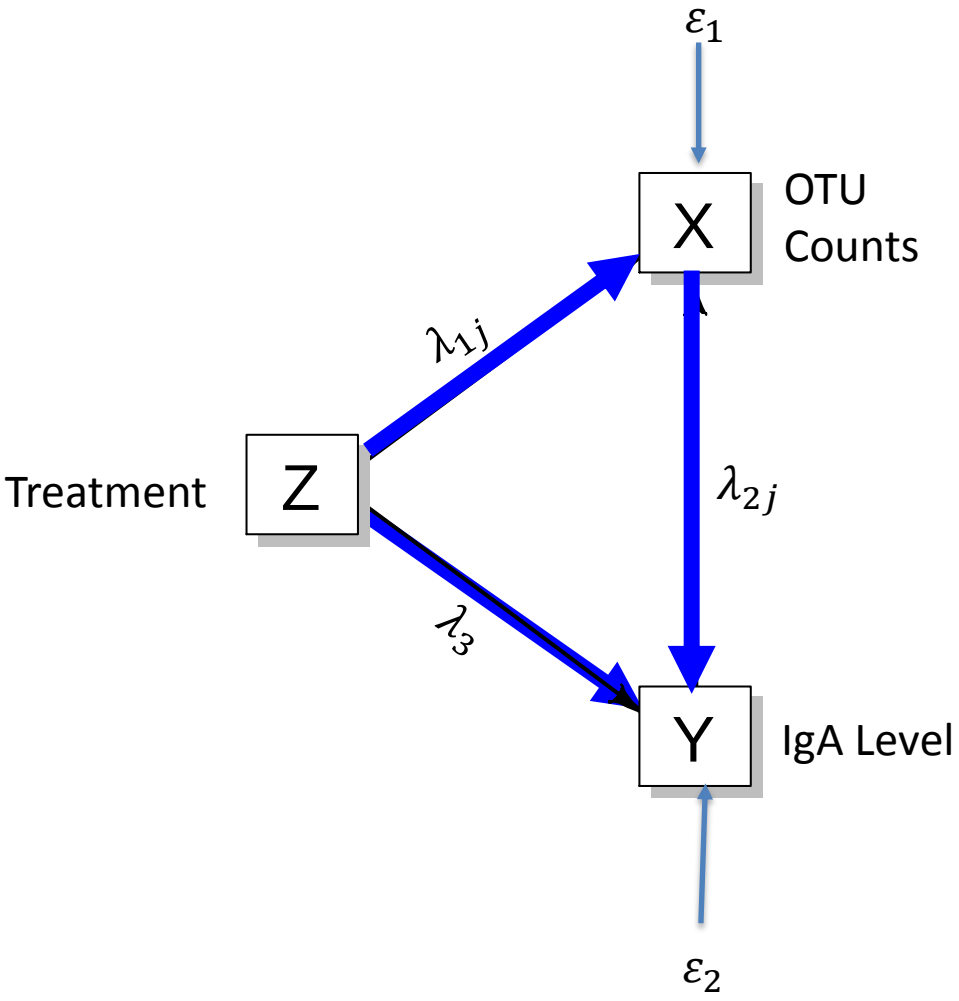
Part 2:
Structural Equation Model

SEM

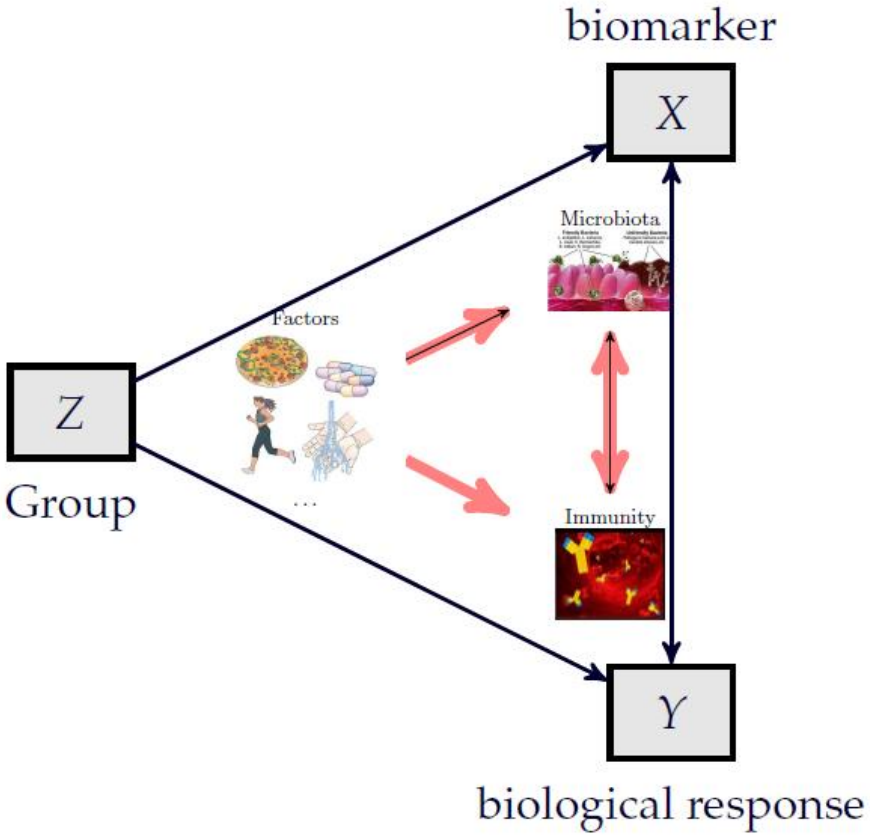
- Dependent variable= endogenous variable.
- Independent variable= exogenous variable.
- An endogenous variable in one equation can be exogenous in another equation.
- Instead of minimizing functions of observed and predicted individual values, difference between the sample covariances and the predicted covariances are minimized in SEM.
- The treatment total effect is split into a direct and an indirect effect components.

SEM: Simple Model

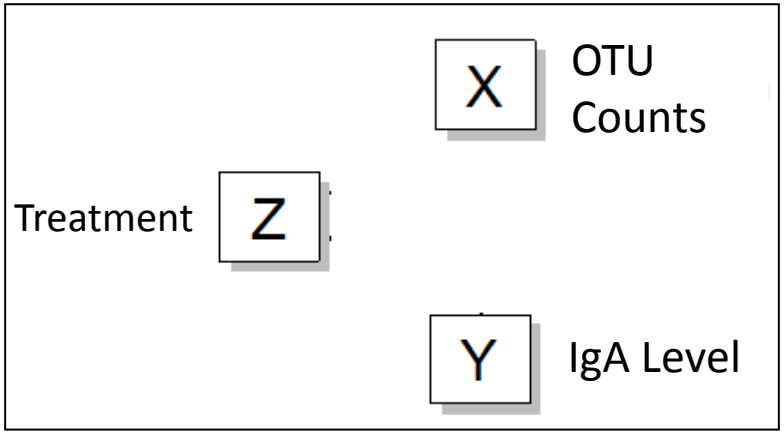
PATH DIAGRAM



Joint Model

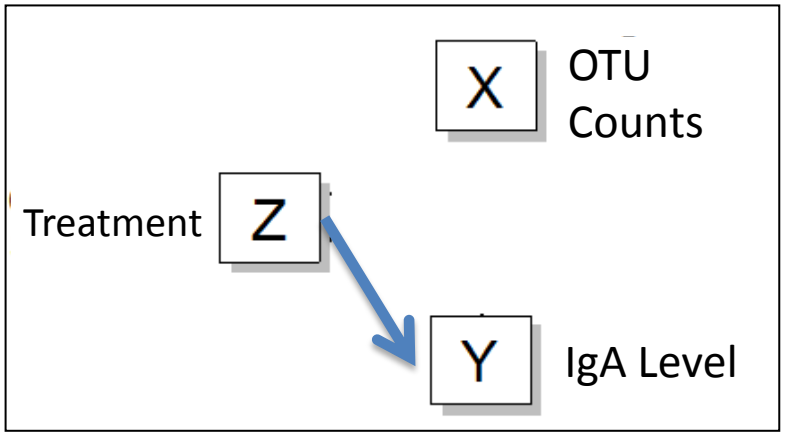


Potential Path Models to Explain the Relation



Independent model (model 1)

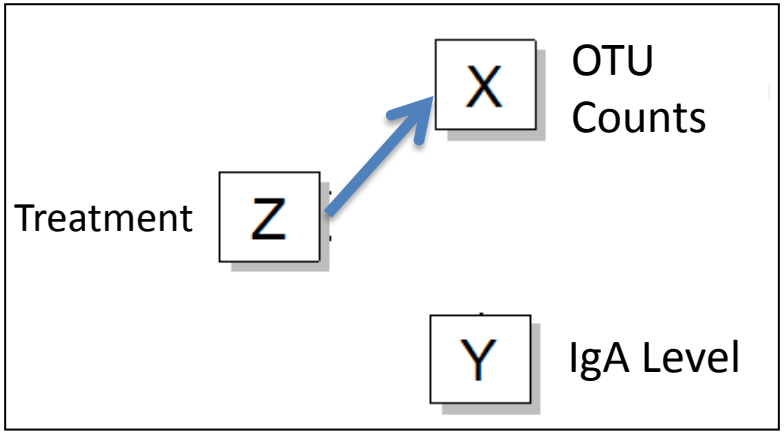
- No relation at all



Single effect model (model 2)

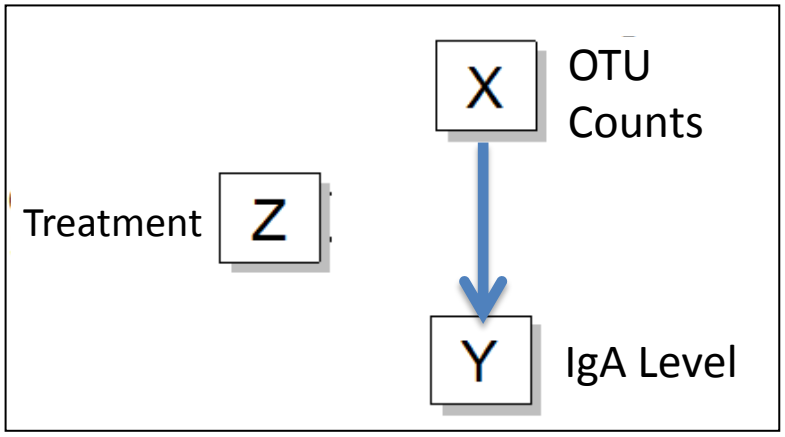
- Direct effect between Z and Y

Potential Path Models to Explain the Relation



Single effect model (model 3)

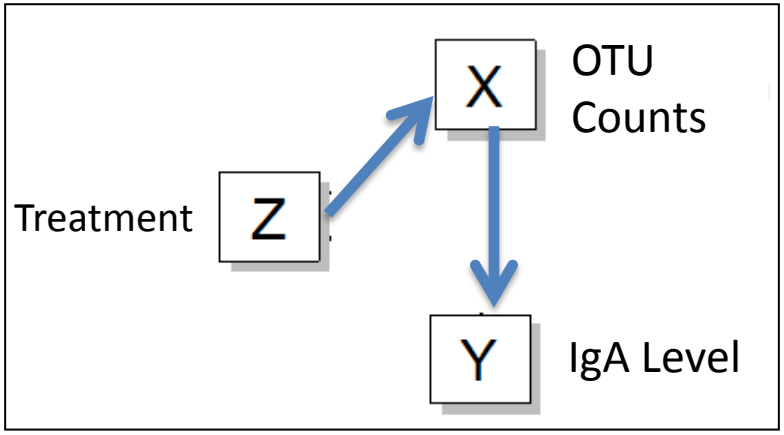
- Direct effect between Z and X



Single effect model (model 4)

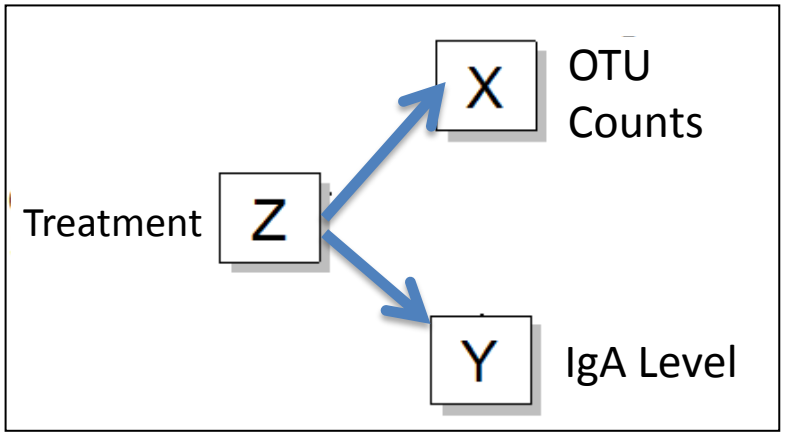
- Direct effect between X and Y

Potential Path Models to Explain the Relation



Indirect effect model (model 5)

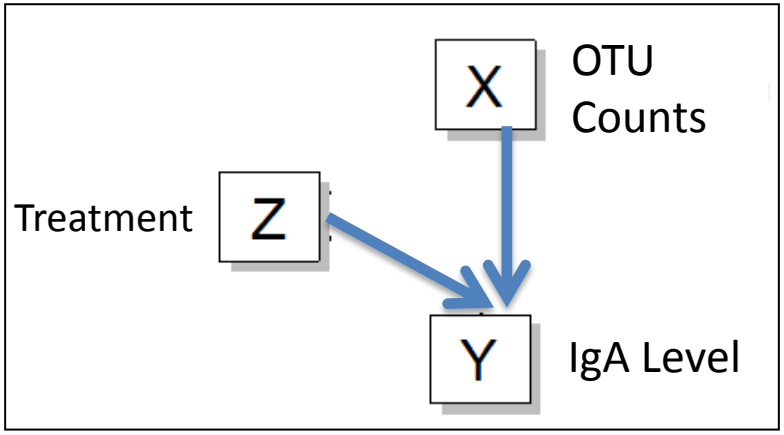
- Indirect effect between Z and Y (complete mediation)



common effect model (model 6)

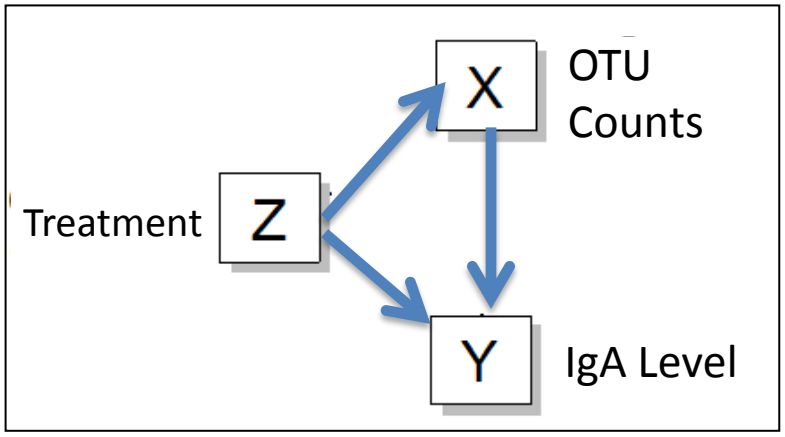
- Z affects X as well as Y

Potential Path Models to Explain the Relation



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

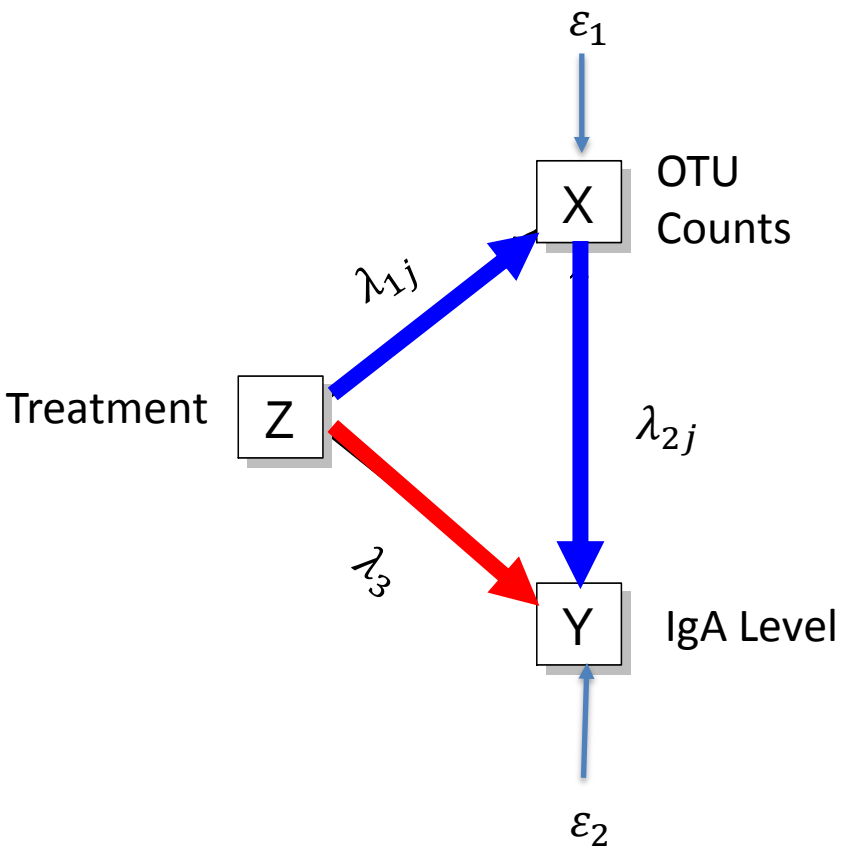
Partial Mediation Model Formulation

The structural equations model corresponding to the partial mediation model is given by

$$X_{ij} = \lambda_{1j} Z_i + \varepsilon_{1i},$$

$$Y_i = \lambda_3 Z_i + \lambda_{2j} X_{ij} + \varepsilon_{2i}.$$

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim N \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \varphi$$



Direct and Indirect Effects

- **Direct effect:** influence of Treatment on the IgA Level that is unmediated with by the OTU Counts

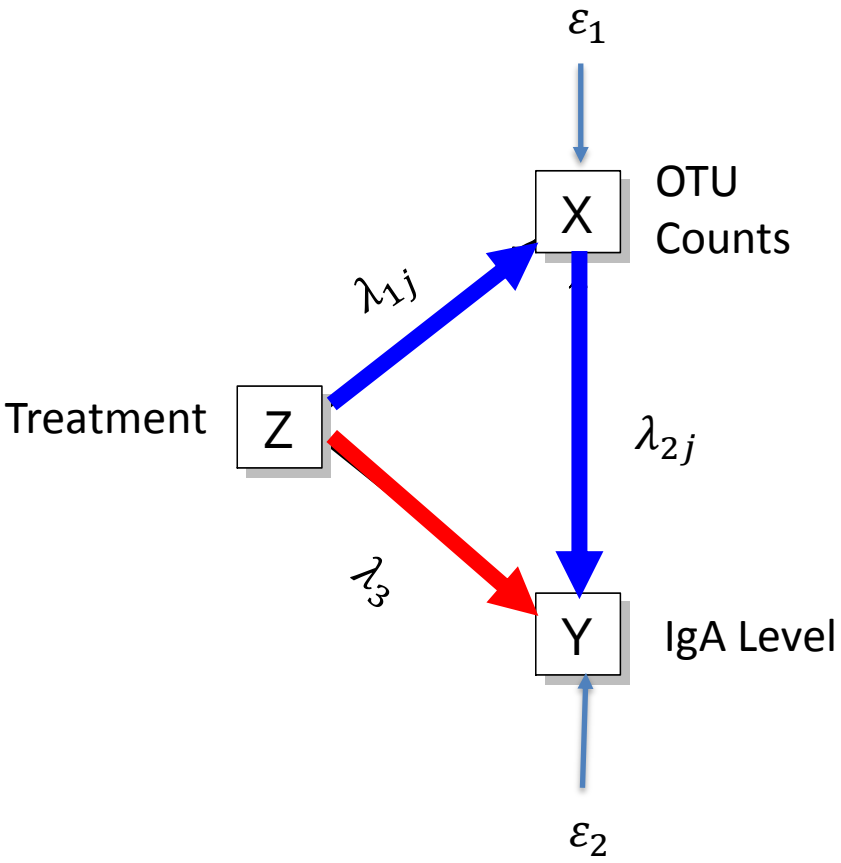
$$\lambda_3$$

- **Indirect effect**: the effect of the Treatment is mediated by the OTU Counts

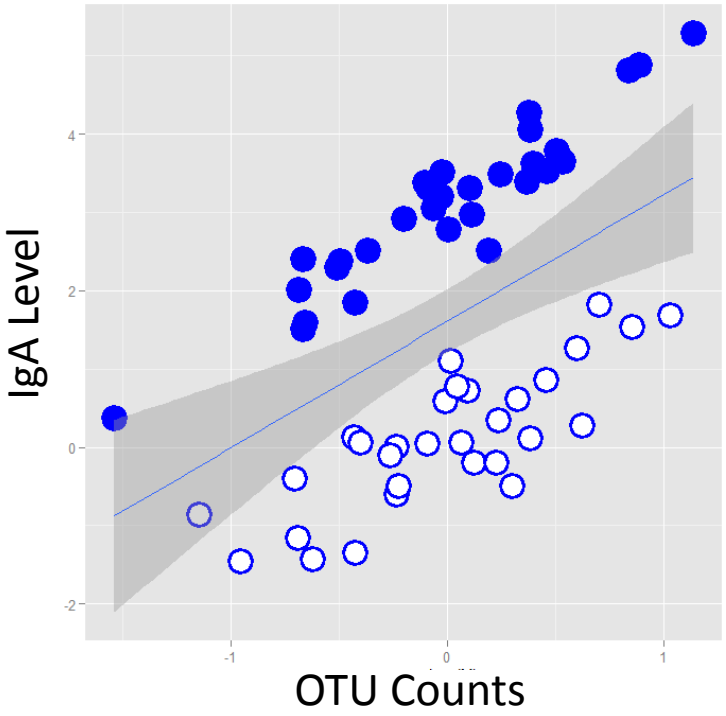
$$\lambda_{1j} * \lambda_{2j}$$

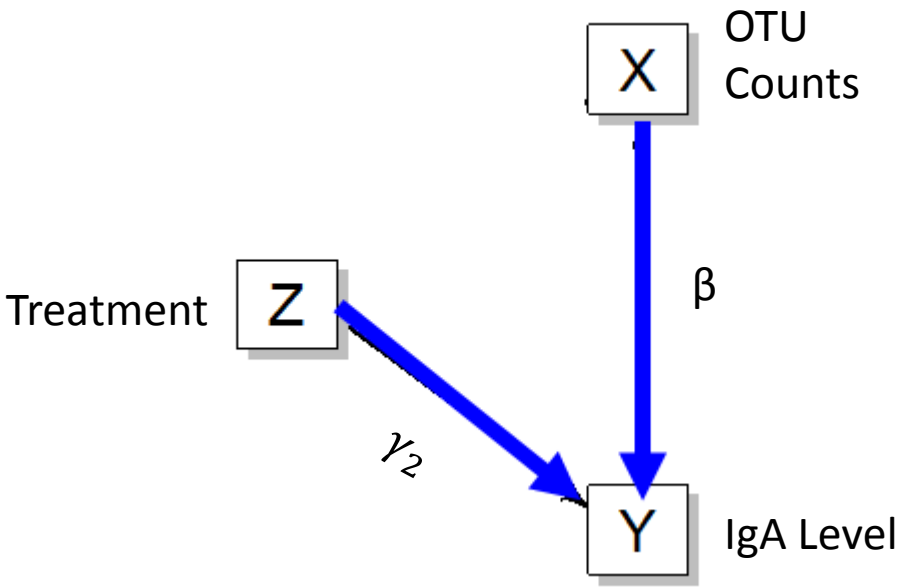
- **Total effect**: sum of the direct and indirect effects.

$$\lambda_{1j} \times \lambda_{2j} + \lambda_3$$



Direct Effect



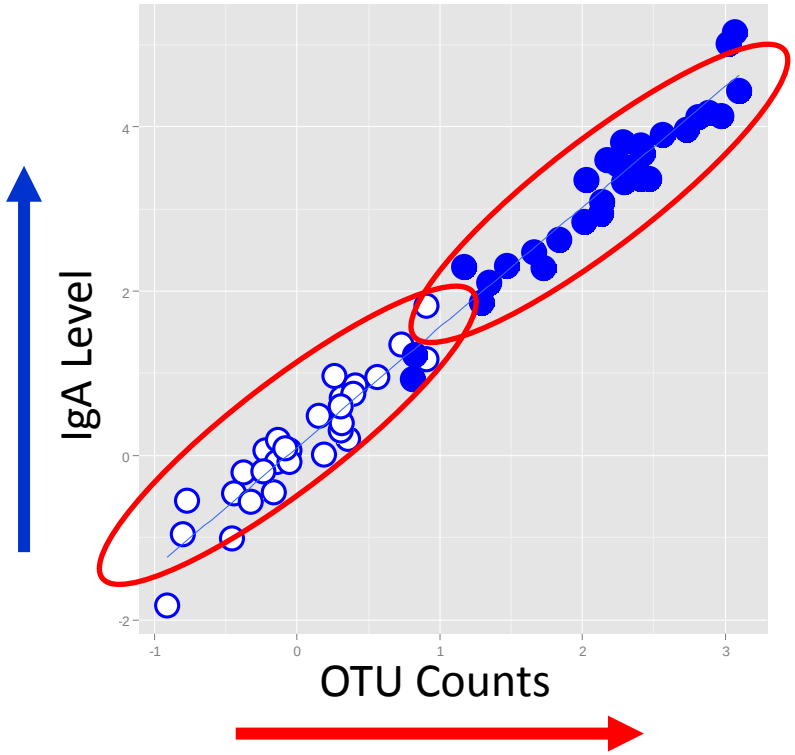


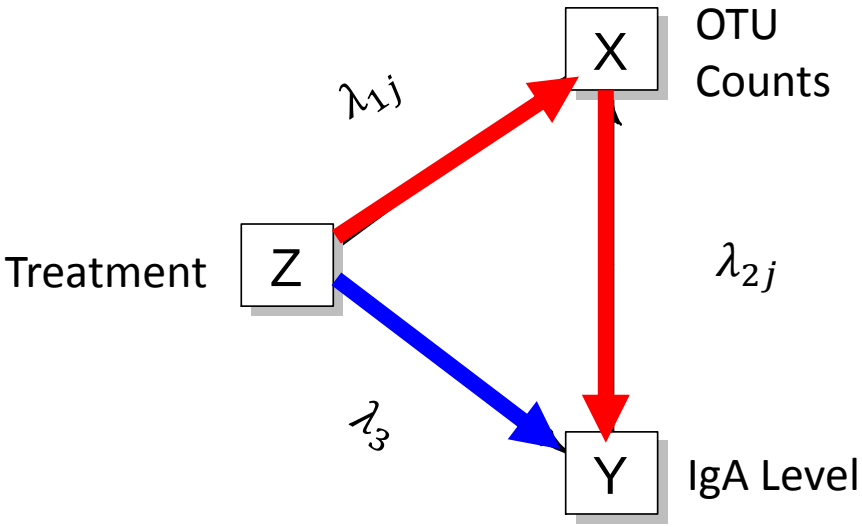
Direct effect of Treatment on IgA γ_2

Direct effect of OTU on IgA β

IgA is a collider, X and Z mutually cause the effects on Y.

Indirect Effect



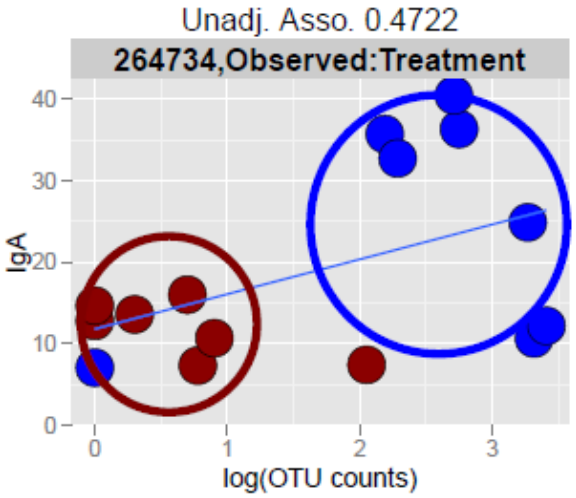


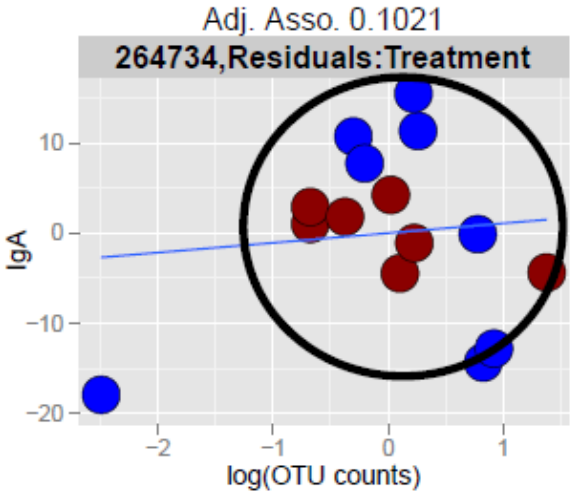
Direct effect+ Indirect effect

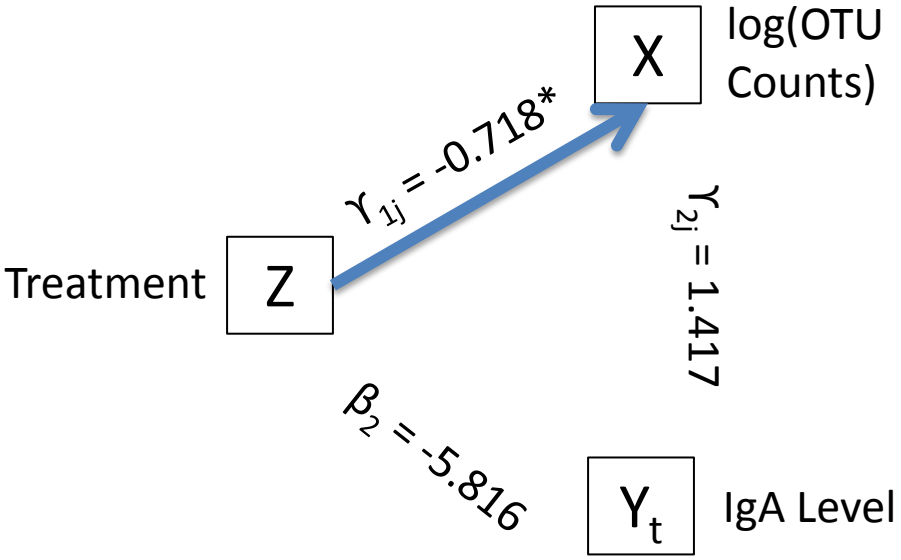
Z and X on Y Z via X on Y

Results: Partial Mediation Model

OTU 264734



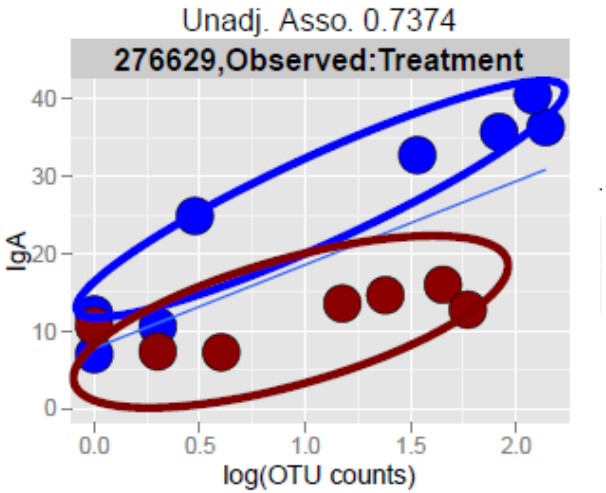


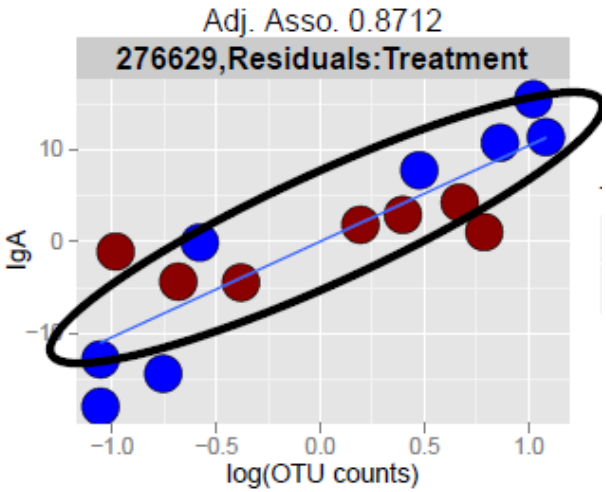


implies model 3 should be fitted.

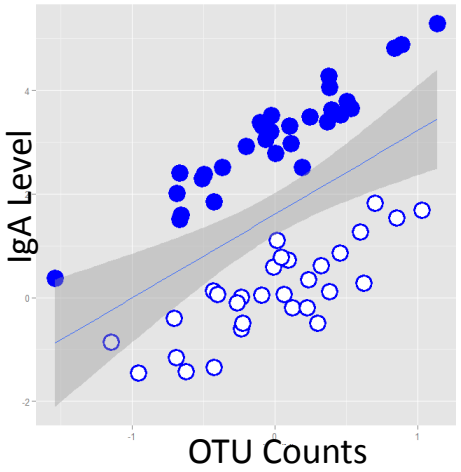
Results: Partial Mediation Model

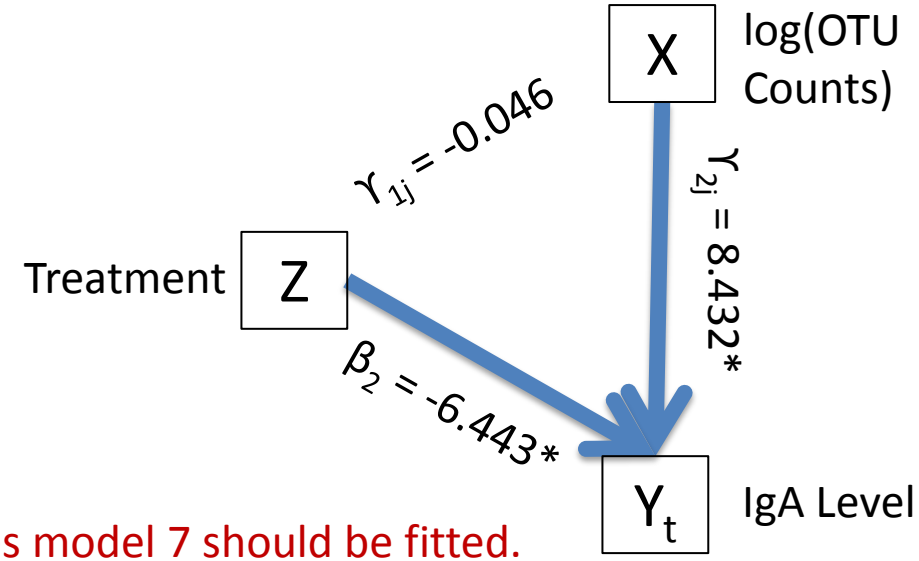
OTU 276629





Direct Effects

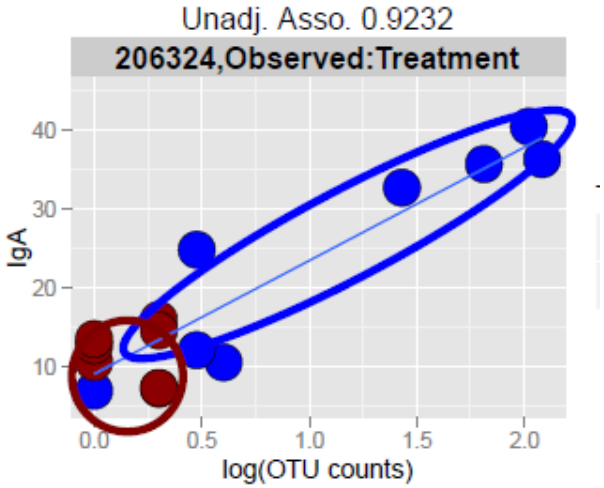


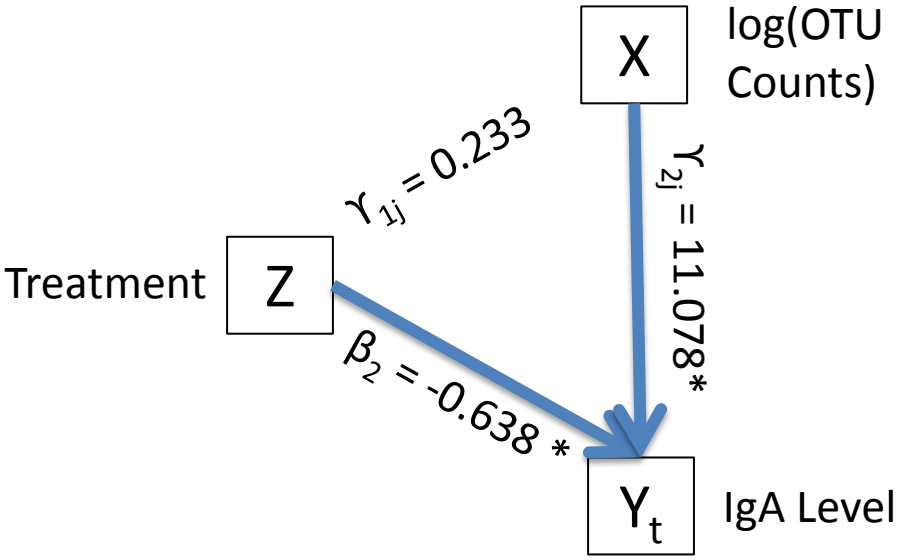


implies model 7 should be fitted.

Results: Partial Mediation Model

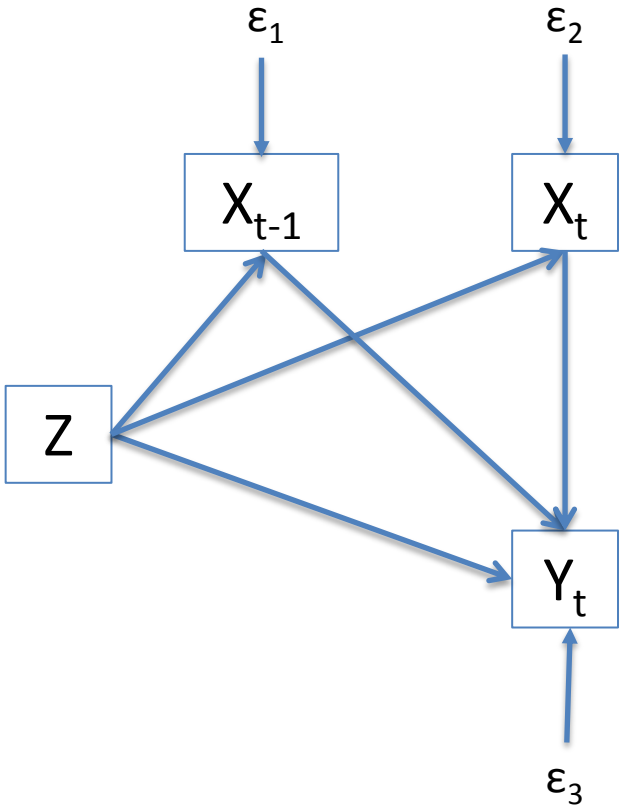
OTU 206324

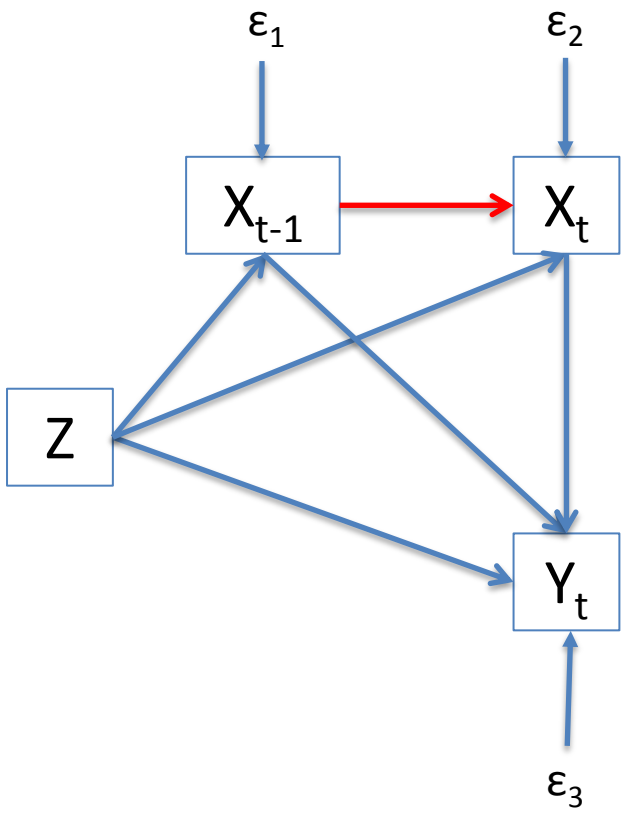




implies model 7 should be fitted.

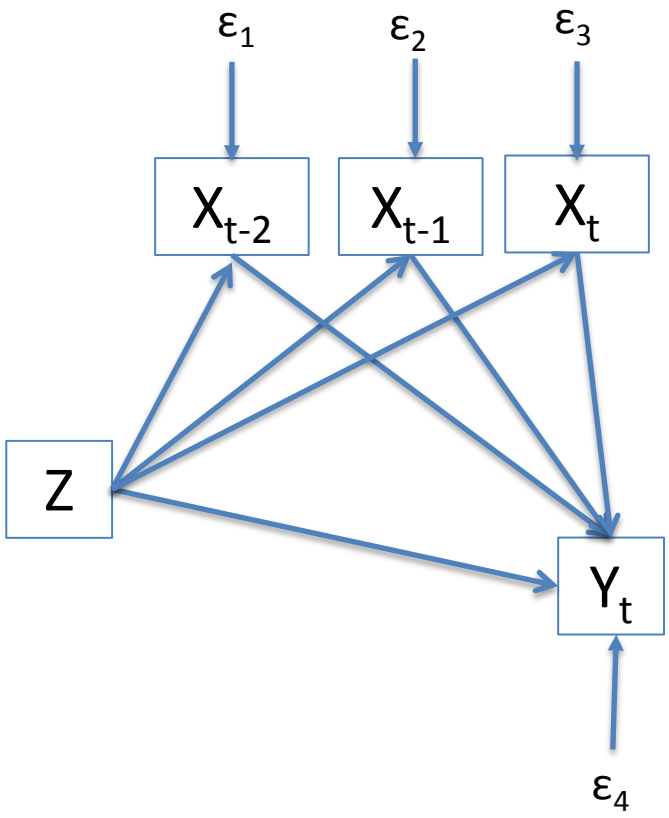
SEM: Complex Models – with 1 lag

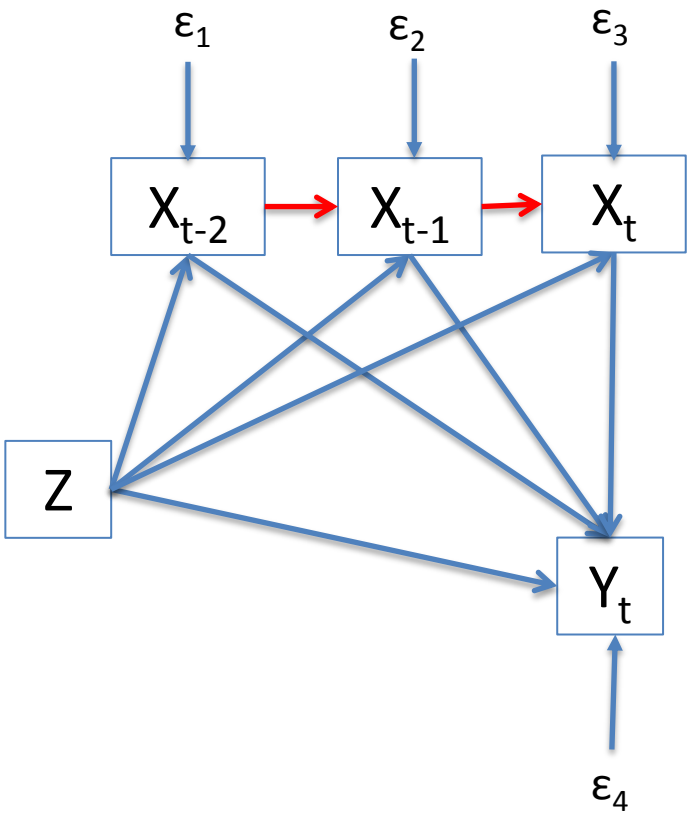




Independent Errors

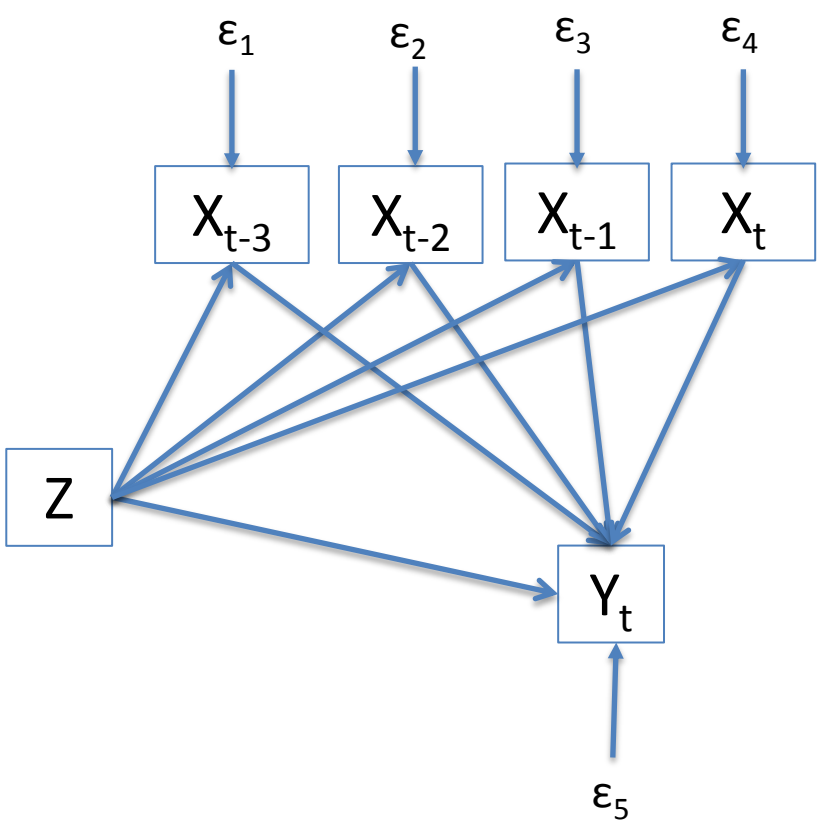
SEM: Complex Models – with 2 lags

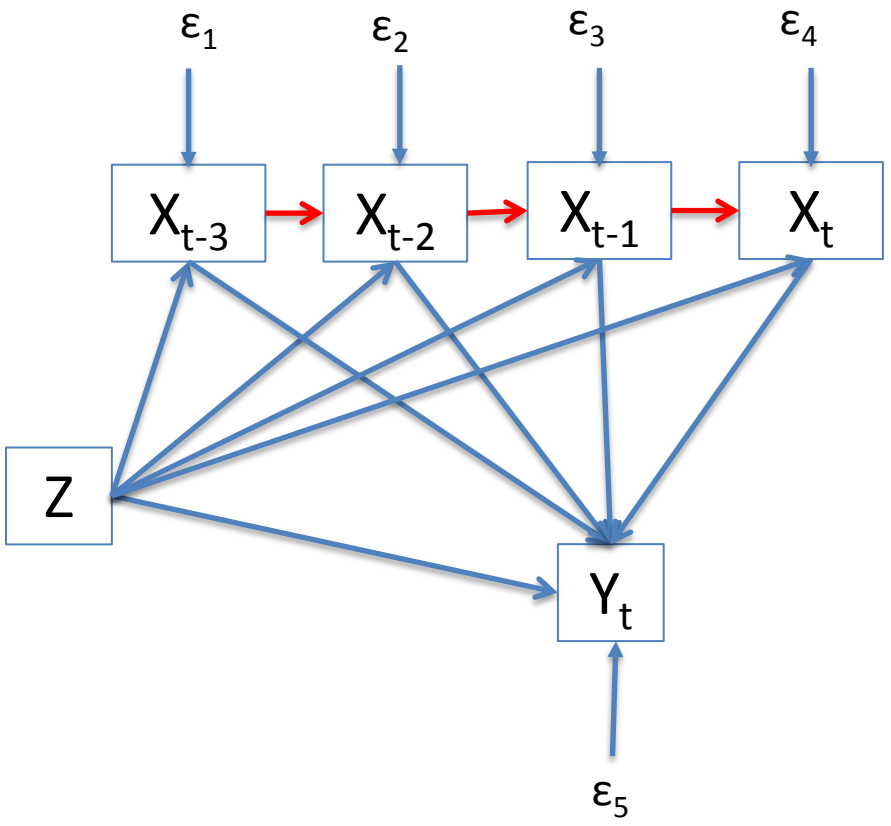




Independent Errors

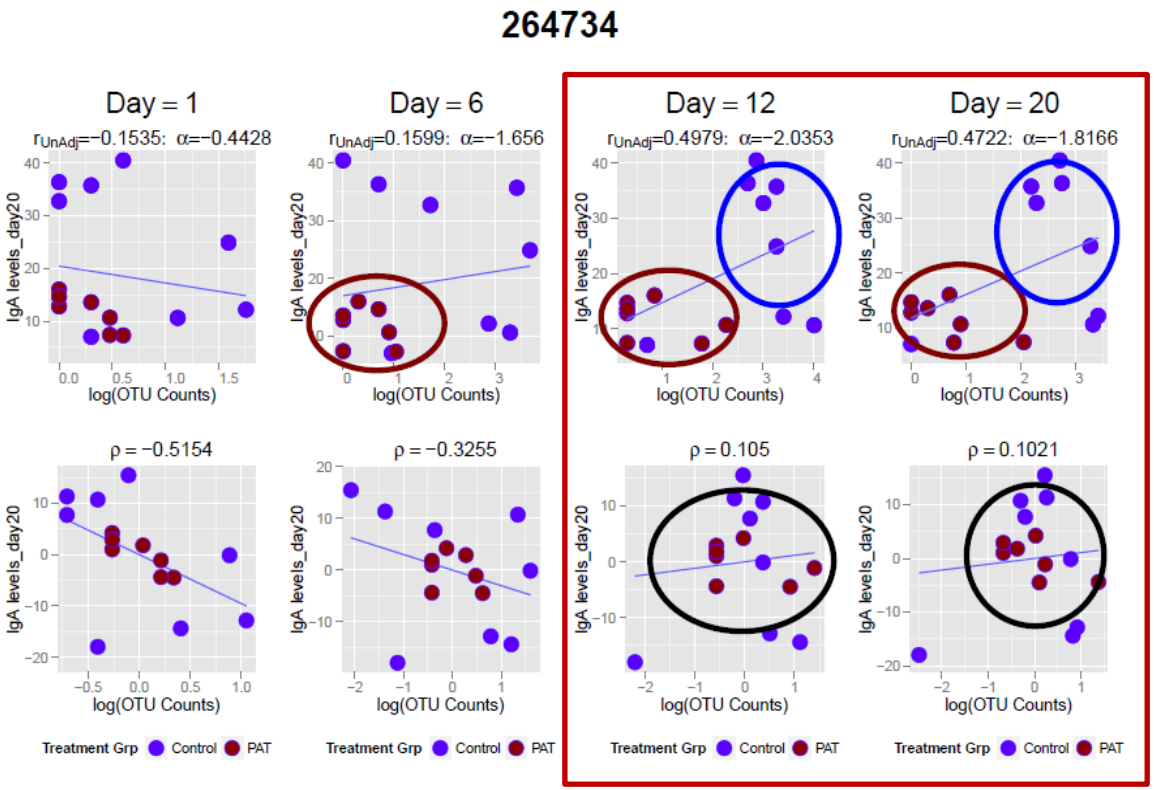
SEM: Complex Models – with 3 lags



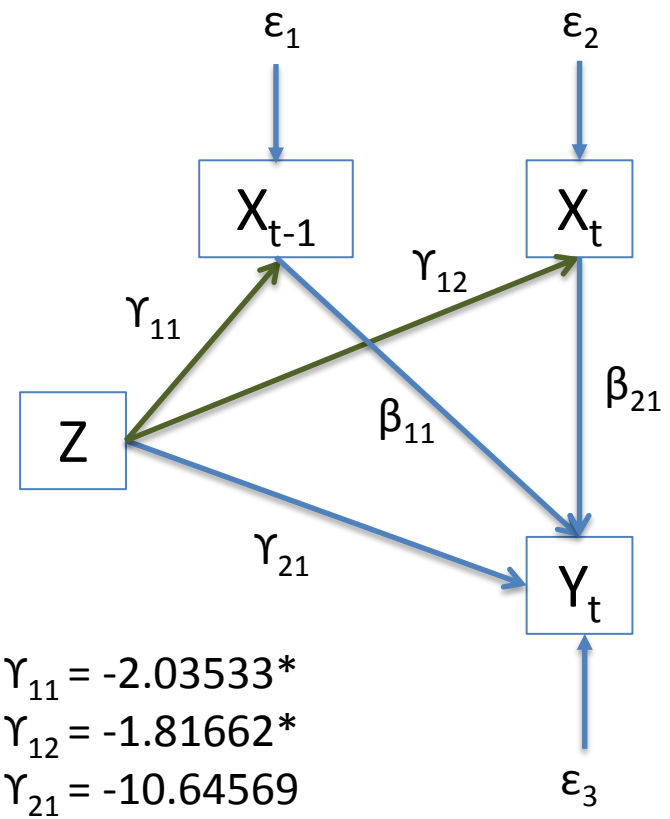


Independent Errors

Results: lag 1

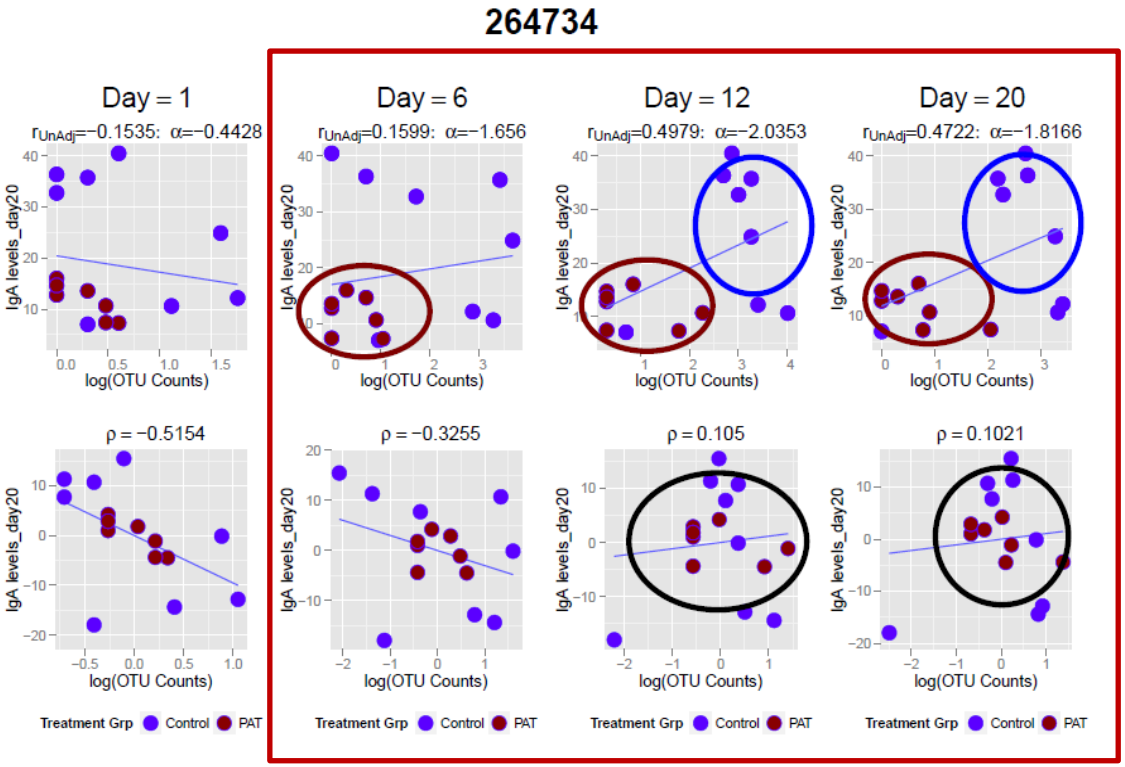


Independent Errors

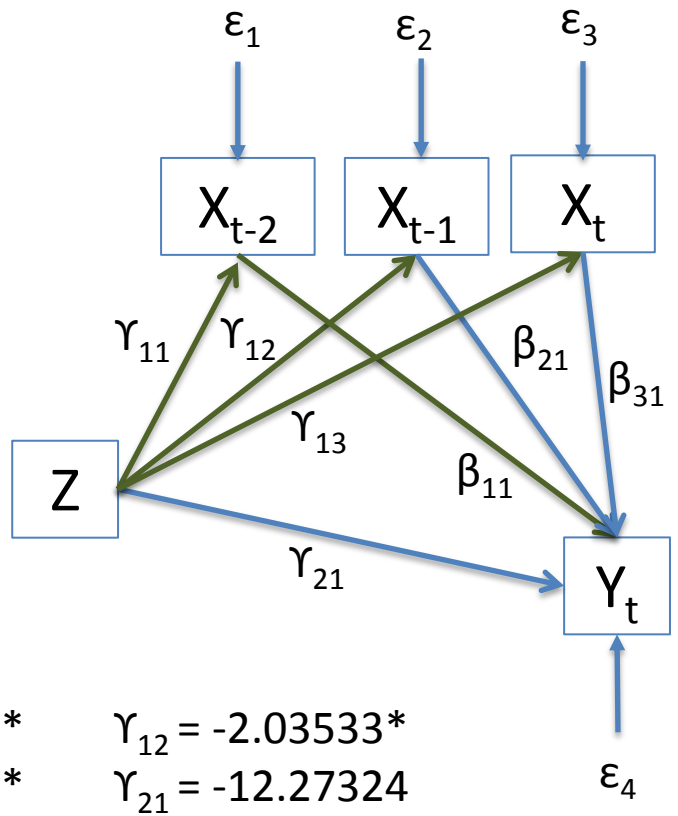


γ_{11} = -2.03533*
γ_{12} = -1.81662*
γ_{21} = -10.64569
β_{11} = 0.72849
β_{21} = 0.60755

Results: lag 2



Independent Errors



Υ_{11} = -1.65599* Υ_{12} = -2.03533*

Υ_{13} = -1.81662* Υ_{21} = -12.27324

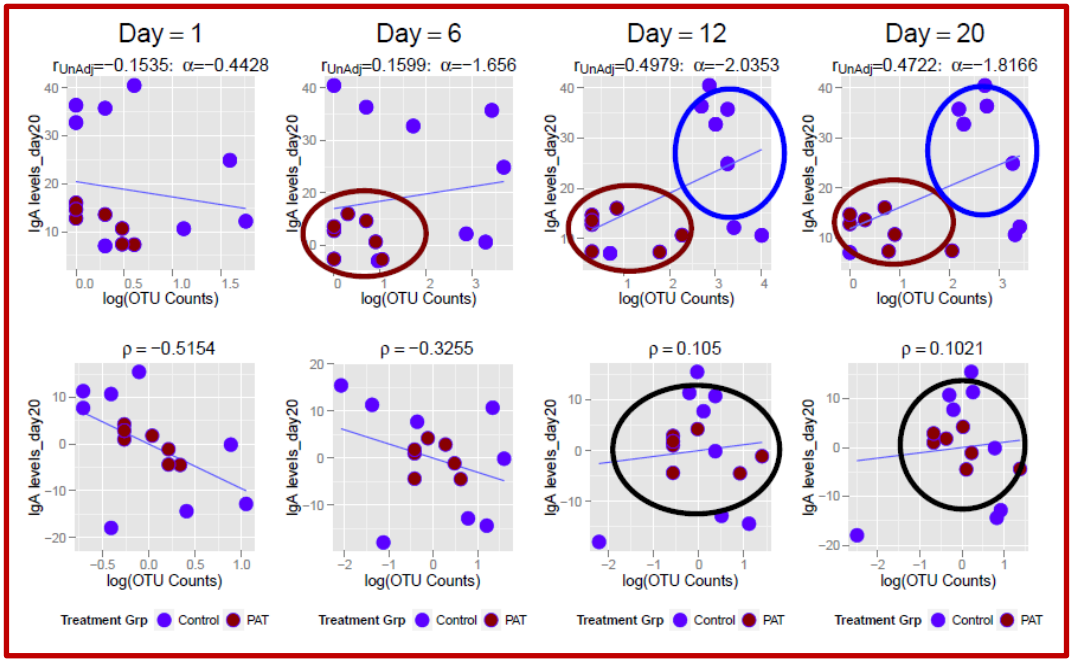
β_{11} = -5.36971*

β_{21} = 4.87526 (sig at 0.1)

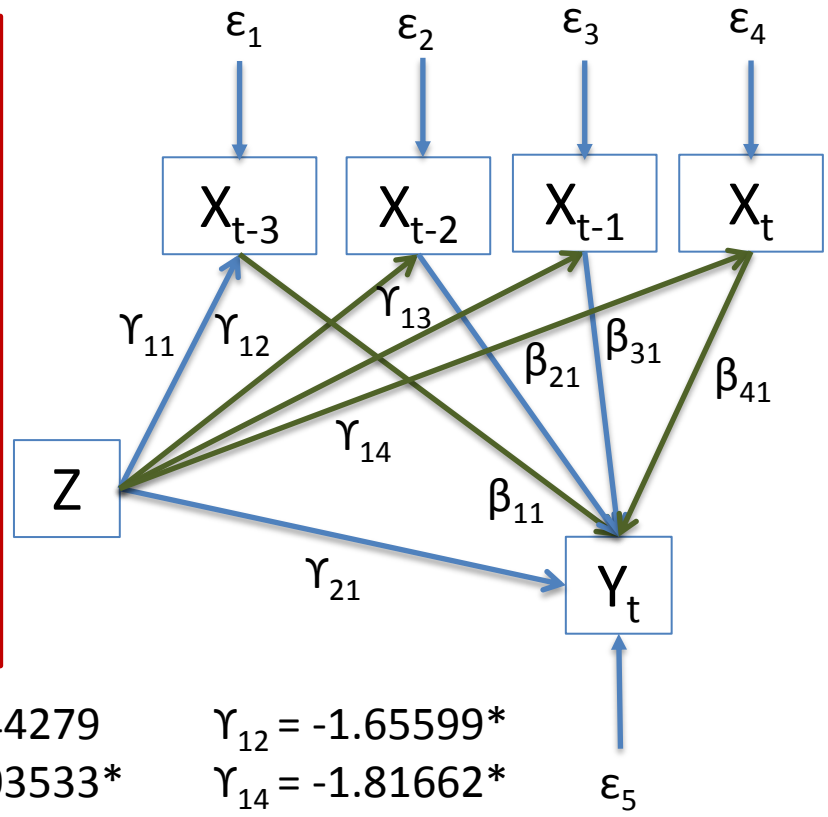
β_{31} = -0.03947

Results: lag 3

264734



Independent Errors



Υ_{11} = -0.44279 Υ_{12} = -1.65599*

Υ_{13} = -2.03533* Υ_{14} = -1.81662*

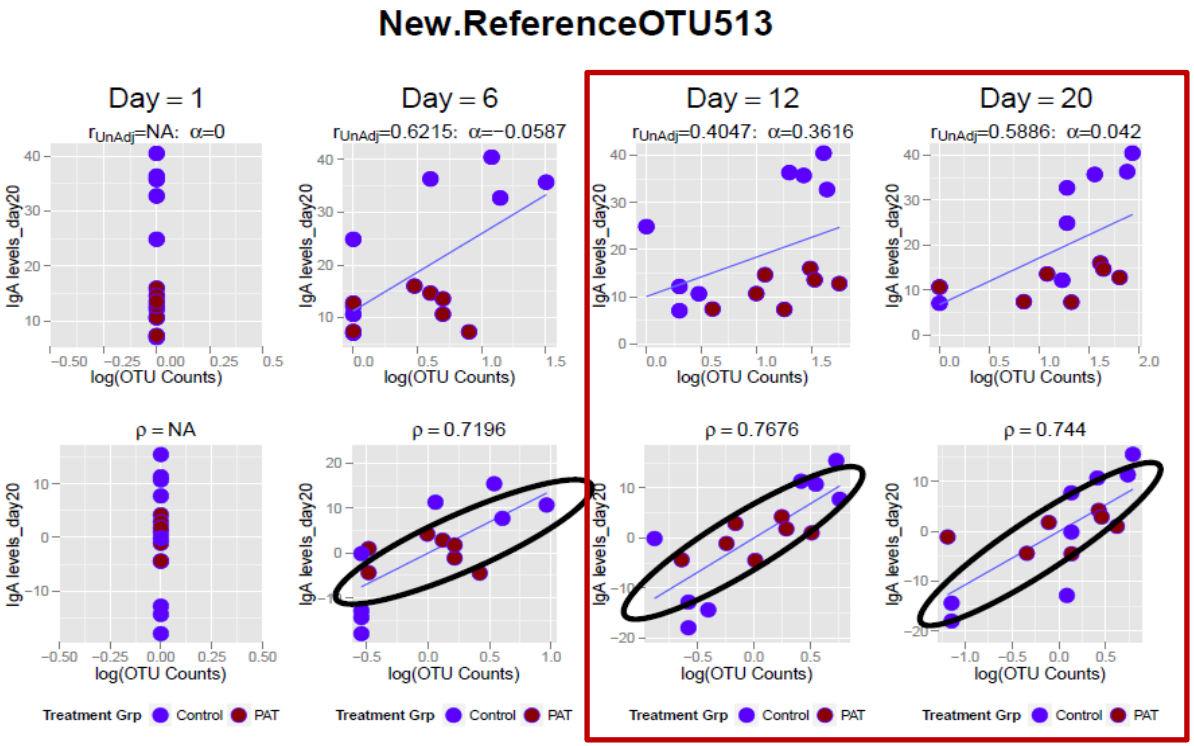
Υ_{21} = -8.14798

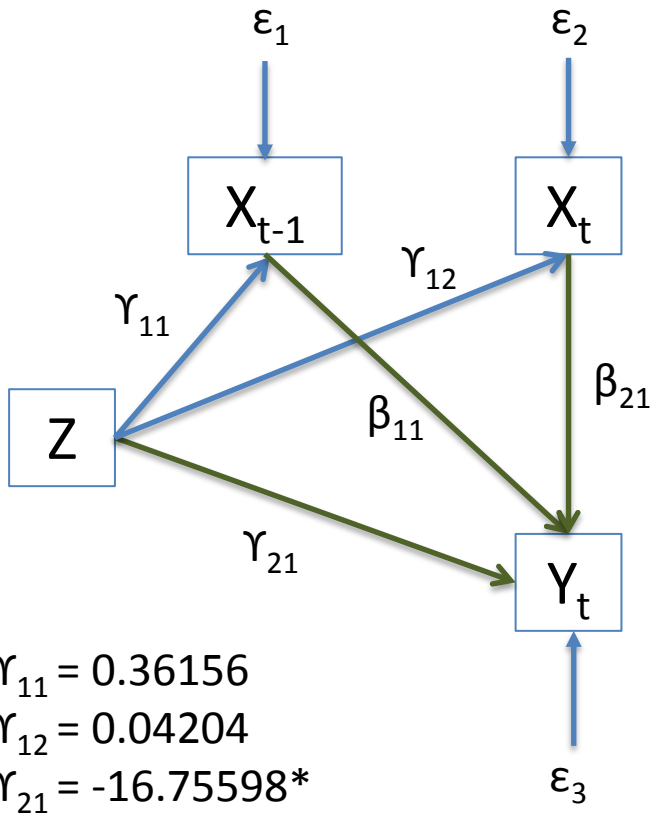
β_{11} = -14.28852* β_{21} = -1.71524

β_{31} = 3.12032 β_{41} = 4.34901*

Results: lag 1

Independent Errors, t =20





$\gamma_{11} = 0.36156$

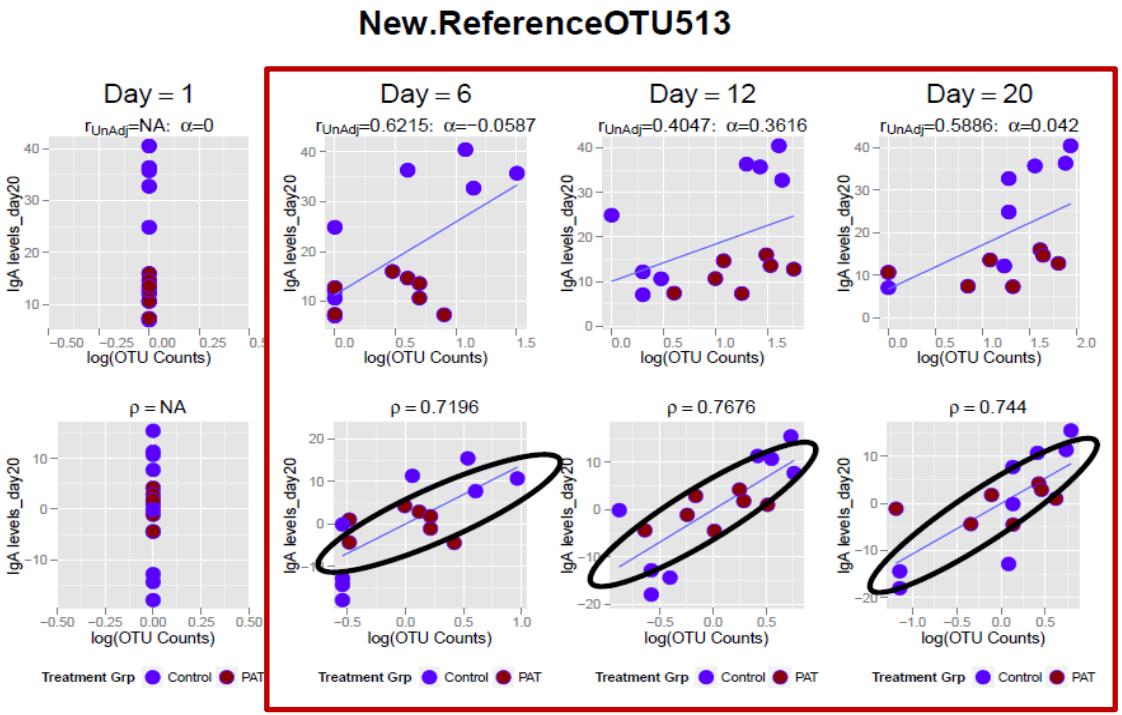
$\gamma_{12} = 0.04204$

$\gamma_{21} = -16.75598^*$

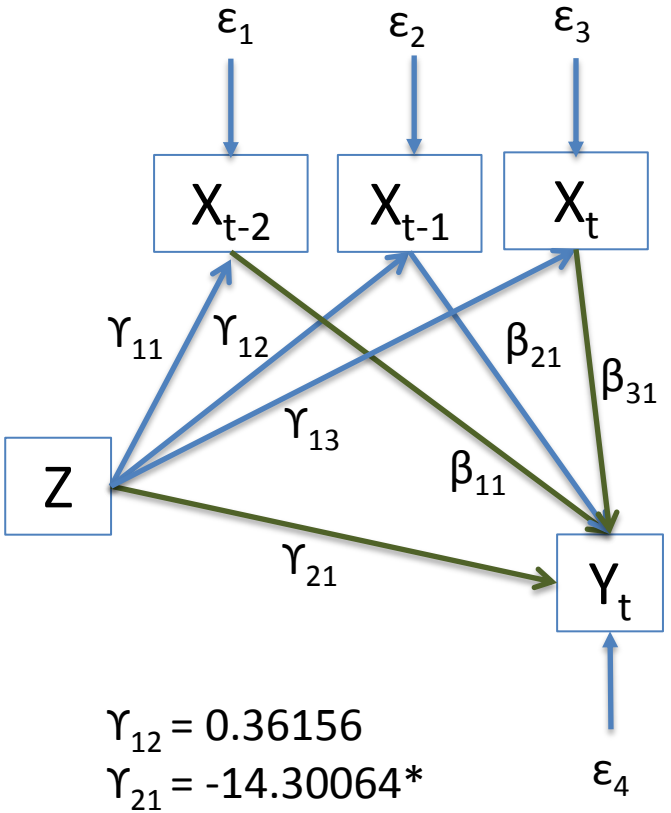
$\beta_{11} = 8.99858^*$

$\beta_{21} = 6.43147^*$

Results: lag 2



Independent Errors



Υ_{11} = -0.05868
Υ_{13} = 0.04204
β_{11} = 7.70733*
β_{21} = 3.39927
β_{31} = 6.93994*

Υ_{12} = 0.36156
Υ_{21} = -14.30064*

Conclusions

- The SEM approach allows to chose different types of OTUs depending on decomposition of the total effect.
- It allows to account for multiple sources of data.
- Computational issues related the (high) dimension of the data or complexity of the model.

- Explore more in this direction.
- Development of models that use more than one OTU.

Thank You for Your Attention..!!