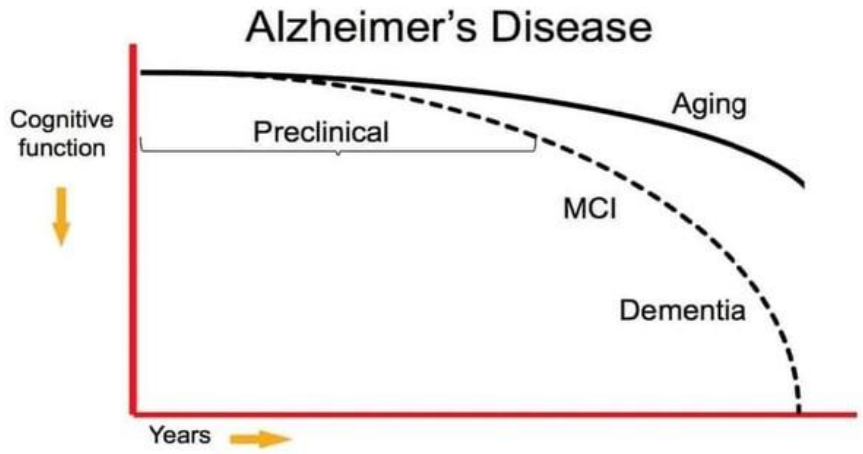
Accelerating Alzheimer's Research: a modular framework for exploring Bayesian disease progression models

Oana Petrof, Dave Lunn, Aris Perperoglou

Disease age modeling in Alzheimer's disease (AD)

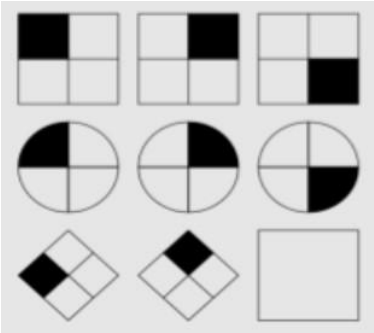
- Alzheimer's disease is a neurological disease causing progressive cognitive impairment
- Onset disease age often unknown
- Long delay between early signs and severe disease progression
- AD is characterized by a monotonic decline
- Non-linear disease trajectory



- https://assignmentpoint.com/understanding-preclinical-alzheimers-disease-alterations/

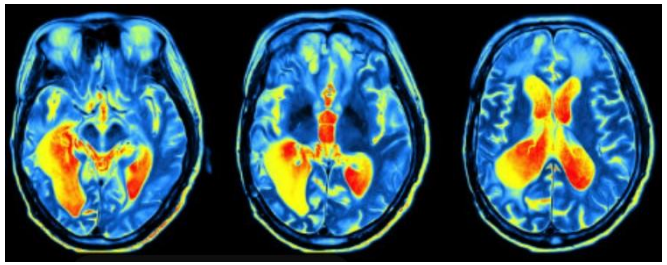
COGNITIVE TESTS

ADAS-COG 13
Alzheimer Disease Assessment Scale-Cognitive



ADNI
ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

PET



- tau PET scans
  - p-tau 181
  - p-tau 217
  - total tau (t-tau)
- amyloid PET scans
  - Ab42
  - Ab40
- molecular biomarkers
  - NfL (Neurofilament Light Chain)
  - GFAP (Glial Fibrillary Acidic Protein)

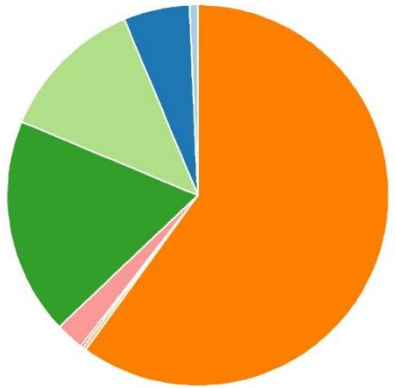
DEMOGRAPHICS

Gender

Age

Years of education

BMI



GENETIC INFORMATION

APOE **genotyping**

GSK

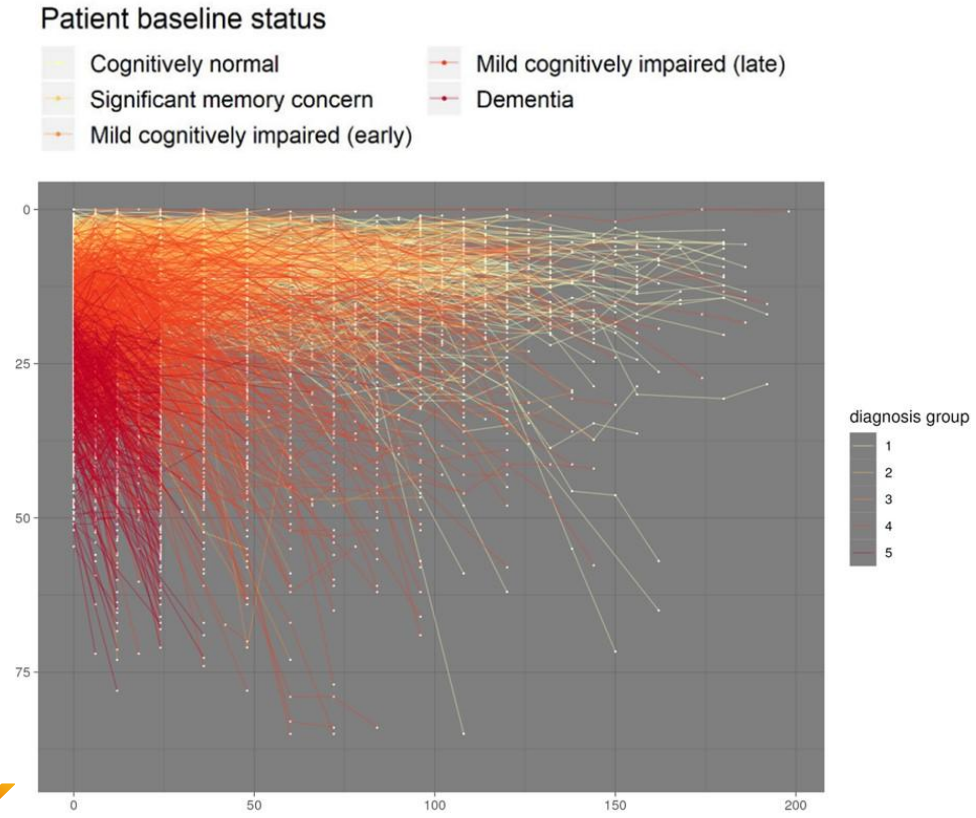
Purpose of this analysis

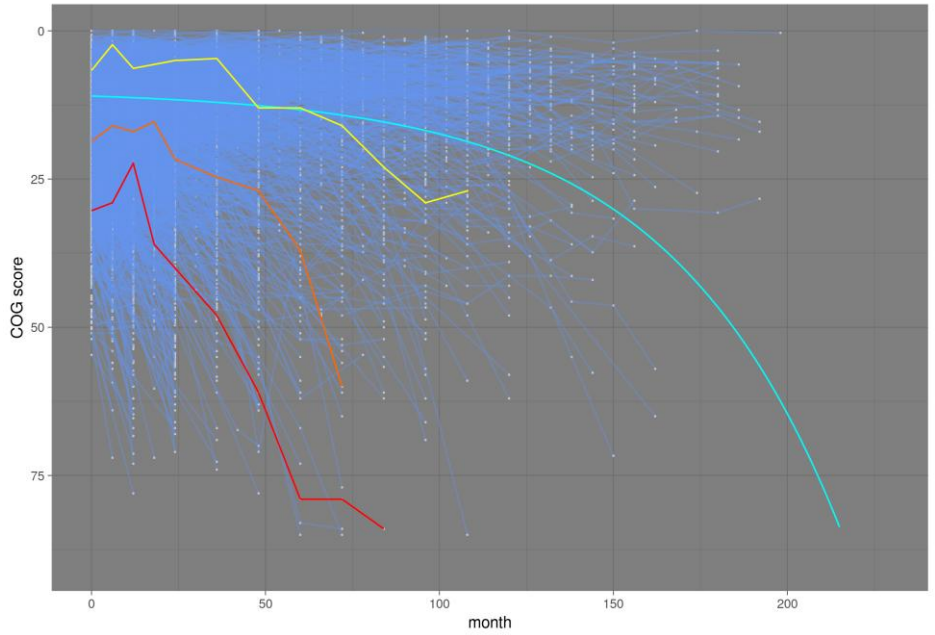
Main Objective: Develop a joint model incorporating key Alzheimer's disease endpoints: ADAS-Cog 13, MMSE, and CDR-Sum of Boxes to identify and quantify demographic and biomarker predictors of Alzheimer's disease progression.

Today's Presentation Focus: Demonstrate a flexible way to break the model into two stages, to facilitate covariate exploration and later the joint modelling of endpoints.

Endpoint of interest

- ADAS-cog 13 (0 - 85-point score) as measure of disease progression
- Higher scores indicate greater cognitive dysfunction
- Continuous endpoint
- The patients are observed at different points along that trajectory
- Align them on a common disease trajectory

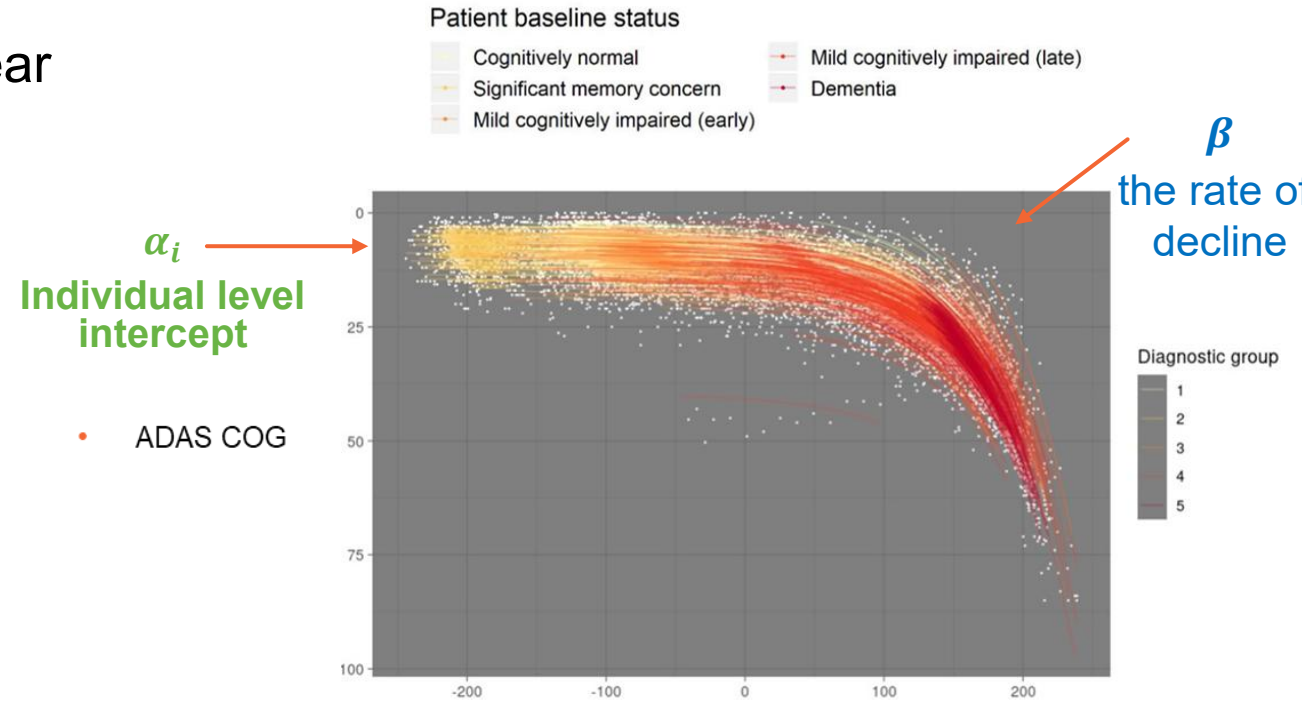




Disease age model

- Consider the Bayesian hierarchical non-linear mixed-effects disease age model
- i indexes “patients” ,
- j indexes “visits”,
- g indexes the “disease group”

$$y_{ij} \sim N\left(\mu_{ij}, \sigma_y^2\right)$$
$$\mu_{ij} = \alpha_i + e^{(\beta(t_{ij}+\theta_i))}$$
$$\theta_i \sim N\left(\gamma_{g(i)}, \sigma_{g(i)}^2\right)$$

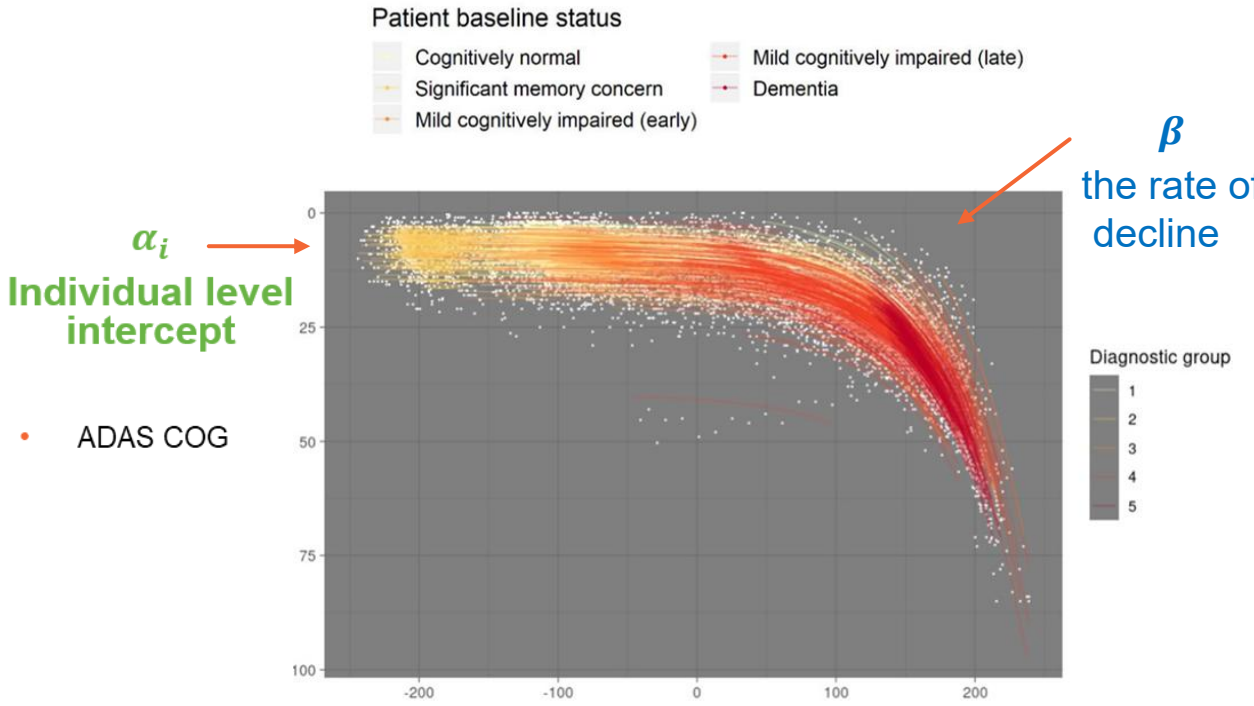


- Each patient belongs to group $g(i)$ according to his/her cognitive status at baseline.
- Introducing patient-level “time-shifts” to align patients to a common population disease trajectory

- Raket, L. L. (2020). Statistical disease progression modeling in Alzheimer disease. *Frontiers in big Data*, *3*, 24.

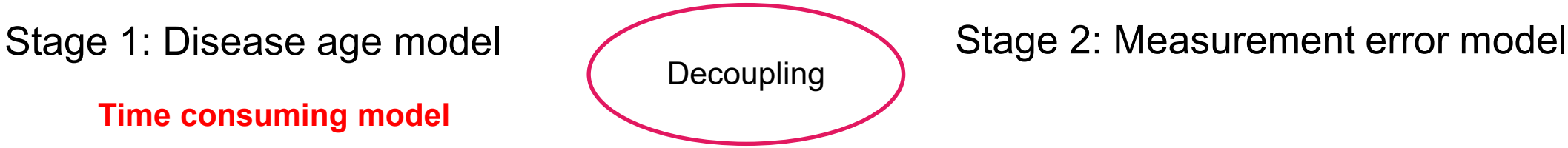
Disease age model

- **Long computation time (~ 6 hours)**
- **Difficult to do a proper covariate exploration given the large number of covariates to be explored**



Modularization approach

- ***Proposal:*** **In principle, we can efficiently explore a vast range of covariate models using just the outputs from a single fit (without covariates) of the time-consuming model.**



Modularization approach

Stage 1: Disease age model without covariates

Recoupling

Stage 2: Measurement error model the outcome is not observed, but estimated with uncertainty

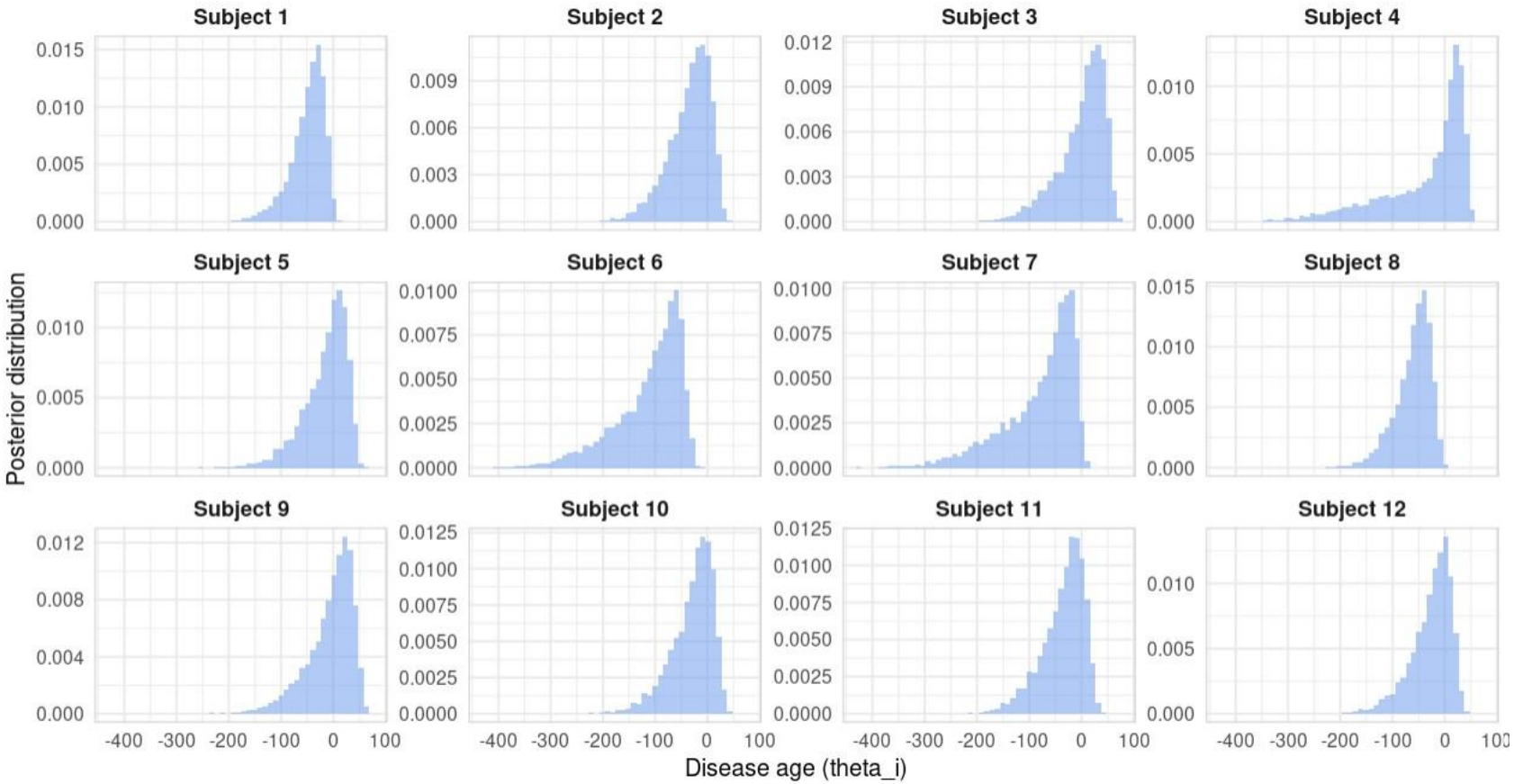
Disease ages including uncertainty

Include predictors in the modelling of the disease ages

- Inference is on the joint model

GSK

Results from stage 1: disease age posterior distributions

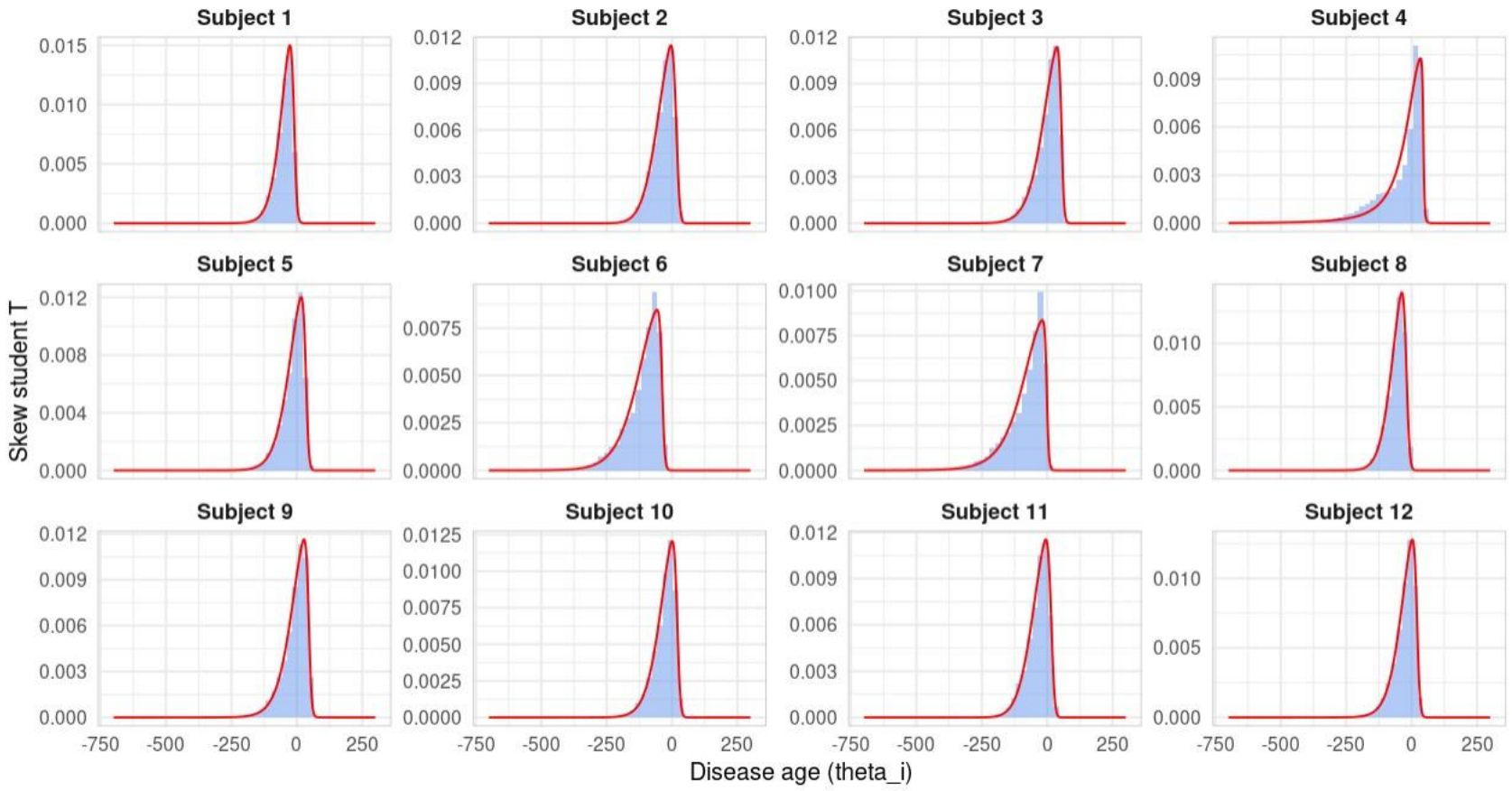


Stage 2: Modularization approach

- In a second stage analysis, a posterior distribution for each patient describing the disease age.
- Ideally, we want to use not only one point estimate (mean/median) for each disease age, but also the *uncertainty* around it.
- *A skew student t distribution (SST)* - a flexible distribution with 4 parameters: the location, the scale, the skewness and the degrees of freedom (Azzalini & Capitanio, 2003).

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society Series B: Statistical Methodology, 65(2), 367-389.

Stage 2: Modularization approach



- Fit Skew T distribution to individual disease age posterior samples.
- Extract and save four subject-specific parameters for subsequent modeling stages.

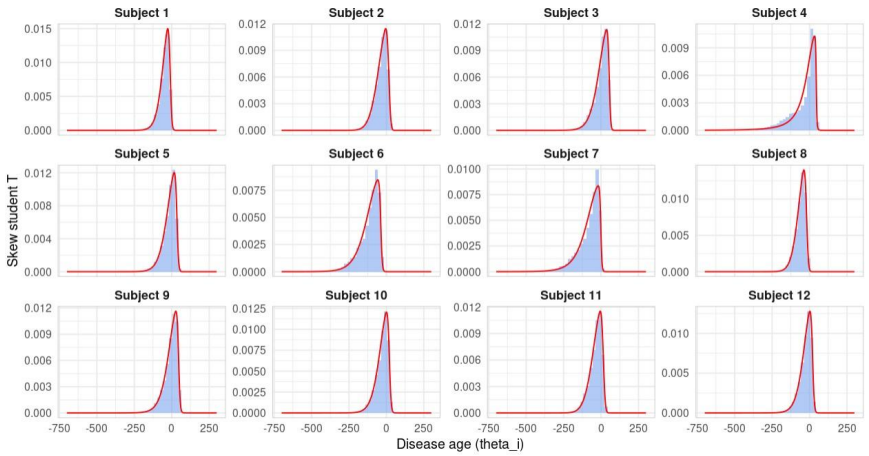
We cannot assign two priors to one node in BUGS / JAGS / NIMBLE

- From the first model stage, each patient has a posterior distribution describing his/her disease age.

$$\boldsymbol{\theta_i} \sim \text{SST}(x_i, \omega_i, \alpha_i, \nu_i)$$

- Our objective is to find covariates, in X_i, say, that explain heterogeneity in $\boldsymbol{\theta_i}$ - disease age of every patient

$$\boldsymbol{\theta_i} \sim \boldsymbol{N(B_i * X_i, \sigma_\theta)}$$



We cannot assign two priors to one node in BUGS / JAGS / NIMBLE

- From the first model stage, each patient has a posterior distribution describing his/her disease age.

$$\boldsymbol{\theta_i} \sim \text{SST}(x_i, \omega_i, \alpha_i, \nu_i)$$

- Our objective is to find covariates, in X_i, say, that explain heterogeneity in $\boldsymbol{\theta_i}$ - disease age of every patient

$$\boldsymbol{\theta_i} \sim \boldsymbol{N(B_i * X_i, \sigma_\theta)}$$

$$\boldsymbol{L_{\theta_i} = f_{\theta_i}(\theta_i, \omega_i, \alpha_i, \nu_i)}$$

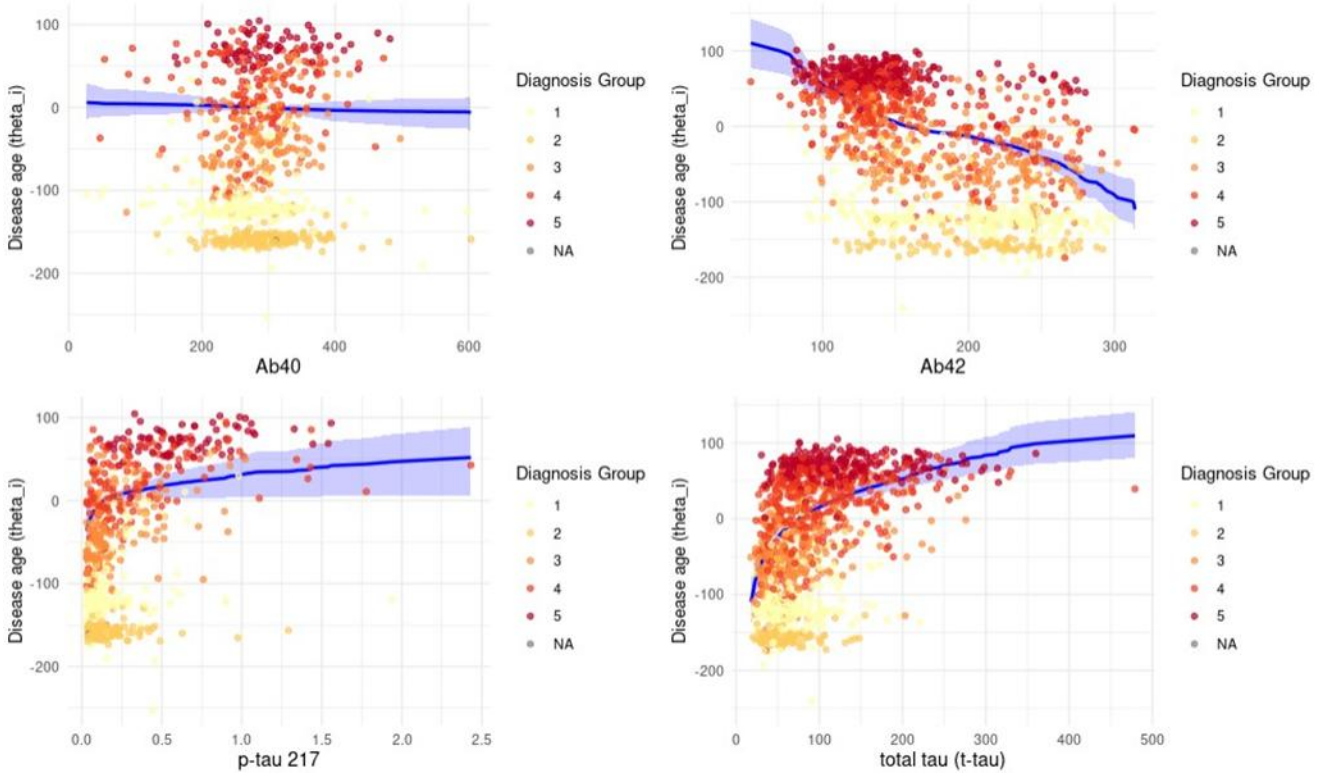
$$\boldsymbol{0s \sim dpois\,(\lambda_i)}$$

$$\boldsymbol{\lambda_i = large - f_{\theta_i}(\theta_i, \omega_i, \alpha_i, \nu_i)}$$

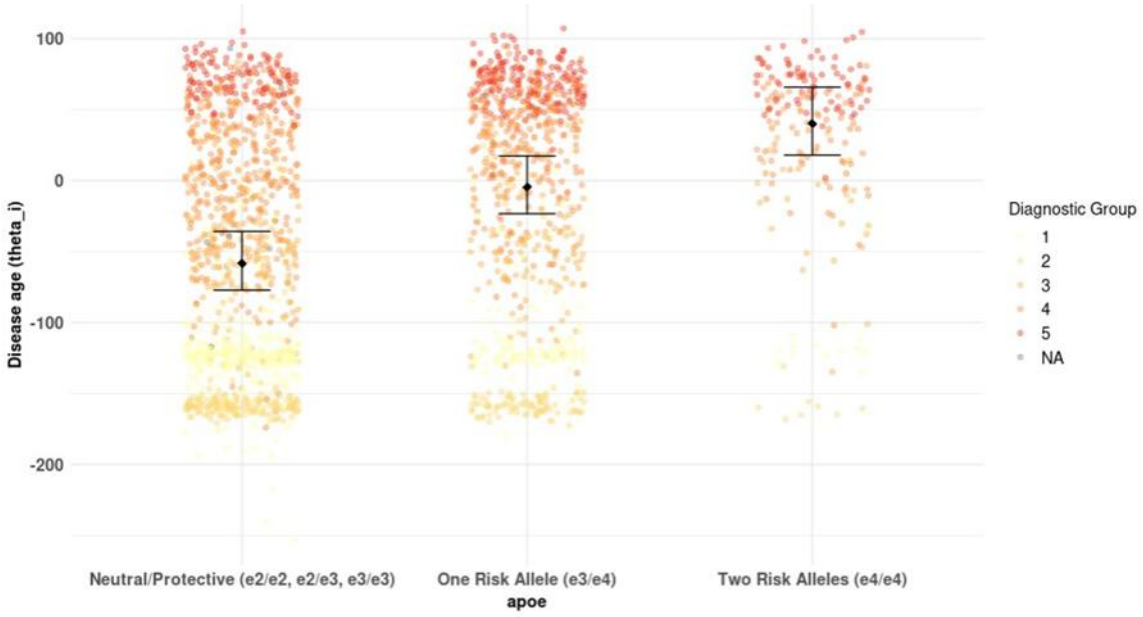
$$\boldsymbol{\theta_i \sim N(B_i * X_i, \sigma_\theta)}$$

- Using the zeros trick, by using the pdf of $\boldsymbol{\theta_i}$, the new expression written similarly as a likelihood term, moves the $\boldsymbol{\theta_i}$ mathematically to the right-hand side of the equation.
- We are going to use this mechanism to exploit this re-written expression, to allow us to specify two priors for $\boldsymbol{\theta_i}$.
- In this way, the model can reconcile these 2 equations into the same modelling strategy.

Biomarkers



- Illustration of covariates fit
- Advantage of the modularization: few minutes to fit the stage 2 model
- Used quantile normalization to fit the model
- Fit a linear effect
- Transform back on the original scale
- The effect looks nonlinear
- Use multiple imputation which relies on MVN distribution for all biomarkers



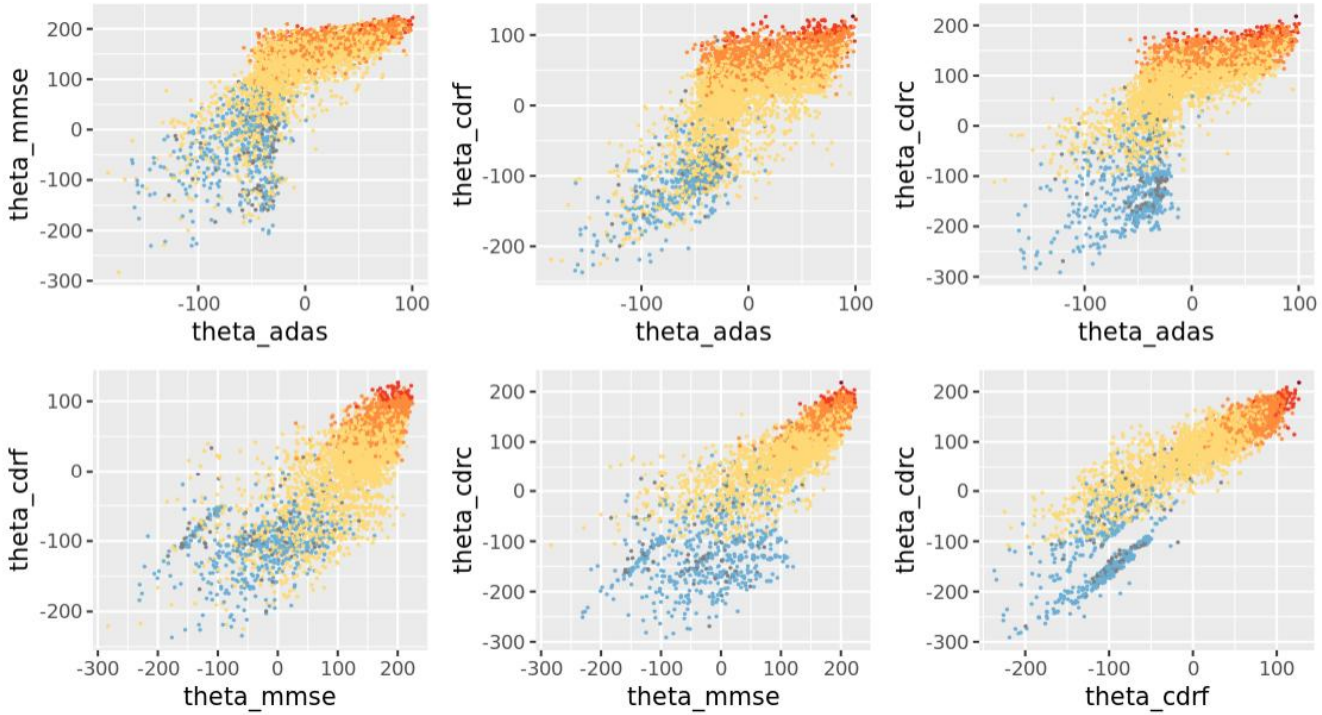
GSK

Take home message:

- **Modularization** – flexible method to run complex models:
  - **Single endpoint model**
    - Flexible way to run a covariate analysis using two stages
    - Stage 1 – time consuming model which we fit only once
    - Stage 2 – covariate analysis runs in a couple of minutes compared to hours
    - Predictions for future subjects' disease ages based on their ADAS-cog 13 values and personal profile (age, gender, biomarker values, etc.)

Take home message:

- **Modularization** – flexible method to run complex models:
  - **Joint modelling** of several endpoints
    - Use the results from single endpoint models to explore joint endpoint models assuming a MVN distribution for the disease age results from stage 1.



- **Flexible approach which could be applied to other disease areas**

Thank you!

Special thanks: Dave Lunn!