Analytic Approaches: Longitudinal Modeling and Quantifying Change

Tara Madhyastha

Principal Research Scientist, Amazon Web Services

- **Analytic Approaches: Longitudinal Modeling and Quantifying Change**
- What is Longitudinal Modeling?
- Common Statistical Models of Longitudinal Change
- Developmental Questions and Applications
- Available (and Missing) Software Packages and Tools
- Troubleshooting and Gaps in the Literature

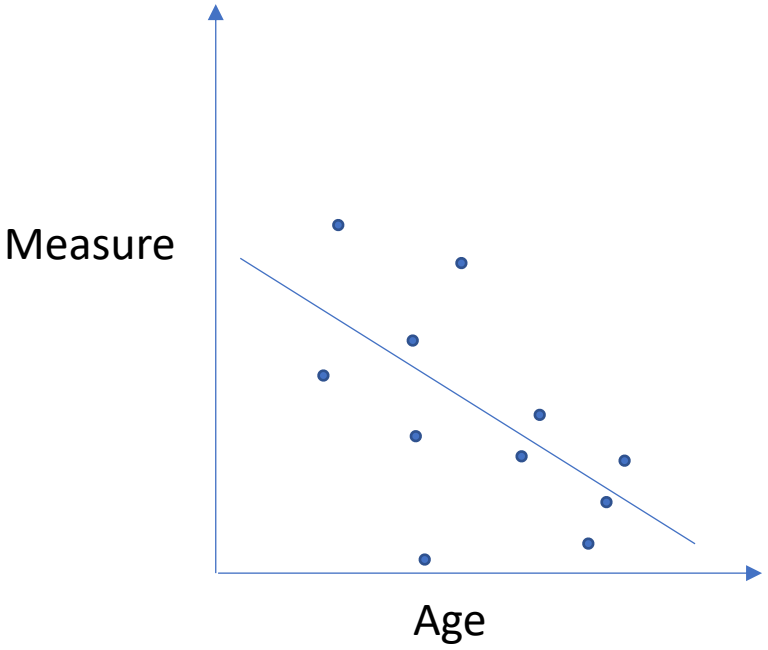
Outline

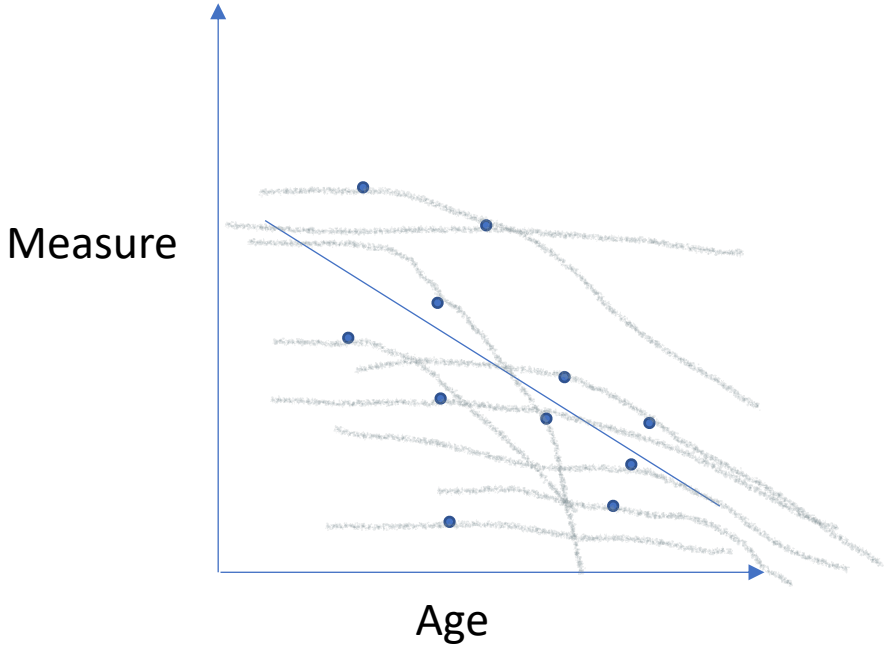
- Cross Sectional vs Longitudinal Design
- Developmental Questions and Applications
- Common Statistical Models of Longitudinal Change
- Available (and Missing) Software Packages and Tools
- Troubleshooting and Gaps in the Literature

Cross-sectional vs Longitudinal Design

- Cross sectional study looks at participants of different ages at the same time
- Longitudinal study follows participants across multiple time points

Cross-sectional vs Longitudinal Design





Ingredients for Measuring Longitudinal Change

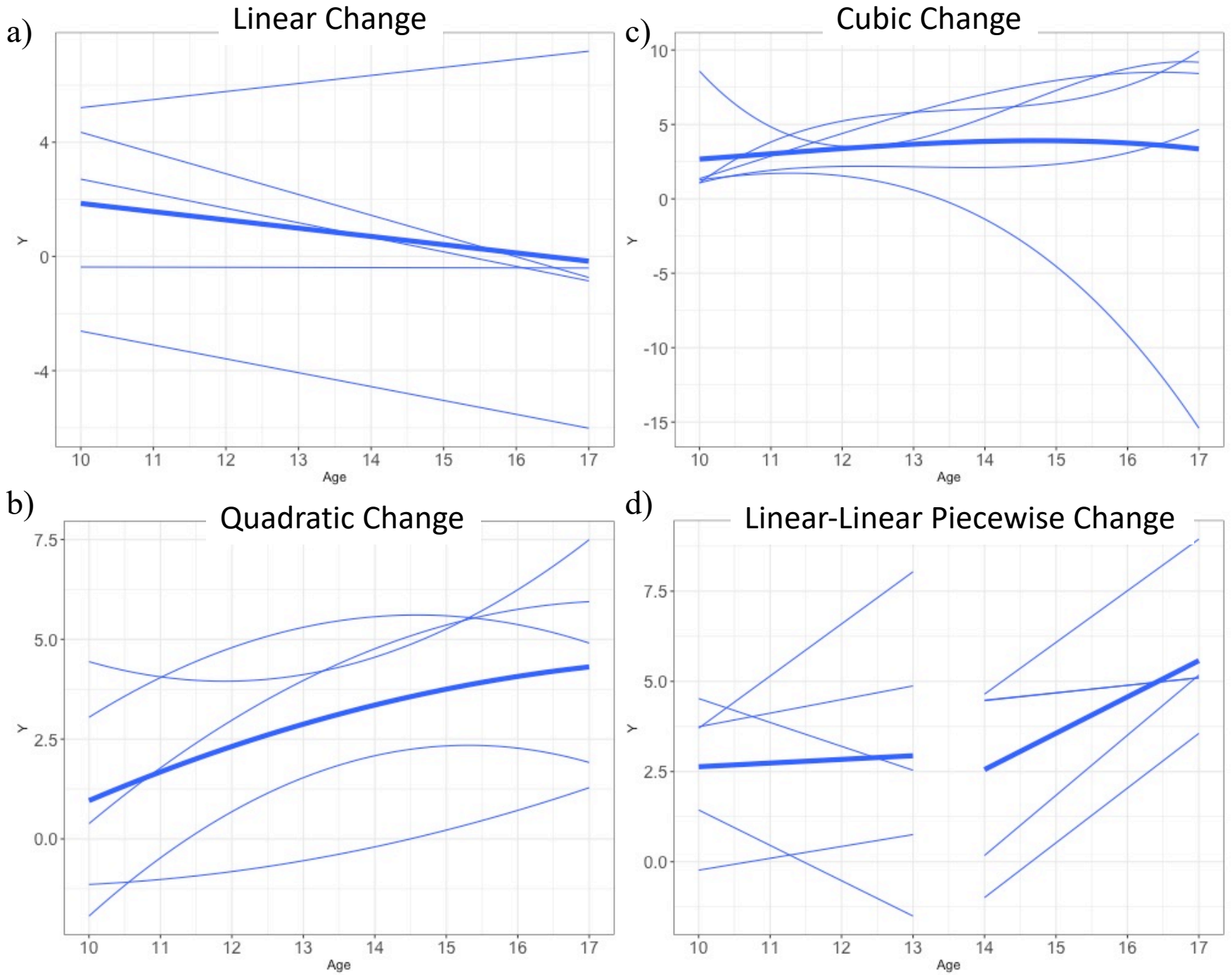
- Longitudinal data
- A sensible measure for time
- Enough occasions to examine the correct form of growth
  - 3 occasions = linear
  - more = better
- The right statistical models

Considerations for Designing a Longitudinal Study

- Recruitment
  - Sampling
  - Drop out
  - Exclusion
- Acquisition
  - Scanner drift
  - Motion
- Analysis
  - How will you model change?

More About Time

- The timing/frequency/spacing of measurement should be driven by a theory about change
  - Too far apart might miss the form of change entirely
  - Too close might be dominated by reactivity
- Coding of time should make sense
- The measurement (ideally) should not change over time
  - Scanner drift, processing algorithms
  - Measurement constructs


a)
Linear Change
c)
Cubic Change
b)
Quadratic Change
d)
Linear-Linear Piecewise Change
Y
Age
10
11
12
13
14
15
16
17

Questions that we might want to ask about change

- How does change differ across people?
- What is the relationship between change and time-invariant predictors?
- What is the relationship between change in an outcome and change in a predictor?
- Are there latent groups of individuals with different change trajectories?
- Does trajectory of change predict outcomes?

Questions that we might want to ask about change

- How does change differ across people?
  - Does change in FC within different networks differ across individuals with anxiety?
- What is the relationship between change and time-invariant predictors?
  - Do genetic factors modify individual trajectories?
- What is the relationship between change in an outcome and change in a predictor?
  - Is change in levels of stress correlated with change in FC within different networks?
- Are there latent groups of individuals with different change trajectories?
  - Do groups with hypothesized underlying differences``fall out" as different longitudinal trajectories?
- Does trajectory of change predict outcomes?
  - Can we look at longitudinal FC change before some onset of symptoms to determine who will become affected?

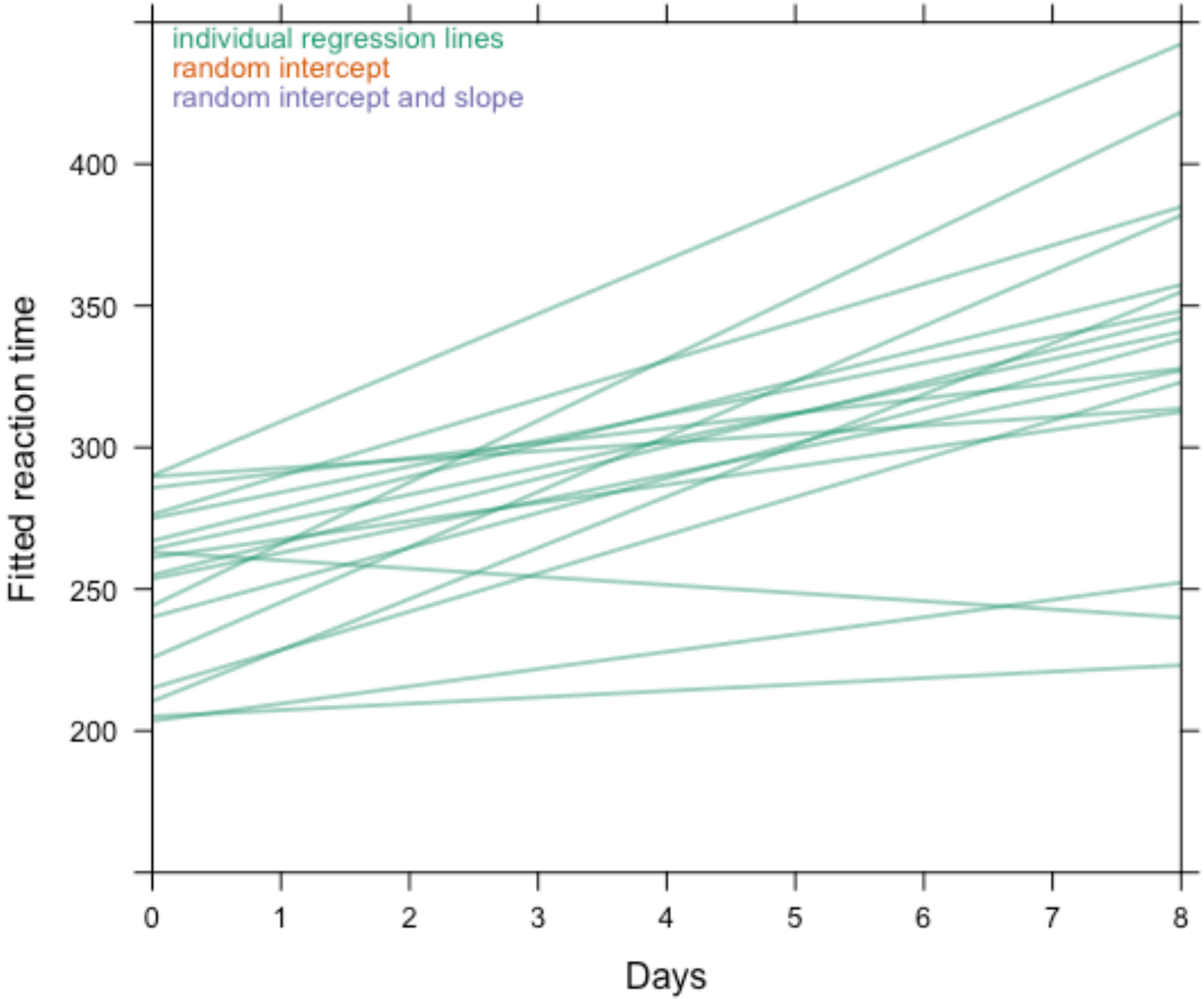
Statistical frameworks that allow us to answer these questions

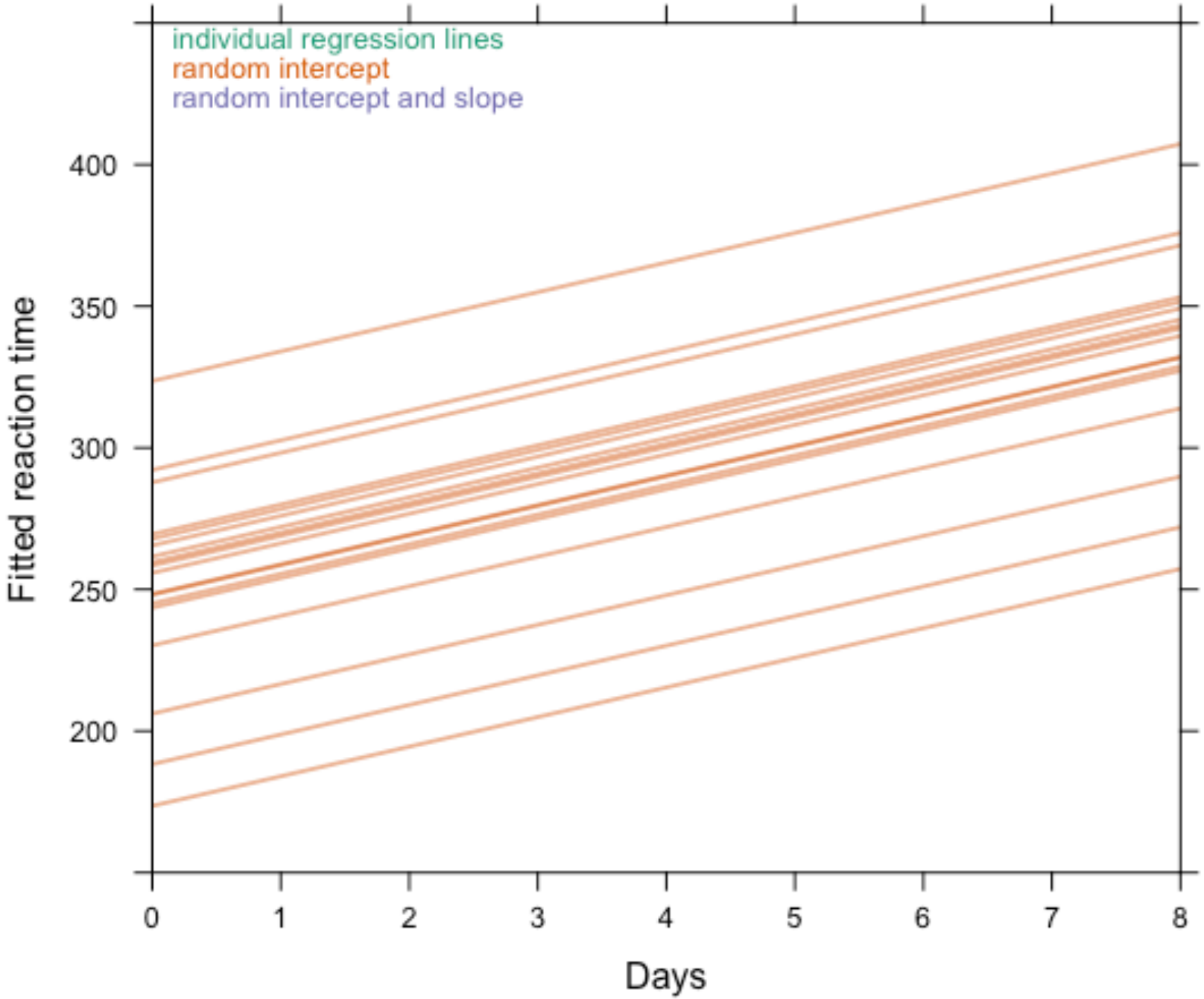
- generalized mixed linear modeling
  - multilevel models (hierarchical linear growth models)

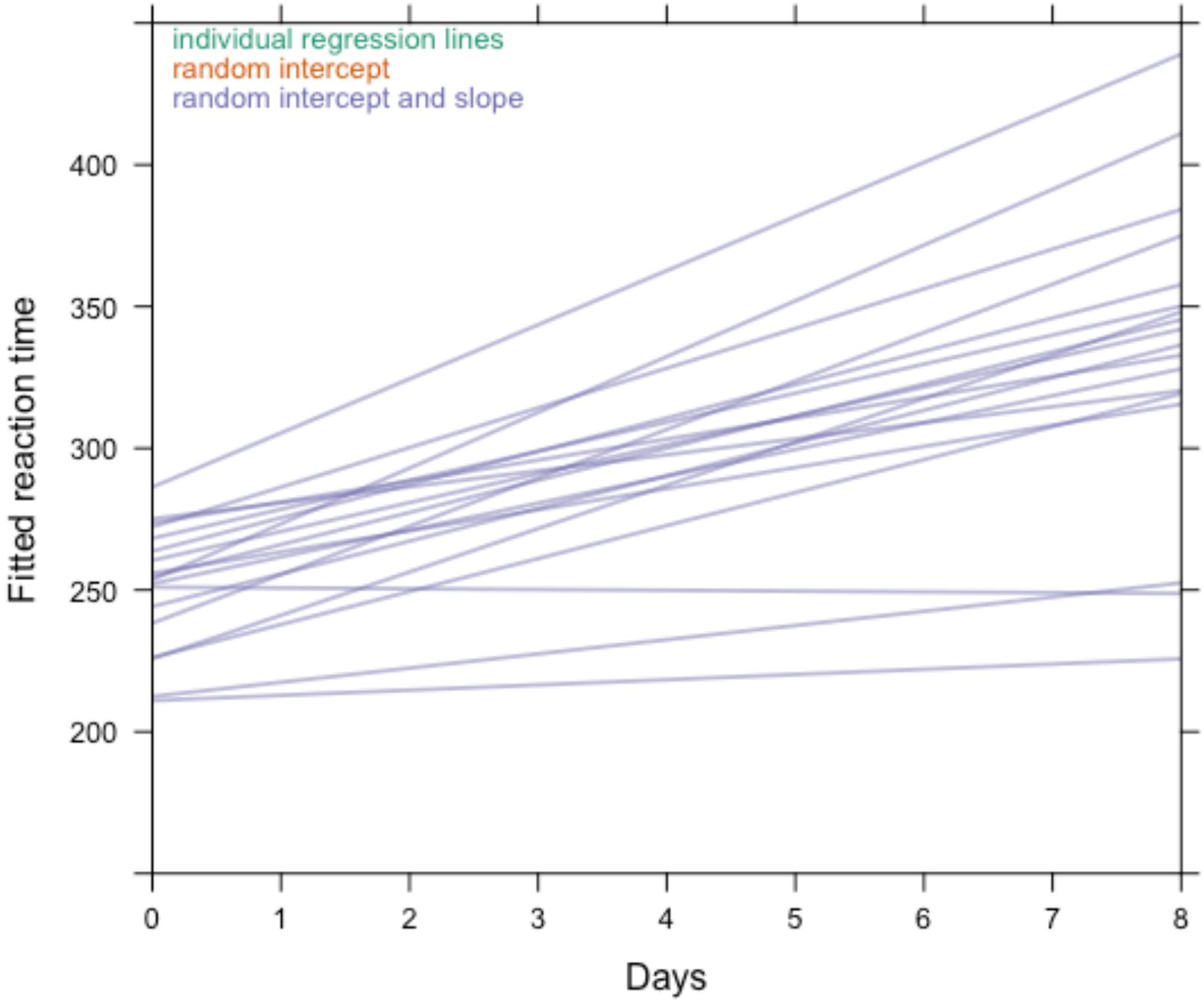
- latent growth curve modeling (a more flexible superset of multilevel models)

Mixed Effects Models

	Fixed Effects	Random Effects
Examples	Male, Female Treatment, Control	Subjects with fMRI data Subjects with multiple measurements
Assume	All groups sampled	Not all groups sampled (groups come from a probability distribution)
We care about	These groups only	Generalizing from this sample to a larger population
We model	The population mean	The variance (within and between subject)


individual regression lines
random intercept
random intercept and slope
Fitted reaction time
400
350
300
250
200
0
1
2
3
4
5
6
7
8
Days


individual regression lines
random intercept
random intercept and slope
Fitted reaction time
400
350
300
250
200
0
1
2
3
4
5
6
7
8
Days


individual regression lines
random intercept
random intercept and slope
Fitted reaction time
400
350
300
250
200
0
1
2
3
4
5
6
7
8
Days

Questions that we might want to ask about change

- How does change differ across people?
- What is the relationship between change and time-invariant predictors?
- What is the relationship between change in an outcome and change in a predictor?
- Are there latent groups of individuals with different change trajectories?
- Does trajectory of change predict outcomes?

Questions you can answer with two time points

- General Linear Models can answer
  - How much is between-person variability at first time associated with between-person variability at second time?
  - How much does a covariate at first time, controlling for brain measure, predict individual differences at second time?

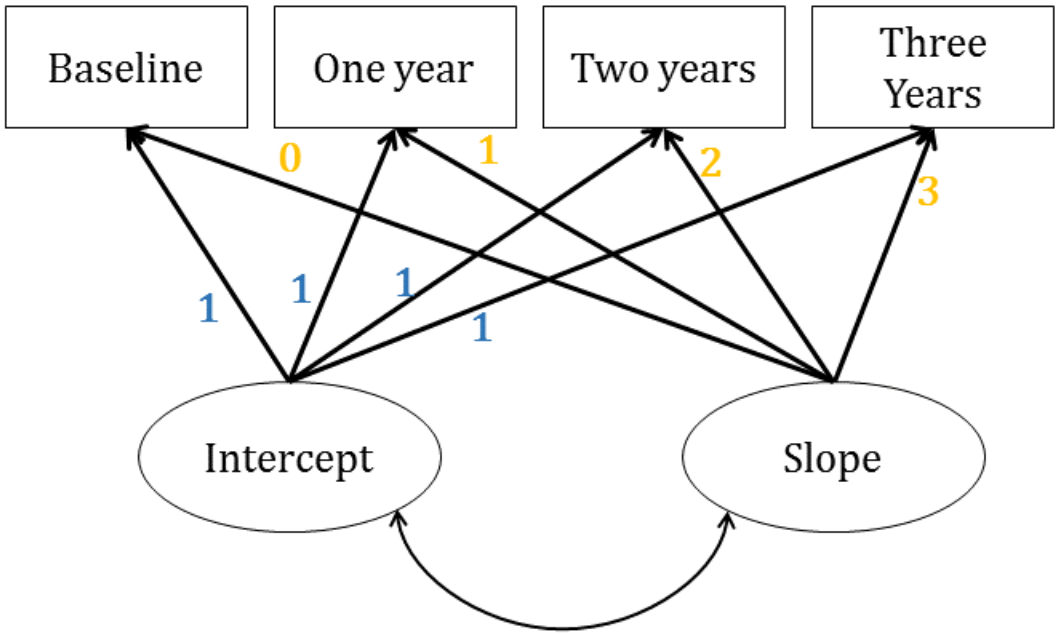
Questions you can answer with three or more time points

- Growth Curve Models
  - Observations for individuals are predicted by time (very flexible)
  - Intercept, slope are random effects
  - How do individuals differ in change over time
- Latent Growth Curve Models
  - Intercepts and slopes are latent variables, and the effect of time is is estimated from these latent variables
  - You can look at relationships of growth parameters
  - You can look at relationships of different growth processes!

Important features of latent growth curve modeling

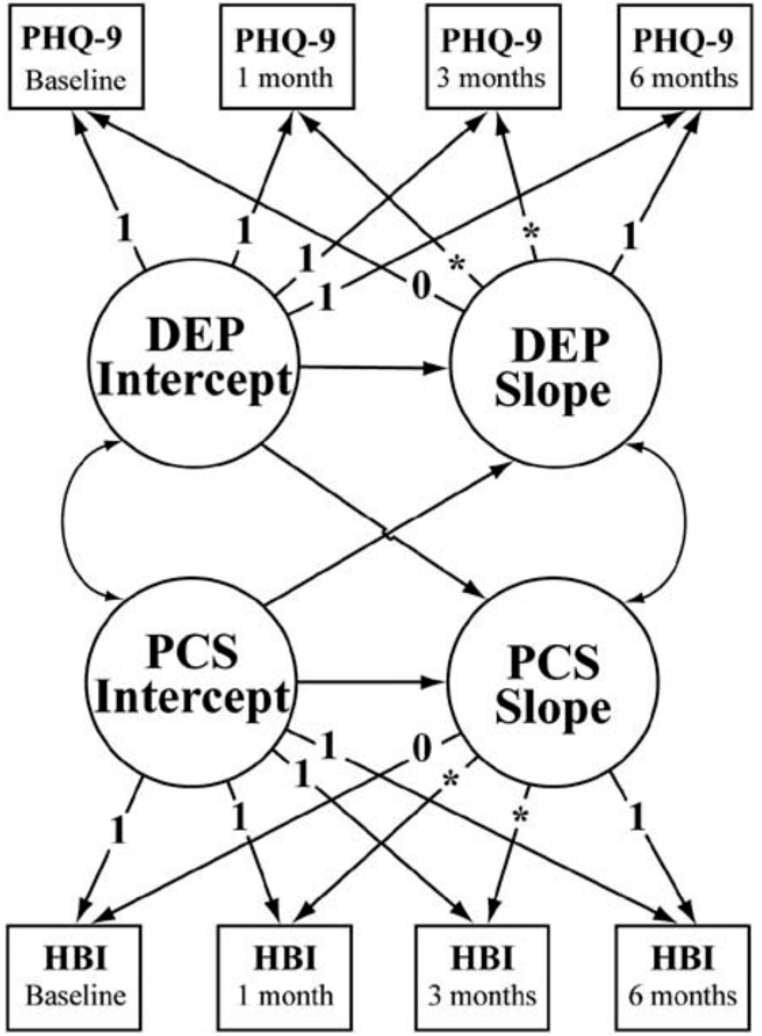
- latent variables (indicators and latent ``constructs'')
- can model the relationships among trajectories of change in multiple constructs
- can use change in a construct as a predictor of a different outcome

Latent Growth Model



Latent Growth Model

Parallel Process Growth Model



depressive symptoms

post-concussive symptoms

Brain Inj. 2017;31(13-14):1736-1744.

Statistical Model	Between-group mean differences	Rank-order change	Between individual change	Within individual change	Predictors of within-individual change over time	Correlations between growth processes	Latent Groups	Change as a predictor
Independent Samples T-test	X							
ANOVA	X							
(Multiple) Regression, General Linear Modeling	X	X						
Repeated Measures ANOVA	X	X	X					
Auto-regressive panel models	X	X	X					
Latent change score	X	X	X	X	X	X		X
Multilevel growth models	X	X	X	X	X			
Latent growth curve	X	X	X	X	X	X		X
Growth mixture model	X	X	X	X	X	X	X	X

State of the art in fMRI analysis software

- FSL
  - Person-level statistics (task effects)
  - Person-level estimates combined to estimate mean effect
  - Person-level estimates are combined
  - FLAME – random intercepts but not random slopes
- SPM
  - Similar limitations to FSL
- AFNI
  - 3dLME - Implements wrappers around R `nlme` library
  - Addresses limitations of GLM
- Sandwich Estimator
  - Accounts for within-subject correlation in longitudinal data

Assumptions (GLM)

- Correlation between repeated measurements for the same subject is constant
- Sphericity – variance of differences between any two time points is the same
- Random intercepts but not random slopes - i.e., individual differences in change over time across multiple measurement occasions
- Difficult to deal with missing data

fMRI Comparison Chart

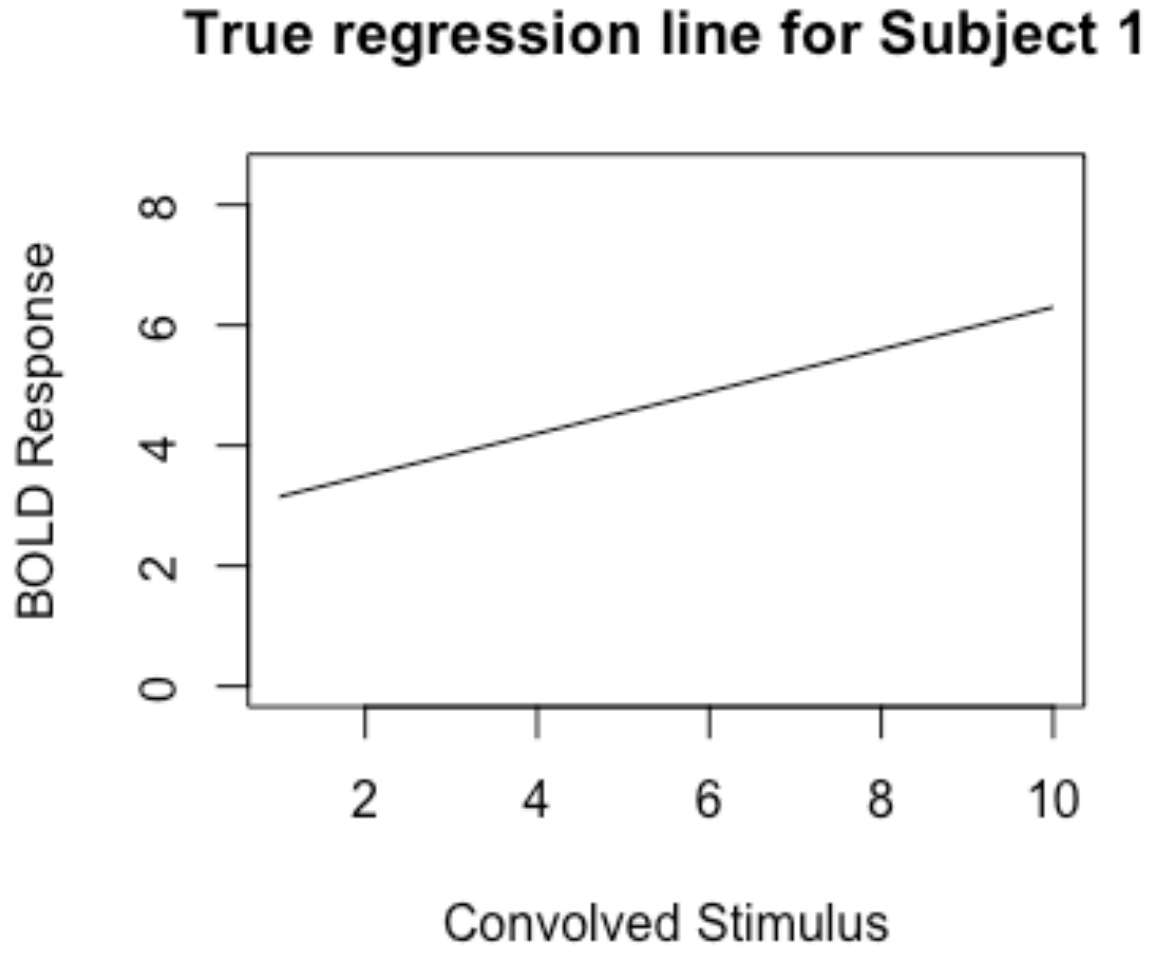
fMRI Analysis Software	Missing Data	Voxel-wise model comparison	Between-group mean differences	Rank-order change	Between individual change	Within individual change	Predictors of within-individual change over time	Correlations between growth processes	Latent Groups	Change as a predictor
FSL			**x**	**x**	**x**					
SPM			**x**	**x**	**x**					
AFNI			**x**	**x**	**x**					
AFNI (3dNLME)	**x**	**x**	**x**	**x**	**x**	**x**	**x**			**x**

First Level Model (Single Subject)

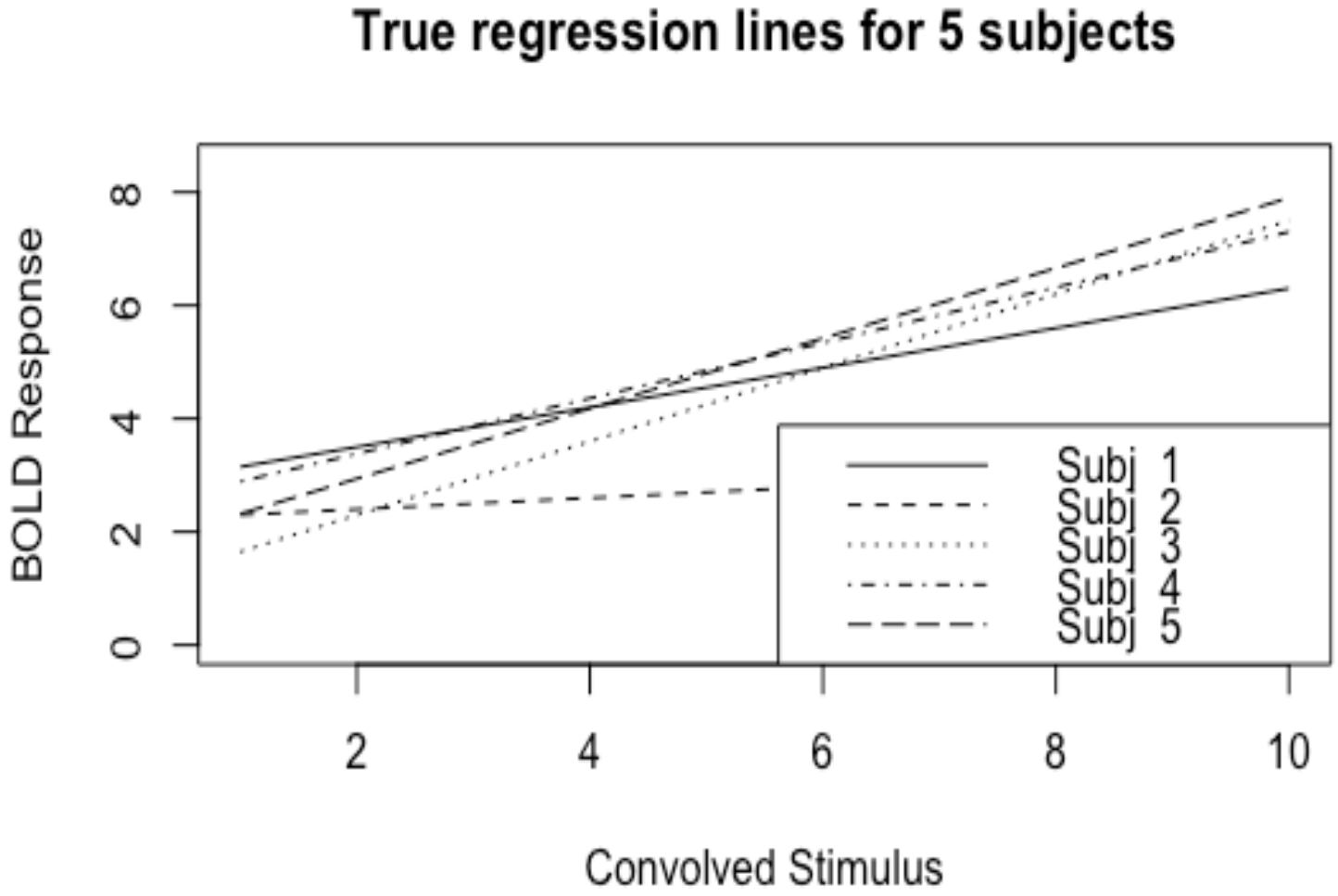
- $Y = \beta_0 + X\beta_0 + e$
- Y is the fMRI signal at a single voxel
- X is the stimulus EV (let us assume only one)
- β_0 is the intercept
- β_1 is the slope parameter, or the effect of the stimulus
- e is the error

```
x<-1:10
beta.0<-2.8
beta.1<-0.35
y.fit <-beta.0 + x*beta.1
plot(c(1,10), c(0,8.5),type="n", main="True regression
line for Subject 1", xlab="Convolved Stimulus", ylab="BOLD
Response”)
lines(x,y.fit)
```

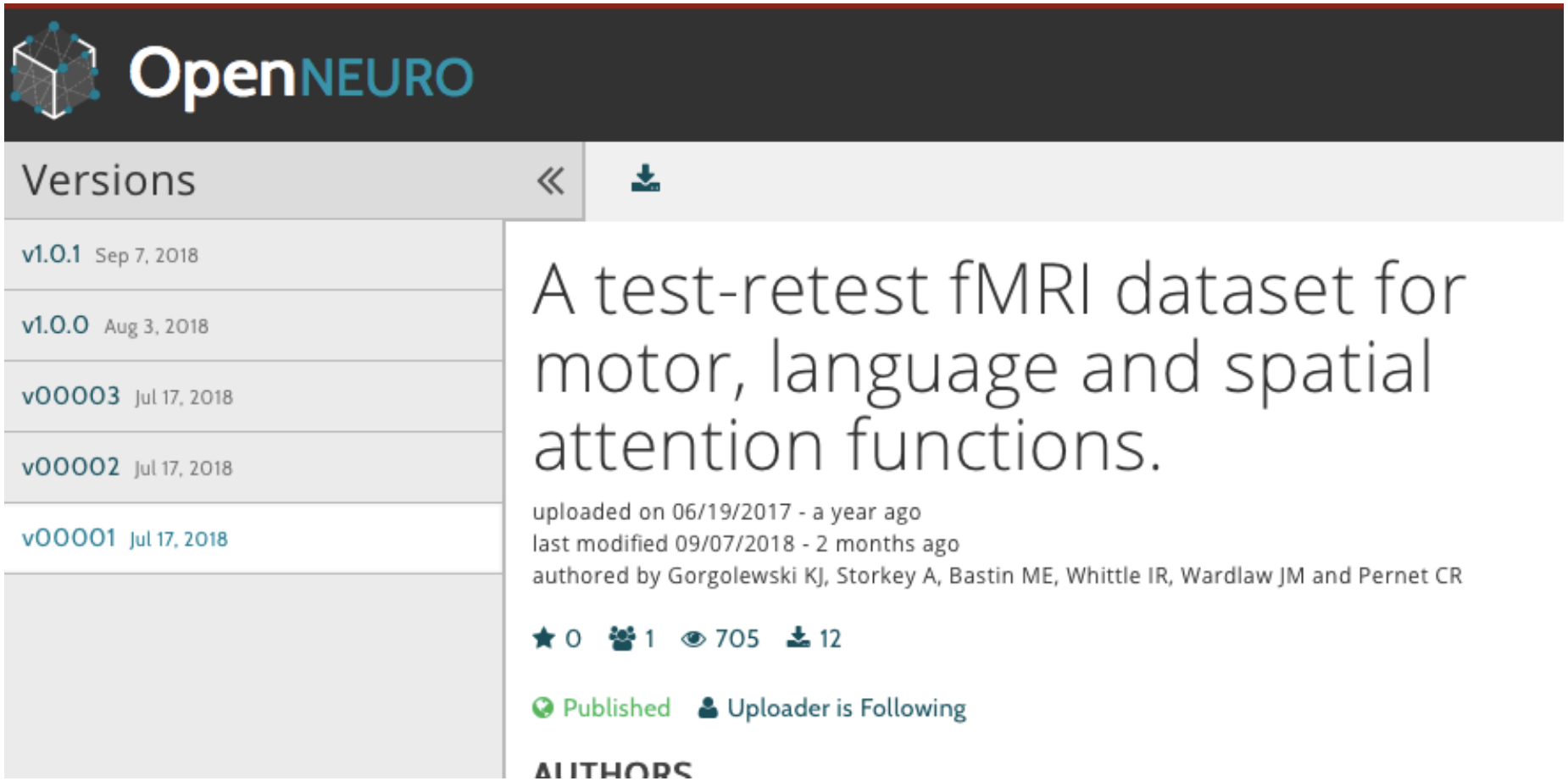
First Level Model (Single Subject)



First Level Model (Multiple Subjects)



An Actual Task



10 subjects, 2 occasions, finger/hand/mouth movement interleaved with a fixation cross

FSL FLAME

First level analysis

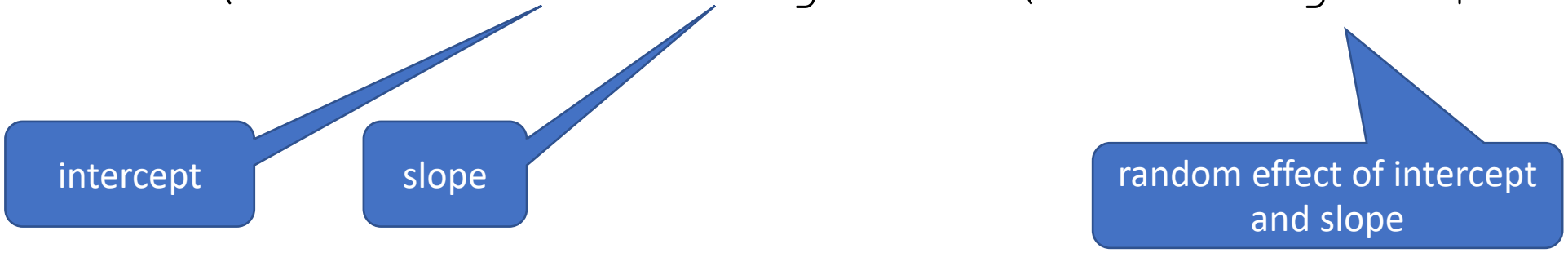
- For each subject, estimate intercept and slope, and within-subject variance

Second level analysis

- Estimate group mean intercept and slope
- But it isn’t just an average
- It is a Bayesian approach a little like weighting the individual estimates, where the weights are a function of the variance

A Random Effect Model in R (lmer)

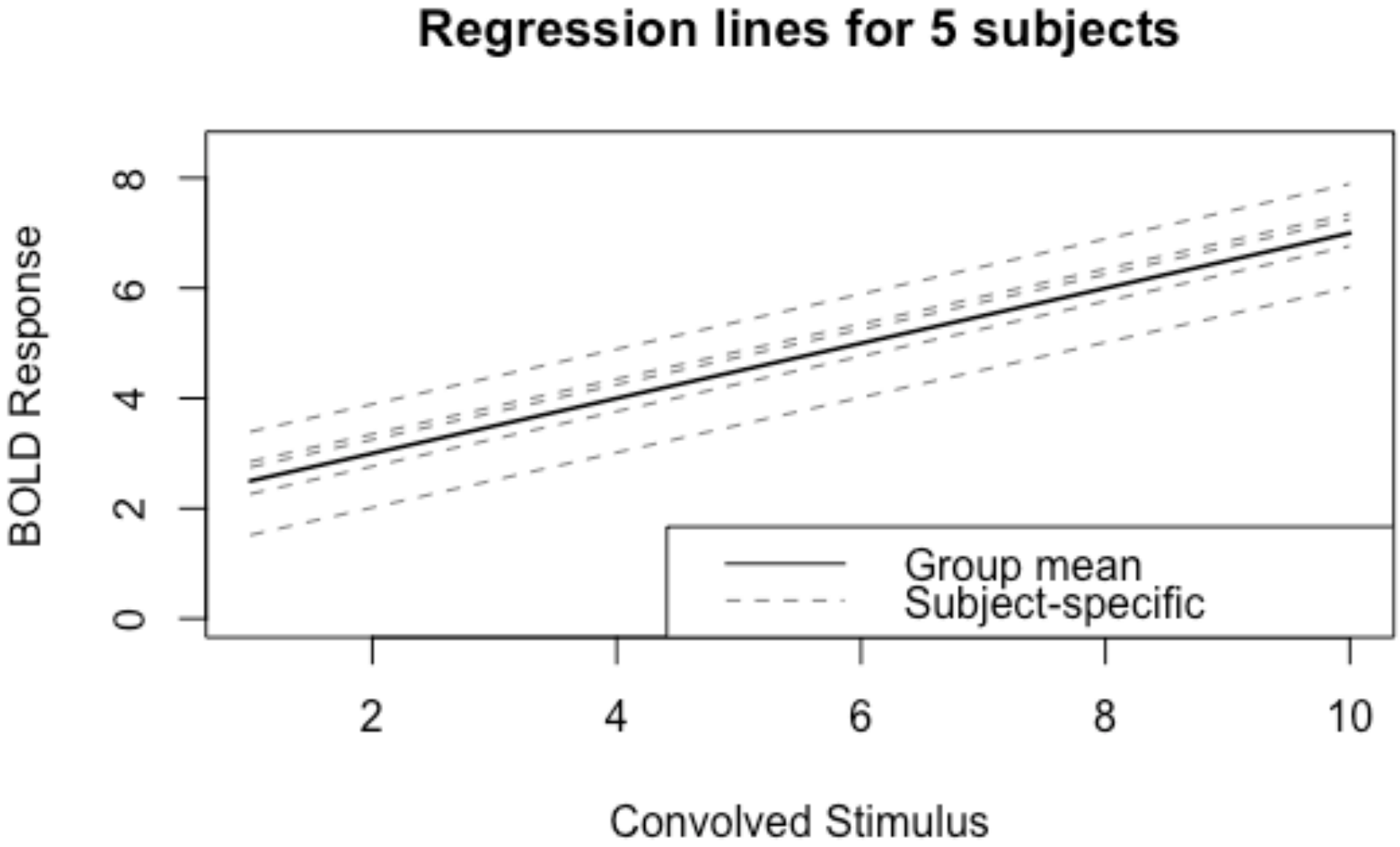
```
mod<-lmer(BRAIN ~ 1 + Finger + (1 + Finger |subject))
```



- Unlike FSL FLAME, this does not estimate the subject specific intercepts and slopes
  - Group means for intercept and slope
  - Between-subject variance in intercept and slope (random effects)

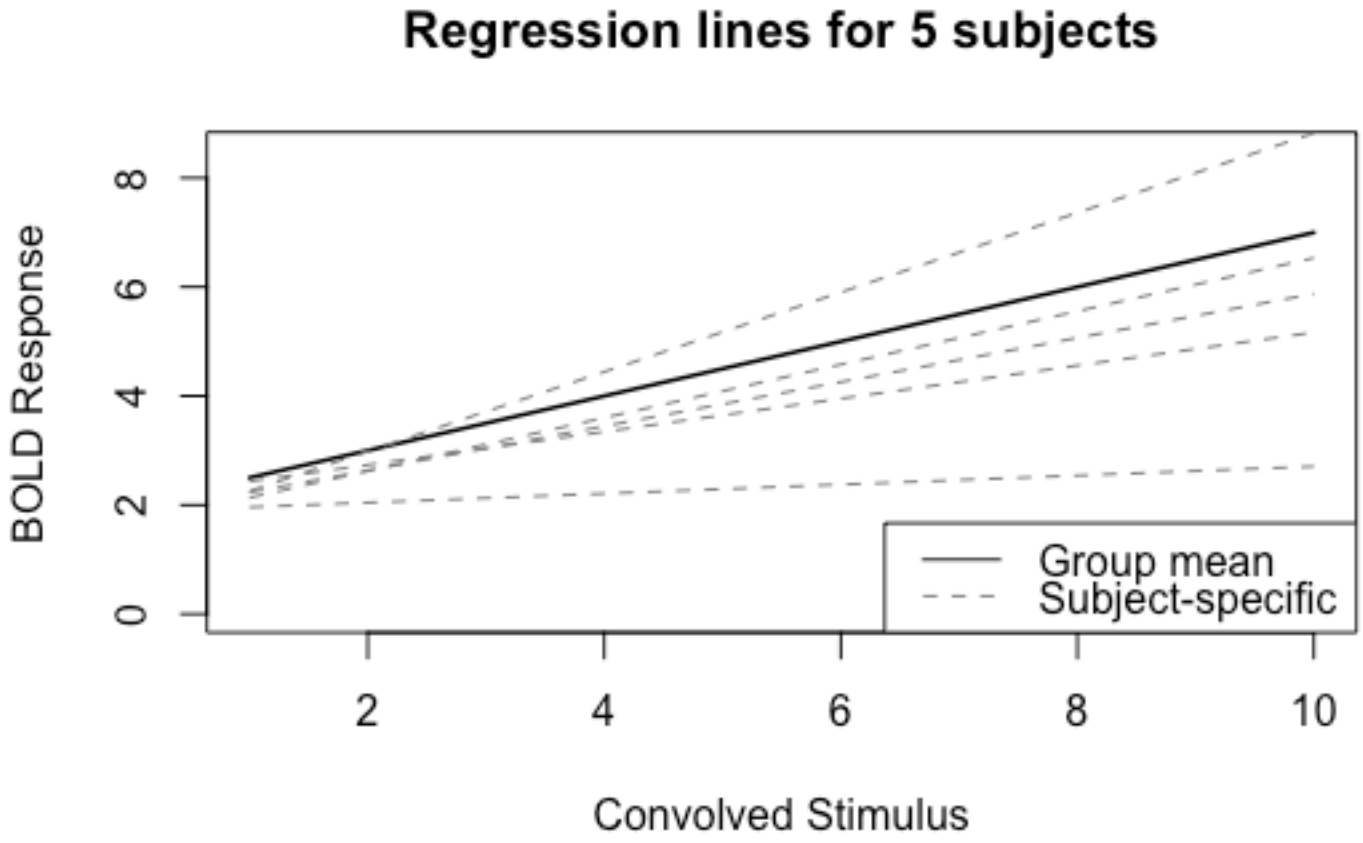
Random Intercept Only

```
mod1<-lmer(BRAIN ~ 1 + Finger + (1|subject))
```



Random Intercept and Slope

```
mod1<-lmer(BRAIN ~ 1 + Finger + (1 + Finger|subject))
```



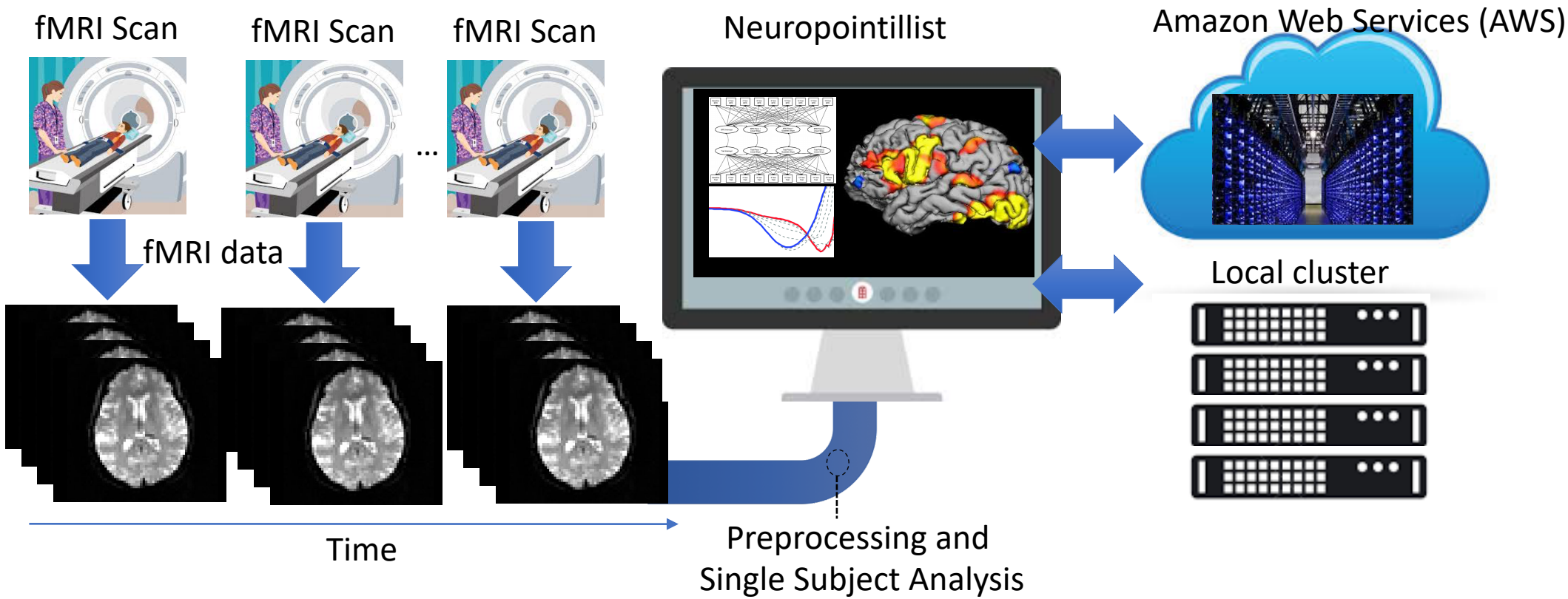
Testing Models

```
mod1<-lmer(BRAIN ~
Finger+Foot+Lips+WhiteMatter+X+Y+Z+RotX+RotY+Rot
Z + (1|idnum) + (1|time), REML = FALSE)

mod2<-lmer(BRAIN ~
Finger+Foot+Lips+WhiteMatter+X+Y+Z+RotX+RotY+Rot
Z + (1 + Finger+Foot+Lips|idnum) + (1|time),
REML = FALSE)

anova(mod1, mod2)
```

Neuropointillist Overview



Approach

Data Sets

Subj1_time1.nii.gz
Subj2_time1.nii.gz
…
SubjN_time1.nii.gz

Subj1_time2.nii.gz
Subj2_time2.nii.gz
…
SubjN_time2.nii.gz

+

Data Set Labels

SubjID,time,age
1,1,10
2,1,11
…
N,1,11

SubjID,time,age
1,2,11
2,2,12
…
N,2,12

+

Mask file

Voxel Data

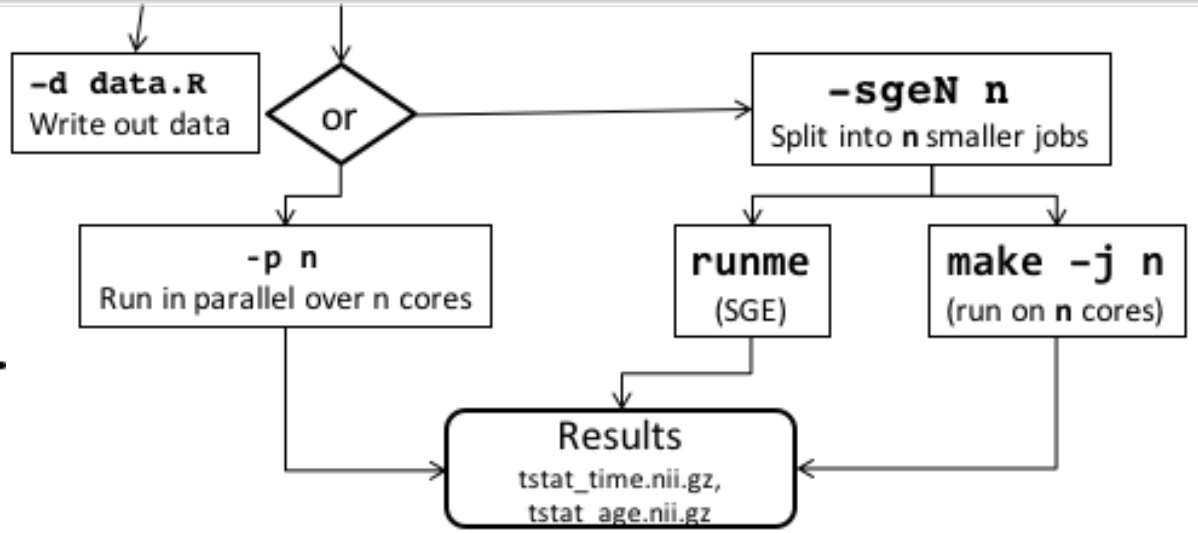
V_1	V_2	V_3	…	V_{mask}
3.8	3.6	3.5	…	1.5
4.2	4.6	4.3	…	2.0
…				
3.5	4.2	4.6	…	2.5

Design Matrix

SubjID	time	age
1	1	10
2	1	11
…		
N	2	12

model.R

```
load(nlme)
processVoxel(v){
        mod <- lme(v~time+age+time*age)
        [more code]
        return(list(tstat_time, tstat_age, tstat_interaction)
}
```



These Kinds of Models Take Time

Time required is (approx) time to run one model X number of voxels in the mask

- You can calculate time to run a model for one voxel using system.time
- Multiply by # of voxels

- You can also use –sgeN to generate lots of jobs, and `/usr/bin/time` to time just one
- Then adjust the number of jobs.

ibic.github.io/neuropointillist

Receipt

Neuropointillist Tutorial

neuropointillist

Flexible modeling of neuroimaging data in R, point by point

View on GitHub

Neuropointillist Tutorial

This is a tutorial for using neuropointillist to run some examples. This assumes that you have already followed the directions in Installation.

Setting your PATH variable

After you have downloaded and installed the neuropointillist programs in a directory, you need to add this directory to your PATH variable. Suppose that you have downloaded the neuropointillist package into `~/neuropointillist`. Assuming you are running the bash shell, edit your PATH as follows:

```
export PATH=$PATH:~/neuropointillist
```

This code will make it so that when you type `npoint` or `npointrun` at the command line, your shell

Considerations

- Neuropointillist does not read/write cifti files
  - This is pretty minor to add (cifti toolbox)
- Cluster correction will be on the surface
  - Permutation testing will work- ETAC is farther away

Resources

- Recent Neuropointillist workshop
- Tutorial on fMRI analysis using GLM
- Longitudinal analysis
  - Model Selection for Longitudinal Data
  - Code to illustrate the above (in R)
  - Data to run the code (Rdata file)
- Special issue: Methodological Challenges in Developmental Neuroimaging: Contemporary Approaches and Solutions
  - Edited by Michelle Byrne, Kate Mills, Jennifer Pfeifer, Nick Allen
  - Volume 33, **Pages 1-224 (October 2018**