# Longitudinal Model Selection in Applied Research: Tips and Tricks for Matching Model Choice to Theory

Ethan M. McCormick, Ph.D.

Postdoctoral Fellow (Dr. Rogier Kievit)
Donders Institute at the RadboudUMC

@McCormickNeuro
ethan.mccormick@radboudumc.nl
https://mccormickneuro.github.io/

FIT'NG Together,
March 15, 2022

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

# Acknowledgements: A Village

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Universiteit Leiden

UNIVERSITY OF MARYLAND

UNIVERSITY OF OREGON

Erasmus University Rotterdam

MONASH University

HARVARD UNIVERSITY

NIH National Institute of Mental Health

NIH National Institute on Drug Abuse

NSF

erc European Research Council

Radboudumc

DONDERS INSTITUTE

NIH National Institute of Mental Health

# Selecting a Longitudinal Modeling Approach

- Overwhelming number of ways to model repeated measures data
  - Often told "use theory", but not *how* to do that
- Different frameworks*
  - Mixed-effects & structural equation models
  - A lot of historical baggage
- Need a conceptual decision tree to adjudicate between model choices
  - Access to training in many methods



King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., & Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. Developmental Cognitive Neuroscience, 33, 54-72.

# Longitudinal Primer + Codebook

The Hitchhiker's Guide to Longitudinal ...



Published with bookdown

## The Hitchhiker's Guide to Longitudinal Models


Mills, & Jennifer H. Pfiefer

The following document is a code companion to The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods, https://osf.io/bn6yu/.

Some general notes about this code companion:

- We believe in the importance of using real data in our examples of longitudinal models. However, some of the models we discuss can not yet be fit using publicly-available neuroimaging data (most



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Longitudinal Modeling in Pre-collected Data

- Matching models to theory is often complicated by the need to use pre-existing data sources
  - Other people have made decisions for you

- But with more datasets becoming publicly available, there may be options in the future

**Using large, publicly available data sets to study adolescent development: opportunities and challenges**

Rogier A. Kievit[2,3], Ethan M. McCormick[2,3,a], Delia Fuhrmann[1,4,a], Marie K. Deserno[5,6,a] and Amy Orben[1,a]



Kievit, R. A., McCormick, E. M., Fuhrmann, D., Deserno, M. K., & Orben, A. (2022). Using large, publicly available data sets to study adolescent development: opportunities and challenges. Current Opinion in Psychology, 44, 303-308.

# Longitudinal Modeling in Pre-collected Data

- Matching models to theory is often complicated by the need to use pre-existing data sources
  - Other people have made decisions for you
- But with more datasets becoming publicly available, there may be options here in the future
- Sometimes we have to work within constraints
  - Occasionally can turn these into a strength
  - Might be able to take intentional subsamples



McCormick, E. M. (2021). Multi-Level Multi-Growth Models: New opportunities for addressing developmental theory using advanced longitudinal designs with planned missingness. *Developmental Cognitive Neuroscience*, *51*, 101001.

# Matching Models and Theory

- Often, we fit standard models because "that's how they are done" without interrogating the assumptions different model specifications impose
  - Our goal is to peel back those assumptions and ask if they make sense for testing a particular theory
  - Also (hopefully) help make better theories

- But…perfect should not be the enemy of the good
  - Sometimes we must make compromises
  - Important to be transparent about those



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Outline for the Rest of the Talk

- 3 use cases
  - Assessment designs and longitudinal model selection
    - Cohort vs accelerated studies
  - Uncovering sensitive periods
    - Linear and nonlinear effects
  - Understanding brain-behavior relationships
    - Exogenous and endogenous predictors
- Conclusions and Next Steps

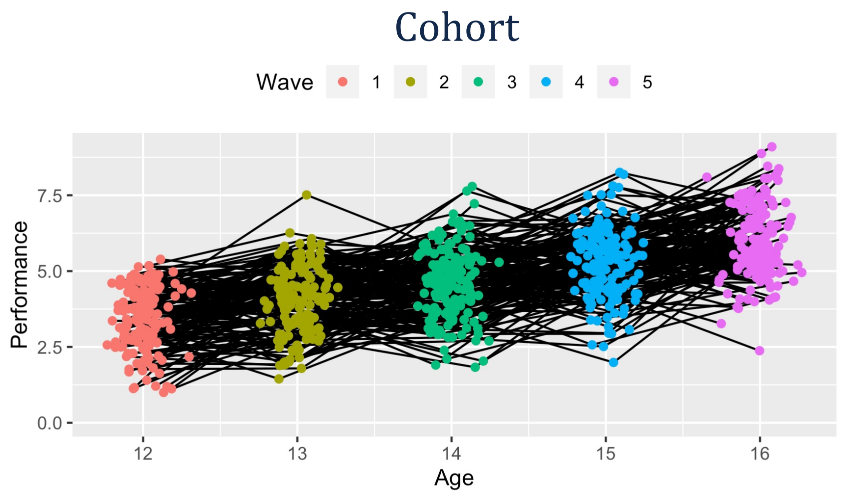


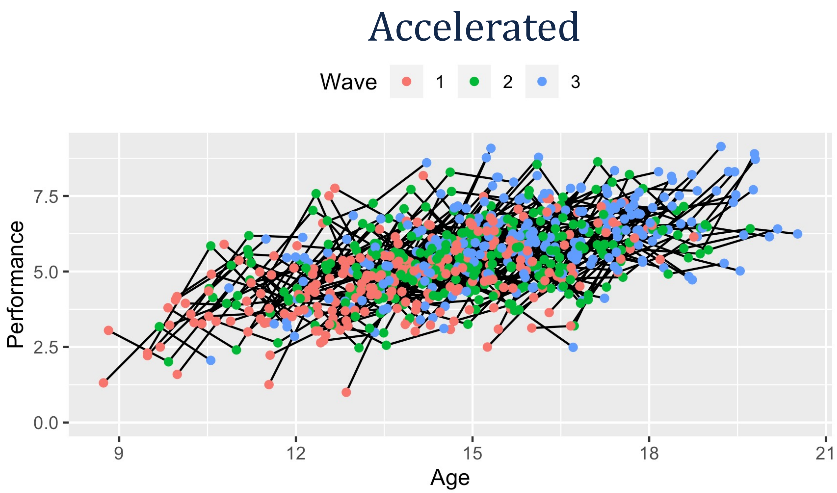
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Assessment designs and longitudinal model selection

# Use-Case 1: Assessment Designs

- Cohort- vs Accelerated Longitudinal Designs






McCormick, E. M. (2021). Multi-Level Multi-Growth Models: New opportunities for addressing developmental theory using advanced longitudinal designs with planned missingness. *Developmental Cognitive Neuroscience*, *51*, 101001.

# Use-Case 1: Assessment Designs

- Cohort- vs Accelerated Longitudinal Designs
  - Historically,
    - Cohort designs → SEM methods
    - Accelerated designs → MEM-based methods



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

## Mixed-Effect Data & Model

- Linear Model:

$$y_{ti} = \underbrace{\gamma_{00} + \gamma_{10} Time_{ti}}_{fixed\ effects} +$$

$$\underbrace{u_{0i} + u_{1i} Time_{ti}}_{random\ effects} + r_{ti}$$

- Time is an observed covariate
  - Model treats it like *any other* covariate
  - Allows for individually-varying values

**Executive Function Data: Long Format**

| id | wave | dlpfc |
|---|---|---|
| 1 | 0 | -0.184 |
| 1 | 1 | 1.129 |
| 1 | 2 | -0.840 |
| 1 | 3 | 0.472 |
| 2 | 0 | 0.801 |
| 2 | 1 | 1.129 |
| 2 | 2 | 0.801 |
| 2 | 3 | 1.457 |
| 3 | 0 | 0.472 |
| 3 | 1 | 1.129 |
| 3 | 2 | 0.144 |
| 3 | 3 | 0.144 |



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

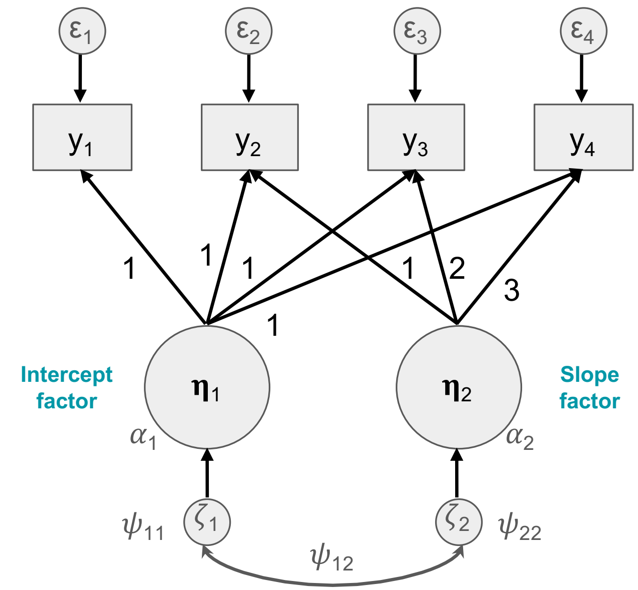
## SEM Data & Model

- Linear Model:

$$\boldsymbol{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

- Time is hard-coded into the factor-loading matrix
  - Repeated measures are separate variables
  - Need to group variables into time-bins*




**Executive Function Data: Wide Format**

| id | dlpfc1 | dlpfc2 | dlpfc3 | dlpfc4 |
|---|---|---|---|---|
| 1 | -0.184 | 1.129 | -0.840 | 0.472 |
| 2 | 0.801 | 1.129 | 0.801 | 1.457 |
| 3 | 0.472 | 1.129 | 0.144 | 0.144 |
| 4 | 0.472 | 0.472 | 0.472 | 0.472 |
| 5 | -0.840 | 2.114 | 2.442 | 2.442 |



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 1: Assessment Designs

- Modern methods
  - Can allow for individually-varying assessment in SEMs with definition variables through Mplus or OpenMx

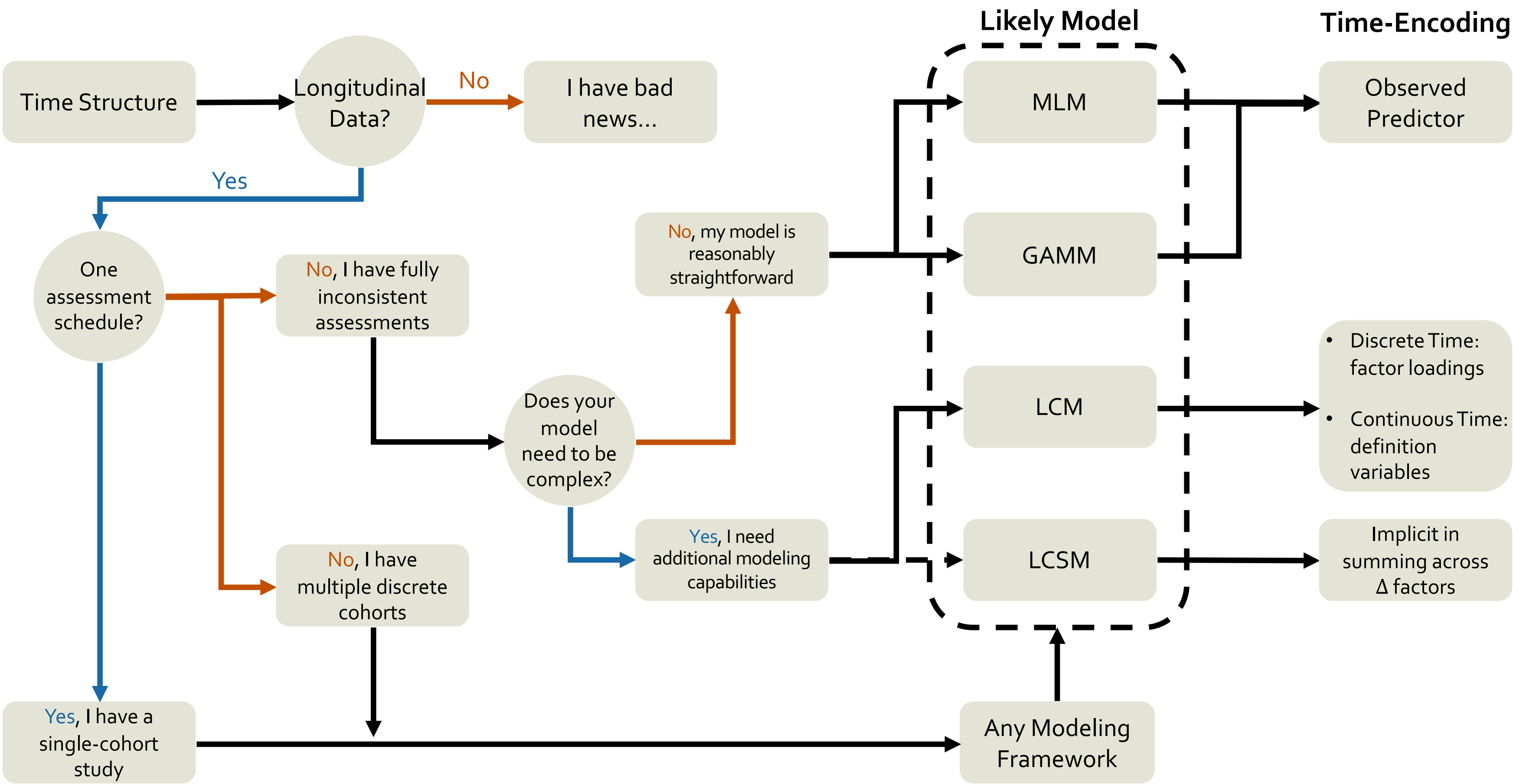
- But in general, accelerated designs are still modeled primarily with mixed-effects models

- Main theoretical distinction: how much age-variation can you compress before the model is mis-specified?



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 1: Assessment Designs Heuristic Map






McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Uncovering Sensitive Periods

# Use-Case 2: Sensitive Periods

- Sensitive periods
  - A window of development where the organism is differentially- (usually hyper-) sensitive to a particular environmental* input
  - Especially abundant in the peri-natal period

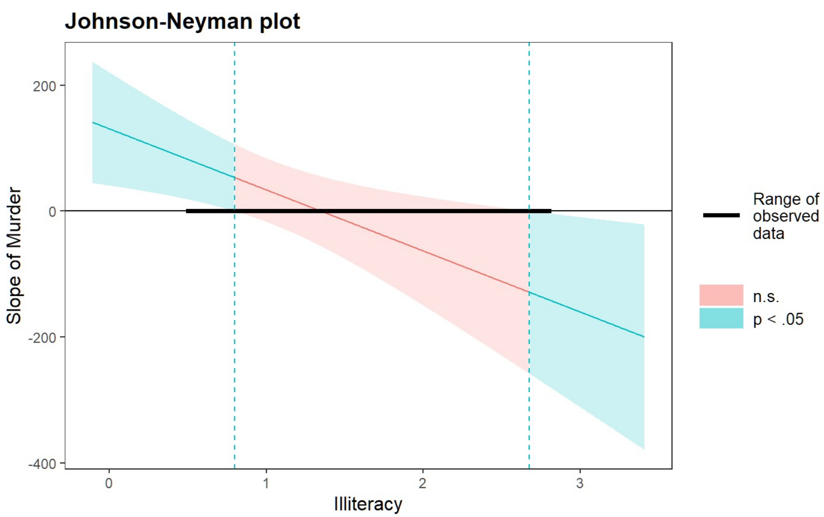
- Main theoretical question: Do we believe that the effects of predictors should be consistent across development?
  - Practical concerns: how to include in our models

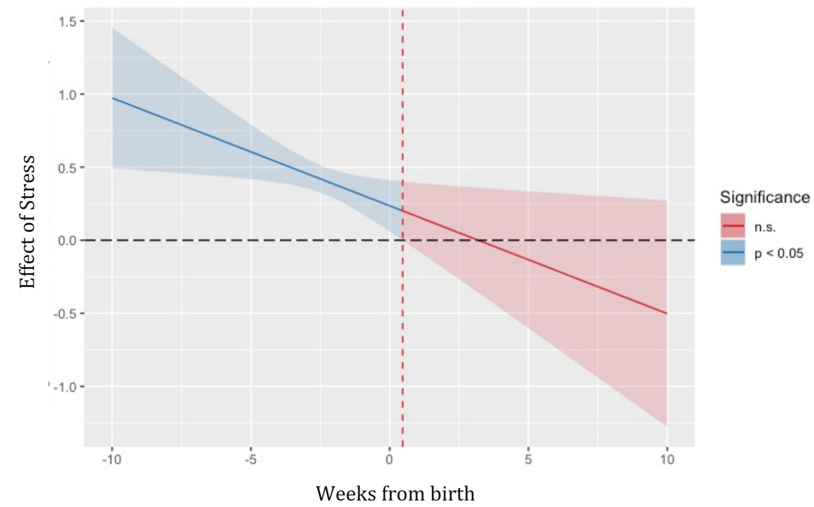


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 2: Modeling Interactions for Sensitive Periods

- Modeling sensitive periods are fundamentally about interactions (e.g., maternal stress and PFC pruning)
  - Main effect: Holding age constant, there is an effect of maternal stress ($\beta_{stress}$)
  - Interaction: The effect of maternal stress depends (in part) on the age of the offspring ($\beta_{stress\ x\ age}$)






McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 2: Modeling Interactions for Sensitive Periods

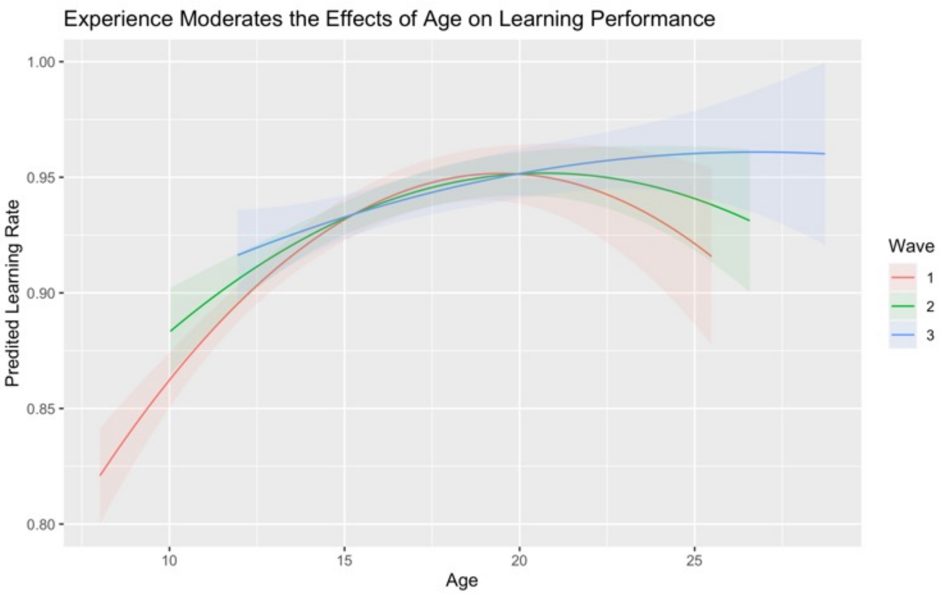
- Simple tests for sensitive periods
  - In the prior example, we have a bilinear interaction ($\beta_{stress}$ changes linearly over age)
  - Quadratic/cubic effects
    - The effect of age itself changes across age in some fashion



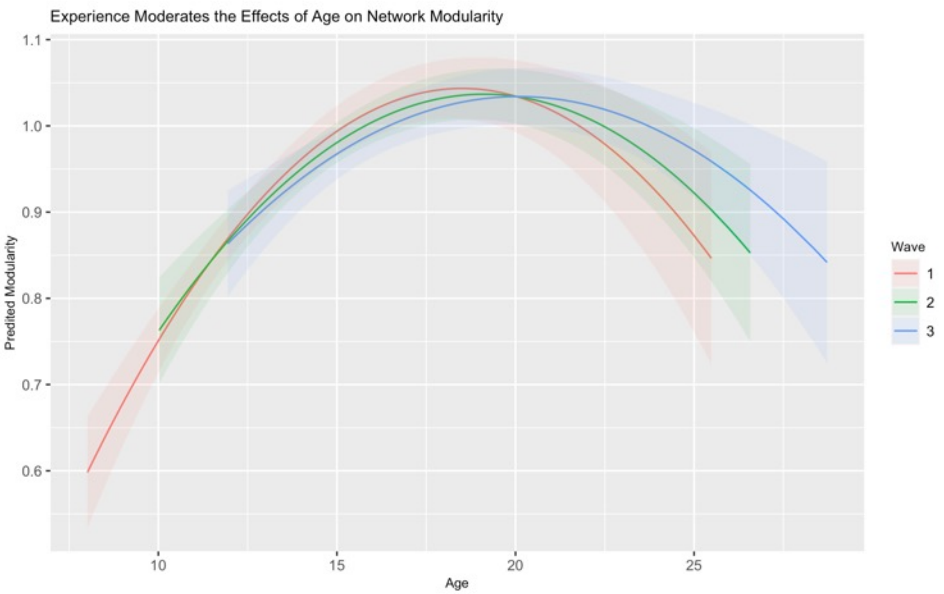
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Interaction Models

## Learning Rate




## Network Modularity






McCormick, E. M., Peters, S., Crone, E. A., & Telzer, E. H. (2021). Longitudinal network re-organization across learning and development. *NeuroImage*, *229*, 117784.

# Use-Case 2: Modeling Interactions for Sensitive Periods

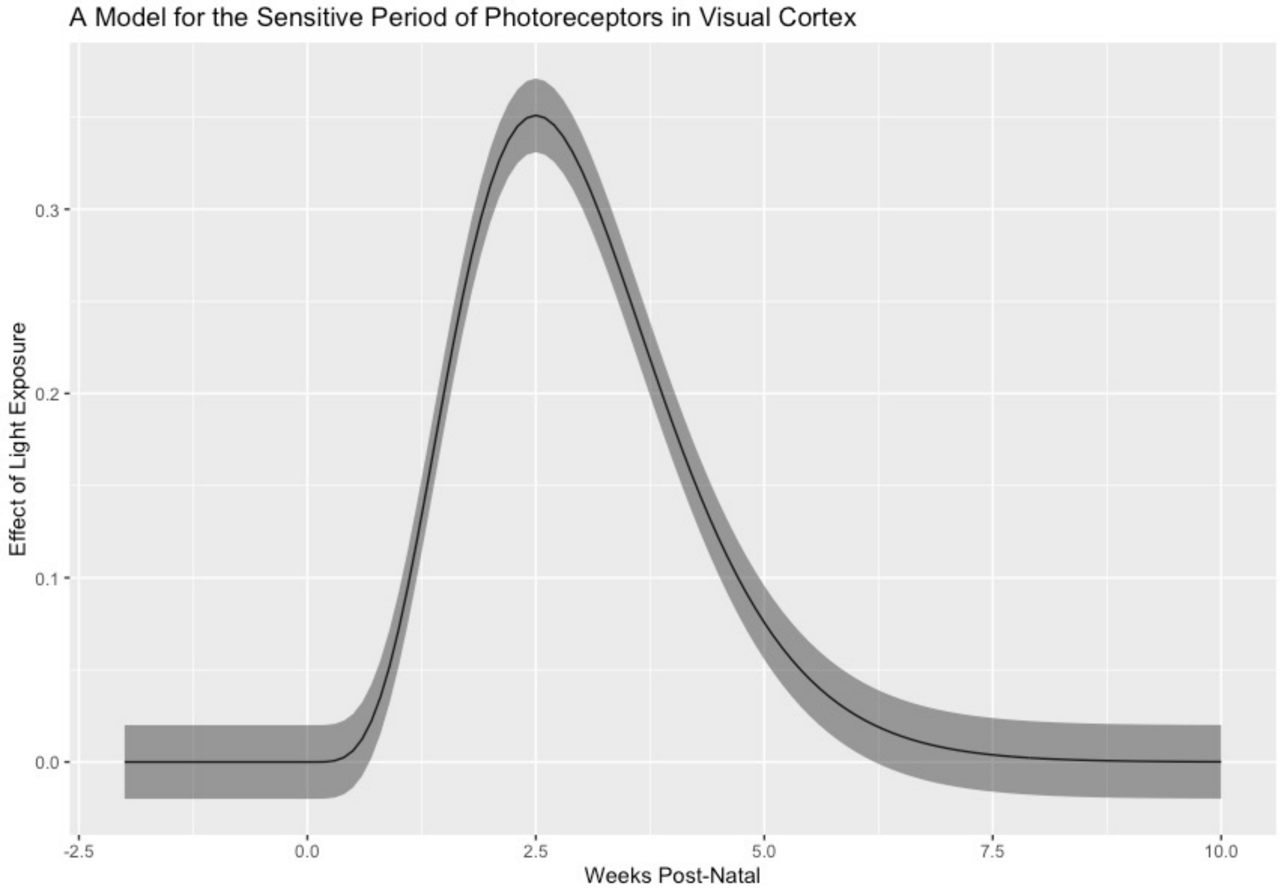
- Simple tests for sensitive periods
  - In the prior example, we have a bilinear interaction ($\beta_{stress}$ changes linearly over age)
  - Quadratic/cubic effects
    - The effect of age itself changes across age in some fashion
  - Generally requires that we pre-specify the sensitive period relationship
    - Can be limiting if the effects are truly nonlinear



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 2: Nonlinear Interactions for Sensitive Periods

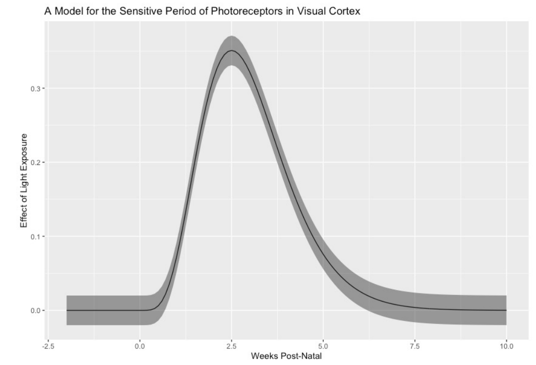
- Nonlinear tests for sensitive periods



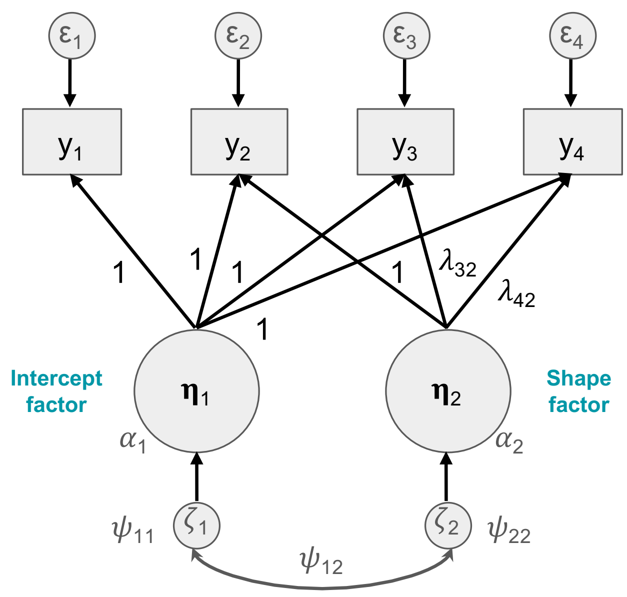



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 2: Nonlinear Interactions for Sensitive Periods




- Nonlinear tests for sensitive periods
  - Generalized additive models
    - Spline-based approach
    - Allows naturally for nonlinear interactions
    - Fixed versus random effects
  - Latent Basis models
    - Free-factor loadings
    - Including predictors of shape factor






McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 2: Nonlinear Interactions for Sensitive Periods

- Nonlinear tests for sensitive periods
  - Often require many more timepoints to model
    - Can benefit from accelerated designs with a fixed effect

- Differential Coupling
  - Require us to move away from univariate (i.e., single-outcome models)
  - How do two (or more) constructs travel together through time?



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Understanding brain-behavior relationships

# Use-Case 3: Brain-Behavior Relationships

- Brain-Behavior Relationships
  - Fundamentally about testing causal explanations for *why* we see particular developmental trajectories

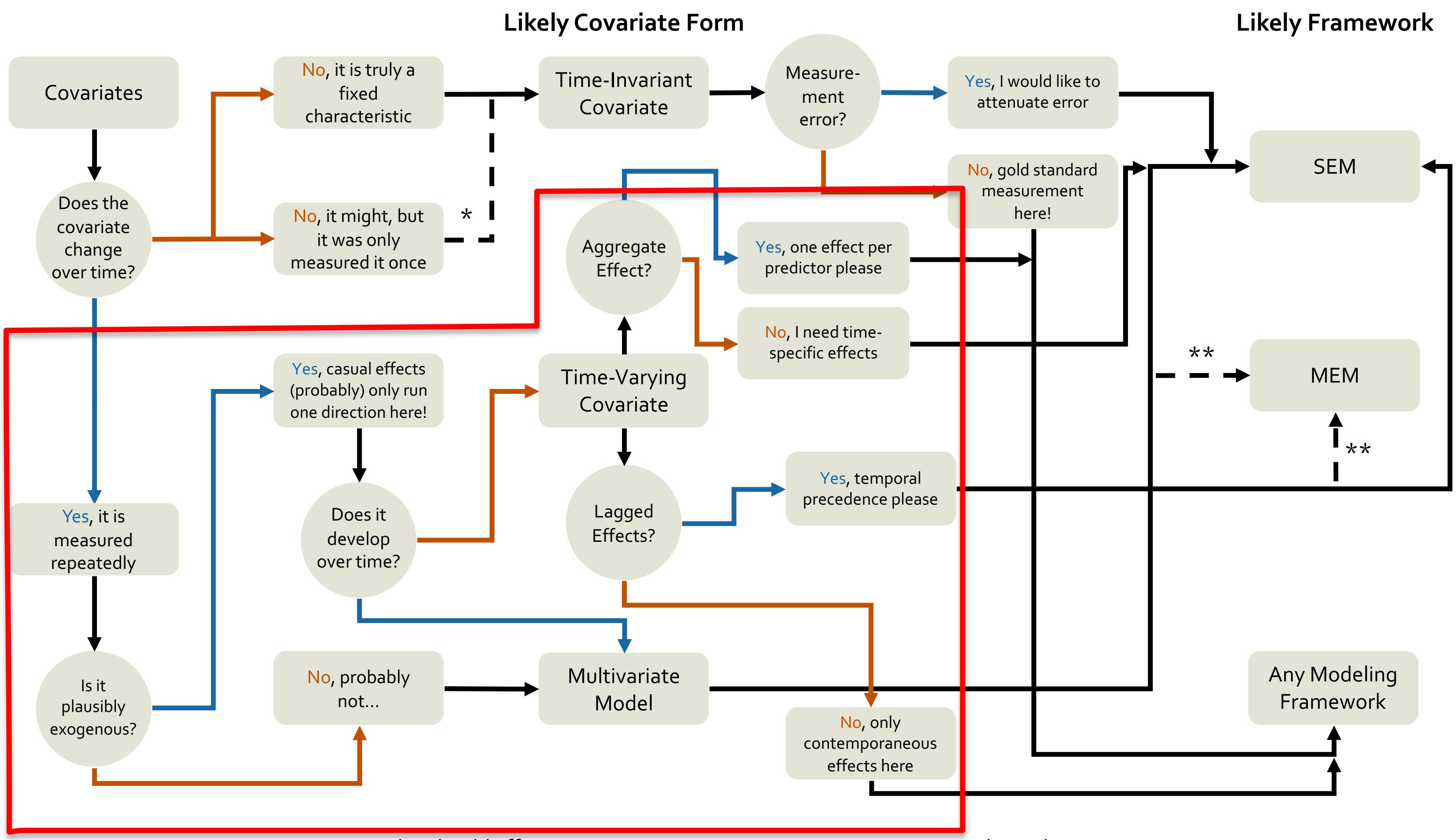
  - Time-invariant Effects
    - Explain stable between-person differences

  - Time-varying Effects
    - Can explain within-person processes
    - Can be exogenous or endogenous



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 3: Brain-Behavior Relationships Heuristic Map




* this should affect your interpretations

** a more complicated option



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

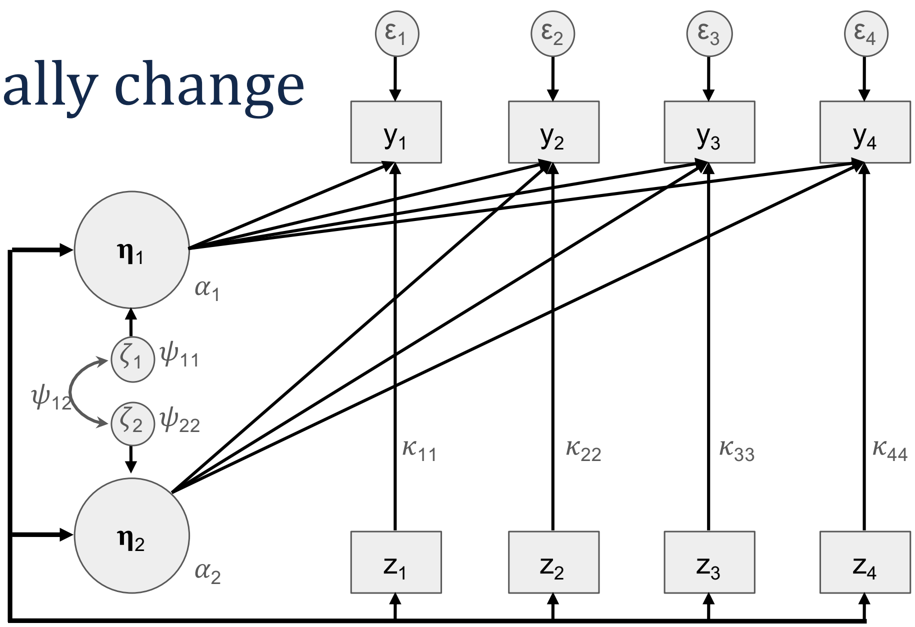
# Use-Case 3: TVC-Models

- Exogenous Time-varying Covariates
  - Only one causal direction
    - TVC "fixed and known"
  - Typically, a univariate model
    - MEM or SEM will work
  - TVC does not systematically change

MLM Equation

$$y_{ti} = \gamma_{00} + \gamma_{10}Time_{ti} + \gamma_{20}TVC_{ti} + u_{0i} + u_{1i}Time_{ti} + r_{ti}$$

SEM Diagram



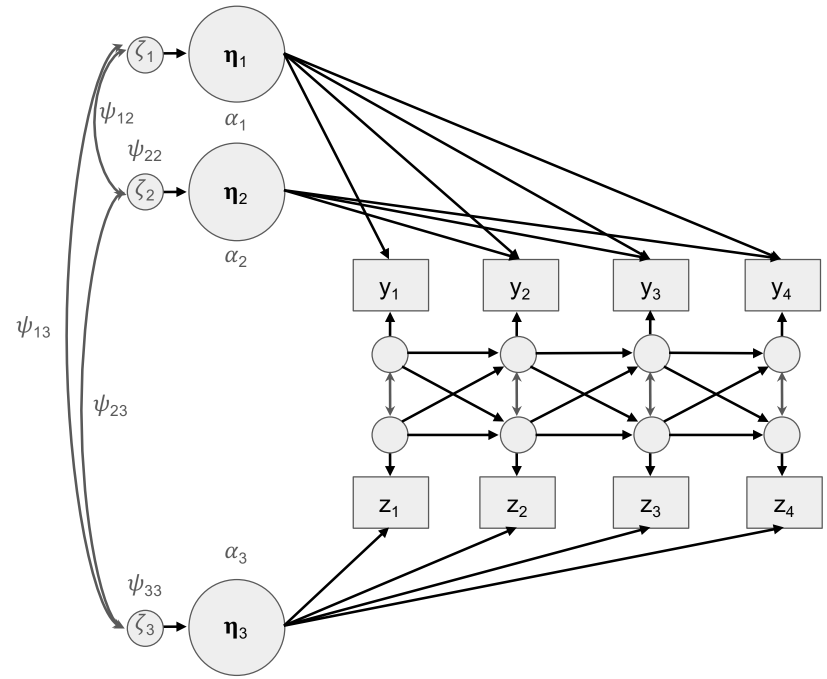



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Use-Case 3: Multivariate Models

- Endogenous (Multivariate) Time-varying Effects
  - Reciprocal relationships over time
    - TVC assumed measured with error
  - Real limitations for MEMs
    - SEMs predominate
  - Different functional forms

SEM Diagram






Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*(5), 879.

# Use-Case 3: Within- and Between-Person Effects

- TICs only explain between-person variance, but TVCs can explain both within- and between variance
  - To separate out these effects, we need to do a little extra work
- MEMs
  - Centering of predictors by extracting the means
- SEMs
  - Structured residuals



Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583-619.

# Conclusions and Next Steps

# Conclusions and Next Steps

- Longitudinal model selection is complex and there are rarely 100% “right” answers
  - We should strive to be better but know that there are inherent limitations
- Understanding model assumptions gives us more reliable estimates of the effects of interes
  - Constrains our conclusions
- "Working out the terms of moral justifica unending task.”
  - Also works for matching models to theory



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker’s Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Conclusions and Next Steps

- More in-depth exploration of these topics: Preprint
  - Including links to lots of primary sources
- Hands on code applications of these models: Online codebook companion
  - Primarily in R but some (future) resources for Mplus/SAS
- Additional training in advanced topics: Flux Pre-conference (details to come)



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

# Especial thanks to these lovely people

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Universiteit Leiden

UNIVERSITY OF MARYLAND

UNIVERSITY OF OREGON

MONASH University

HARVARD UNIVERSITY

Erasmus University Rotterdam

NIH National Institute of Mental Health

NIH National Institute on Drug Abuse

NSF

erc European Research Council

Radboudumc

DONDERS INSTITUTE

NIH National Institute of Mental Health

# Questions?

@McCormickNeuro

ethan.mccormick@radboudumc.nl

https://mccormickneuro.github.io/