Mixed-Effect Regression for the Clinical Sciences

Keith Lohse, PhD, PStat

Department of Health, Kinesiology, and Recreation

Department of Physical Therapy and Athletic Training

University of Utah

keith.lohse@health.utah.edu

@keith_lohse

HEALTH
UNIVERSITY OF UTAH

Why this workshop?

HEALTH
UNIVERSITY OF UTAH

Mixed-effect regression

- Many of us are probably comfortable with mixed-factorial *ANOVA* where we have between-subjects and within-subjects *factors*.
- Many of us are probably familiar *ordinary least squares* regression using the *general linear model*.
- Some of us probably recognize that these analyses are even different (i.e., ANOVA is a special case of OLS GLM).
- Fewer of us are probably familiar mixed-effect regression as an analytical technique.

Mixed-effect regression is new(er)

- If you have heard of it, you have probably heard of some of the advantages that mixed-effects regression over ANOVA (e.g. its ability to handle missing data).

- However, because mixed-effects regression is relatively new (compared to ANOVA),
  - It is not taught in a lot of statistics programs,
  - It has less documentation for a non-specialist audience,
  - It is mostly applied in specialty fields,
  - It is often poorly reported in published literature ("*mixed-muddles*" – S. Senn)

In this workshop, I want to...

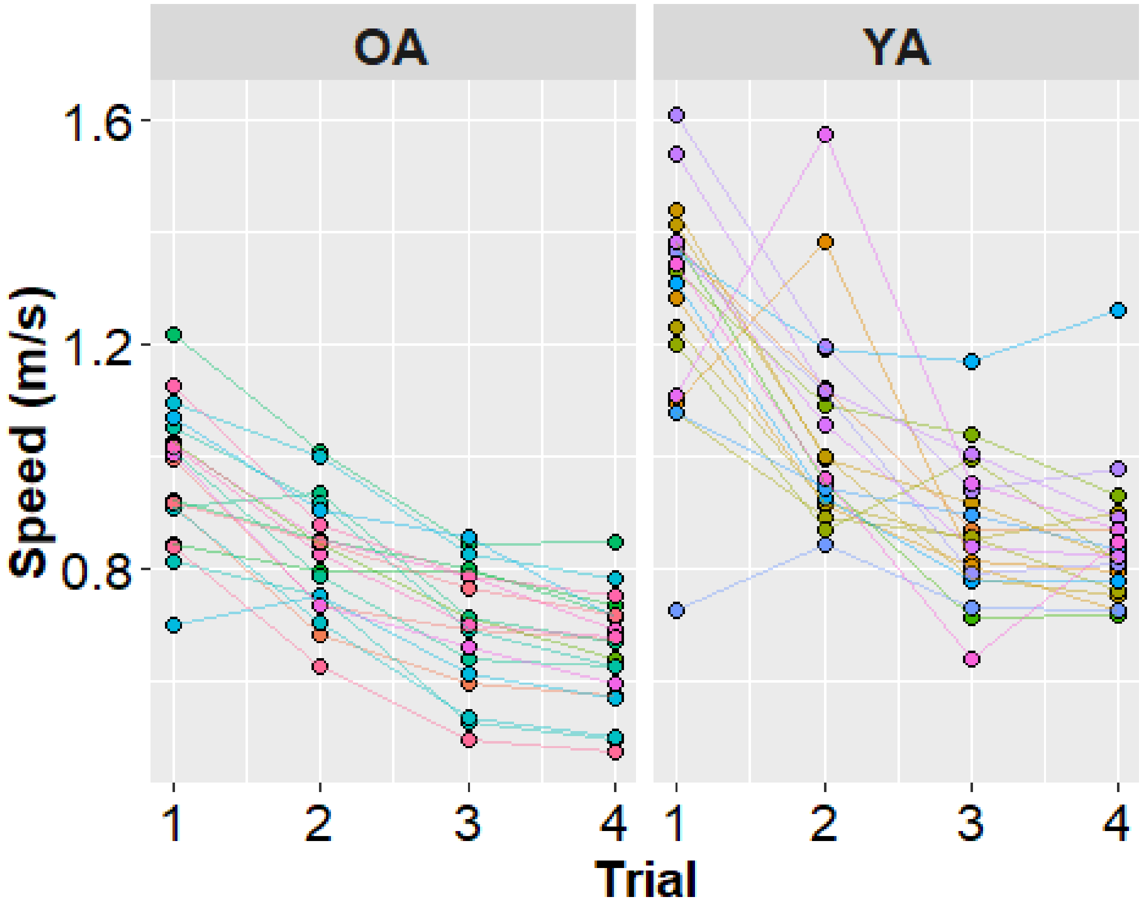
- Provide an introduction to linear mixed-effect regression.
- Discuss its strengths and weaknesses, especially relative to factorial ANOVA.
- Give you data files and code to implement these models in the open programming language R.
- Give us all the opportunity to work on some of your own data.

Timeline

- Monday, 8-10am (10-11am BYOD)
  - Introduction to mixed-effects models.
  - Mixed-model analogues to factorial ANOVA
- Wednesday, 8-10am (10-11am BYOD)
  - Mixed-models for truly longitudinal designs.
- Friday, 8-10am (10-11am BYOD)
  - Mixed-models for time-series data in factorial designs.

What is mixed-effect regression?

HEALTH
UNIVERSITY OF UTAH



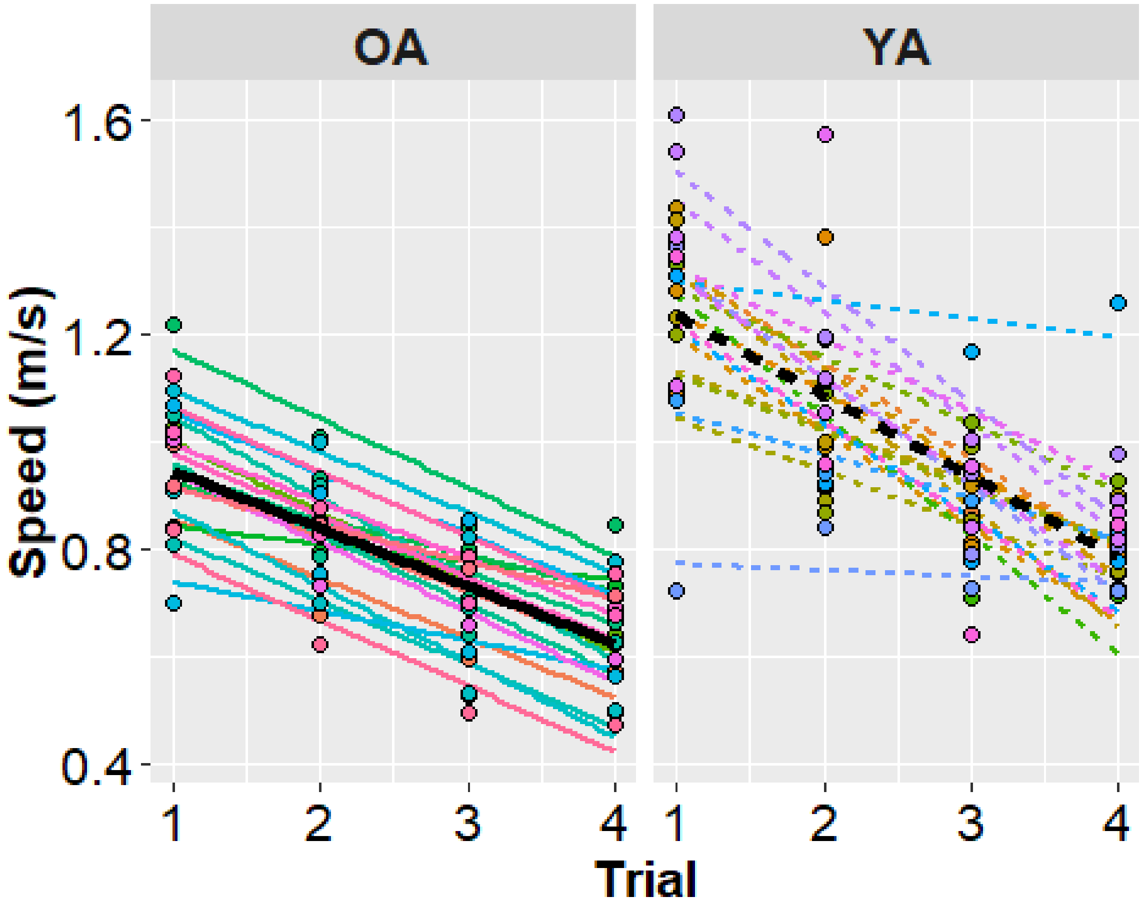
OLS - Regression

$$y_i = \beta_0 + \beta_1(Time_i) + \epsilon_i$$

LME - Regression

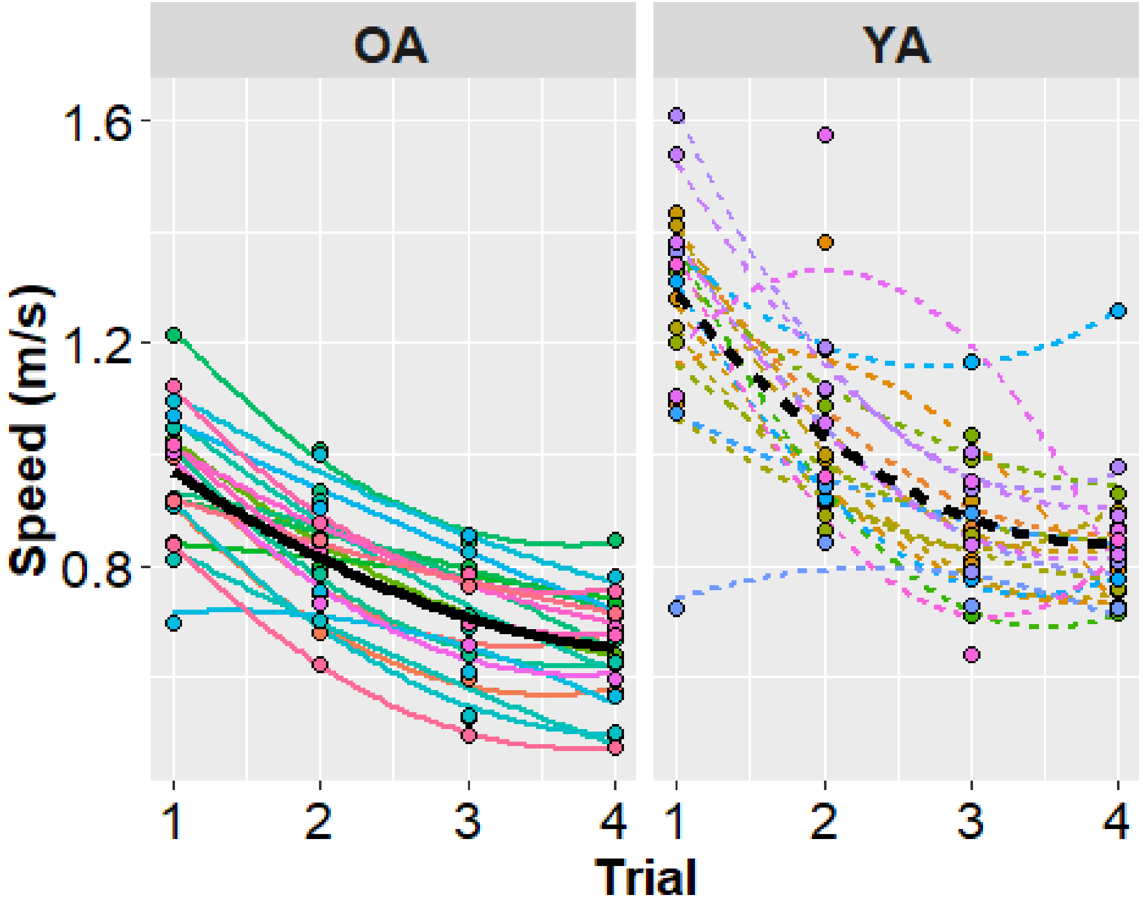
$$y_{ij} = \beta_0 + \beta_1(Time_{ij}) + U_{0j} + U_{1j}(Time_{ij}) + \epsilon_{ij}$$

$$y_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(Time_{ij}) + \epsilon_{ij}$$



LME - Regression

$$y_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(Time_{ij}) + \epsilon_{ij}$$



LME - Regression

$$\cancel{y_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(Time_{ij}) + \epsilon_{ij}}$$

$$y_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(Time_{ij}) + (\beta_2 + U_{2j})(Time_{ij}^2) + \epsilon_{ij}$$

The Mixed-Effects Model:

$$y_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(Time_{ij}) + \epsilon_{ij}$$

- The *MODEL* includes fixed effects and random effects.
- Fixed-Effects are the group-level β's, these effects parallel the traditional main-effects and interactions that you have probably encountered in other statistical analyses.
- Random-Effects are the participant-level U_j's that remove statistical dependency from our data. (This is bit of a simplification, but you can think of not including the appropriate random-effects like running a between-subjects ANOVA when you should be running a repeated-measures ANOVA.)
- The *ERRORS*, or more specifically Random Errors, are the difference between our *MODEL*'s predictions and the actual *DATA*, ϵ_{ij}'s.

Contrasting mixed-factorial ANOVA and mixed-effect regression.

HEALTH
UNIVERSITY OF UTAH

1. Modeling Outcomes Over Time

RM ANOVA

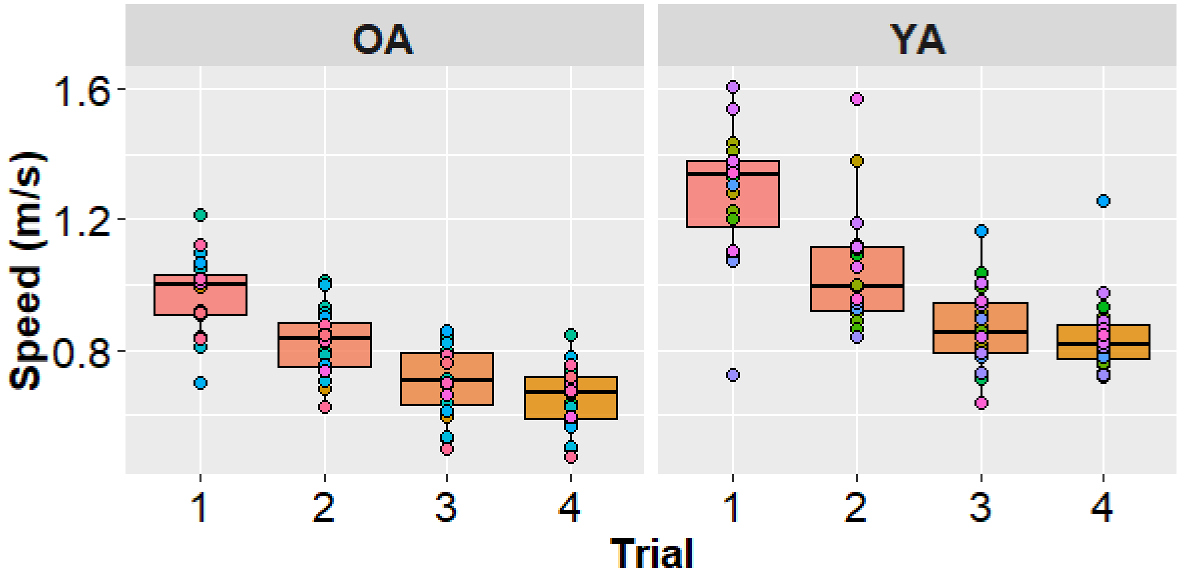
- Addresses questions about mean-differences, after between-subject variance is removed.
- Discrete timepoints are treated as categorical, with only the mean at each timepoint formally considered.

Mixed-Effect Regression

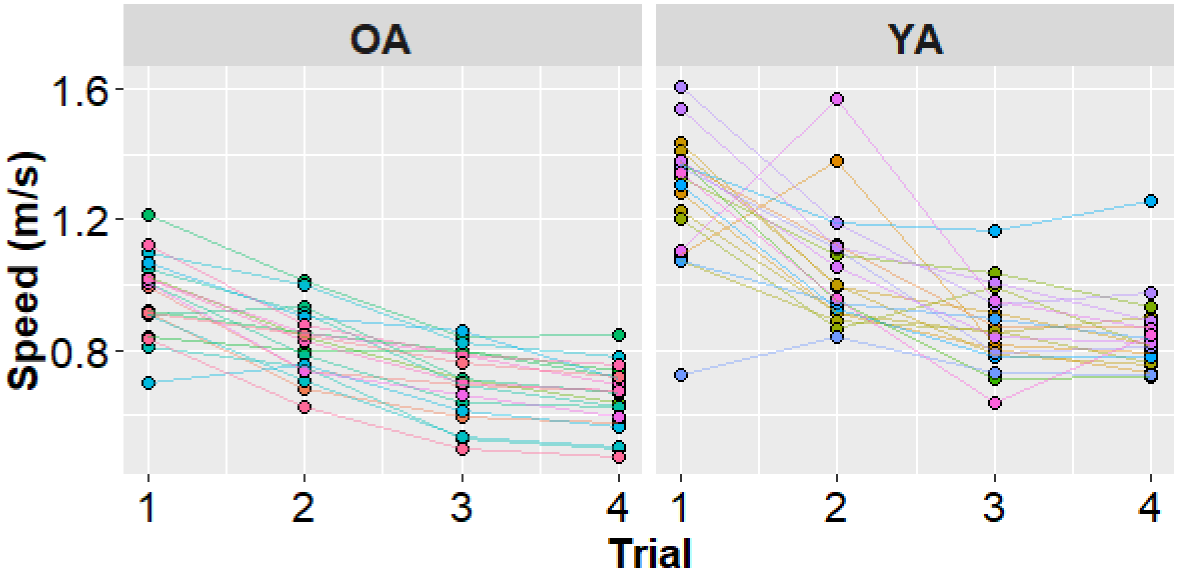
- Time is modeled explicitly as a trajectory for each individual.
- The shape of the trajectory is determined by fitting progressively more complex mathematical functions.

1. Modeling Outcomes Over Time

RM ANOVA



Mixed-Effect Regression



2. Variability in Time

RM ANOVA

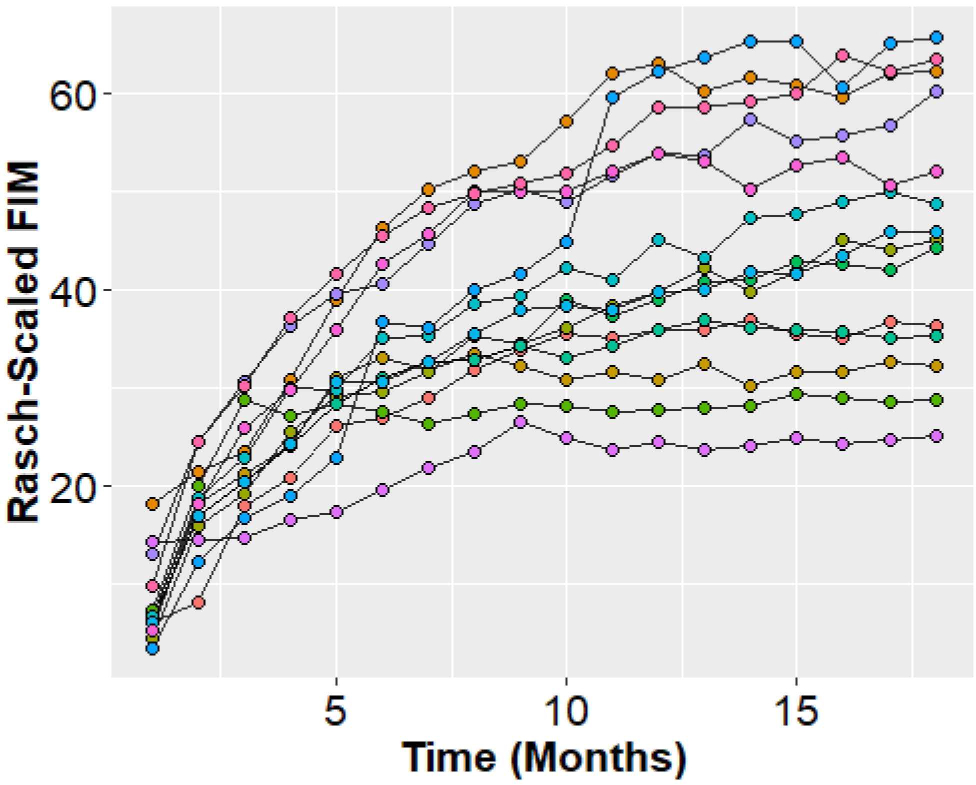
- Assumes common, identically timed data collections.
- This can lead to increased variability in the discrete time-points that is really due to variation in when data were collected.

Mixed-Effect Regression

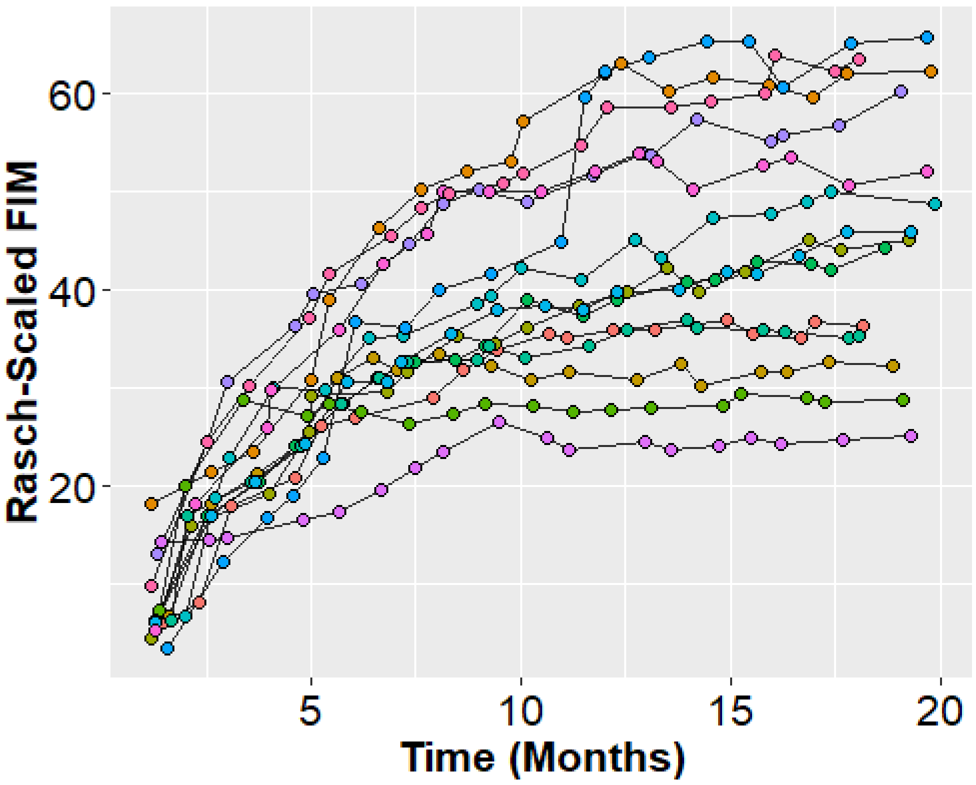
- Can accommodate variability in spacing of time points and in the actual timing of individual data collection.
- The model can also account for increased heterogeneity of the data over time, but residuals still need to be homogeneous.

2. Variability in Time

RM ANOVA



Mixed-Effect Regression



3. Data Missing on the Outcome

RM ANOVA

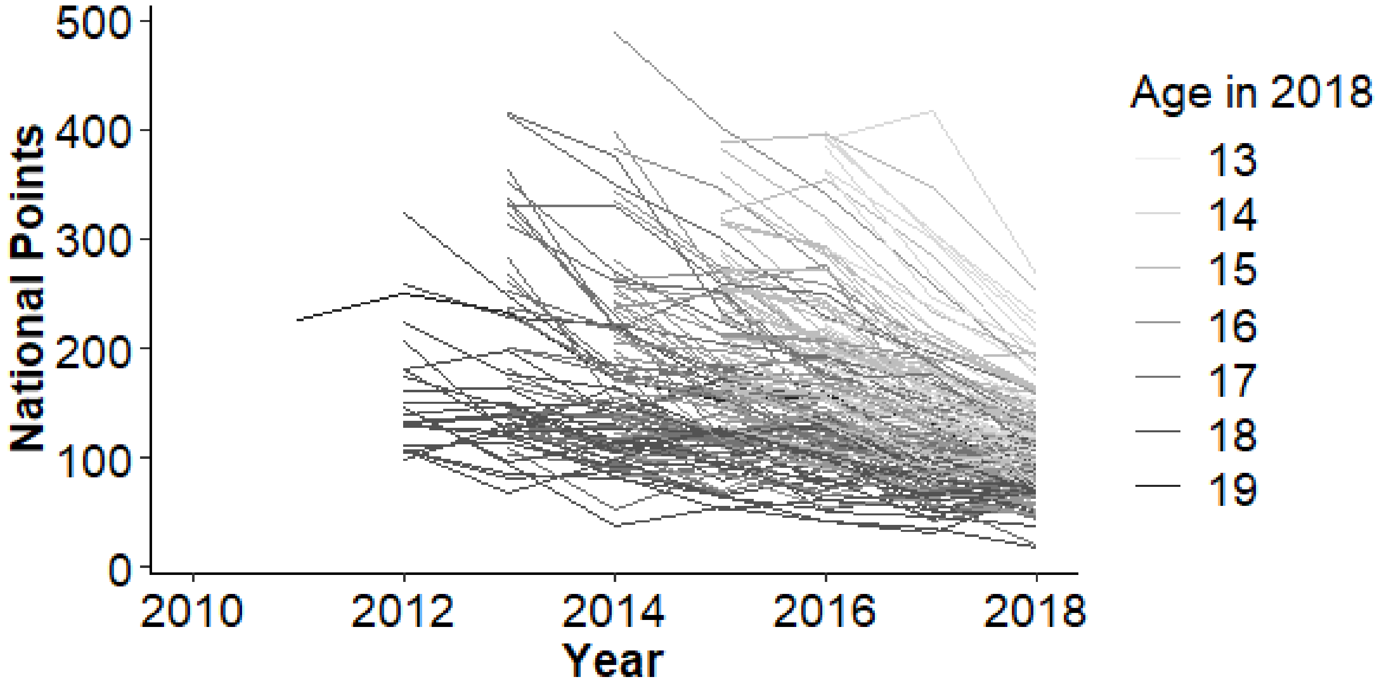
- Missing data generally cannot be accommodated.
- If data are *missing at random* (MAR), multiple imputation can be used for estimation.
- If data are *missing not at random* (MNAR), listwise deletion will reduce statistical power and potentially introduce bias.

Mixed-Effect Regression

- Data that are MAR can be accommodated without exclusion or imputation.
- Data that are MNAR can be fit, but factors associated with missingness need to identified and included in the model.
- If data are missing around key-moments, this may lead to poor model fit/selection.

3. Data Missing on the Outcome

Mixed-Effect Regression



[Lohse, Chen, & Kozlowski, 2020]

4. Data Missing on Explanatory Variables

RM ANOVA

- Missing between-person data in explanatory variables/covariates cannot be accommodated.

- Cases need to either dropped from the model or imputed.

Mixed-Effect Regression

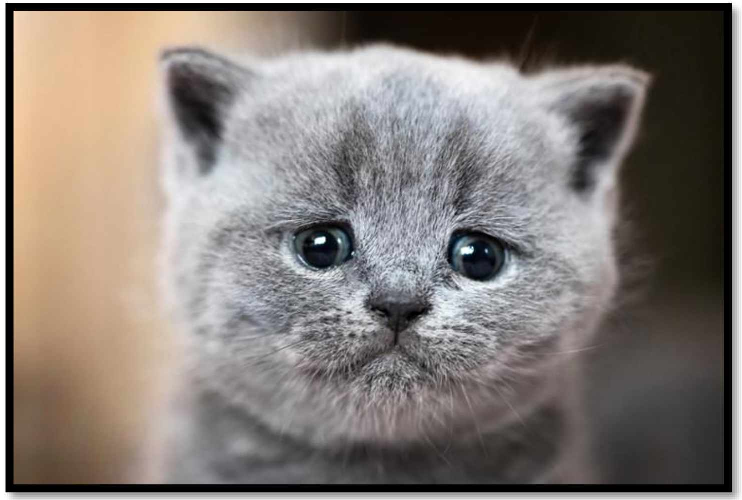
- Missing between-person data in explanatory variables/covariates cannot be accommodated.

- Cases need to either dropped from the model or imputed.

4. Data Missing on Explanatory Variables

RM ANOVA

Mixed-Effect Regression



5. Including Covariates that Change over Time

RM ANOVA

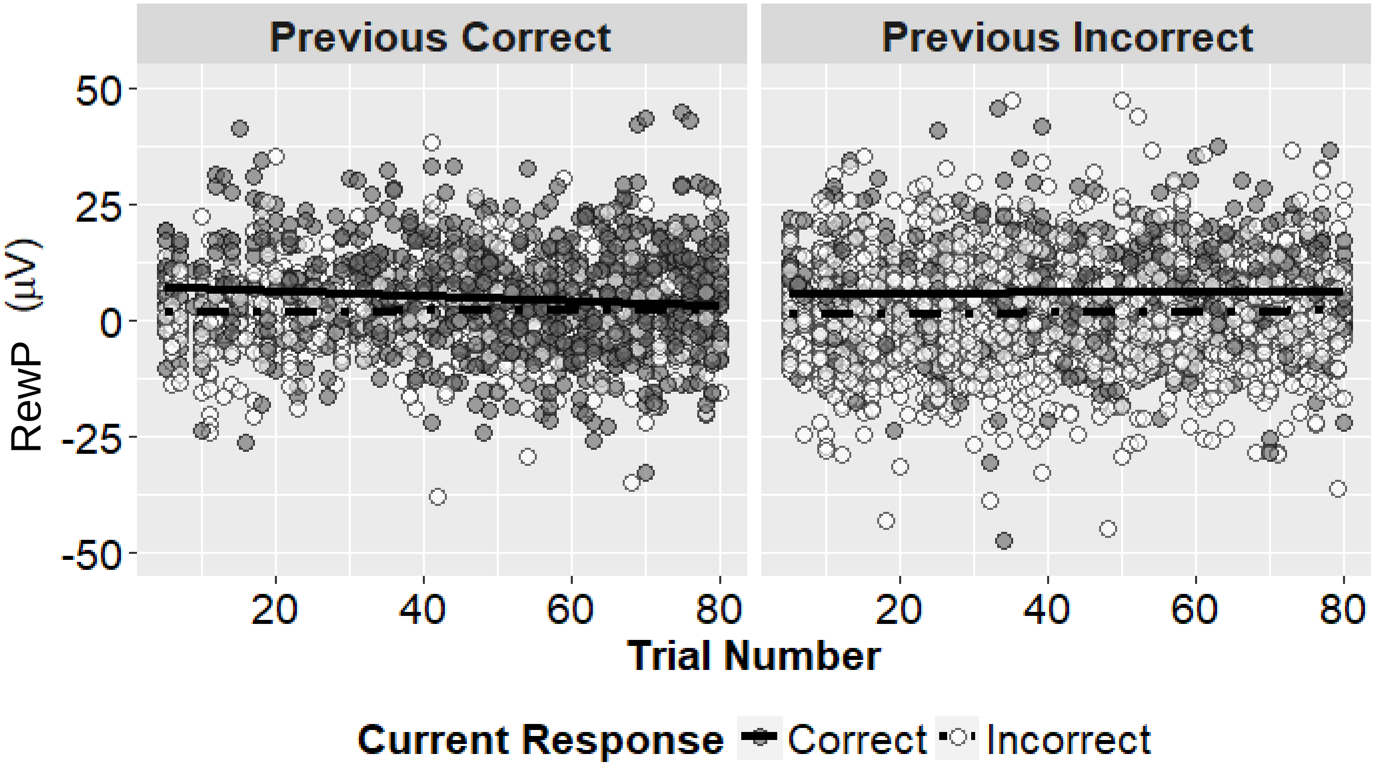
- Time-varying covariates cannot be included in an RM ANOVA model.

Mixed-Effect Regression

- Time varying covariates can be included, but you need to careful about collinearity and variance at both the between- and within-subject levels.

5. Including Covariates that Change over Time

Mixed-Effect Regression



[Lohse, Miller, Daou, Valerius, & Jones, 2020]