

Review of Steps in Unconditional Longitudinal Modeling

For all outcomes:

1. Empty Model; Calculate ICC
2. Decide on a metric of time
3. Decide on a centering point
4. Estimate means model and plot individual trajectories

If your outcome shows systematic change:

5. Evaluate fixed and random effects of time
6. Still consider possible alternative models for the residuals (**R** matrix)

If your outcome does NOT show ANY systematic change:

5. Evaluate alternative models for the variances (**G+R**, or **R**)

1. Empty Means, Random Intercept Model

- Not really predictive, but is a useful parsimonious baseline
 - Fit of “worst” longitudinal model to start building from
 - Partitions variance into between- and within-person variance
- Calculate **ICC** = between / (between + within variance)
 - = Average correlation between occasions
 - = Proportion of variance that is between persons
 - Effect size for amount of person dependency due to mean differences
- Tells you where the action will be:
 - If most of the variance is **between-persons in the random intercept (at level 2)**, you will need **person-level** predictors to reduce that variance (i.e., to account for inter-individual differences)
 - If most of the variance is **within-persons in the residual (at level 1)**, you will need **time-level** predictors to reduce that variance (i.e., to account for intra-individual differences)

2. Decide on the Metric of Time

- “Occasion of Study” as Time:
 - Can be used generically for many purposes—is my preferred default
 - Can still include age, event time as time-invariant predictors of change
- “Age” as Time:
 - Is equivalent to time-in-study if same age at beginning of study
 - Implies age convergence → that people only differ in age regardless of when they came into the study (BP effects = WP effects)
- “Distance to/from an Event” as Time:
 - Is appropriate if a distinct process is responsible for changes
 - Also implies convergence (BP effects = WP effects)
 - Only includes people that have experienced the event
- Make sure to use exact time regardless of which “time” used

3. Decide on a Centering Point

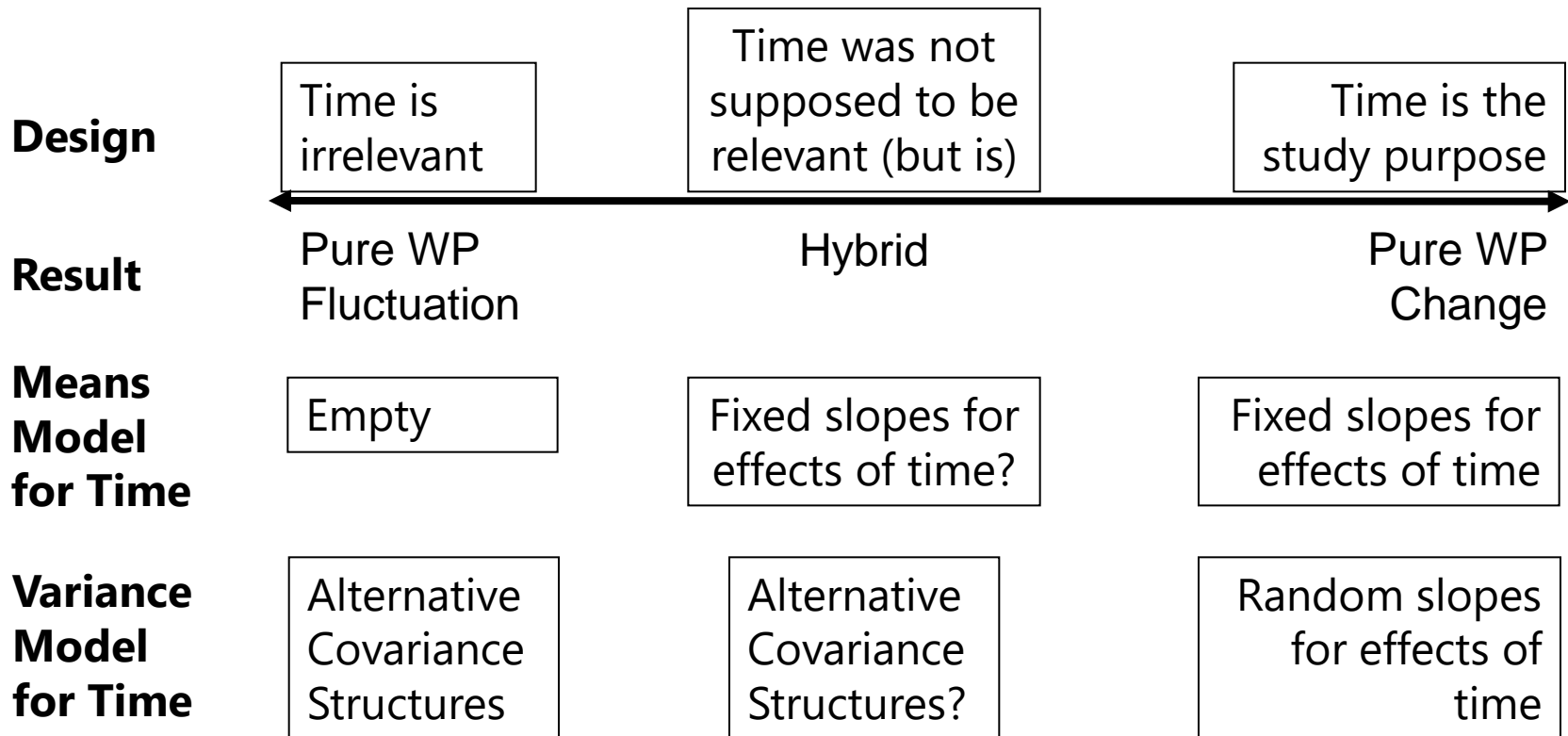
- How to choose: At what occasion would you like a snap-shot of inter-individual differences?
 - Intercept variance represents inter-individual differences at that particular time point (that you can later predict!)
- Where do you want your intercept?
 - Re-code time such that the centering point = 0
 - Use ESTIMATE statements to get predictions at other times
- Different versions of time = 0 will produce statistically equivalent models with re-arranged parameters
 - i.e., conditional level and rate of change at time 0

4. Plot Saturated Means and Individuals

- If time is balanced across persons:
 - Estimate a saturated means model to generate means
- If time is NOT balanced across persons:
 - Create a rounded time variable to estimate means model ONLY
 - Still use exact time/age variable for analysis!
- Plot the means – what kind of trajectory do you see?
- Please note: ML/REML estimated means per occasion may NOT be the same as the observed means (i.e., as given by PROC MEANS). The estimated means are what would have been obtained *had your data been complete* (assuming MAR), whereas observed means are not adjusted to reflect any missing data (MCAR). Report the ML/REML estimated means.

What if I have no change?

- Longitudinal studies are not always designed to examine systematic change (e.g., daily diary studies)
- In reality, there is a continuum of fluctuation to change:



5. and 6. for **Systematic Change**: Evaluate Fixed and Random Effects of Time

Model for the Means:

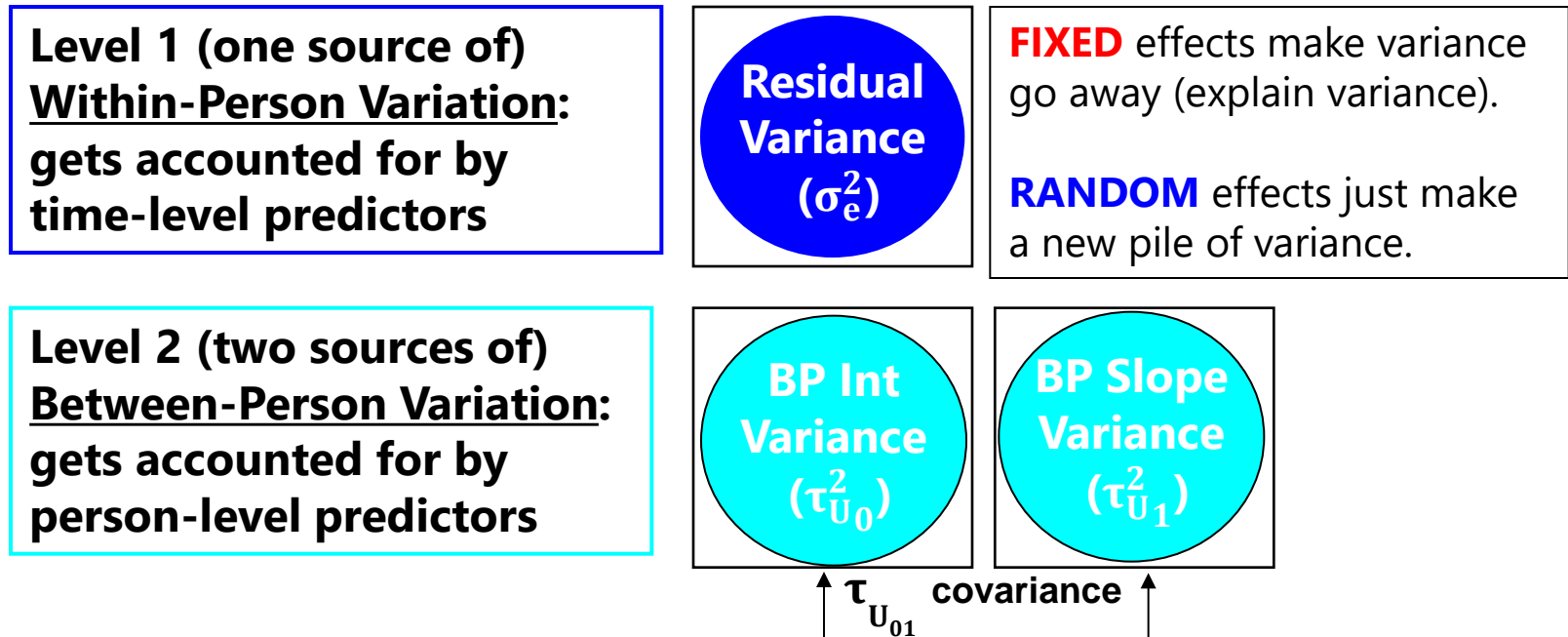
- What kind of fixed effects of time are needed to create a function with which to parsimoniously represent the observed means across occasions?
 - Linear or nonlinear? Continuous or discontinuous?
 - Polynomials? Pieces? Log time? Truly nonlinear curves?
 - Use obtained p -values to test significance of fixed effects

Model for the Variance (focus primarily on G):

- What kind of random effects of time are needed:
 - To account for individual differences in aspects of change?
 - To describe the variances and covariances over time?
 - Do the residuals show any covariance after accounting for random effects?
 - Use REML $-2\Delta LL$ tests to test significance of new effects (or ML if big N)

Random Effects Models for the Variance

- Each source of correlation or dependency goes into a new variance component (or pile of variance) until each source meets the usual assumptions of GLM: normality, independence, constant variance
- **Example 2-level longitudinal model:**



Now we get to add predictors to account for each pile!

5. for **NO Systematic Change**: Evaluate Alternative Covariance Structures

Model for the Means:

- Be sure you don't need any terms for systematic effects of time
- If not, keep a fixed intercept only

Model for the Variance (focus primarily on **R):**

- How many parameters do you need to predict the original data?
- I recommend the hybrid: Random Intercept in **G** + Structure in **R**
 - Separates BP and WP variance
 - Likely more parsimonious than just **R**-only model
- Compare alternative models with the same fixed effects
 - Nested? REML $-2\Delta LL$ test for significance
 - Non-nested? REML AIC and BIC for "supporting evidence"

Alternative Covariance Structure Models

- Models for fluctuation typically include only a covariance structure, and at most a random intercept (random slopes for time won't help in the absence of systematic change)

Between-Person Random Intercept in G + Within-Person Structure in R

Level 1 (one source of) Within-Person Variation:

Gets accounted for by time-level predictors

Residual Variance
(σ_e^2)

Level 2 (one sources of) Between-Person Variation:

Gets accounted for by person-level predictors

BP Int Variance
($\tau_{U_0}^2$)

TOTAL Structure in R

All sources of variation and covariation are held in one matrix, but if dependency is predicted accurately then it's ok.

Total Variance
(σ_T^2)

Why spend so much effort on unconditional models of time? Here is the reasoning...

- The fixed effects of time are what the random effects of time are varying around...
- The random effects of time form the variances that the person-level predictors will account for...
- The effects of person-level predictors are specified as a function of the time effects already in the model...
- The effects of time-varying predictors are supposed to account for variance not accounted for by the model for time...
- What fixed and random time effects of time you include in the model dictate what is to be predicted. **Get time right first!**

Example Sequence for Testing Fixed and Random Polynomial Effects of Time

1. Empty Means, Random Intercept → to calculate ICC
2. Fixed Linear, Random Intercept → check fixed linear p -value
 - Effect size via pseudo- R^2 : Fixed linear time should reduce level-1 σ_e^2 relative to #1
3. Random Linear → check $-2\Delta LL(df \approx 2)$ for random linear variance
 - No effect size via pseudo- R^2 (reallocated part of level-1 σ_e^2 to new level-2 pile instead)
 - Now two kinds of level-2 BP differences to be predicted later
4. Fixed Quadratic, Random Linear → check fixed quadratic p -value
 - Effect size via pseudo- R^2 : Fixed quadratic time should reduce level-1 σ_e^2 relative to #3
5. Random Quadratic → check $-2\Delta LL(df \approx 3)$ for random quadratic variance
 - No effect size via pseudo- R^2 (reallocated part of level-1 σ_e^2 to new level-2 pile instead)
 - Now three kinds of level-2 BP differences to be predicted later
6. (continue with additional fixed and random terms as needed)

Example Sequence for Testing Two Piecewise Slopes (S1, S2) for Effects of Time

1. Empty Means, Random Intercept → to calculate ICC
2. Fixed S1, Fixed S2, Random Intercept → check fixed slopes p -values
 - Effect size via pseudo- R^2 : Fixed time slopes should reduce level-1 σ_e^2 relative to #1
3. Random S1, Fixed S2 → check $-2\Delta LL(df \approx 2)$ for random S1 variance
 - No effect size via pseudo- R^2 (reallocated part of level-1 σ_e^2 to new level-2 pile instead)
 - Now two kinds of level-2 BP differences to be predicted later
4. Random S1, Random S2 → check $-2\Delta LL(df \approx 3)$ for random S2 variance
 - No effect size via pseudo- R^2 (reallocated part of level-1 σ_e^2 to new level-2 pile instead)
 - Now three kinds of level-2 BP differences to be predicted later
5. (continue with additional fixed and random terms as needed, such as adding quadratic slopes for each piece)