Analysis plan

1 Choosing the right outcome model

There are thus three types of mean models one can consider:

- Follow-up only: models only the follow-up response, ignoring baseline.
- Change analysis models the difference in outcome from baseline to follow-up. First, the researcher would compute differences $(y_{i,j} - y_{i,1})$ and use these for analysis.
- Regressions models would model the follow-up outcomes, adjusting for the baseline value of the outcome.

2 Choosing the right predictor model

One needs to examine the relationship between the outcome and the predictor. When time is the only predictor. Check how the outcome is changing over time. In case, there are other predictors, you need to check those as well. There are two types of predictors: 1) Baseline-only predictors (like age, treatment assignment variable, disease status, etc) and 2) time-varying predictors (like biomarker measurements, or even disease status if it changes over time, etc). We often only include the baseline values in the analysis, even for the biomarkers. It depends on the research qn. For example, one may ask 1) Having a certain level of cholesterol level at the beginning of the trial, how a certain drug is affecting the SBP level. Then baseline cholesterol level should be the predictor. Over a short period of time cholesterol level does not change too much. 2) But if the trial goes on for several years and the drug or several other factors may even impact the cholesterol level, we should include this variable as a time-varying predictor.

$$\begin{aligned} \text{Baseline model} \quad & y_{i,j} = \alpha + \beta x_i + \epsilon_{i,j} \\ \text{Time-varying predictor} \quad & y_{i,j} = \alpha + \beta x_{i.j} + \epsilon_{i,j} \end{aligned}$$

It again depends on the dataset and the research question to choose the right model.

3 Error model

If the outcome is continuous, the most common distribution is normal. To specify a normal distribution, we need two parameters, one for the mean and another for the variance. The expectation will be $E(\epsilon_{i,j}) = 0$. But for variance, there are several possibilities. We need to ask the following questions:

- Is there any predictor controlling the variance? In that case, we must consider a mixed model and adjust for those predictors with random effects.
- Model the residual variance directly and appropriately. In this case, first, examine the 'observation times' both from the data or the study description and decide which covariance structures are appropriate. In the case of a balanced design, certain covariances are appropriate like autoregressive, moving average, unstructured dense covariance, unstructured sparse covariance, compound symmetry, etc. in the case of discrete observation times. When observation times are continuous, exponential or Gaussian kernel-based covariances are more appropriate.
- There may be missing data. However, as long as the 'ignorable' missing data assumption holds it should not impact the choice of covariance.

4 Inference

- Testing or comparing different covariance structures and/or mean structures to select a reasonable model. Selection of the mean model usually only requires plotting the data. Selection of the right covariance structure requires testing or comparing.
- Testing the regression coefficients directly using Wald or LRT tests.
- Contrast-based one-degree of freedom inference.

5 Things to be discussed

- Handling missing data.
- Nonparametric modeling of the longitudinal data