

## Project for MS/PhD Participants

### Phase III

Please use the following disclaimer in the first slide, and call it drug X instead of Brodalumab throughout the presentation.

Disclaimer of data: The data is simulated and is not real data from clinical study. They are for educational and exercise purpose only.

Disclaimer: The design of the trial is created based on publicly available open-source information in immunology therapeutic area. It is for educational purpose only. The presentations reflect the views of the speakers based on their understanding of the open source information and simulated data, and have not been evaluated by University of connecticut. The materials cannot be used for promotional activities. They are not intended to diagnose, treat, cure or prevent any disease.

The simulated data contains complete data. It means that every patient has data collected for endpoint at visit 6. In a real clinical trial, there are usually patients who drop out in the middle of the study, which creates missing data problems. There are three commonly seen missing mechanisms:

MCAR (Missing completely at random): The probability of missing is the same for all patients.

MAR (Missing at random): Most missingness is not completely at random. But if the probability of missing only depends on available information in the analysis model (e.g. baseline characters or the response data in previous visits before visit 6), then it is called missing at random.

MNAR (Missing not at random): Missingness that depends on unobserved predictors. For example, if the treatment causes discomfort or toxicity, then a patient is more likely to drop out.

In this analysis, we will focus on PASI score data (ADPA dataset). Please simulate missing data according to the following mechanism, and analyze data using different approaches.

1. Please randomly select 10% patients from the overall population and set their PASI scores at Visit 6 to be missing. This is a case for missing completely at random (MCAR).
2. Analyze the data using PASI75 (binary data of responder or non-responder) by the following three methods:
  - a. Include only completers. Patients who drop out and have missing data will be excluded for analysis. Analyze by logistic regression with treatment and sex (from ADSL data).
  - b. Impute the patients who have missing data at visit 6 to be non-responders (i.e. patients who have missing data or who have visit 6 data but didn't reach PASI75 will be both treated as non-responder). Analyze by logistic regression with treatment and sex.
  - c. Compare the results with the one from complete data set. Analyze by logistic regression with treatment and sex.
3. Repeat the exercise in 1 and 2 for 20% and 30% of missingness and compare the results.

4. Set the probability of missing to be correlated with the change from baseline for PASI score. Starting from visit 4, if the PASI score for visit 3 is higher than baseline or more than 90% of baseline score (<10% improvement), then the probability of missing at visit 4 is 30%, otherwise the probability of missing is 5%. If visit 4 is missing, then all subsequent visits are missing. Repeat this for visit 5 based on visit 4 values, and for visit 6 based on visit 5 value. This is a case for missing at random (MAR).
5. Repeat the exercise in 2 for data generated in 4.
6. Bonus point: Create MNAR mechanisms that you are interested.
  - a. Hint 1: For example, there could be an underlying subject-level random variable called “toxicity index”, whose mean increases with increasing dose (and mean for the active control arm is the same as the higher dose of test drug), with subject random variability. The probability of missing is positively correlated with the underlying “toxicity index”. You may choose to create your own index based on dose, or use some AE data in the ADAE dataset (Note that AE occurrence date is relative to study start date in this dataset, and the visits are equally spaced in the ADPA data). Then repeat the exercise above.
  - b. Hint 2: Another example could be that the probability of missing can depend on the value of response at visit 6. The less improvement, the more likely to be missing.