# Homework 4

*Biost 540*

**General Instructions**:

- When working in groups, elect one member as the leading member who will be responsible for uploading the homework assignment solutions. Please note that all members in the group are expected to equally contribute to the assignment!
- Be sure to show work for all problems! `R` code (and output) should not appear in the main body of the homework; however, the code should appear at the end of the assignment as an Appendix. It should be possible for someone to use the code to reproduce any figures or numeric results.

## Part A: Sixcity Data - Missing Data

1. Perform an exploratory data analysis to describe the missingness pattern in the dataset.
2. Using the complete dataset from Homework 3 fit a **conditional** model for respiratory function with only the main effects of smoking status and age.
3. Using the data with missingness fit a **conditional** model for respiratory function with only the main effects of smoking status and age.
4. Using the data with missingness fit a **conditional** model for respiratory function with only the main effects of smoking status and age – but on imputed datasets. Specifically:
   - Obtain imputations using the data in long format and account for the clustering of the data at the patient level and the method "2l.bin".
   - Obtain imputations using the data in wide format and the method "logreg". Be sure to assess convergence of the imputation procedure.
   - For both imputation strategies fit the models over imputed datasets, pool the results and obtain the estimated coefficients.
5. Compare the results from items 2), 3) and 4). What do you conclude?

## Part B: IMPS Data

The `imps` data set is data from National Institute of the Mental Health Schizophrenia Collaborative Study, where the effect of chlorpromazine, fluphenazine, or thioridazine treatment on the overall severity of the schizophrenia disorder is of interest. More information about the dataset can be found in the documentation for the dataset:

```
library(wgeesel)
data(imps)
?imps
```

1. Perform an exploratory data analysis to describe the missingness pattern in the dataset.
2. Fit a **marginal** model for indicator of high severity of schizophrenia (i.e. `Y`) with main effects for time, indicator of treatment, and sex using all available data.
3. Fit a **marginal** model with IPW GEE using the same outcome and predictors as (2) assuming the following missing models:
   - Missingness depends on time and treatment
   - Missingness depends on time, treatment, sex, and most recent previous outcome
4. Compare the results from (2), (3a), and (3b). What do you conclude?