**PUBH 7485/8485**

**Simulation Exercise**

The primary response for this study is the CD4 count at two years with the goal of giving combinations of treatments that will result in the greatest average CD4 count at two years. The data were generated as described in the class notes. In particular, note the following

* All participants begin without any treatment resistance.
* If a participant has not taken any treatment previously, there is no probability of becoming resistant.
* Once a participant becomes resistant, they remain resistant to treatment at subsequent time periods.
* A person’s log CD4 count at a particular time point may depend on prior treatment history, prior log CD4 history, resistance history (including the current value), and possibly interactions.
* To simplify things, whether or not a person becomes newly resistant to treatment at a given timepoint only depends on prior treatment history.

These data are meant to represent results from an observational study. Therefore, how the covariates evolve over time and how treatment was assigned to patients over time is not known to you but you can be assured that the treatment assignment followed the sequential randomization assumption.

The goal here is to conduct several simulated clinical trials with a large number of participants (say 100,000) to evaluate different treatment sequences (see below). For example, suppose we want to study the treatment sequence (0, 0, 1, 1), that is, do not give treatment at the first two time points but give treatment at the last two time points. In particular, you should

1. Simulate baseline log CD4 count for all participants.
2. Assign baseline treatment based on the strategy being studied (i.e., no treatment in our example).
3. Simulate (a) resistance status and (b) log CD4 at 6 months.
4. Assign treatment at 6 months based on the strategy being studied (i.e., no treatment in our example).
5. Simulate (a) resistance status and (b) log CD4 at 12 months.
6. Assign treatment at 12 months based on the strategy being studied (i.e., no treatment in our example).
7. Simulate (a) resistance status and (b) log CD4 at 18 months.
8. Assign treatment at 18 months based on the strategy being studied (i.e., no treatment in our example).
9. Simulate (a) resistance status and (b) log CD4 at 24 months.

Now that you have data on outcomes at 24 months for 100,000 trial participants following a particular treatment sequence, you can compute any summary measure you would like (e.g., mean of CD4 count).

In order to simulate the data in steps 1, 3, 5, 7, and 9, you will need a statistical model. The observational data that you have can be used to fit such a statistical models.

Using Monte Carlo methods, estimate for the following treatment combinations

Where is the potential CD4 count at 2 years had individuals been treated using combination

If you have time, you can implement a bootstrap procedure to estimate the uncertainty (e.g., standard error, 95% CIs for each of these estimates). In particular, to obtain B bootstrap estimates, you would sample with replacement from the original, observational dataset and reconduct the 100,000 person clinical trial to obtain the estimate of .