**Lab G-formula**

**P1822 – Statistical Methods for Causal Inference**

**Linda Valeri**

We analyze data from an observational study enrolling smokers, an example presented in Chapter 13 of the Causal Inference handbook by Hernan and Robins. Interest lies in investigating the causal effect of smoking cessation on weight gain in the span of 10 years.

The **nhefs** dataset is posted on CANVAS along with the codebook and R code.

Note that different models can be used and different choices regarding confounding adjustment can be made. I encourage you to work together and using slightly different choices than what done in the code.

1. Inspect the exposure (qsmk) and outcome (wt82\_71) distributions, consider whether transformations are needed.
2. Have a look at the covariates in the codebook and justify the choice of inclusion in the confounders set
3. Fit the outcome regression model unadjusted for confounders and interpret the coefficient for the smoking cessation variable
4. Fit the outcome regression model adjusting for confounding and interpret the coefficient for the smoking cessation variable
5. Write the g-formula for this observational study
6. Estimate the standardized outcome in the observed sample, in the hypothetical scenario when everyone quits smoking and in the hypothetical scenario when no one quits smoking
7. Obtain estimate, standard error, and confidence intervals for the marginal causal effect of smoking cessation on weight gain.