Causal Group Project-Preliminary Project Description

Group Members: Shelley Facente, Steph Holm, Lizzy Kinnard, Veronica Pear

3/26/2019

Target Population/Inclusion Criteria of Study

These data are from the National Longitudinal Survey of Youth 1997. This was recruited as a nationally representative cohort of youth age 12-16 (initial n=9000) in 1997. These youth have since been followed longitudinally. The target population is youth in the United States.

Causal Question:

What is the effect of having been bullied prior to age 12 on incidence of drug use in adolescence or adulthood?

Target Causal Parameter:

The target causal parameter is the difference in the counterfactual probability of drug use, if all kids were bullied prior to age 12, and the counterfactual probability of drug use, if all kids were not bullied prior to age 12:

$$\psi^F(P_{U,X}) = P_{U,X}(Y_1 = 1) - P_{U,X}(Y_0 = 1) = E_{U,X}(Y_1) - E_{U,X}(Y_0)$$

where Y_a denotes the counterfactual outcome under an intervention to set bullying status A = a.

Data Description:

Key variables:

A: Bullying before the age of 12 (asked in 1997)

Y: Incident drug use ("cocaine or other hard drugs") after 1997

W: Race/ethnicity, year of birth, sex, BMI, sexual orientation (gleaned from post-1997 data), learning disability, not living with both biological parents, lived through hard times, citizenship (all Ws were measured at baseline unless otherwise noted)

Sample size: 7,783

Marginal distribution of exposure and outcome:

Variable	No	Yes
Bullied < 12	6267	1516
Incident drug use	6434	1349

Anticipated Challenges:

There are many potential Ws in our dataset, and in order to make this manageable we will need to determine the appropriate variables to include in our analysis. We have already verified that there is sufficient data in each of the strata (as found in the data description), so while we were originally concerned about this, we think it will be fine. Our major challenge will likely be that this is longitudinal data, which we have only dealt with briefly in class; we will need to pay careful attention to time-ordered variables and make sure they are appropriately handled in our causal model.