

What Factors Predict Progression to Severe Obesity in Young Adults?

GH 811 In-class Exercise

October 22, 2019

Background

Individuals with severe obesity have elevated risks of death and are responsible for a disproportionate share of the health care costs. Intervening in the progression to severe obesity through lifestyle change, medical management and other interventions is critical to reducing the health and financial burden of the obesity epidemic in the United States. To maximize the effectiveness and efficiency of such interventions, it is important to understand how many people in the population progress to severe obesity and why some progress and others do not. Information of this nature can be used to better target interventions to those individuals most likely to benefit.

Your goal in this exercise will be to investigate the determinants of progression to severe obesity (BMI above 35 kg/m²) among young adults with obesity (BMI 30-35 kg/m²) using data from the National Health and Nutrition Examination Survey (NHANES), the gold standard survey program for health surveillance in the United States.

In your data, obesity status in young adulthood is determined using data on self-reported weight at age 25, which was collected in the weight history module of the NHANES. Obesity status in mid-adulthood is determined using clinically measured height and weight for individuals ages 27 and above at the time of survey.

You will investigate the probability of progression to severe obesity using multivariate logistic regression analysis. Covariates included in the dataset include demographic, socio-economic and behavioral factors, including age at survey, gender, race/ethnicity, educational attainment and smoking status (never, former and current). The evidence you generate from this analysis will be used to help identify young adults at particularly high risk for progression to severe obesity that stand to benefit the most from early lifestyle intervention, medical management and surgery.

Data

The data for this exercise are available on the course website. Variable names and codes are described below. The data include respondents from NHANES III, conducted between 1988 and 1994 and from the continuous NHANES, conducted biennially between 1999 and 2010. The sample includes a total of 19,990 observations.

Codebook

| Variable | Description | Values |
|------------|------------------------------------|---|
| age | | |
| male | DV | 0=female; 1=male |
| bmi_m | BMI at survey (kg/m ²) | Min=15.0; Max=72.3 |
| bmi25 | BMI at age 25 (kg/m ²) | Min=15.0; Max=40.3 |
| smoke | smoking status | never=0; former=1; current=2 |
| edu_recode | Educational attainment | 0=less than HS; 1=HS; 2=more than HS |
| nhwhite | DV | 1=non-Hispanic White |
| hisp | DV | 1=Hispanic |
| black | DV | 1=Black |
| other | DV | 1=Other |
| prog | DV | 0=did not progress; 1=progressed |
| bmi_cat | BMI at survey, categorical | 0=underweight; 1=normal wt; 2=overweight; 3=obese I; 4=obese II |
| bmi25_cat | BMI at 25, categorical | 0=underweight; 1=normal wt; 2=overweight; 3=obese I; 4=obese II |
| srh_poor | poor self-rated health (DV) | 1=poor self-rated health |
| mortstat | mortality status (DV) | 1=died in interval between survey and end of follow-up in 2011 |

Presentation

You have **until the end of class** to complete your analysis. At the end of the exercise, you will present your findings to the class in a power point presentation. Each group will be allotted 10 minutes to present. Your analysis should contain the following elements:

1. Visually explore the data in 2-3 graphs using ggplot2. These graphs may explore overall patterns in the data and/or patterns by age, gender or other characteristics. Your graphs may take the form of scatter plots, histograms, density plots, boxplots, or others of your choosing. Label your graphs Figure 1, Figure 2, etc. Make sure the graphs contain titles and axis labels.
2. Generate a new dataset restricted to the subset of individuals who were in the obese I category at age 25 (hint: dplyr may be useful here). Produce a table of descriptive statistics displaying socio-demographic and behavioral characteristics of individuals who progress to severe obesity vs. those who do not progress. Label this Table 1.
3. Create a table reporting the number and proportion of respondents in each BMI category at survey among those who were obese I at 25. Label this Table 2.

4. Investigate the predictors of progression to severe obesity using multivariate logit regression. Progression is defined as moving from obese I to obese II between age 25 and survey. People who did not progress include those who maintained weight in the obese I range over time as well as those who lost weight between age 25 and survey. Display your findings in a table, labeled Table 3. At a minimum, the table should include odds ratios and 95% CI for each variable in the model.
5. In your concluding slides, discuss the limitations of your data and study. If you could collect additional data, what other variables would you like to show to shed light on this question? How would you use such additional data?

R tips

- Read in `ggplot2` and `dplyr` libraries using `library()`. You only need to install these packages if you haven't done so previously.
- One way to subset your data is using the `filter()` verb in the `dplyr` package.
- A few useful graph types in `ggplot2`: `geom_point()`, `geom_density()`, `geom_boxplot()`
- How to re-code variables as factor variables: `data$male<-as.factor(data$male)`. This will be useful for producing graphics with correct legends and for entering categorical variables into your logit model.
- A useful reference on implementing logit regression in R:
<http://www.ats.ucla.edu/stat/r/dae/logit.htm>