DSCI 610

HW 3: NHANES Survey Data and Study Designs

By: Cody York

1. (4 pts) Briefly illustrate the sample design for NHANES survey data.

The sample design for NHANES survey data is organized into four stages.

* 1. In stage 1, the study will select the primary sampling unit (PSU) of the data. In this case, the study is looking at counties or small groups of countries with probabilities proportional to their size.
  2. In stage 2, the study will select selection segments within the PSU. These are groups of blocks containing households.
  3. In stage 3, the study will select specific households from stage 2.
  4. In stage 4, the study will select individuals within the household from stage 3.

1. (2 pts) Download the demographics data from the most recent available survey cycle.

I downloaded the demographic dataset from <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx>[? Component=Demographics&Cycle=2017-2020](./I%20downloaded%20the%20demographic%20dataset%20from%20https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx%3FComponent=Demographics&Cycle=2017-2020%20for%20the%20pre-pandemic%20years%20of%202017-2020.) for the pre-pandemic years of 2017-2020. The file was named P\_DEMO.XPT.

1. (2 pts) Download two laboratory data files that you are interested in from the most recent available survey cycle.

The laboratory files were downloaded from <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx? Component=Laboratory&Cycle=2017-2020.> I chose the Complete Blood Count with 5- Part Differential in Whole Blood (P\_CBC.XPT) and Hepatitis B: Core antibody, Surface antigen, and Hepatitis D antibody (P\_HEPBD.XPT) datasets from the laboratory results.

1. (8 pts) Import all three data sets into R and combine a subset of variables into an analysis dataset. Save the dataset as .rds file. You can use R markdown to for this question

This question is solved in the attach R markdown file.

1. (4 pts) Create at least two descriptive statistics tables for two key variables from your analysis dataset. You can use the same R markdown file in question 4 to create the summary tables.

This question is solved in the attach R markdown file.

1. (3+4+4 pts) Briefly illustrate the design aspects of BNT162b2 mRNA Covid-19 Vaccine trial. Assume a hypothetical scenario where you would like to design an intervention trial for a health condition in a target population. State a brief design protocol for your trial, similar to the method section discussed in BNT162b2 mRNA Covid-19 Vaccine trial. Create a flow-chart demonstrating randomization of the intervention, follow-up and completion, similar to Figure 1 in BNT162b2 mRNA Covid-19 Vaccine trial.

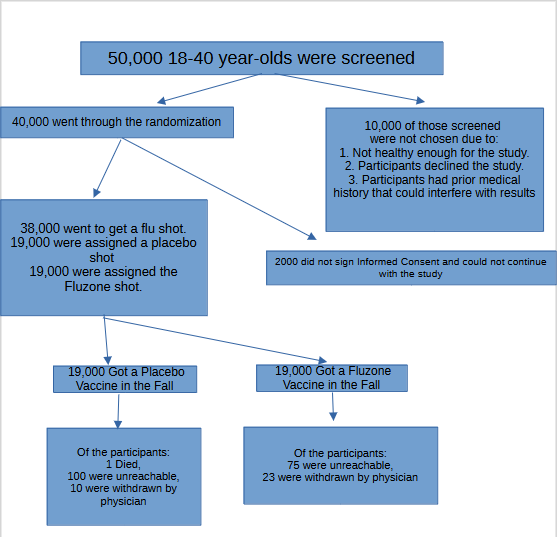
The design aspects of the trial are that individuals 16 years or older were assigned two does of either placebo or the vaccine 21 days apart. The trial was multinational, placebo-controlled and an observer-blinded trial.

For my design protocol:

In an ongoing national, placebo-controlled, observer-blinded trial where we randomly assigned 18 through 40 year-olds to get a flu shot during the fall season of either a placebo shot or a seasonal influenza vaccine from Fluzone. The primary endpoints were the efficacy of the seasonal influenza vaccine against laboratory confirmed

influenza strains and safety of the shot.

For the Flow Chart:



1. (3 pts) Identify the study design for the “BMI and future risk for COVID-19 infection and death across sex, age and ethnicity” study. What are the aspects of the design considered for this study?

The aspects of design were that this is a cohort study of 500,000 50-81 year-olds that were recruited between 2006 and 2010 that were recruited. The exposure of interest was BMI and data was provided for the SARS\_CoV\_2 tests. The participants were tested and determinations were made based on whether they had died or survived Covid-19.

1. (3 pts) Assume that you are considering designing a study to investigate the association between an exposure and a health condition that may occur due to the exposure. Design a prospective cohort study and show your study protocol with a flow chart. Be specific about the target population, exposure, outcome, and confounding variables (in any).

5000 men and women from Indiana aged 20 through 50 were recruited for a study that had been exposed to HIV. The exposure of interest was rashes on the body. The men and women were required to see a health practitioner regarding their exposure and data on their symptoms were reported to us. Data that was most of interest to us was whether they had got HIV and if the HIV developed enough to cause rashes. We would also collect data on the future outlook of the participants.

1. (3 pts) Now, assume that you do not have the budget in terms of the time or money to conduct a prospective study in question 3. Design a case control study and show your study protocol with a flow chart. Be specific about the cases, controls, and a strategy for collecting information on the exposure.

In this Case Control Study we are going to focus on the affects of exposure to known carcinogens and expected carcinogens to see if the amount of exposure required to cause cancer can be determined or if a carcinogen does not seem to be leading to cancer. This will be a generational study that will monitor the exposed individuals until death to determine if the exposure had any ill effects on their medical life. We will be finding approximately 100,000 individuals male or female and of any age less than 50. that have had more than average exposure to known carcinogens or more than average exposure to our potential carcinogens. We will get the informed consent of these participants that we will require them to have regular checkups with their doctor and any health data of interest to the study will be reported to use for further analysis. This study will be ongoing for the rest of their lives until age 80 and as long as they continue to participant they will receive compensation for their trouble.

Flow Diagram:

