**Part 1 submission**

**Background:**

There is well-known deadly linked between diabetes and tuberculosis. Diabetes was shown to be a risk factor for acquisition of latent tuberculosis infection (Barron, Shaw, Bullard, Ali, & Magee, 2018). There is a significant higher prevalence of hypertension in TB cases group compared to non-TB control group (Chung et al., 2014). There are no studies to examine the association between hypertension and latent tuberculosis.

**Research Question:**

Is hypertension a risk factor of acquiring latent tuberculosis infection after adjusted for some chronic health conditions (dyslipidemia, heart diseases, diabetes, kidney diseases, and hepatitis)

**Description of data and how it collected**

The National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations.

The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country. The sample for the survey is selected to represent the U.S. population of all ages. To produce reliable statistics, NHANES over-samples persons 60 and older, African Americans, and Hispanics.

The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

Findings from this survey will be used to determine the prevalence of major diseases and risk factors for diseases. Information will be used to assess nutritional status and its association with health promotion and disease prevention.

Health interviews are conducted in respondents’ homes. Health measurements are performed in specially-designed and equipped mobile centers, which travel to locations throughout the country. There are three main parts to this survey: the doorstep screener, the home interview and the health examination.

The doorstep screener: An NHANES interviewer will come to a home to talk to subjects about the survey. The interviewer will then ask persons few questions to see if they qualify to participate. This purpose of this step is to identify eligible participants.

The home interview: questions about your health, disease history, and diet were asked.

The health examination: All the health exam tests are performed in the Mobile Examination Center (MEC**)**that contains high-tech medical equipment.

**Where I get it**

I examined many cycles of NHANSES data and see that only cycle 2011-2012 tested participants with tuberculin skin test or QuantiFERON-TB test. So, I used this cycle to answer my question. The number of observation that I can have is around 5000.

I will use demographic data, examination data (tuberculosis, body measures, blood pressure), laboratory data, and questionnaire data.

I get these data from the link below:

<https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2011>

Table. Variables which I consider to be in my analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Type of variables | Variables’ name | The meaning of variables | Data set |
| Outcome | TBDRUIND | Tuberculin skin test results | TBX\_G |
|  | BPQ050A | Now taking prescribed medicine for HBP | BPX\_G |
| Main exposure | BPXSY1  BPXSY2  BPXSY3  BPXSY4 | Systolic Blood pressure for four consecutive measurements | BPX\_G |
| BPXDI1  BPXDI2  BPXDI3  BPXDI4 | Diastolic Blood pressure measurements for four consecutive measurements | BPX\_G |
|  | RIAGENDR | Gender |  |
| Potential covariates, and confounders | RIDAGEYR | Age in years | DEMO\_G |
| RIDRETH3 | Race-ethnicity | DEMO\_G |
| DMDBORN4 | Country of birth (Born in 50 U.S. states or Washington, DC and Born in other countries | DEMO\_G |
| DMDCITZN | Citizenship statusdia | DEMO\_G |
| DMDMARTL | Marital status | DEMO\_G |
| DMDYRSUS | Number of years the participants lived in US | DEMO\_G |
| DMDEDUC2 | Highest grade or level of education completed by adults 20 years and older | DEMO\_G |
| INDHHIN2 | Annual household income | DEMO\_G |
| DMDHHSIZ | Number of people in participant’s household | DEMO\_G |
| BMXBMI | BMI | BMX\_G |
| LBDHDD | Direct HDL cholesterol (mg/dL) | HDL\_G |
| LBXTR | Triglyceride (mg/dL) | TRIGLY\_G |
| LBDLDL | LDL – cholesterol (mg/dL) | TRIGLY\_G |
| LBXTC | Total cholesterol (mg/dL) | TCHOL\_G |
| BPQ100D | Now taking prescribed medicine for cholesterol | BPQ\_G |
| LBDGLTSI | Two hour glucose (mmol/L) |  |
| LBDGLUSI | Fasting Glucose (mmol/L) |  |
| GTDCODE | Incomplete OGTT (two hour glucose) comment code |  |
| LBXGH | HbA1C |  |
| DIQ050 | Taking insulin now |  |
| DIQ070 | Take diabetic pills to lower blood sugar |  |
| LBDHI | HIV test | Skip, only 19 people positive |
| URXUMS | Albumin, Urine (mg/L) |  |
| URXUCR | Creatinine (mg/dL) |  |
| URDABSdiCT | Albumin Creatinine ratio (mg/g) |  |
| LBXHA | Hepatitis A antibody |  |
| LBXHBS | Hepatitis B surface antibody |  |
| LBXHBC | Hepatitis B core antibody |  |
| LBDHBG | Hepatitis B surface antigen |  |
| LBDHD | Hepatitis D (anti-HDV) |  |
| SSRIBA | RIBA Test result for anti HCV |  |
| LBDHCV | Hepatitis C antibody |  |
| LBXHCR | Hepatitis C RNA |  |
| LBDHEG | Hepatitis E IgG antibody |  |
| LBDHEM | Hepatitis E IgM antibody |  |
| ALQ101 | Had a least 12 alcohol drinks/ 1yr | ALQ\_X |
| ALQ120Q | How often drink alcohol over past 12 months | ALQ\_X |
| CDQ001  CDQ002  CDQ003  CDQ004  CDQ005  CDQ006  CDQ009D  CDQ009E  CDQ009F  CDQ009G | The variables needed for Grade 1 Angina or Grade 2 Angina | CDQ\_X |
| SMQ040 | Do you now smoke cigarettes | SMQ\_X |
|  | TBQ040 | Ever told you had active TB | TBQ\_X |
|  | TBQ060 | Lived in a house TB sick person | TBQ\_X |
|  | MCQ160b | Ever told had congestive heart failure | MCQ\_G |
|  | MCQ160c | Ever told had coronary heart disease | MCQ\_G |
|  | MCQ160d | Ever told had angina/angina pectoris | MCQ\_G |
|  | MCQ160e | Ever told had heart attack | MCQ\_G |
|  | MCQ160f | Ever told had stroke | MCQ\_G |
|  |  |  |  |

**What relation I look for in data?**

I expect that there is a positive association between latent tuberculosis infection and hypertension.

**How to analyze data?**

The outcome is binary which is TB infection (Yes/No). My main exposure is categorical variables. Other predictors can be either continuous or categorical variables.

The characteristics of the population will be denoted with proportions for categorical variables and mean (SD) for continuous variables.

The prevalence of TB infection in US is low, less than 5%. We can consider it as rare condition. I used logistic regression to estimate prevalence odds ratio which is equivalent to prevalence rate ratio. The univariate and multivariate logistic regression will be performed.

**REFERENCES**

Barron, M. M., Shaw, K. M., Bullard, K. M., Ali, M. K., & Magee, M. J. (2018). Diabetes is associated with increased prevalence of latent tuberculosis infection: Findings from the National Health and Nutrition Examination Survey, 2011–2012. *Diabetes research and clinical practice, 139*, 366-379.

Chung, W.-S., Lin, C., Hung, C., Chu, Y., Sung, F., Kao, C., & Yeh, J. (2014). Tuberculosis increases the subsequent risk of acute coronary syndrome: a nationwide population-based cohort study. *The international journal of tuberculosis and lung disease, 18*(1), 79-83.