BMIN503

Distinguishing some relationships of asthma in the US from 2017-March 2020

Distinguish some relationships of asthma in the US from 2017- March 2020such as prevalence, incidence, and social determinants of health by using data from the NHANES 2017- March 2020 questionnaire on medical conditions

1. Questions to explore
   1. What is the active asthma rate?: Asthma prevalence: Ever been told you have asthma + Had asthma attack in past year
      1. Emergency care visit for asthma/past yr
         1. Covariates: age groups, gender, race, education level, citizenship status, marital status, poverty ratio, insurance status
            1. See if difference for social factors
      2. Hospital utilization and access to care (P\_HUQ)
   2. Is there an association between social factors and emergency room visit for asthma?
   3. Is there an association between emergency room visit for asthma and hospital utilization? Access to care?

NHANES 2017-2018 Questionnaires:

Health Insurance: <https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_HIQ.htm>

Hospital Utilization & Access to Care: <https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_HUQ.htm>

Medical Conditions: <https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_MCQ.htm>

Demographics: <https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_DEMO.htm>

Articles for research: https://link.springer.com/article/10.1186/s40733-016-0029-3

* Both male and females 1-150 years

|  |  |  |  |
| --- | --- | --- | --- |
| **Initial Variable Names** | **Question** | **Variable Type** | **Sample Count from website** |
| mcq010 | Ever been told you have asthma | Binary | 1325 (yes) |
| mcq025 | Age when first had asthma | Categorical | 1304 (total) |
| mcq035 | Still have asthma | Binary | 1296 (total) |
| mcq040 | Had asthma attack in past year | Binary | 822 (total) |
| mcq050 | Emergency care visit for asthma/past yr | Binary | 825 (total) |
| **Initial Variable Names (DEMO\_J)** | **Question** | **Variable Type** | **Sample Count** |
| riagendr | Gender | Binary | 9254 |
| **ridageyr** | Age in years at screening | Continuous | 9254 |
| ridreth1 | Race/Hispanic origin | Categorical |  |
| ridreth3 | Race/Hispanic origin w/ NH Asian | Categorical |  |
| dmdcitzn | Citizenship status |  |  |
| dmdeduc3 | Education level- children/youth 6-19 | Categorical |  |
| dmdeduc2 | Education level- adults 20+ | Categorical |  |
| dmdmartl | Marital status | Categorical |  |
| indhhin2 | Annual household income | Categorical |  |
| indfmin2 | Annual family income | Categorical |  |
| indfmpir | Ratio of family income to poverty | Categorical |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Do we need to weight data? (needed for publication)

<https://wwwn.cdc.gov/Nchs/Nhanes/2017-2018/P_HUQ.htm> (possibly use to link)

https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Questionnaire&Cycle=2017-2020

look at occupation and smoking

pesticide use- proxy for exacerbation (environmental exposures)

hiq032a = private

hiq032b = medicare

hiq032c = medi-gap

hiq032d = Medicaid

hiq032e = chip

hiq032h = state-sponsored

hiq032i = other government

<https://pubmed.ncbi.nlm.nih.gov/28138394/>

<https://pubmed.ncbi.nlm.nih.gov/30680222/>

r package survey: survey data weighted, nhanes explains how to use, survey design

glm- can have several

library(haven)

MCQ\_J <- read\_dta("~/Box Sync/Fall 2021/Data Science For Biomedical Informatics/Final Project/BMIN503\_Final\_Project/MCQ\_J.dta")

P\_DEMO <- read\_dta("~/Box Sync/Fall 2021/Data Science For Biomedical Informatics/Final Project/BMIN503\_Final\_Project/P\_DEMO.dta")

P\_HIQ <- read\_dta("~/Box Sync/Fall 2021/Data Science For Biomedical Informatics/Final Project/BMIN503\_Final\_Project/P\_HIQ.dta")

P\_HUQ <- read\_dta("~/Box Sync/Fall 2021/Data Science For Biomedical Informatics/Final Project/BMIN503\_Final\_Project/P\_HUQ.dta")

#Redefine variables for MCQ\_J/medical conditions

MCQ\_J2 <- MCQ\_J %>%

dplyr::mutate(mcq010 = factor(mcq010, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq035 = factor(mcq035, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq040 = factor(mcq040, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq050 = factor(mcq050, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160b = factor(mcq160b, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160c = factor(mcq160c, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160d = factor(mcq160d, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160e = factor(mcq160e, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160f = factor(mcq160f, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160m = factor(mcq160m, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160g = factor(mcq160g, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160k = factor(mcq160k, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq170k = factor(mcq170k, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160o = factor(mcq160o, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq160l = factor(mcq160l, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq500 = factor(mcq500, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq220 = factor(mcq220, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq371a = factor(mcq371a, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq371b = factor(mcq371b, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq371c = factor(mcq371c, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(mcq371d = factor(mcq371d, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::select(seqn, mcq010, mcq035, mcq040, mcq050, mcq160b, mcq160c, mcq160d, mcq160e, mcq160f, mcq160m, mcq160g, mcq160k, mcq160o, mcq160l, mcq500, mcq220, mcq371a, mcq371b, mcq371c, mcq371d)

#Redefine variables for P\_DEMO/demographics

P\_DEMO2 <- P\_DEMO %>%

dplyr::mutate(ridageyr = factor(ridageyr)) %>%

dplyr::mutate(riagendr = factor(riagendr, levels = c(1, 2), labels =c("Male", "Female"))) %>%

dplyr::mutate(ridreth3 = factor(ridreth3, levels = c(1, 2, 3, 4, 5, 7),

labels = c("Mexican American", "Other Hispanic", "Non-Hispanic White", "Non-Hispanic Black",

"Non-Hispanic Asian", "Other Race- Including Multiracial"))) %>%

dplyr::mutate(dmdeduc2 = factor(dmdeduc2, levels = c(1, 2, 3, 4, 5),

labels = c("< 9th grade", "9th-11th grade", "High school graduate/GED",

"Some college or AA degree", "College graduate or above"))) %>%

dplyr::mutate(dmdmartz = factor(dmdmartz, levels = c(1, 2, 3),

labels =c("Married/Living with Partner", "Widowed/Divorced/Separated", "Single"))) %>%

dplyr::mutate(indfmpir = factor(indfmpir, levels = c(0 - 4.98, 5),

labels =c("Ratio of family income to poverty = 0 - 4.98", "Ratio of family income to poverty = 5"))) %>%

dplyr::select(seqn, riagendr, ridreth3, dmdeduc2, dmdmartz, indfmpir)

#Redefine variables for P\_HIQ/health insurance

P\_HIQ2 <- P\_HIQ %>%

dplyr::mutate(hiq011 = factor(hiq011, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(hiq032a = factor(hiq032a, levels = c(1), labels = c("Covered by private insurance"))) %>%

dplyr::mutate(hiq032b = factor(hiq032b, levels = c(1), labels = c("Covered by Medicare"))) %>%

dplyr::mutate(hiq032c = factor(hiq032c, levels = c(1), labels =c("Covered by Medi-Gap"))) %>%

dplyr::mutate(hiq032d = factor(hiq032d, levels = c(1), labels =c("Covered by Medicaid"))) %>%

dplyr::mutate(hiq032e = factor(hiq032e, levels = c(1), labels =c("Covered by CHIP"))) %>%

dplyr::mutate(hiq032h = factor(hiq032h, levels = c(1), labels =c("Covered by state-sponsored health plan"))) %>%

dplyr::mutate(hiq032i = factor(hiq032i, levels = c(1), labels =c("Covered by other government insurance"))) %>%

dplyr::mutate(hiq260 = factor(hiq260, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(hiq210 = factor(hiq210, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::select(seqn, hiq011, hiq032a, hiq032b, hiq032c, hiq032d, hiq032e, hiq032h, hiq032i, hiq260, hiq210)

#Redefine variables for P\_HUQ/Hospital Utilization & Access to Care

#should I use HUD062 or just HUQ062????????

P\_HUQ2 <- P\_HUQ %>%

dplyr::mutate(hud062 = factor(hud062, levels = c(1, 2, 3, 4), labels =c("Within the past year (anytime less than 12 months ago)", "within the last 2 years (1 year but less than 2 years ago)", "within 3 years but less than 5 years", "within 5 years but less than 10 years"))) %>%

dplyr::mutate(huq010 = factor(huq010, levels = c(1, 2, 3, 4, 5), labels = c("Excellent", "Very good", "Good", "Fair", "Poor"))) %>%

dplyr::mutate(huq030 = factor(huq030, levels = c(1, 2, 3), labels = c("Yes", "There is no place", "There is more than one place"))) %>%

dplyr::mutate(huq051 = factor(huq051, levels = c(0, 1, 2, 3, 4, 5, 6, 7, 8), labels =c("None", "1", "2 to 3", "4 to 5", "6 to 7", "8 to 9", "10 to 12", "13 to 15", "16 or more"))) %>%

dplyr::mutate(huq071 = factor(huq071, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::mutate(huq090 = factor(huq090, levels = c(1, 2), labels =c("Yes", "No"))) %>%

dplyr::select(seqn, hud062, huq010, huq030, huq051, huq071, huq090)

#Link datasets by SEQN

MCQ\_J$seqn <- as.numeric(MCQ\_J$seqn)

P\_DEMO$seqn <- as.numeric(P\_DEMO$seqn)

P\_HIQ$seqn <- as.numeric(P\_HIQ$seqn)

P\_HUQ$seqn <- as.numeric(P\_HUQ$seqn)

nhanes2017 <- Reduce(function(x, y) merge (x = x, y = y, all = TRUE, by = "seqn"), list(MCQ\_J2, P\_DEMO2, P\_HIQ2, P\_HUQ2))

#only use selected variables (cleaned data) #need to remove NAs???? Also, NAs for many variables

nhanes2017 <- nhanes2017 %>%

dplyr::select(seqn, mcq010, mcq035, mcq040, mcq050, mcq160b, mcq160c, mcq160d, mcq160e, mcq160f, mcq160m, mcq160g, mcq160k, mcq160o, mcq160l, mcq500, mcq220, mcq371a, mcq371b, mcq371c, mcq371d, riagendr, ridageyr, ridreth3, dmdeduc2, dmdmartz, indfmpir, hiq011, hiq032a, hiq032b, hiq032c, hiq032d, hiq032e, hiq032h, hiq032i, hiq260, hiq210, hud062, huq010, huq030, huq051, huq071, huq090)

#How many NA's for each variable?

sapply(nhanes2017, function(x) sum(is.na(x)))

#View of some variables

count\_unique <- rapply(nhanes2017,function(x) length(unique(x))) #There are 24,457 subjects in dataset

table(nhanes2017$riagendr) #males = 7721 & females = 7839

#table(nhanes2017$huq062) # "Combination of HUQ061 and HUQ062"

table(nhanes2017$mcq010) #number of "ever been told you have asthma" = 1325

table(nhanes2017$mcq040) #number of "Had asthma attack in past year" = 359

table(nhanes2017$mcq035) #number of "still have asthma" = 825

table(nhanes2017$mcq050) #number of "ER visit for asthma/past yr" = 167

summary(nhanes2017$ridageyr) #ages

#Demographics for Emergency care visit for asthma/past yr (mcq050 = outcome), predictors = age, sex, insurance status, race, etc.

#outcome is categorical, predictor is categorical (sex), age = continuous

library(dplyr)

subset.nhanes2017 <- nhanes2017 %>% group\_by(mcq050 == "Yes")

glm(data = subset.nhanes2017, mcq050 ~ riagendr, family = binomial)

lm(data = nhanes2017, I(mcq050 == "Yes") ~ I(riagendr == "Male") + I(riagendr == "Female"))

library(nhanesA)

library(plyr)

library(dplyr)

library(survey)

varlist <- c("DEMO", "MCQ", "HIQ", "HUQ")

varlist\_years <- paste0(rep(varlist, each = 1), c("\_P"))

list\_all <- sapply(varlist\_years, function(x) {data.frame(nhanes(x))})

for(i in 1:length(list\_all)) {

assign(names(list\_all)[i], list\_all[[i]])

}

for (i in 1:length(varlist)) {

assign(varlist[i], plyr::rbind.fill(mget(grep(varlist[i], ls(), value = T))))

}

rm(list = grep("\_[J]", ls(), value = T))

nhanes.data <- full\_join(get(varlist[1]), get(varlist[2]), get(varlist[3]),

get(varlist[4]), by = "SEQN")

rm(list = ls()[-which(ls() == "nhanes.data")])

nhanes2017 <- nhanes.data %>%

dplyr::select(id = SEQN, gender = RIAGENDR, age = RIDAGEYR,

race = RIDRETH3, education = DMDEDUC2, income = INDHHIN2,

marital-status = DMDMARTZ, poverty = INDFMPIR, ever-asthma = MCQ010,

have-asthma = MCQ035, asthma-attack = MCQ040, ER-asthma = MCQ050,

chf = MCQ160b, chd = MCQ160c, angina = MCQ160d, MI = MCQ160e,

stroke = MCQ160f, ever-thyroid = MCQ160m, thyroid = MCQ170m,

emphysema = MCQ160g, ever-chronic-bronchitis = MCQ160k,

chronic-bronchitis = MCQ170k, copd = MCQ160o, ever-liver = MCQ160l,

liver = MCQ170l, cancer = MCQ220, losing-weight = MCQ371A,

exercise = MCQ371B, reduce-salt = MCQ371C, reduce-fat = MCQ371D,

insurancebinary = HIQ011, privateinsurance = HIQ032A,

medicare = HIQ032B, medi-gap = HIQ032C, medicaid = HIQ032D,

chip = HIQ032E, state-sponsored = HIQ032H, other-government = HIQ032I,

havemedicare = HIQ260, noinsurance-year = HIQ210,

combination-doctor = HUD062, healthcondition = HUQ010,

routine-healthcare = HUQ030, number-healthcare =HUQ051,

overnight-hospital = HUQ071, mentalhealth-visit = HUQ090)

nhanes.practice <- nhanes.data %>%

select(id = SEQN, gender = RIAGENDR, age = RIDAGEYR,

race = RIDRETH1, education = DMDEDUC2, income = INDHHIN2)

P\_MCQ2$id <- as.numeric(P\_MCQ2$id)

P\_DEMO2$id <- as.numeric(P\_DEMO2$id)

P\_HIQ2$id <- as.numeric(P\_HIQ2$id)

P\_HUQ2$id <- as.numeric(P\_HUQ2$id)

everasthma, asthma, asthmaattack, erasthma,

chf, chd, angina, mi, stroke, everthyroid, thyroid, lungdx,

everliver, liver, cancer, losingweight, exercise, reducesalt,

reducefat

gender, age, race, education, maritalstatus, poverty

insurancebinary, privateinsurance, medicare, medigap, medicaid,

chip, statesponsored, othergovernment, havemedicare, noinsuranceyear

combinationdoctor, healthcondition, routinehealthcare,

numberhealthcare, overnighthospital, mentalhealthvisit