**Last Updated: September 28, 2022, by Aarsh Batra (aarshbatra@uchicagotrust.org)**

**0. Get Data**

**Broad Purpose:** This directory contains the files that describe and implement calculation of

the PM2.5 vs life expectancy dose-response function that is implied by the

methodology and data of the Global Burden of Disease project. The following

instructions are for calculating the impact on global life expectancy.

a. Download GBD mortality data

Go to http://ghdx.healthdata.org/gbd-results-tool

GBD Estimate = Cause of death or injury, Measure = Deaths,

Location = Global

Age = select all (by this I mean that select all available categories)

Sex = Both,

Metric = Rate,

Cause = Lower respiratory infections; Tracheal, bronchus, and lung cancer;

Ischaemic heart disease; Stroke; Chronic obstructive pulmonary disease;

Diabetes mellitus type 2 (also double check the latest GBD paper as it may add/remove certain disease channels that are linked to PM2.5)

Click "Download CSV"

Save to Data/Raw

b. Download GBD relative risks data

Go to http://ghdx.healthdata.org/record/ihme-data/gbd-2019-relative-risks

(or access from main data tool page by clicking "Additional Resources" tab)

Click "Files" tab

Download "Relative risks: particulate matter air pollution"

Save to Data/Raw

**1a.** **Build/clean\_GBD\_relative\_risks.do**

input: dataset at level of disease channel x PM concentration, with relative risk for each age group, downloaded from GBD

purpose: formatting and encoding

output: Intermediate/GBD\_2016\_relative\_risk.dta

**1b. Build/clean\_GBD\_mortality\_rates.do**

input: dataset at level of disease channel (x country, optionally), with observed mortality rates (deaths per 100k) for each age group, downloaded from GBD

purpose: formatting and labeling

output: Intermediate/GBD\_2016\_mortality\_rates.dta, \_country.dta

**1c. Build/clean\_worldbank\_population.do**

purpose: clean country names to be merged with GBD data, for doing life table method by country

output: Intermediate/worldbank\_2016\_population\_dset.dta

2. Analysis

**2a. Analysis/calc\_mortal\_rate\_from\_PM\_by\_age.do**

relies on: 1a, 1b

input: disease x PM conc relative risks, disease x PM conc mortality rates, all-cause mortality rates

purpose:

setup code chunk:

(1) format disease x PM conc RR and disease-level MR data, to create a disease x PM x age dataset of mortality rates (called mainfile). Hypothetical "adjusted" mortality rates at various PM values are calculated by multiplying actually mortality rates by RR(hypothetical PM)/RR(actual global pop-weighted PM).

main code chunk:

(2) Line 140: Use relative risks to calculate PAF=1-1/RR for each disease channel x PM conc x age.

(3) Line 144: Use PAFs and disease x PM x age mortality rates to calculate what the mortality rate from PM would be for each PM conc, via each disease channel, at each age

(4) Line 148: Sum across disease channels to get the all-cause mortality rate attributable PM at each age

adjust\_actual code chunk:

(5) Calculate what all-cause mortality rates would be under each counterfactual PM conc. So for PM=x, calculate [actual\_mr=actual all-cause mortality rate] + [adjusted\_mortal\_rate=adjusted mortality rate from relevant disease channels for x, calculated in step (1)] – [mortal\_rate=actual mortality rate from relevant disease channels]

output:

Intermediate/calculated\_mortality\_rates/pm\_mortality\_rates\_by\_age.xlsx, \_country.xlsx

Intermediate/GBD\_2016\_allcause\_mr.dta, \_country.dta

**2b. Analysis/lifetable\_global.do, \_bycountry.do**

relies on: 1c, 2a

purpose: these files turn mortality rates into life expectancy losses via the life table method.