Replication Readme For "After the Dust Settles: The Infant Health Impacts of Dust Storms"

Published in the *Journal of the Association of Environmental and Resource Economists*

Benjamin A. Jones (bajones@unm.edu)

May 2020

This file contains information on the programs and data used to generate the figures and tables in the manuscript. The specific techniques used to collect and construct the data are described in the manuscript. The following files have been uploaded to the Dataverse archive:

Files:

"Readme_Dust_Data_Replication.pdf"	(this Readme document)
"Dust.do"	Stata Do file to generate figures and tables
"Dust_data.dta"	Stata data file

The Stata data file contains variables from the following data sources, which are in the public domain and which I have a right to distribute with proper citation:

1. National Weather Service Storm Events Database (for data on dust storm events). These data can be downloaded here: ftp://ftp.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/
2. NOAA National Centers for Environmental Information (for data on temperature and precipitation).[1] These data can be downloaded here: ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/by_year/
3. US EPA AirData network (for resultant wind speed and wind direction data).[2] These data can be downloaded here: https://aqs.epa.gov/aqsweb/airdata/download_files.html
4. US Census Bureau Small Area Income and Poverty Estimates Program (for annual data on county-level poverty rates). These data can be downloaded here: https://www.census.gov/programs-surveys/saipe/data/datasets.html
5. US Census Bureau American Community Survey 1-Year Estimates (for annual education data at the county-level). These data can be downloaded here: https://data.census.gov/cedsci/?q=Educational%20Attainment&t=Educational%20Attainment
6. National Weather Service Archived Watches/Warnings at Iowa State University (for data on dust storm warnings). These data can be downloaded here: https://mesonet.agron.iastate.edu/request/gis/watchwarn.phtml

[1] Note that the Stata data file does not include the actual bins for temperature or precipitation as used in the manuscript. Rather, I only include mean monthly temperature and mean monthly precipitation (which allows for an approximate replication of the results). This is because the bins were constructed specifically for each pregnancy and, it is possible, that one could back-out the length of each pregnancy in a given county if the bins were provided, which cannot be allowed since the birth data are restricted.

[2] The Stata data file only includes average monthly wind speed and wind direction data for the same reasons as above (namely that sharing data on the number of days during pregnancy that wind speed/direction fell into a given bin might reveal the length of each pregnancy in a county, which violates my data use agreement for the birth outcome data).

The following data are not included in the Stata data file because they are restricted or I do not have the right/permission to share them publically.

1. US CDC National Center for Health Statistics (NCHS) Natality Data (for data on birth outcomes and mother characteristics with identification of the county of birth/residence). These data can be requested from NCHS by completing a Project Review Form that is then submitted, along with supporting material, to NCHS for review. The review typically takes 2-3 months. Instructions for how to do this are provided here: https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm. Upon approval, researchers will need to sign a Data Use Agreement (DUA) that prohibits the sharing of the restricted data to any individual, group, or organization that is not listed in the DUA along with other requirements. The same natality data that I use, but without information on the county of birth/residence can be downloaded here: https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm
2. I do not have the right to distribute the monthly gridded satellite PM2.5 data used in the manuscript, which I obtained from the Atmospheric Composition Analysis Group website: http://fizz.phys.dal.ca/~atmos/martin/?page_id=140. To the best of my knowledge, these data are should be obtained directly by researchers from this website. I use monthly data from the North American Regional Estimates (v4.NA.02.MAPLE), which can be downloaded from the original author's ftp site here: ftp://stetson.phys.dal.ca/Aaron/V4NA02/Monthly/. I encourage anyone who wants to replicate my analysis to download this monthly PM2.5 data.

The Stata data file uploaded to Dataverse is daily over 2010-2017 for dust storm ever counties. To construct the final dataset used in the manuscript, one would need to obtain the restricted NCHS natality data and merge it with the uploaded Stata data file. Then, one would need to appropriately calculate the number of dust storm events occurring for each pregnancy (using the dust storm event data included in the uploaded Stata file) and the number of pregnancy days where temperature, precipitation, and wind speed/direction fell into the bins listed in the manuscript (also using the weather data included in the uploaded Stata file). For the PM2.5 analysis in the manuscript, one would need to download the PM2.5 data from above and then merge it, month-by-month, with the dust storm event data in the uploaded Stata file. Note that the ACS and SAIPE data are annual and not daily.