README – Data and Codes

Canary in a Coal Mine:
Infant Mortality and Tradeoffs
Associated with Mid-20th Century Air Pollution

Karen Clay, Carnegie Mellon University and NBER
Joshua Lewis, University of Montreal
Edson Severnini, Carnegie Mellon, IZA and NBER

The analysis was carried out using the statistical software STATA, version 14. The STATA file "*Canary_Codes.do*" contains all the commands needed to obtain the figures and tables in the paper. The file clearly marks the codes associated with each figure/table, and include further explanations for a smooth run. For example, for Figure 1, with two panels, we include:

```
*******************************************************************************
*******************************************************************************
***FIGURE 1
*******************************************************************************
*******************************************************************************

****************************************
***PANEL A
****************************************
codes…
graph export graphs/Fig_1a.pdf, replace

****************************************
***PANEL B
****************************************
codes…
graph export graphs/Fig_1b.pdf, replace
```

Observe that the two parts of the figure will be saved in the subfolder "graphs." In fact, as indicated in the beginning of the dofile, when running the codes, please change the directory to whatever folder the data and codes have been saved. Also, create four subfolders within the main folder: "data," "dofiles," "logfiles," and "graphs" – this is crucial to run the codes smoothly. Finally, create a subfolder in the "data" subfolder named "eventstudy."

For tables, the codes save all relevant estimates in the file "Canary_output.xml" within the subfolder "logfiles." To see the estimates, just open the .xml file in Excel. Each column contains the estimates corresponding to a table, panel, and column. For example, the label T3_PB_c2 indicates estimates for Table 3, Panel B, column 2. For appendix tables, there is an "A" before

the "T," e.g., ATA5_PC_c3, which corresponds to Appendix Table A.5, Panel C, column 3. For summary statistics tables and additional information in tables with the main estimates, the statistics are reported in the main logfile "Canary_logfile.log," also saved in the subfolder "logfiles." Just scroll down to the marks associated with that table. For example, for Appendix Table A.2, go to:

```
********************************************************************************
********************************************************************************
***APPENDIX TABLE A.2
********************************************************************************
********************************************************************************

****************************************
***PANEL A
****************************************
codes…

****************************************
***PANEL B
****************************************
codes…

****************************************
***BOTTOM ROWS
****************************************
codes…
```

All relevant statistics are reported right after the codes. To create a table, just copy and paste them in Excel.

Regarding the data, there are six STATA files with the extension *.dta*, which must be saved in the subfolder "data." The two main files are:

- *eventstudy_treat30miles_90_11_7.dta*: data sources mentioned in the dictionary of variables below;
- *imr_final_balanced.dta*: data sources mentioned in the dictionary of variables below.

The other four auxiliary files are:

- *HSUS_Electricity2.dta*: data for Figure 1a from Gartner (2006), *Historical Statistics of the United States*, Table Db218-227. Electric utilities-power generation and fossil fuel consumption by energy source: 1920-2000;

- *coalbysector.dta*: data for Figure 1b United States Bureau of Mines, *Minerals Yearbook* (various years);
- *dist_newhydro_after1962.dta*: data from Severnini (2014);
- *tspEPA_1957_62_semibalanced.dta*: annual summary data on pollution provided by the US Environmental Protection Agency (EPA) through a Freedom of Information Act (FOIA) request.

All variables are labeled in the main .dta files, so to get a quick description of the variables, just type "describe" in STATA once the file is open. In any case, here is the dictionary of variables with the data source for each variable in parentheses – when the source is implied by previous related variables, we do not repeat the source.

For the dataset *eventstudy_treat30miles_90_11_7.dta*:

-- **state_fips**: Federal Information Processing Standard (FIPS) state codes
-- **county_fips**: County FIPS
-- **plant_id**: Power Plant ID
-- **idcountyplant**: ID for a plant-county pair
-- **year**: Year of the information in the sample
-- **cinityropr**: Power plant initial year of operation (FPC, 1947, 1948-62)
-- **imr**: Infant mortality rate (per 1,000 live births) (Bailey et al., 2016)
-- **lnimr**: log(infant mortality rate)
-- **births_rs**: Number of births by county of residence (Bailey et al., 2016)
-- **ccapacity**: Coal-fired electricity generating capacity (FPC, 1947, 1948-62, 1963)
-- **f_coalq**: Coal consumption in electricity generation (in 1,000 tons) (FPC, 1947, 1948-62)
-- **generation**: Net electricity generation (million kilowatt-hours) (FPC, 1947, 1948-62)
-- **cap30mile_hydro**: Hydroelectric capacity within 30 miles of county centroid (in 100s of MW) (FPC, 1963)
-- **scap1940**: Dummy for BELOW median coal generating capacity in 1940
-- **bcap1940**: Dummy for ABOVE median coal generating capacity in 1940
-- **laty**: Country centroid latitude
-- **lony**: Country centroid longitude
-- **yprcp**: Total precipitation (in millimeters) (NOAA)
-- **ytav**: Average temperature (in degrees Celsius) (NOAA)
-- **yddb10**: Degree days below 10 degrees Celsius (calculation based on NOAA data)
-- **ydda29**: Degree days above 29 degrees Celsius (calculation based on NOAA data)
-- **dist_miles**: Distance between power plant and county centroid (calculated using ArcGIS based on data from FPC, 1963)
-- **dist0_10**: Distance between power plant and county centroid -- 0-10 miles
-- **dist10_20**: Distance between power plant and county centroid -- 10-20 miles
-- **dist20_30**: Distance between power plant and county centroid -- 20-30 miles
-- **dist30_40**: Distance between power plant and county centroid -- 30-40 miles
-- **dist40_50**: Distance between power plant and county centroid -- 40-50 miles
-- **dist50_60**: Distance between power plant and county centroid -- 50-60 miles
-- **dist60_70**: Distance between power plant and county centroid -- 60-70 miles
-- **dist70_80**: Distance between power plant and county centroid -- 70-80 miles

-- **dist80_90**: Distance between power plant and county centroid -- 80-90 miles
-- **sulf**: Sulfates (in 1000s ug/m-3) based on our coal consumption data & the AP3 model (Muller & Mendelsohn, 2007, 2009)"
-- **pop**: Population (Haines & ICPSR, 2010)
-- **pop1940**: Population in 1940
-- **lnpop1940**: log(population in 1940)
-- **popf15_44**: Female population age 15-44 (Bailey et al., 2016)
-- **spopurb**: Share of urban population (Haines & ICPSR, 2010)
-- **emp**: Total employment (Haines & ICPSR, 2010)
-- **emp1940**: Total employment in 1940
-- **lnemp1940**: log(total employment in 1940)
-- **mfg**: Manufacturing employment (Haines & ICPSR, 2010)
-- **smfg**: Share of manufacturing employment
-- **mfg1940**: Manufacturing employment in 1940
-- **lnmfg1940**: log(Manufacturing employment in 1940)
-- **light**: Share of households with electric lighting (Haines & ICPSR, 2010)
-- **light1940**: Share of households with electric lighting in 1940
-- **railroads**: Railroad mileage within county in 1911 (Donaldson & Hornbeck 2016)
-- **inst1944**: Predicted interstate highway from 1944 plan (Michaels 2008)
-- **control**: County 30-90 miles from power plants
-- **treat**: County 0-30 miles from a power plant
-- **post**: Power plant operating
-- **treatpost**: Interaction treat x post
-- **sc40treatpost**: Interaction treatpost x BELOW median coal generating capacity in 1940
-- **bc40treatpost**: Interaction treatpost x ABOVE median coal generating capacity in 1940
-- **ltreatpost**: Interaction treatpost x BELOW median household electricity access in 1940 (electricity access is proxied by access to electric lighting from Haines & ICPSR, 2010)
-- **htreatpost**: Interaction treatpost x ABOVE median household electricity access in 1940
-- **streatpost**: Interaction treatpost x SMALL Plant (< 75MW)
-- **btreatpost**: Interaction treatpost x LARGE Plant (≥ 75MW)
-- **treatdown_post**: Interaction treatpost x DOWNWIND (90◦arc from plant centroid)
-- **treatup_post**: Interaction treatpost x UPWIND (270◦arc from plant centroid)
-- **treat20post**: treatpost w/ alternative treat 0-20 miles from a power plant
-- **sc40treat20post**: Interaction treat20post x BELOW median coal generating capacity in 1940
-- **bc40treat20post**: Interaction treat20post x ABOVE median coal generating capacity in 1940
-- **ltreat20post**: Interaction treat20post x BELOW median household electricity access in 1940
-- **htreat20post**: Interaction treat20post x ABOVE median household electricity access in 1940
-- **treat40post**: treatpost w/ alternative treat 0-40 miles from a power plant
-- **sc40treat40post**: Interaction treat40post x BELOW median coal generating capacity in 1940
-- **bc40treat40post**: Interaction treat40post x ABOVE median coal generating capacity in 1940
-- **ltreat40post**: Interaction treat40post x BELOW median household electricity access in 1940
-- **htreat40post**: Interaction treat40post x ABOVE median household electricity access in 1940
-- **treat60post**: treatpost w/ alternative treat 0-60 miles from a power plant
-- **treat60postsulf**: Interation treat60post x sulf
-- **evtime**: Year - Power plant initial year of operation
-- **evtimet11b**: 11 years before plant opening

-- **evtimet10b**: 10 years before plant opening
-- **evtimet9b**: 9 years before plant opening
-- **evtimet8b**: 8 years before plant opening
-- **evtimet7b**: 7 years before plant opening
-- **evtimet6b**: 6 years before plant opening
-- **evtimet5b**: 5 years before plant opening
-- **evtimet4b**: 4 years before plant opening
-- **evtimet3b**: 3 years before plant opening
-- **evtimet2b**: 2 years before plant opening
-- **evtimet1b**: 1 year before plant opening
-- **evtimet0**: Year power plant starts operation
-- **evtimet1a**: 1 year after plant opening
-- **evtimet2a**: 2 years after plant opening
-- **evtimet3a**: 3 years after plant opening
-- **evtimet4a**: 4 years after plant opening
-- **evtimet5a**: 5 years after plant opening
-- **evtimet6a**: 5 years after plant opening
-- **evtimet7a**: 7 years after plant opening
-- **sc40evtimet0**: Interaction evtimet0 x BELOW median coal generating capacity in 1940
-- **sc40evtimet1a**: Interaction evtimet1a x BELOW median coal generating capacity in 1940
-- **sc40evtimet2a**: Interaction evtimet2a x BELOW median coal generating capacity in 1940
-- **sc40evtimet3a**: Interaction evtimet3a x BELOW median coal generating capacity in 1940
-- **sc40evtimet4a**: Interaction evtimet4a x BELOW median coal generating capacity in 1940
-- **sc40evtimet5a**: Interaction evtimet5a x BELOW median coal generating capacity in 1940
-- **sc40evtimet6a**: Interaction evtimet6a x BELOW median coal generating capacity in 1940
-- **sc40evtimet7a**: Interaction evtimet7a x BELOW median coal generating capacity in 1940
-- **bc40evtimet0**: Interaction evtimet0 x ABOVE median coal generating capacity in 1940
-- **bc40evtimet1a**: Interaction evtimet1a x ABOVE median coal generating capacity in 1940
-- **bc40evtimet2a**: Interaction evtimet2a x ABOVE median coal generating capacity in 1940
-- **bc40evtimet3a**: Interaction evtimet3a x ABOVE median coal generating capacity in 1940
-- **bc40evtimet4a**: Interaction evtimet4a x ABOVE median coal generating capacity in 1940
-- **bc40evtimet5a**: Interaction evtimet5a x ABOVE median coal generating capacity in 1940
-- **bc40evtimet6a**: Interaction evtimet6a x ABOVE median coal generating capacity in 1940
-- **bc40evtimet7a**: Interaction evtimet7a x ABOVE median coal generating capacity in 1940
-- **levtimet0**: Interaction evtimet0 x BELOW median household electricity access in 1940
-- **levtimet1a**: Interaction evtimet1a x BELOW median household electricity access in 1940
-- **levtimet2a**: Interaction evtimet2a x BELOW median household electricity access in 1940
-- **levtimet3a**: Interaction evtimet3a x BELOW median household electricity access in 1940
-- **levtimet4a**: Interaction evtimet4a x BELOW median household electricity access in 1940
-- **levtimet5a**: Interaction evtimet5a x BELOW median household electricity access in 1940
-- **levtimet6a**: Interaction evtimet6a x BELOW median household electricity access in 1940
-- **levtimet7a**: Interaction evtimet7a x BELOW median household electricity access in 1940
-- **hevtimet0**: Interaction evtimet0 x ABOVE median household electricity access in 1940
-- **hevtimet1a**: Interaction evtimet1a x ABOVE median household electricity access in 1940
-- **hevtimet2a**: Interaction evtimet2a x ABOVE median household electricity access in 1940
-- **hevtimet3a**: Interaction evtimet3a x ABOVE median household electricity access in 1940

-- **hevtimet4a**: Interaction evtimet4a x ABOVE median household electricity access in 1940
-- **hevtimet5a**: Interaction evtimet5a x ABOVE median household electricity access in 1940
-- **hevtimet6a**: Interaction evtimet6a x ABOVE median household electricity access in 1940
-- **hevtimet7a**: Interaction evtimet7a x ABOVE median household electricity access in 1940

For the dataset *imr_final_balanced.dta*:

-- **state_fips**: Federal Information Processing Standard (FIPS) state codes
-- **county_fips**: County FIPS
-- **year**: Year
-- **imr**: Infant mortality rate (per 1,000 live births) (Bailey et al., 2016)
-- **lnimr**: log(infant mortality rate)
-- **dimr**: Infant mortality rate - change 1940-1930
-- **births_rs**: Number of births by county of residence (Bailey et al., 2016)
-- **deaths_under_1yr**: Number of infant deaths by county of residence (Bailey et al., 2016)
-- **cap30mile**: Coal-fired generating capacity within 30 miles of county centroid (100s of MW) (FPC, 1947, 1948-62, 1963)
-- **cap30mile_1962**: Coal-fired capacity within 30 miles of county centroid (100s of MW) in 1962
-- **dcap30mile62_38**: Coal capacity within 30 miles of county centroid (100s of MW) - change 1962-1938
-- **coal30mile**: Coal consumption within 30 miles of county centroid (100K Tons) (FPC, 1947, 1948-62)
-- **cap50mile**: Coal-fired generating capacity within 50 miles of county centroid (100s of MW)
-- **cap100mile**: Coal-fired generating capacity within 100 miles of county centroid (100s of MW)
-- **cap30mile_hydro**: Hydroelectric capacity within 30 miles of county centroid (100s of MW) (FPC, 1963)
-- **cap50mile_hydro**: Hydroelectric capacity within 50 miles of county centroid (100s of MW)
-- **scap1940**: Dummy for BELOW median coal generating capacity in 1940
-- **bcap1940**: Dummy for ABOVE median coal generating capacity in 1940
-- **b1950cap30mile**: Interaction cap30mile x dummy for BEFORE 1950
-- **a1950cap30mile**: Interaction cap30mile x dummy for AFTER 1950
-- **scap1940cap30mile**: Interaction cap30mile x BELOW median coal generating capacity in 1940
-- **bcap1940cap30mile**: Interaction cap30mile x ABOVE median coal generating capacity in 1940
-- **sc40b50cap30mile**: Interaction scap1940cap30mile x dummy for BEFORE 1950
-- **bc40b50cap30mile**: Interaction bcap1940cap30mile x dummy for BEFORE 1950
-- **sc40a50cap30mile**: Interaction scap1940cap30mile x dummy for AFTER 1950
-- **bc40a50cap30mile**: Interaction bcap1940cap30mile x dummy for AFTER 1950
-- **scap1940cap50mile**: Interaction cap50mile x BELOW median coal generating capacity in 1940
-- **bcap1940cap50mile**: Interaction cap50mile x ABOVE median coal generating capacity in 1940
-- **scap1940cap100mile**: Interaction cap100mile x BELOW median coal generating capacity in 1940

-- **bcap1940cap100mile**: Interaction cap100mile x ABOVE median coal generating capacity in 1940
-- **cap30mileL2wlight**: Interaction cap30mile x BELOW median household electricity access in 1940 (electricity access is proxied by access to electric lighting from Haines & ICPSR, 2010)
-- **cap30mileH2wlight**: Interaction cap30mile x ABOVE median household electricity access in 1940
-- **cap50mileL2wlight**: Interaction cap50mile x BELOW median household electricity access in 1940
-- **cap50mileH2wlight**: Interaction cap50mile x ABOVE median household electricity access in 1940
-- **cap100mileL2wlight**: Interaction cap100mile x BELOW median household electricity access in 1940
-- **cap100mileH2wlight**: Interaction cap100mile x ABOVE median household electricity access in 1940
-- **amdiarrcap30milelightL2w**: Interaction cap30mileL2wlight x ABOVE median infant diarrhea deaths in 1930 (USBC, 1930)
-- **cap30mileL2wlightH2wcoalstove**: Interaction cap30mileL2wlight x ABOVE median use of coal cookstoves in 1940 (Haines & ICPSR, 2010)
-- **cap30milecapdop10l**: Interaction cap30mile x dummy for plant operating for <10 years
-- **cap30milecapdop10m**: Interaction cap30mile x dummy for plant operating for ≥10 years
-- **laty**: Country centroid latitude
-- **lony**: Country centroid longitude
-- **yprcp**: Total precipitation (in millimeters) (NOAA)
-- **ytav**: Average temperature (in degrees Celsius) (NOAA)
-- **yddb10**: Degree days below 10 degrees Celsius (calculation based on NOAA data)
-- **ydda29**: Degree days above 29 degrees Celsius (calculation based on NOAA data)
-- **lat**: Country centroid latitude
-- **lon**: Country centroid longitude
-- **yprcpm**: Total precipitation (in millimeters) (NOAA) - average of t-4 to t
-- **ytavm**: Average temperature (in degrees Celsius) (NOAA) - average of t-4 to t
-- **yddb10m**: Degree days below 10 degrees Celsius (NOAA) - average of t-4 to t
-- **ydda29m**: Degree days above 29 degrees Celsius (NOAA) - average of t-4 to t
-- **dist30_1960**: Distance between county centroid and power plant of 30MW or larger in 1960
-- **n0**: No. obs before coal plant openings - n0=25: no plant within county in the sample
-- **pop**: Population (Haines & ICPSR, 2010)
-- **pop1940**: Population in 1940
-- **lnpop1940**: log(population in 1940)
-- **spopurb**: Share of urban population (Haines & ICPSR, 2010)
-- **ppopurb**: Percentage of urban population
-- **dppopurb**: Percentage of urban population - change 1940-1930
-- **swhite**: Share of white population (Haines & ICPSR, 2010)
-- **pwhite**: Percentage of white population
-- **dpwhite**: Percentage of white population - change 1940-1930
-- **hschool**: Share of population 25yo+ with a high school degree (Haines & ICPSR, 2010)
-- **phschool**: Percentage of population 25yo+ with a high school degree
-- **emp**: Total employment (Haines & ICPSR, 2010)

-- **lnemp**: log(total employment)
-- **emp1940**: Total employment in 1940
-- **lnemp1940**: long(total employment in 1940)
-- **dlnemp**: log(total employment) - change 1940-1930
-- **mfg**: Manufacturing employment (Haines & ICPSR, 2010)
-- **lnmfg**: log(manufacturing employment)
-- **smfg**: Share of manufacturing employment
-- **pmfg**: Percentage of manufacturing employment
-- **dpmfg**: Percentage of manufacturing employment - change 1940-1930
-- **mfg1940**: Manufacturing employment in 1940
-- **lnmfg1940**: log(Manufacturing employment in 1940)
-- **dmfg**: Manufacturing employment - change 1940-1930
-- **mfgwages_bls90**: Manufacturing payroll per worker (1990 USD) (Haines & ICPSR, 2010)
-- **lnmfgwages_bls90**: log(manufacturing payroll per worker)
-- **dmfgwages_bls90**: Manufacturing payroll per worker (1990 USD) - change 1940-1930
-- **retwages_bls90**: Retail payroll per worker (1990 USD) (Haines & ICPSR, 2010)
-- **lnretwages_bls90**: log(retail payroll per worker)
-- **dretwages_bls90**: Retail payroll per worker (1990 USD) - change 1940-1930
-- **mrhouse_bls90**: Median dwelling rent (1990 USD) (Haines & ICPSR, 2010)
-- **lnmrhouse_bls90**: log(median dwelling rent)
-- **dlnmrhouse_bls90**: Median dwelling rent (1990 USD) - change 1940-1930
-- **N_mrhouse**: Number of obs with information on median dwelling rent
-- **light**: Share of households with electric lighting (Haines & ICPSR, 2010)
-- **light1940**: Share of households with electric lighting in 1940
-- **railroads**: Railroad mileage within county in 1911 (Donaldson & Hornbeck 2016)
-- **inst1944**: Predicted interstate highway from 1944 plan (Michaels 2008)
-- **farmelec**: Percentage of farm households with electricity (Haines & ICPSR, 2010)
-- **runwat**: Percentage of households with running water (Haines & ICPSR, 2010)
-- **elecgasstove**: Percentage of households with modern (electric + gas) stoves (Haines & ICPSR, 2010)

References

Bailey, Martha, Karen Clay, Price Fishback, Michael Haines, Shawn Kantor, Edson Severnini, Anna Wentz, and Inter-university Consortium for Political & Social Research ICPSR. 2016. "*U.S. County-Level Natality and Mortality Data, 1915-2007*." Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A "Market Access" Approach." *Quarterly Journal of Economics*, 131(2): 799– 858.

FPC, U.S. Federal Power Commission. 1947. "*Steam-Electric Plant Construction Cost and Annual Production Expenses, 1938-1947*." Washington, DC: U.S. Federal Power Com- mission.

FPC, U.S. Federal Power Commission. 1948-62. "*Steam-Electric Plant Construction Cost and Annual Production Expenses (Annual Supplements)*." Washington, DC: U.S. Federal Power Commission.

FPC, U.S. Federal Power Commission. 1963. "*Principal Electric Power Facilities in the United States (Map)*." Washington, DC: U.S. Federal Power Commission.

Haines, Michael R., and Inter-university Consortium for Political & Social Research ICPSR. 2010. *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, icpsr.org.

Michaels, Guy. 2008. "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System." *Review of Economics and Statistics*, 90(4): 683–701.

Muller, Nicholas Z, and Robert Mendelsohn. 2007. "Measuring the damages of air pollution in the United States." *Journal of Environmental Economics and Management*, 54(1): 1–14.

Muller, Nicholas Z, and Robert Mendelsohn. 2009. "Efficient pollution regulation: getting the prices right." *American Economic Review*, 99(5): 1714–39.

NOAA, National Oceanic & Atmospheric Administration. "*Global Historical Climatology Network*." www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily.

USBC, U.S. Bureau of the Census. 1930. "*Mortality Statistics*." US Department of Commerce, Bureau of the Census.