Replication Files READ ME

“Temperature, Disease and Death in London:

Analyzing Weekly Data for the Century from 1866-1965”

W. Walker Hanlon Casper Worm Hansen Jake Kantor

This document describes the replication files used to produce the analysis in the paper referenced above. The replication files are organized into several folders, which we discuss in order, starting with the raw data files and proceeding to the results.

Data Folder

The raw data folder contains the following files:

***aggregate\_mortality\_weekly\_data*** *–* This file is based on the Registrar Generals “Weekly Returns of Births Deaths and Marriages in London”. The data cover the full sample period, 1866-1965, though not all variables are available in every year. The file contains the following fields:

* *date –* Generated from the week and year information in the original data
* year
* week – note that the data are reported weekly for between 51 and 53 weeks a year. Weeks at the beginning or end of a year may include a few days that actually fell into the previous or next year.
* *births*
* *totaldeaths*
* *thickfog\_report –* this contains an indicator for whether thick fog was reported in the city in a particular day. See Hanlon (2019), *“London Fog: A Century of Pollution and Mortality, 1866-1965”* for further details on this variable
* *temperature –* mean weekly temperature. This is not the main temperature variable used in the paper. This and the following three weather variables are reported at Greenwich (before WWII) or Kew Gardens (after WWII). They are used in robustness exercises only.
* *pressure* – this is used to calculate absolute humidity in robustness exercises
* *humidity –* relative humidity
* *rainfall* – in inches
* *infants01 –* infant deaths
* *decade* – decades, enumerated from the 1860s (1) to 1960s (10).

***by\_age\_mortality\_weekly\_data –*** *This file contains mortality by age group for the various age groups reported in the Registrar General’s weekly reports. The variables are:*

* *year, week*
* *age\_all, age\_0\_1, etc.* – the present the number of deaths in the indicated age groups. Note that the reported age groupings change over time, so not all ages are available for all years.

***digestive\_mortality\_weekly\_data*** -- This file presents weekly digestive disease mortality for 1870-1965. Because of the way digestive deaths were reported, it is possible to construct this series in a consistent way than other causes of death. This series includes diarrhea & dysentery, cholera, typhoid, enteritis and gastritis.

* *year, week*
* *digestive\_all –* total digestive deaths
* *digestive\_0\_1 –* digestive deaths among infants age 0-1
* *digestive\_1\_5* – digestive deaths among children ages 1-5

***modern\_mortality\_weekly\_data*** – this file contains weekly mortality data from 1981-2006 as well as temperature data from Oxford for those years.

* *week*
* *year*
* *oxford\_max\_temp* – maximum weekly temperature observed at Oxford
* *oxford\_min\_temp* – minimum weekly temperature observed at Oxford
* *deaths –* total deaths in London in the week

***mortality\_by\_cause\_weekly\_data* –** This file contains mortality data broken down by cause of death and age group for 1870-1939. The causes of death have been grouped into categories. You can view the individual causes of death that fall into each category in the “auxiliary\_files” folder, “causes\_by\_category” file. This file contains the following variables:

* *date, year, decade, week*
* *category*  - this contains the cause of death category
* *age\_all ­-*  deaths at all ages
* *age\_0\_1, age\_1\_5, etc. –* these variables report deaths in various age categories. Note that age categories changed over time, so some age groupings are missing for some years.
* *births*
* *thickfog\_report –* weeks when heavy fog was reported in the city
* *rainfall –* in inches

***oxford\_temp\_weekly\_data –*** this file provides weekly maximum and minimum temperatures measured at the Radcliffe Observatory, Oxford, UK. These weekly observations are built using the daily temperature data available from the Met’s MIDAS dataset. The Oxford station number (DCNN) is 4522. The daily data are mapped to weeks based on the days included in each week of the Registrar General’s reports. For the mapping from days into weeks, see “auxiliary\_files/day\_to\_week\_tt.csv”. This file contains the following variables:

* *year, week*
* *oxford\_max\_temp* – maximum weekly temperature observed at Oxford
* *oxford\_min\_temp* – minimum weekly temperature observed at Oxford

***subdigestive\_mortality\_weekly\_data*** – this file provides mortality data for series within the digestive category for 1870-1939. The variables are:

* *date, year, decade, week*
* *new\_cause* – this is the more detailed cause of death, which contains the following options: Cholera, Diarrhoea and Enteritis (including dysenter and gastritis), Typhoid fever, and Other Fever (which includes simple continued fever, etc.)
* *age\_all –* deaths across all age groups
* *thickfog\_report, pressure, humidity, rainfall –* as described above

Do Files

This folder contains the STATA .do files that prepare and analyze the data. The two main files are:

**0\_prepare\_datasets.do** – this file processes the raw data into the intermediate datasets used in the analysis.

**1\_analysis\_initiation\_file.do** – this file initiates the analysis files that generate the tables and figures found in the paper and appendix

Note: To use either of these do files, you need to update the global at the top of the file to reflect your filepath

Intermediate datasets

This folder contains the intermediate datasets that are produced from the raw data found in the Data folder and which feed into the analysis. Each of these datasets can be understood by referencing the portion of **0\_prepare\_datasets.do** that produces it from the raw data.

Results

This folder contains the results produced by running **1\_analysis\_initiation\_file.do**, organized into folders corresponding to the tables and figures found in the draft and appendix.

Auxiliary Files

The Auxiliary Files folder contains some additional information that can be helpful for understanding our main data files.

***causes\_by\_category*** – contains a listing of the various causes that are included in each of the aggregated cause-of-death categories used in the cause-of-death analysis (the Excel version also includes the number of deaths falling into each of the original causes for observations up to WWII).

***day\_to\_week\_tt* –** provides a mapping from days to the weeks used in the Registrar General’s weekly reports.

Temp

This is a folder for holding temporary files produced during intermediate steps of the analysis or data preparation.