Online Appendix To:
"Killer Cities: Past and Present"

W. Walker Hanlon — UCLA and NBER

Yuan Tian — UCLA

January 7, 2015

This document describes the data used in the paper "Killer Cities: Past and Present" and presents additional details related to the analysis. It begins with an overview and instructions related to the replication files that we have provided. The next section includes an additional discussion of some of the related literature. The third section provides additional details related to the construction of the data. Finally, we provide some additional results not included in the main draft.

1 Overview of the replication files

The data file contains two separate data sets and two Stata .do files, one each for the analysis of English data and Chinese data. Each Stata analysis code can be run after altering the global filepath at the top to indicate the folder's location. Each analysis file produces a set of regression results which are saved in the Results folder in TEX format.

The Stata data file for the English data contains the following variables:

- *loc* – the district name

- *year* – the start year of the decade that the mortality data comes from
- *asr_total* – total age standardized mortality
- *seaport* – an indicator variable for whether the district included a seaport
- *seaport_tonnage* – value of total imports through the seaport in millions of pounds sterling
- *air_frost* – the average number of air frost days in the location based on modern data
- *di_shr* – the share of heavily polluting industries in local production
- *ln_dist_pop* – the natural log of district population (in tens of thousands)
- *density* – the population density in thousands of persons per square kilometer

The Stata data file for the Chinese data contains the following variables:

- *loc* – the district (or county) name
- *asr_total* – total age standardized mortality
- *di_shr* – the share of heavily polluting industries in local production
- *ln_dist_pop* – the natural log of district population (in tens of thousands)
- *density* – the population density in thousands of persons per square kilometer
- *north* – an indicator variable for whether the district is in a province that is northern of Huai River
- *coast* – an indicator variable for whether the district is in a coastal province

2 Additional related literature

In addition to the studies cited in the main text, this study is related to a substantial literature documenting the relationship between pollution and mortality using data from developed countries (e.g., Samet *et al.* (2000), Pope III *et al.* (2002), Chay *et al.* (2003), etc.). Less is known about pollution in developing countries, where pollution levels can be substantially higher than those observed in developed countries today. However, recent studies have begun to explore this issue (Jayachandran (2009), Almond *et al.* (2009), Jia (2012), Ebenstein (2012), Greenstone & Hanna (2014), etc.). The major challenge faced by this literature is a lack of reliable data on pollution levels, pollution causes, and outcomes. For similar reasons, our understanding of historical pollution patterns is also somewhat limited, though recently this topic has been studied more intensively (see, e.g., Troesken & Clay (2011), Barreca *et al.* (2014), Clay *et al.* (2014)).

There is also a natural connection to work on the Environmental Kuznets Curve (Grossman & Krueger (1995)), which suggests that pollution will respond to rising income by following a predictable inverted U-shaped path. Support for this idea rests in part on the observed evolution of pollution levels in early industrializing countries, including England. This point is highlighted by Zheng & Kahn (2013), who write, "A long-run urban environmental history for developed nations suggests that local environmental problems could improve in Chinese cities." Similarly, Vennemo *et al.* (2009) argue that, "China's development with respect to environmental pollution appears to be following a path that is similar to the one established by more industrialized countries when they were at earlier stages of development."

3 Data construction

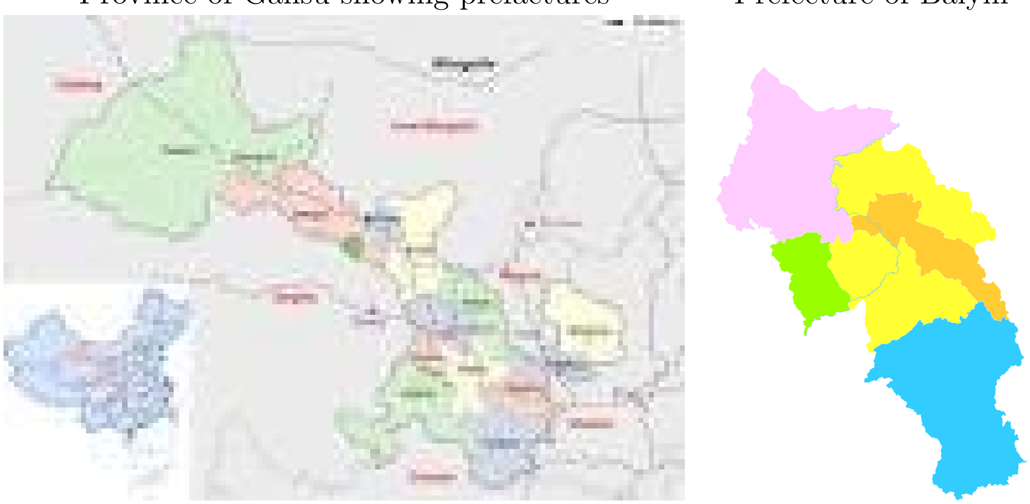
This section provides some additional details about the methods used to construct the data sets provided in the downloadable file. We begin with a discussion of the geographic units used in our data. Next, we discuss the construction of the age-standardized mortality variable. We then discuss the construction of the pollution variable.

3.1 Geographic units

In China, local administrative units are organized in several levels. Below the national level is the province. The next level down is the prefecture, followed by the district or county. We use the district or county level. This division is illustrated for one province, Gansu, and prefecture, Baiyin, in Figure 1.

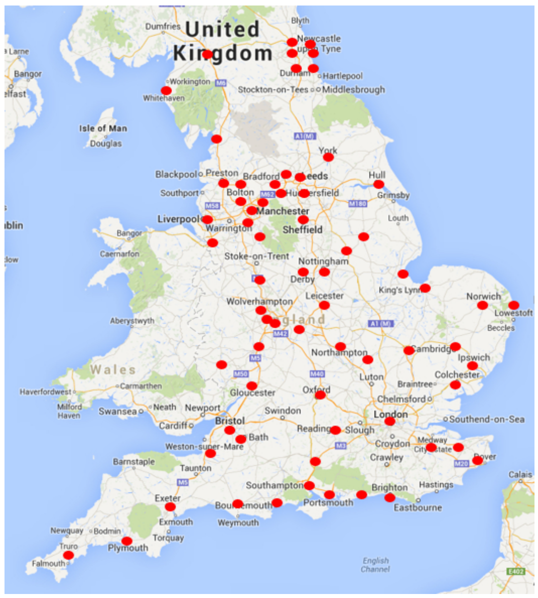
Figure 1: Geographic units in the Chinese data

Province of Gansu showing prefactures Prefecture of Baiyin



For England, the country is divided into counties which are then subdivided into over 600 Registrar's Districts. Our analysis includes the 64 districts corresponding to the cities for which occupation data were reported in the 1861 Census of Population, with the exception of London, a substantial outlier, and a few cities for which we had difficulty constructing the full set of explanatory variables. Figure 2 provides a map of the cities included in our analysis database.

Figure 2: Locations included in the English data



The occupation data used in this study was collected for the city level. City boundaries do not necessarily correspond to the boundaries of Registrar's Districts; they may be larger or smaller. However, we believe that the city's industrial composition is a good indicator of the level of pollution experienced by the residents of the principal district of the city.

3.2 Age-standardized mortality

Both the Chinese and English data report mortality by age category. Using this information we have constructed age-standardized mortality rates using the following formula:

$$Mortality_i^{std} = \sum_g^G MR_{gi} PS_g \,,$$

where MR_{gi} is the mortality rate in age group g in county i and PS_g is the share of population in age group g in the whole country.

4 Constructing the pollution measure

To construct our pollution measure, we require a list of heavily polluting industries. This list is obtained from the Chinese government, which classified industries as either heavily polluting or less polluting.[1] The list used for the Chinese data is in Table 4.

captionList of heavily polluting industries for Chinese data

	Heavily polluting industries		**Less polluting industries**
1	Mining	16	Food
2	Brewing	17	Beverage
3	Textile	18	Tobacco
4	Leather making	19	Apparel
5	Paper making	20	Wood processing/furniture
6	Petroleum processing, coking, nuclear fuel	21	Printing
7	Chemical products manufacturing	22	Sport/musical instrument
8	Pharmacy(medical)	23	Metal products manufacturing
9	Chemical fiber manufacturing	24	Mechanical manufacturing
10	Rubber products manufacturing	25	Recycling
11	Plastic products manufacturing	26	Other power generation
12	Nonmetallic mineral products manufacturing	27	Water supply
13	Ferrous metal melting and processing	28	Forestry/farming
14	Nonferrous metal melting and processing	29	Other industries
15	Thermal power		

This list has been mapped to the set of industries available in the English data from Hanlon & Miscio (2014). Table 1 provides a list of the polluting and less polluting

[1] In the document issued by Chinese Environment Protection Bureau([2003] No.10) named *About Inspection of Environmental Qualification of Companies that are Applying for Listing and Refinancing*, the heavy-polluting industries are: metallurgical, chemical, petrochemical, coal, thermal power, building materials, paper making, brewing, pharmaceutical, fermentation, textile, leather and mining.

industries mapped to the industrial categories available in the English data.

Table 1: List of heavily polluting and less polluting industries for English data

Heavily polluting industries	**Less polluting industries**
Chemicals & drugs	Apparel
Earthenware & bricks	Food processing
Leather goods	Instruments & jewelry
Metal and machinery manuf.	Oils & soaps
Mining and related activity	Shipbuilding
Paper manufacturing	Tobacco processing
Brewing & distilling	Vehicle production
Textiles	Wood & furniture
Gas, electricity & water utilities	

In China, we have data on sales by polluting industries and all industries in a location i. We use this to construct a pollution measure for location i according to the following formula:

$$PollnRevShr_i = \frac{\text{Sales revenue in polluting industries}_i}{\text{Sales revenue in all industries}_i}.$$

In England, we have data on employment by industry. We use this to construct a pollution measure for location i according to the following formula:

$$PollnEmpShr_i = \frac{\text{Employment in polluting industries}_i}{\text{Employment in all industries}_i}.$$

5 Additional results

Table 2 provides some additional results for China. Column 1 presents results where we have included district-level GDP per capita into our baseline regression. As expected, GDP per capita is negatively related to mortality. Including this variable diminishes somewhat the relationship between population or population density and mortality, but these relationships remain statistically significant. In Column 2 we add in the average years of schooling in the district. Education also has a negative impact on mortality, a finding that matches previous results (e.g., Lleras-Muney (2005)). Including this variable further reduces the relationships between population and mortality and between population density and mortality. Overall, these results suggest that the relationship between population and mortality and between population density and mortality are largely driven by the fact that more populated and more densely populated districts also have higher income and more education, two factors that reduce mortality. Finally, note that when these additional controls are included the impact of pollution actually increases and becomes more statistically significant.

Table 2: Additional results for China including education and GDP per capita

	DV: Age-standardized mortality	
Log Population	-0.220***	-0.121
	(0.0825)	(0.0789)
Population density	-0.188**	-0.123
	(0.0817)	(.0814)
Polluting industry shr.	0.537*	0.529*
	(0.266)	(.261)
GDP per capita	-0.0311***	-0.0243***
	(0.0064)	(0.0063)
Avg. years of schooling		-0.386***
		(0.145)
Other controls	Yes	Yes
Observations	221	221
R-squared	0.341	0.369

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors in parenthesis. Additional control variables: indicator for provinces north of the Huai River, indicator for coastal provinces.

References

Almond, Douglas, Chen, Yuyu, Greenstone, Michael, & Li, Hongbin. 2009. Winter Heating or Clean Air? Unintended Impacts of China's Huai River Policy. *American Economic Review*, **99**(2), 184–90.

Barreca, Alan, Clay, Karen, & Tarr, Joel. 2014 (February). *Coal, Smoke, and Death: Bituminous Coal and American Home Heating*. NBER Working Paper No. 19881.

Chay, Kenneth, Dobkin, Carlos, & Greenstone, Michael. 2003. The Clean Air Act of 1970 and Adult Mortality. *Journal of Risk and Uncertainty*, **27**(3), 279–300.

Clay, Karen, Lewis, Joshua, & Severnini, Edson. 2014 (November). *Benefits and Costs of Electricity Pre-Clean Air Act*. Working paper.

Ebenstein, Avraham. 2012. The Consequences of Industrialization: Evidence from Water Pollution and Digestive Cancers in China. *Review of Economics and Statistics*, **94**(1), 186–201.

Greenstone, Michael, & Hanna, Rema. 2014. Environmental Regulations, Air and Water Pollution, and Infant Mortality in India. *American Economic Review*, **104**(10), 3038–72.

Grossman, Gene M., & Krueger, Alan B. 1995. Economic Growth and the Environment. *The Quarterly Journal of Economics*, **110**(2), pp. 353–377.

Hanlon, W.W., & Miscio, A. 2014 (December). *Agglomeration: A Dynamic Approach*. NBER Working Paper No. 20728.

Jayachandran, Seema. 2009. Air Quality and Early-Life Mortality: Evidence from Indonesias Wildfires. *Journal of Human Resources*, **44**(4), 916–954.

Jia, Ruixue. 2012 (November). *Pollution for Promotion*.

Lleras-Muney, A. 2005. The Relationship Between Education and Adult Mortality in the U.S. *Review of Economic Studies*, **72**(1), 189–221.

Pope III, C, Burnett, RT, Thun, MJ, & et al. 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-Term Exposure to Fine Particulate Air Pollution. *JAMA*, **287**(9), 1132–1141.

Samet, Jonathan M., Dominici, Francesca, Curriero, Frank C., Coursac, Ivan, & Zeger, Scott L. 2000. Fine Particulate Air Pollution and Mortality in 20 U.S. Cities, 19871994. *New England Journal of Medicine*, **343**(24), 1742–1749. PMID: 11114312.

Troesken, W, & Clay, K. 2011. Did Frederick Brodie Discover the World's First Environmental Kuznets Curve? Coal Smoke and the Rise and Fall of the London Fog. *In:* Libecap, G., & Steckel, R. H. (eds), *The Economics of Climate Change: Adaptations Past and Present*. University of Chicago Press.

Vennemo, Haakon, Aunan, Kristin, Lindhjem, Henrik, & Seip, Hans Martin. 2009. Environmental

Pollution in China: Status and Trends. *Review of Environmental Economics and Policy*, **3**(2), 209–230.

Zheng, Siqi, & Kahn, Matthew E. 2013. Understanding China's Urban Pollution Dynamics. *Journal of Economic Literature*, **51**(3), 731–772.