The Impact of Sustained Exposure to Particulate Matter on Life Expectancy: New Evidence from China's Huai River Policy*

Avraham Ebenstein[1], Maoyong Fan[2], Michael Greenstone[3], Guojun He[4], and Maigeng Zhou[5]

January 2016

Abstract:
This paper finds that a 10 μg/m^3 increase in airborne particulate matter (PM_{10}) reduces life expectancy by 0.75 years (95% CI: 0.22, 1.28). This estimate is derived from quasi-experimental variation in PM_{10} generated by China's Huai River Policy, which provides free or heavily subsidized coal for indoor heating during the winter to cities north of the Huai River but not to the south. The findings are derived from a regression discontinuity design based on distance from the Huai River, and are robust to using parametric and non-parametric estimation methods, different kernel types and bandwidth sizes, and adjustment for a rich set of demographic and behavioral covariates. Furthermore, the shorter lifespans are almost entirely due to elevated rates of cardiorespiratory mortality, suggesting that PM_{10} is the causal factor. The estimates imply that bringing all of China into compliance with its Class I standards for PM_{10} would save 4.3 billion life years.

Word Count: 150

Keywords: Pollution, Particulate Matter, Life Expectancy, China
JEL Classifications: Q13, P28, Q28, Q53

---

* The order of authors is alphabetical, and all five authors are co-first authors. Faraz Hayat, Dan Stuart and Harshil Sahai provided excellent research assistance. This study is supported by the China National Science and Technology Pillar Program 2013 (2013BAI04B02) from the Ministry of Science and Technology.
[1] Department of Economics, Hebrew University of Jerusalem, Mount Scopus, Israel
[2] Department of Economics, Ball State University, Muncie IN 47304, USA
[3] Department of Economics, University of Chicago, Chicago IL 60637, USA and National Bureau of Economic Research, Cambridge MA 02138, USA
[4] Division of Social Science, Division of Environment, and Department of Economics, The Hong Kong University of Science and Technology, Hong Kong, China
[5] National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

I. Introduction

Is airborne particulate matter today's greatest environmental risk to human health? Among the 5.9 billion people who live in countries where particulate matter smaller than 10 micrometers (PM_{10}) readings are available, 4.5 billion are currently exposed to PM_{10} concentrations that are at least twice the concentration that the World Health Organization (WHO) considers safe. The health effects of inhaling PM_{10} have been widely studied and found to be robustly associated with elevated risk of heart disease, stroke, and lung cancer (Pope 1989, Dockery et al. 1993, Wellenius et al. 2005). However, questions regarding the health effects of PM_{10} and its appropriate regulation continue to be of tremendous scientific and policy relevance, because it is apparent that the existing evidence has not convinced countries to adopt and enforce tough emission standards. Current policies regarding PM_{10} also influence the probability that the world will face disruptive climate change, because the combustion of fossil fuels that cause PM_{10} also cause greenhouse gas emissions.

At least three limitations have plagued the existing evidence linking health to air pollution, especially at the concentrations that prevail in many of today's developing countries. First, the literature is almost entirely comprised of observational studies, comparing populations across locations with varying exposure to pollution. These studies are likely to confound air pollution with unobserved determinants of health that are correlated with pollution exposure (e.g., income, hospital quality, water pollution, etc.) as has been empathized in a recent *Science* article (Domenici et al. 2014). Second, the available evidence is largely based on examinations of populations exposed to the modest levels of PM_{10} that are commonly observed in developed countries where reliable pollution and health data are more readily available (Pope et al. 2009). PM_{10} concentrations in many developing countries (e.g., India and China) are five to ten times higher

than in developed countries; consequently, the existing evidence has little empirical relevance for these countries if there is a nonlinear relationship between health and pollution. Third, the most important questions about pollution center on the impacts of sustained exposure (e.g., lifetime exposure). However, there have been few opportunities to measure long-run exposure to air pollution.[1] As a consequence, the existing literature focuses on shorter run variation in PM_{10} exposure and often examines outcomes (e.g. hospitalization, infant outcomes) that are only indirectly related to longer run outcomes like life expectancy.[2]

This paper estimates the effect of sustained exposure to PM_{10} on life expectancy with recent data from China and, in so doing, addresses each of the previous literature's limitations. First, the quasi-experimental research design is based on China's Huai River Policy. The policy was instituted during the 1950s when economic resources were allocated through central planning, and dictated that areas to the north of the Huai River received free or highly subsidized coal for indoor heating. This led to the construction of a coal-powered centralized heating infrastructure only in cities north of the Huai River, and no equivalent system in cities to the south; the legacy of that policy persists even today.[3] Consequently, the findings are derived from a regression discontinuity (RD) design based on distance from the Huai River.

[1] Notable exceptions include Dockery (1993), Pope et al. (2002), and Pope et al. (2009). However, these studies have at least one of the following limitations: (a) exploit observational variation in PM (b) use small samples (c) assume no selective migration (d) focus on low levels of pollution found in the US.

[2] Note that the literature in China is almost exclusively composed of studies examining short-term increases in morbidity or mortality due to temporal variation in pollution (Wu et al. 2010, Chen et al. 2012, Rich et al 2015, Wu et al. 2010). This is also characteristic of much of the literature in developed countries. For example, Chay and Greenstone (2003) find a robust positive relationship between total suspended particulates and infant mortality in the U.S.

[3] China began to install centralized heating systems in the 1950s with assistance from the Soviet Union. Resource constraints made national coverage infeasible, and so Chinese Premier Zhou Enlai proposed the Qin-Huai line as the southern cutoff point. This divide roughly traces China's Qin Mountains and Huai River near the 33 degree latitude line, which also corresponds to China's January temperature 0^0 Celsius contour. In the late 1970s and early 1980s, the centralized heating systems were dramatically expanded as a result of China's economic reforms, but the policy was kept in place. All buildings (including factories, workplaces and civilian buildings) in urban areas north of the Qin-Huai line were provided with centralized and heavily-subsidized winter heating.

Second, the average PM_{10} concentration in China during the examined period is 102 $\mu g/m^3$ (i.e., more than 5 times the WHO standard), so the results are informative for the 1.36 billion people currently living in China, as well as the several billion other people that are exposed to high PM_{10} concentrations around the world. Further, the analysis is conducted with the most comprehensive data file ever compiled on health and pollution in a developing country.

Third, the Huai River Policy produces sustained differences in PM_{10} concentrations between the areas north and south of the river that date back to China's first measurement of airborne particulate matter in the early 1980s, and persist even today. The combination of these differences and the restrictions on internal migration due to the *hukou* system (which have become less binding over time) provide a unique opportunity to observe the effects of sustained differences in PM_{10} exposure on the life expectancy of nearly fixed populations.

The analysis indicates that PM_{10} exposure causes people to live substantially shorter and sicker lives at the concentrations present today in China and other developing countries. We estimate that the Huai River Policy generates an increase in PM_{10} exposure of 41.6 $\mu g/m^3$ (95% CI: 11.6, 71.6) and a decline of 3.1 years (95% CI: 1.3, 4.9) in life expectancy just to the north of the river. Further, the elevated mortality rates are concentrated among cardiorespiratory causes of death, while there is little evidence of a difference in mortality rates for causes that are not plausibly related to air pollution. More broadly, the results suggest that long-term exposure to an additional 10 $\mu g/m^3$ is associated with a 0.75 year (95% CI: 0.22, 1.28) decline in life expectancy.[4] As a back

[4] The effects of living north of the Huai River on PM_{10} and life expectancy are derived from non-parametric application of the regression discontinuity design based on distance from the Huai River. We follow Calonico et al. (2014a) to determine the optimal bandwidths to estimate the discontinuity in each variable. The instrumental variables, or two-stage least squares, estimate of the effect of PM_{10} on life expectancy can also be generated by taking the ratio of the estimated discontinuity in life expectancy to the estimated discontinuity in PM_{10}, with both estimated by local linear regression. The instrumental variable estimate uses the optimal bandwidth chosen for life expectancy for both PM_{10} and life expectancy, so it does not simply equal the ratio of the estimates of the effects of the Huai River Policy on these variables. See Section III for details.

of the envelope calculation, this estimate suggests that bringing all of China into compliance with its *own* Class I standard of 40 μg/m^3 standard would save 4.3 billion life years, but even a more modest improvement in which the country was brought into compliance with China's Class II standard of 70 μg/m^3 standard would save 1.3 billion life years.[5]

We probed the robustness of the results in several ways. First, the estimates are qualitatively unchanged by choices about parametric versus non-parametric estimation methods (Gelman and Imbens 2014), kernel type and bandwidth size. Second, they are also statistically unchanged by adjustment for a rich set of demographic and behavioral covariates. Relatedly, we cannot reject that predicted life expectancy, based on all available covariates except PM_{10}, is equal north and south of the river. Third, in a placebo exercise, we find that the discontinuity in pollution and life expectancy at the Huai River is only evident at the river itself, and not at geographic displacements of the actual boundary. Using artificial boundaries at 1 degree latitude intervals away from the river, we are unable to find a statistically significant increase in pollution or decrease in life expectancy.

This paper is related to Chen et al. (2013), which also exploits the Huai River Policy to examine the health impact of total suspended particulates (TSPs) using data from the 1990s but improves on it in several critical ways. First, this study examines the health consequences of the smaller particles that comprise PM_{10}, rather than the broader category of TSPs. Environmental regulation throughout the world is focused on PM_{10} and having direct evidence on the health consequences of exposure is highly policy relevant. Second, the analysis is conducted with

[5] This estimate is generated by applying the estimated impact of PM_{10} on life expectancy with population exposure estimates for PM_{10} taken from An et al. (2013). They estimated exposure in each of 11 regions in China. We calculate the number of life years that would be saved by reducing PM_{10} to the national standards levels (40 μg/m^3 or 70 μg/m^3) by each region. If a region's PM_{10} exposure is already lower than the standards, a value of zero is assigned. The total life years saved are the sum of life years saved across all regions.

significantly more accurate measures of mortality from a more recent period (2004-2012). In particular, the nationally representative mortality data are generated from a coverage population that is roughly 8 times larger than in Chen et al. (2013), and were collected following China's rapid economic growth after the 1990s. Third, Chen et al. (2013) present results from parametric RD approaches that have subsequently been found to have undesirable statistical properties (Gelman and Imbens 2014). This study presents results using both parametric and non-parametric estimation strategies and considers the sensitivity of the results to changes in functional form, bandwidth, and kernel choices. Fourth, the Chinese CDC collected data on health-relevant behaviors in 2010, allowing for adjustment for smoking rates, dietary patterns, and other potential risk factors for cardiorespiratory mortality that were unavailable for the previous study.

The rest of the paper proceeds as follows. Section II describes the data sources, while Section III outlines the empirical strategy. Section IV presents the results and Section V provides some conclusions.

II. Data Sources

The heart of the analysis is based on two data files that together provide location-specific information on life expectancy, health, and air pollution. The mortality and life expectancy data come from the Chinese Center for Disease Prevention and Control's (CDC) Disease Surveillance Points (DSP) survey. The DSP is a remarkably high-quality nationally representative survey; it provides detailed cause-of-death data (verified by verbal autopsy) for a coverage population of over 77 million people at 161 separate locations for each year between 2004 and 2012. The cause of death information recorded in the autopsies is used to assign all deaths to either cardiorespiratory causes of death (i.e., heart, stroke, lung cancers, and respiratory illnesses) that are plausibly related

to air pollution exposure or non-cardiorespiratory causes (i.e., cancers other than lung, injuries, and all other causes).[6] The latter are used as a placebo-style test.

The air pollution data were compiled by combining print and electronic resources for six main pollutants over the period 1981-2012. These data are available in the online appendix and represent, to our knowledge, the most comprehensive archive of Chinese air pollution data ever assembled. To estimate the impacts of long-run exposure to pollution, the location-level panel data are collapsed to a 154 observation, location-level, cross-sectional data set; the panel data are collapsed in this manner because the Huai River regression discontinuity design is fundamentally a cross-sectional design.

Two other data sets are used to adjust the estimates for potential confounders. Measures of health-relevant behavior, such as smoking prevalence and dietary patterns, are taken from a CDC survey conducted in 2010 and the paper's results are adjusted for them. This survey is almost ideal for this study's purposes, since it was designed to capture patterns in behavior specifically for the coverage population at each DSP location. We use the 2010 census as a secondary data set which enables us to adjust the estimates for demographic covariates included in the census and not available from the CDC survey (e.g. share minority, average years of education).[7]

III. Econometric Model

We use two approaches to estimating the relationship between PM_{10} and health outcomes. The first approach is a "conventional" strategy that uses ordinary least squares to fit the following

[6] The specific diseases included in cardiorespiratory mortality are listed in Table S1.

[7] Note that the adjustment for covariates is reported for the parametric estimation. In the non-parametric estimation, we have also estimated models where we focus on residual life expectancy (after accounting for observables). Since the covariates are similar on either side of the river, we primarily report non-parametric results that are based on actual life expectancy, but non-parametric results using residual life expectancy are available upon request.

equation to the cross-sectional data file:

$$Y_j = \beta_0 + \beta_1 PM_j + X_j\Gamma + \varepsilon_j \quad (1)$$

where *j* references a city or location in China. PM_j is the particulate matter (PM_{10}) concentration in city j, X_j is a vector of the observable characteristics of the city that might influence health outcomes other than air quality, and ε_j is a disturbance term. The dependent variable Y_j is either a measure of mortality rates in city j or its residents' life expectancy, which is a simple function of age-specific mortality rates.

The coefficient β_1 measures the effect of PM_{10} exposure on mortality, after controlling for the available covariates. Consistent estimation of β_1 requires that unobserved determinants of mortality do not covary with PM_j after adjustment for X_j. Thus, the "conventional" approach rests on the assumption that linear adjustment for the limited set of variables available in the census removes all sources of confounding. Previous research has raised substantive concerns about the validity of this assumption (Chay et al. 2003, Domenici et. al 2014). Further, pollution concentrations are measured with error and it is well know that classical measurement error will attenuate the coefficient associated with PM_{10}.

The second approach leverages the regression discontinuity (RD) design implicit in the Huai River Policy to measure its impact on PM_{10} concentrations and life expectancy.[8] This paper's RD design exploits the Huai River Policy that provides free or heavily subsidized coal for indoor heating north of the river and no subsidies to the south. Specifically, we separately test whether the Huai River Policy caused discontinuous changes in PM_{10} and life expectancy to the north of the river. The respective necessary assumptions are that any unobserved determinants of PM_{10} or

[8] The RD design was developed more than five decades ago and has been used successfully to test the causal nature of relationships in a wide range of fields including psychology, education, statistics, biostatistics, and economics (Cook and Campbell 1979, Imbens and Lemieux 2008, Card et al. 2009, Lee and Lemieux 2010).

mortality change smoothly as they cross the river. If the relevant assumption is valid, adjustment for a sufficiently flexible polynomial in distance from the river or local linear regressions on either side of the river will remove all potential sources of bias and allow for causal inference.

In practice, we estimate the following parametric equations to test for the impacts of the Huai River Policy:

$$PM_j = \alpha_0 + \alpha_1 N_j + f(L_j) + X_j\gamma + u_j \quad (2a)$$

$$Y_j = \delta_0 + \delta_1 N_j + f(L_j) + X_j\varphi + \varepsilon_j \quad (2b)$$

where j references a city or location in China. PM_j is the average annual ambient concentration of PM_{10} in city j over the period 2004-2012 and Y_j is a measure of city j's mortality or life expectancy at birth. N_j is an indicator variable equal to 1 for locations that are north of the Huai River line, $f(L_j)$ is a 3rd order polynomial in degrees north of the Huai River interacted with N_j (chosen based on goodness of fit criteria), and X_j is a vector of the demographic and city characteristics, other than air quality, that are associated with mortality rates.

An alternative estimation strategy for the RD approach involves non-parametric identification of (2a) and (2b). For example, consider the following set-up for estimation by local linear regression:

$$PM_j = \alpha_0 + \alpha_1 N_j + \alpha_2(L_j - L_N) + \alpha_3 N_j(L_j - L_N) + u_j \quad (3a)$$

$$s.t.\ L_N - h \leq Lj \leq L_N + h$$

$$Y_j = \delta_0 + \delta_1 N_j + \delta_2(L_j - L_N) + \delta_3 N_j(L_j - L_N) + \varepsilon_j \quad (3b)$$

$$s.t.\ L_N - h \leq Lj \leq L_N + h$$

where L_N is the degrees latitude along which the Huai River lies and h is latitude degrees of bandwidth (i.e., the acceptable distance from the discontinuity for sample inclusion). The choice of h involves balancing the conflicting goals of focusing comparisons near the river where the

identification assumption is strongest and having a large enough sample for reliable estimation. In the non-parametric results, we rely on a choice-rule which determines the optimal h as a function of the data (e.g. Imbens and Kalyanaraman 2012, Calonico et al. 2014a).[9]

Below we report on estimation of equations (3a) and (3b) for PM_{10}, life expectancy at birth, and several other outcomes of interest, such as cardiorespiratory mortality. The parameter of interest is α_1 or δ_1, which provide an estimate of whether there is a discontinuity in outcomes at locations just to the north of the Huai River (relative to locations to the south). These estimates are reported using three separate kernels or weighting functions—triangle, uniform, and Epanechnikov.

Importantly, the results in (2) and (3) can each be used to develop estimates of the impact of PM_{10} concentrations on life expectancy. Specifically, if the Huai River Policy only influences mortality through its impact on PM_{10}, then it is valid to treat equation (2a) as the first-stage in a two-stage least squares (2SLS) system of equations. An important appeal of the 2SLS approach is that it produces estimates of the impact of units of PM_{10} on life expectancy, so the results are applicable in other settings (e.g., India and other developing countries with comparable PM_{10} concentrations). The second stage equation is:

$$Y_j = \beta_0 + \beta_1 \widehat{PM}_j + \beta_2 f(L_j) + X_j\varphi + \varepsilon_j \qquad (2c)$$

where $\widehat{PM}_j$ represents the fitted values from estimating (1a) and the other variables are as described above. The 2SLS approach offers the prospect of solving the confounding or omitted variables problem associated with the estimation of the air pollution-health effects relationship and is a solution to the attenuation bias associated with the mismeasurement of PM_{10}.

[9] We also report results from the parametric RD approach described in equations (1a) and (1b) that restricts the sample to locations within 5 degrees latitude of the Huai River. This sample restriction which drops all observations that are distant from the Huai River can be thought of as informal way of implementing our local linear methods, which employ kernel weighting functions that assign greater weight to observations near the discontinuity.

A non-parametric analogue to (2c) can be estimated by taking the ratio of the estimated discontinuity in life expectancy to the estimated discontinuity in PM_{10}, with both estimated by local linear regression. The result is an instrumental variables, Wald estimate of the impact of PM_{10} on life expectancy that is analogous to the 2SLS estimates produced in (2c) and is based on the recommendations of Calonico et al. (2014b) for implementing a "fuzzy RD". Generally, this approach is used to assess the impact of a binary treatment where the probability of treatment rises at some threshold, but being above or below the threshold does not fully determine treatment status. In our context, exposure to PM_{10} increases significantly at the Huai River, but pollution exists south and north of the river, making our context naturally analogous to a "fuzzy RD". They recommend that the "fuzzy RD" coefficient be estimated as the ratio of two "sharp" discontinuities, using the optimal bandwidth for life expectancy for both life expectancy and PM_{10}.[10] They also provide a formula for the standard error of this estimate that is derived using the asymptotic properties of the ratio of the separately estimated discontinuities. See Calonico et al. (2014a) for further details.

IV. Empirical Results

A. Assessing the Validity of the Huai River RD Design

Table 1 reports the summary statistics for several of the key determinants of mortality rates and provides evidence on the validity of the RD design. Columns (1) and (2) report the means along with the standard deviation in cities south and north of the Huai River line. Column (3) reports the mean difference between the North and the South, along with the associated standard error. Column (4) also reports the difference (and standard error) but this time it is adjusted for a

[10] Note that this will imply that the estimates will not exactly match estimates generated optimally for both variables separately.

cubic polynomial in degrees north of the Huai River that is allowed to vary north and south of the river, so that it is a test for a discontinuous change at the Huai River line. Column (5) reports the discontinuous change at the Huai River line using local linear regression to estimate the size of the discontinuity, estimated with a triangular kernel and bandwidth selection method prescribed by Imbens and Kalyanaraman (2012). A direct test of the RD design's identifying assumption that unobservables change smoothly at the boundary is, of course, impossible but it would nevertheless be reassuring if observable determinants change smoothly at the boundary. (This is analogous to the test in randomized control trials that observable determinants of the outcome are independent of treatment status.)

Two key points emerge from this table. First, there are large differences in PM_{10} exposure among Southern and Northern Chinese residents. In contrast, concentrations of nitrogen oxides and sulfur dioxide are statistically equivalent on both sides of the river, after implementation of either the parametric or non-parametric RD approach as reported in columns (4) and (5). These pollutants generally do not travel as far as PM_{10}, so it is not surprising that the only discrete difference in air pollution is for PM_{10}.

Second, there is little systematic evidence of discrete differences in other determinants of life expectancy to the north of the Huai River. Column (3) reports that there are substantial differences in the determinants of mortality rates (e.g., temperature, health behaviors) when one compares all of the North to all of the South. However, these differences generally disappear in columns (4) and (5). Of special interest, there is no evidence of discrete differences in health behaviors, including the fraction of people who smoke regularly, drink heavily, and exercise insufficiently. It is also noteworthy that the differences are generally smaller with the non-parametric RD approach in column (5). We augment the analysis of potential confounders by

predicting life expectancy with *all* the potential covariates from Panels B and C and testing for a discontinuity to the north of the river. These variables collectively explain a substantial portion of the variation in life expectancy (R squared=.34). However, as reported in Table 2, no significant change in predicted life expectancy is found at the Huai River. Overall, we conclude that Table 1 supports the validity of the RD design based on distance from the Huai River.

B. Estimates of the Effect of the Huai River Policy on PM_{10}, Life Expectancy, and Cause-Specific Mortality Rates

We begin the analysis graphically with an assessment of the Huai River Policy's impact on pollution. Figure 2 plots PM_{10} at DSP locations against their degrees north of the Huai River line. The circles represent the average PM_{10} concentration across locations within 1 degree latitude distance bins from the Huai River; each circle's size is proportional to the population at the DSP locations within the relevant bin. The plotted line is generated by using a kernel-weighted local linear regression on either side of the river, which is similar to a non-parametric RD approach (Cleveland 1979). This is estimated with a triangular kernel and bandwidth chosen by the method prescribed by Imbens and Kalyanaraman (2012), with the data separated into 1 degree latitude bins.[11] The figure reveals a discontinuous change in ambient PM_{10} concentrations to the north of the river; it indicates that the Huai River Policy increased PM_{10} concentrations by about 42 $\mu g/m^3$. The plot in Figure 3 reveals a striking discrete decline in life expectancy at the border of roughly 3 years, which mirrors the increase in PM_{10}. Together, these figures reveal a sharp increase in PM_{10}

[11] Note that the magnitude of the discontinuity observed in the figure will almost exactly match the discontinuity estimated by the local linear regression (since the figure is generated with the same technique), but will not be exactly equal since the local linear regression estimate since this estimate is only valid pointwise at the discontinuity and the figure is plotted for a range of points to the south and north of the Huai River line. For example, while we estimate the discontinuity at the Huai as 41.6 $\mu g/m^3$, the figure is generated with a discontinuity of 40.3 $\mu g/m^3$, which is the resulting gap when plotting PM_{10} over the range of points above and below the Huai River with local linear regression.

and a sharp decline in life expectancy at precisely the location where the Huai River Policy went into effect.

A logical placebo test to assess the significance of these findings is to explore whether discontinuities are observed in other regions of China. Figure 4 reports on the estimated discontinuities in PM_{10} and life expectancy at 1 degree latitude intervals North and South of the Huai River across China, as well as at the actual Huai River (which is reported as the 0^0 displacement). The discontinuities are estimated from versions of equation (3a) and (3b) using a triangular kernel and the bandwidth selection approach of Imbens and Kalyanaraman (2012). The figure demonstrates that the only statistically significant discontinuous changes in PM_{10} and life expectancy occur at the actual Huai River. In all other instances, an estimated effect of zero is within the 95% confidence interval. These results provide further evidence that the effects in Figures 2 and 3 are due to the Huai River Policy, rather than an artifact of this application of the RD approach.

Table 2 statistically summarizes the graphical findings in Figures 2 and 3 by reporting estimates (and standard errors) associated with the North indicator variable from fitting equations for several variables of interest using a variety of RD approaches. Columns (1)-(4) apply the parametric RD approach from equations (2a) and (2b). The first two columns use the full sample, while the third and fourth columns restrict the sample to DSP locations within 5 degrees latitude of the Huai River.[12] The estimates in the odd columns are unadjusted, while those in the even columns are adjusted for the covariates in Panels B and C of Table 1. In both samples, we use the AIC goodness of fit criteria for life expectancy to choose the functional form for the polynomial in latitude. We focus on specifications where latitude is interacted fully with our North dummy,

[12] This sample restriction is a "rule of thumb" style approach to introducing the appeal of the non-parametric RD approach within the parametric framework.

so that latitude is allowed to affect outcomes differently north and south of the Huai River. The AIC recommends a cubic polynomial in degrees latitude away from the Huai River in the full sample and a linear term in the restricted sample.

Columns (5)-(7) report on the estimated discontinuity at the Huai River using local linear regression and bandwidth selected by the method proposed by Imbens and Kalyanaraman (2012). The columns report results when using a triangular, Epanechnikov, and uniform bandwidth, respectively. This non-parametric approach has the benefit that no functional form need be imposed on the data. Furthermore, it provides a method to place a greater weight on DSP locations near the Huai River.

Panel A presents evidence of a significant increase in PM_{10} and decrease in life expectancy at the Huai River. At the boundary, PM_{10} rises by 33/33 $\mu g/m^3$ and life expectancy declines by 3.3/3.5 years in the full and restricted samples, respectively, without adjustment for covariates. The decline in life expectancy is driven by a statistically significant increase in cardiorespiratory mortality rates of 42/39 percent at the boundary (Panel B).[13] In contrast, there is little systematic evidence of a meaningful discontinuity in non-cardiorespiratory mortality rates north of the river.

The estimates from the non-parametric approach tend to be of a modestly greater magnitude and somewhat less precise. The estimated increase in PM_{10} north of the river ranges from 42 to 43 $\mu g/m^3$, while the decline in life expectancy ranges from 3.1 to 3.3 years. Again, the estimates suggest that the decline in life expectancy is due to elevated cardiorespiratory mortality rates.

Each of the approaches to implementing the RD design has strengths and weaknesses (Imbens and Lemieux 2008). As such, we perform a rich set of alternative specifications across

[13] Table S1 divides the cardiorespiratory and non-cardiorespiratory cause of death categories into sub-categories and reports on the effects of the North indicator on mortality rates in these subcategories.

both parametric and non-parametric estimation methods to thoroughly explore the sensitivity of our findings to our choice of econometric method. The results are described below and available in greater detail in the SOM. The important finding across all of these approaches is the robustness of the estimated increase in PM_{10} and decline in life expectancy north of the Huai River to this wide variety of approaches to implementing the RD design. Nevertheless, for ease of exposition and to reflect the discomfort with using higher order polynomials in the parametric approach (Imbens and Gelman 2014), we emphasize the column (5) estimates going forward.

C. Estimates of the Effect of PM_{10} on Life Expectancy

Table 3 provides an opportunity reports on the estimated effect of 10 $\mu g/m^3$ of PM_{10} on life expectancy and cardiorespiratory mortality rates from alternative estimation approaches. Columns (1) – (4) report on the conventional OLS approach detailed in equation (1); the full sample is used in the first two columns, while the sample is restricted to the 79 DSP sites within 5 degrees latitude of the Huai River. Columns (5) – (11) report on a variety of applications of the RD IV or 2SLS approach, with the details noted in the column headings and rows at the bottom of the table.

The traditional OLS estimates suggest substantially smaller estimates of the health effects of PM_{10}. Specifically, the OLS estimates suggest that an additional 10 $\mu g/m^3$ of sustained exposure to PM_{10} is associated with a decline in life expectancy of 0.16 to 0.34 years, with the higher end of the range coming from the restricted sample. These estimates would all be judged to be statistically significant by conventional criteria, and given the novel data set we are using, seem to be a contribution in their own right.

The instrumental variables estimates based on the Huai River RD design are substantially

larger. The parametric approach outlined in equation (2c) produces estimates of the effect of an additional 10 $\mu g/m^3$ of PM_{10} on life expectancy that range from -0.85 to -1.07 years in columns (5) through (8); three of the four estimates are statistically significant. The estimated effect from the non-parametric approach ranges from -0.74 to -0.77 years in columns (9) through (11) and would all be judged to be statistically significant by conventional criteria.[14] We emphasize the -0.77 estimate from the triangle kernel to match the choice of emphasized specification in Table 2. The larger magnitude of these IV estimates suggests that some combination of omitted variables (e.g., more polluted areas are richer) and measurement error reduce the magnitude of the OLS estimates, relative to the true effect of PM_{10} on life expectancy.

D. Supplementary Materials

In the appendix accompanying this paper, we explore heterogeneity in the results across different populations, as well as the results' sensitivity to a rich set of robustness checks. Table S4 reveals that the mortality and life expectancy findings hold for men and women.[15] Table S5 and Figure S3 demonstrate that the impacts on cardiorespiratory mortality rates are generally evident over the entire course of the life cycle, not just for the young and old. Tables S6, S7, and S8 explore the sensitivity of the results to different parametric and non-parametric applications of the RD design. Except for parametric RD specifications that are unappealing based on the AIC information criterion (i.e., a goodness of fit test), these estimates are all qualitatively similar to the ones presented in Tables 2 and 3. Further, the reported AIC information criteria test statistics make clear

[14] Recall, we follow the method of Calonico et al. (2014b), which uses the bandwidth determined by the optimal bandwidth for estimating the discontinuity in the outcome variable, which in this case is life expectancy. Thus, these estimates are not simple ratios of the estimated effects of the coefficients associated with the North indicator from the equations for life expectancy and PM_{10} in Table 2.

[15] Table S3 fails to find evidence of a discontinuous change in health-related behaviors by gender to the north of the Huai River.

that the cubic and linear polynomials in latitude (interacted with the North indicator) fit the data best for the life expectancy outcome in the full and restricted samples, respectively; for this reason Tables 2 and 3 rely on these functional form choices.

Tables S9 – S13 explore a series of other issues. Table S9 documents that the PM_{10} results are broadly unchanged by using alternative methods to assign PM_{10} concentrations to DSP locations. Table S10 verifies that using a more restricted sample of DSP locations, in which the sample is composed of only DSP locations very near monitoring stations, only changes the results slightly. Table S11 shows that the estimated health effects of the Huai River Policy are qualitatively unchanged by expanding the sample to the full set of 161 DSP locations from the 154 locations with PM_{10} data. Table S12 demonstrates that restricting the analysis to the mortality and pollution data collected between 2009-2012, after the Chinese monitoring system significantly increased its number of monitoring stations, does not significantly affect the results. This is especially informative because it underscores that the health effects problem is a current one, not just a historical one. In Table S13, we verify that excluding a small number of PM_{10} measurements that were imputed from TSP only slightly affects the results.

A natural concern related to the research design is that the government used the Huai River as the demarcation line for changes in other government policies related to public health. If this were the case, these other policies could confound the estimates of PM_{10} on health, however these concerns seem unfounded for several reasons. First, the Huai River follows the January zero degree average temperature line (Celsius), and this was in fact the basis of its choice as a method to divide the country for free heating. Second, local policies generally hew to administrative boundaries associated with cities and provinces; indeed, the Huai River cuts through several provinces. Third, we compiled some additional variables on city-level policies and other variable that are plausibly

related to health from the China City Statistical Yearbooks; Table S14 demonstrates that the null hypothesis that these variables are equal on both sides of the Huai River generally cannot be rejected.

E. Comparison of Results to Chen et al. (2013)

Chen et al. (2013) also exploit the RD design based on the Huai River Policy but rely on mortality data from the 1990s and measures of total suspended particulates (TSP), instead of this paper's use of 2004-2012 mortality data and PM_{10} data. They find that life expectancies are about 5.5 years lower and TSPs are roughly 184 $\mu g/m^3$ higher, just to the north of the Huai River. More generally, the analysis suggests that sustained exposure to 100 $\mu g/m^3$ of TSP is associated with a reduction in life expectancy at birth of about 3.0 years.

A back-of-the-envelope calculation suggests that this paper's results are broadly consistent with Chen et al.'s (2013) estimate of the effect of airborne particulate matter on life expectancy. Using Chinese data from monitoring stations that track TSP and PM_{10}, we found that PM_{10} accounts for roughly 37.5% of TSP. Thus, the Chen et al (2013) estimates translate into a prediction that 10 units of PM_{10} reduce life expectancies by 0.8 years (3/.375) that is well within sampling error of this paper's estimate of 0.75 years.[16] Overall, it is striking that the application of the Huai River RD design produces the same basic relationship between airborne particulate matter and life expectancy in two different decades.

[16] Note also that if TSPs are composed of roughly 37.5% PM_{10}, this suggests that PM_{10} concentrations have declined by roughly 40% in China since the 1990s. This decline has been approximately equal on the north and south sides of the river in percentage terms. Owing to the higher initial levels in the North, the absolute decline has been larger in the North, potentially explaining why the life expectancy penalty for living north of the river has declined from 5.5 years to 3.4 years.

V. Conclusion

The analysis suggests that the Huai River Policy, which had the laudable goal of providing indoor heat, had disastrous consequences for human health. Specifically, it led to PM_{10} concentrations that are 41.6 $\mu g/m^3$ (95% CI: 11.6, 71.6) or 46% higher in the North and reductions in life expectancies of 3.1 years lower (95% CI: 1.3, 4.9) in the North due to elevated rates of cardiorespiratory mortality. More broadly, the results suggest that sustained exposure to an additional 10 $\mu g/m^3$ is associated with a 0.75 year (95% CI: 0.22, 1.28) decline in life expectancy.

The implications of these results for public health are potentially immense. Over 85% of China's population resides in locations where current PM_{10} levels exceed the country's *own* Class II air quality standard of 70 $\mu g/m^3$ (An et al. 2013). This paper's estimates suggest that bringing these parts of the country into compliance with this air quality standard would save roughly 1.3 billion life years in China. Reducing PM_{10} concentrations such that the entire country met the Class I standard of 40 $\mu g/m^3$ would lead to an even larger gain of 4.3 billion life years.

The risks to health from PM_{10} exposure are not confined to China. Some notable examples from the 2014 WHO Ambient Air Pollution Database include India's (population 1.3 billion) average PM_{10} concentration of 134 $\mu g/m^3$, Pakistan's (population 181 million) 282 $\mu g/m^3$, Bangladesh's (population 157 million) 163 $\mu g/m^3$, Iran's (population 77 million) 127 $\mu g/m^3$, and Mexico's (population 124 million) 79 $\mu g/m^3$. In total, more than 4.5 billion people live in countries with average PM_{10} concentrations that are at least twice the concentration that the WHO considers safe. This paper's results suggest that in many parts of the world there is no greater current environmental risk to health than airborne particulate matter

References

An Xingqin, Hou Qing, Li Nan, Zhai Shixian. 2013. "Assessment of human exposure level to PM_{10} in China", *Atmospheric Environment* 70:376-386.

Calonico, S., Cattaneo, M. D. and Titiunik, R. 2014a. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82: 2295–2326.

Calonico, S., Cattaneo, M. D., and R. Titiunik. 2014b. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *Stata Journal* 14(4): 909-946.

Card, David, Carlos Dobkin, and Nicole Maestas. 2009. "Does Medicare Save Lives?" *Quarterly Journal of Economics* 124(2): 597-636.

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

Chen, Renjie, Haidong Kan; Bingheng Chen; Wei Huang; Zhipeng Bai; Guixiang Song and Guowei Pan. 2012. "Association of Particulate Air Pollution With Daily Mortality." *American Journal of Epidemiology* 175(11):1173-1181.

Chen, Yuyu, Ebenstein, Avraham, Greenstone, Michael & Li, Hongbin. 2013. "Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy". *Proceedings of the National Academy of Sciences* 110(6):12936-41.

Cleveland, W. S. 1979. "Robust locally weighted regression and smoothing scatterplots." *Journal of the American Statistical Association* 74:829-836.

Cook, T.D. and Campbell, D.T. 1979. *Quasi-Experimentation: Design and Analysis for Field Settings*. Rand McNally, Chicago, Illinois.

Dockery DW, Pope, CAIII, Xu, X., Spengler JD, et al. 1993. "An association between air pollution and mortality in six U.S. cities". *New England Journal of Medicine* 329:1753 1759.

Dominici, Francesca, Greenstone, Michael and Cass Sunstein. 2014. "Particulate Matter Matters". *Science* 344(6181):257–259.

Imbens, Guido and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A guide to Practice". *Journal of Econometrics* 142(2):615-635.

Lee, David S. and Alexandre Mas. 2011. "Long-run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961-1999." *Quarterly Journal of Economics* 127(1):333-378.

Lee, David S. and Thomas Lemieux. 2010. “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature* 48(2): 281-355.

Pope, C. Arden. 1989. “Respiratory Disease Associated with Community Air Pollution and a Steel Mill, Utah Valley.” *American Journal of Public Health* 79:623-28.

Pope, C. Arden. 2010. “Epidemiology of Fine Particulate Air Pollution and Human Health: Biologic Mechanisms and Who’s at Risk?” *Environmental Health Perspectives* 108(4):713–723.

Pope, C. Arden, Richard T. Burnett, Michael J. Thun et al. 2002. “Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution.” *Journal of the American Medical Association* 287(9):1132-1141.

Pope, C. Arden, Ezzati, M., & Dockery, D. W. 2009. "Fine-Particulate Air Pollution and Life Expectancy in the United States". *New England Journal of Medicine* 360(4), 376-386.

Rich, D. Q., Kipen, H. M., Huang, W., Wang, G., Wang, Y., Zhu, P., & Ohman-Strickland, P. (2012). "Association Between Changes in Air Pollution Levels During the Beijing Olympics and Biomarkers of Inflammation and Thrombosis in Healthy Young Adults". *Journal of the American Medical Association* 307(19), 2068–2078.

Wellenius, Gregory A, Joel Schwartz, Murray A. Mittleman. 2005. "Air Pollution and Hospital Admissions for Ischemic and Hemorrhagic Stroke Among Medicare Beneficiaries." *Stroke* 36:2549-2553.

World Health Organization (2004). Global health observatory data repository. Technical report.

Wu, Shaowei; Furong Deng; Jie Niu; Qinsheng Huang; Youcheng Liu and Xinbiao Guo. 2010. "Association of Heart Rate Variability in Taxi Drivers with Marked Changes in Particulate Air Pollution in Beijing in 2008." *Environmental Health Perspectives* 118(1), 87-91.

Table 1

Summary Statistics, Means and (Standard Deviations)

	North	South	Difference in Means	Adjusted Difference (polynomial)	Adjusted Difference (local linear)
	(1)	(2)	(3)	(4)	(5)
Panel A: Air Pollution Exposure at China's Disease Surveillance Points					
Particulate Matter (PM_{10})	116.0	90.9	25.1***	32.9***	41.6***
	(30.7)	(27.3)	(5.2)	(11.7)	(15.3)
Sulphur Dioxide (SO_2)	56.3	44.7	11.6***	1.6	-5.1
	(24.2)	(23.6)	(4.0)	(8.5)	(9.3)
Nitrogen Oxide (NO_X)	35.7	31.3	4.3*	-7.7	-9.0*
	(11.2)	(12.9)	(2.3)	(5.8)	(4.9)
Panel B: Climate at the Disease Surveillance Points					
Heating Degrees	5.9	2.7	3.2***	1.0**	1.7
(1,000 degrees)	(1.4)	(1.4)	(0.3)	(0.5)	(1.2)
Cooling Degrees	1.2	2.2	-1.0***	-0.4*	-0.6*
(1,000 degrees)	(0.4)	(0.9)	(0.1)	(0.2)	(0.3)
Annual Rainfall	26.3	54.5	-28.3***	-2.3	-11.4**
(inches)	(9.0)	(11.6)	(1.8)	(3.4)	(4.5)
Panel C: Demographic and Health Behavior Characteristics of China's Disease Surveillance Points					
Per capita Income	11.2	12.7	-1.5	-8.1**	-5.7
(1,000 *yuan,* 2010)	(5.3)	(5.8)	(1.0)	(4.0)	(4.0)
Years of Education	8.4	8.2	0.1	-0.1	-0.3
	(0.6)	(0.8)	(0.1)	(0.5)	(0.5)
Share Minority (%)	7.6	8.8	-1.2	8.4	15.4*
	(16.6)	(20.8)	(2.7)	(5.4)	(8.5)
Coastal City Dummy	0.23	0.26	-0.03	-0.56**	-0.28
	(0.43)	(0.44)	(0.08)	(0.24)	(0.18)
Smoke Regulary (%)	34.7	35.7	-0.9	-0.6	-2.9
	(6.4)	(6.6)	(1.2)	(4.1)	(4.1)
Heavy Drinker (%)	8.3	7.5	0.9	-2.8	-5.0
	(3.7)	(2.9)	(0.6)	(3.1)	(3.8)
Excessive Red Meat (%)	18.4	34.5	-16.0***	0.8	-5.9
	(12.5)	(17.2)	(2.6)	(5.4)	(3.9)
Insufficient Exercise (%)	24.4	20.0	4.4**	-1.8	-1.3
	(13.6)	(9.0)	(2.0)	(7.9)	(6.5)
Panel D: Predicted Life Expectancy at China's Disease Surveillance Points					
Predicted Life	76.1	76.3	-0.2	-1.4	-1.2
Expectancy (years)	(1.5)	(1.8)	(0.3)	(1.1)	(1.0)

Notes : The sample is restricted to DSP locations (n=154) within 150 kilometers of an air quality monitoring station. Pollution measures are calculated as the city's average reading in the years prior to the DSP period. Degree days are the absolute value of the deviation of each day's average temperature from 65^0 F, averaged over the years prior to the DSP period. Predicted life expectancy is calculated by OLS using all the demographic and meteorological covariates shown in Panel B and C. The results in column (4) are adjusted for a cubic in degrees of latitude north of the Huai River boundary, which is allowed to vary north and south of the boundary. In column (5), we report the estimated discontinuity at the Huai River using local linear regression and bandwidth selected by the method proposed by Imbens and Kalyanaraman (2012) using a triangular kernel. All results are weighted by the population at the DSP location. Heteroskedastic-consistent standard errors are reported in parentheses in columns (1)-(4), and conventional local linear regression discontinuity standard errors are reported in column (5). * significant at 10% ** significant at 5%. *** significant at 1%. Source: China Disease Surveillance Points (2004-2012), China Environmental Yearbooks (1981-2012), World Meteorological Association (1980-2012).

Table 2

Regression Discontinuity Estimates of the Impact of the Huai River Policy

	Polynomials				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Pollution and Life Expectancy							
Particulate Matter (PM_{10})	32.9***	27.6***	32.9**	21.7**	41.6***	42.9***	43.3**
	(11.7)	(9.2)	(12.7)	(9.0)	(15.3)	(16.3)	(18.2)
Life Expectancy	-3.3***	-2.3**	-3.5***	-2.1*	-3.1***	-3.2***	-3.3***
at Birth (years)	(1.0)	(1.0)	(0.8)	(1.2)	(0.9)	(1.0)	(1.2)
Panel B: Cause-specific Mortality							
Cardiorespiratory	0.42***	0.30**	0.39***	0.22	0.37***	0.39***	0.40***
(per 100,000, log)	(0.12)	(0.14)	(0.11)	(0.13)	(0.11)	(0.11)	(0.14)
Non-Cardiorespiratory	0.07	0.06	0.12*	0.07	0.13	0.13	0.16
(per 100,000, log)	(0.09)	(0.09)	(0.07)	(0.08)	(0.08)	(0.08)	(0.10)
Observations	154	154	79	79			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Demographic Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes : Each cell in the table represents a separate regression. In columns (1)-(4), we report OLS estimates of the coefficient on a "North of the Huai River" dummy after controlling for a polynomial in distance from the Huai river interacted with a North dummy. The results in columns 2 and 4 include the demographic covariates reported in Table 1. In columns (5)-(7), we report the estimated discontinuity at the Huai River using local linear regression and bandwidth selected by the method proposed by Imbens and Kalyanaraman (2012) for different kernel weighting methods. Heteroskedastic-consistent standard errors are reported below the coefficients in columns (1)-(4) and conventional local linear regression discontinuity standard errors are reported in columns (5)-(7). * significant at 10% ** significant at 5%. *** significant at 1%.

Table 3

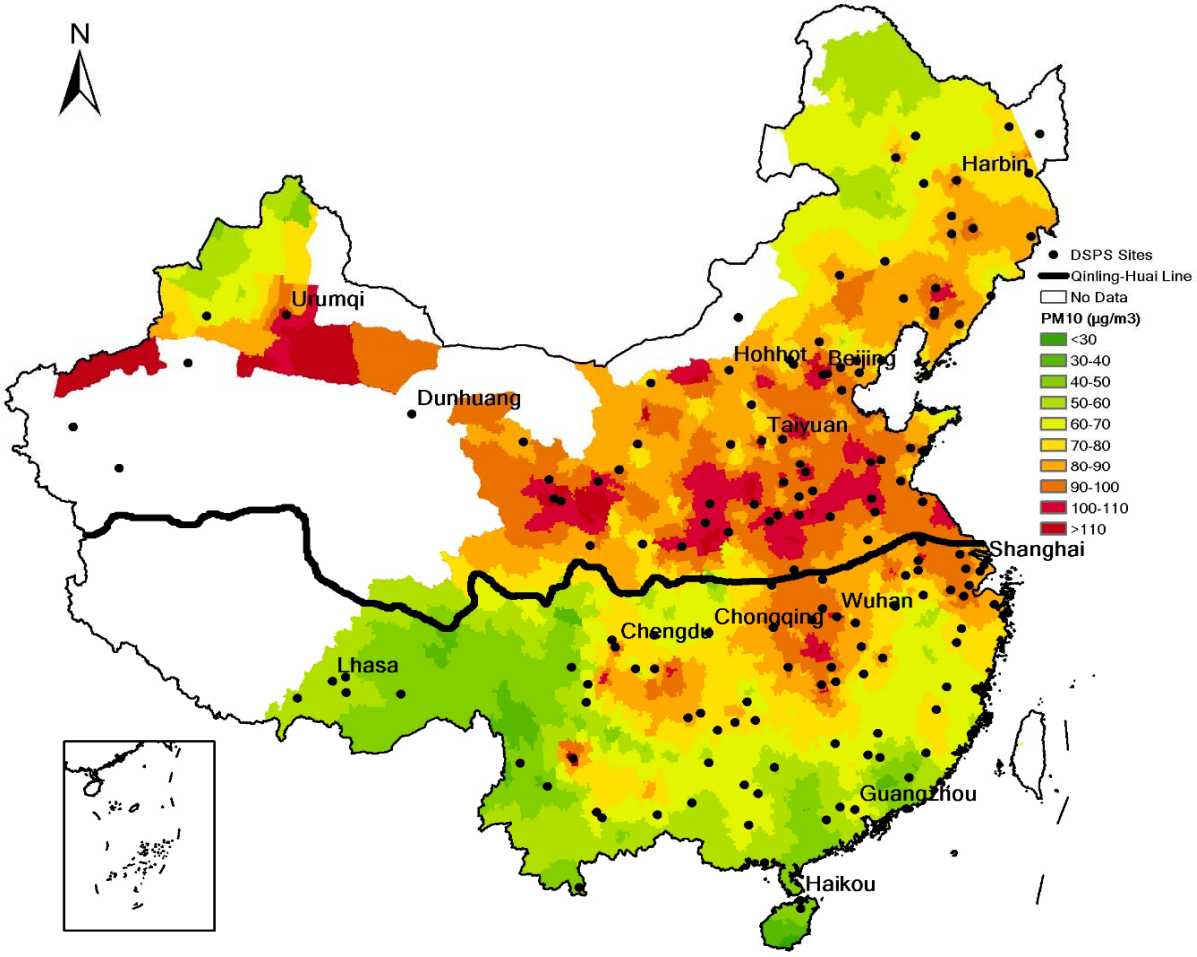
Comparing OLS and Regression Discontinuity Estimates of the Impact of 10 $\mu g/m^3$ of PM_{10} on Health Outcomes

	OLS		OLS within 5 Degrees		2SLS Polynomial RD		2SLS within 5 Degrees		Local Linear Regressions ("Fuzzy" RD)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Life Expectancy at Birth (years)	-0.16*	-0.17**	-0.34***	-0.24**	-1.02*	-0.85	-1.07**	-0.98*	-0.75***	-0.74***	-0.77**
	(0.08)	(0.08)	(0.07)	(0.10)	(0.55)	(0.54)	(0.46)	(0.55)	(0.27)	(0.26)	(0.30)
Cardiorespiratory (per 100,000, log)	0.02***	0.02**	0.04***	0.02**	0.13**	0.11*	0.12**	0.10*	0.10**	0.10**	0.10**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.06)	(0.06)	(0.05)	(0.06)	(0.04)	(0.05)	(0.05)
Observations	154	154	79	79	154	154	79	79			
Controls	No	Yes	No	Yes	No	Yes	No	Yes			
Kernel									Triangle	Epanech.	Uniform

Notes : Each cell in the table represents a separate regression. All estimates report the impact of In columns (1) and (4), we report OLS estimates of the association between PM_{10} and the listed outcome. In columns (5)-(8), we report the 2SLS IV estimates using "North to Huai River" as the instrunmental variable. The results in columns (5)-(8) also include a cubic polynomial in degrees latitude from the Huai River interacted with a North dummy variable. In columns (9)-(11), we estimate the impact of PM_{10} on the listed outcomes treating distance from the Huai River as the forcing variable and PM_{10} as the treatment variable, with the Huai River representing a "fuzzy" discontinuity in the level of pollution exposure. Note that this specification generates the point estimate as the Wald ratio of the discontinuity in PM_{10} to the discontinuity in life expectancy, using the optimal bandwidth chosen for life expectancy to estimate both (Calonico et al. 2014). Results are all presented in terms of the health impact of an additional 10 $\mu g/m^3$ of long-term PM_{10} exposure. Heteroskedastic-consistent standard errors are reported below the coefficients in columns (1)-(8) and conventional local linear regression discontinuity standard errors are reported in columns (9)-(11). * significant at 10% ** significant at 5%. *** significant at 1%.

Figure 1

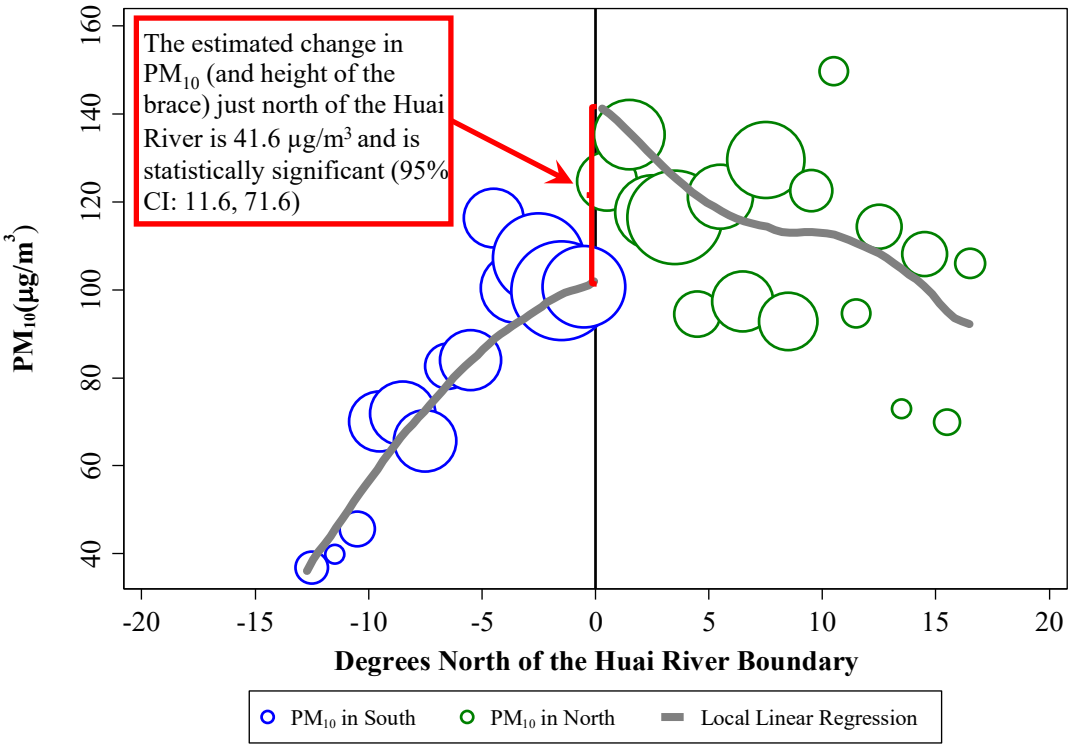
Pollution in China and the Huai River/Qinling Mountain Range



Notes : The cities shown are the locations of the Disease Surveillance Points. Cities north of the solid line were covered by the home heating policy. The figure coloring is generated by interpolating PM_{10} levels at the 12 nearest pollution monitoring stations to create a high resolution grid of pollution throughout China (.1 degree latitude cell width). Areas are left in white which are not within acceptable range of a station.

Figure 2

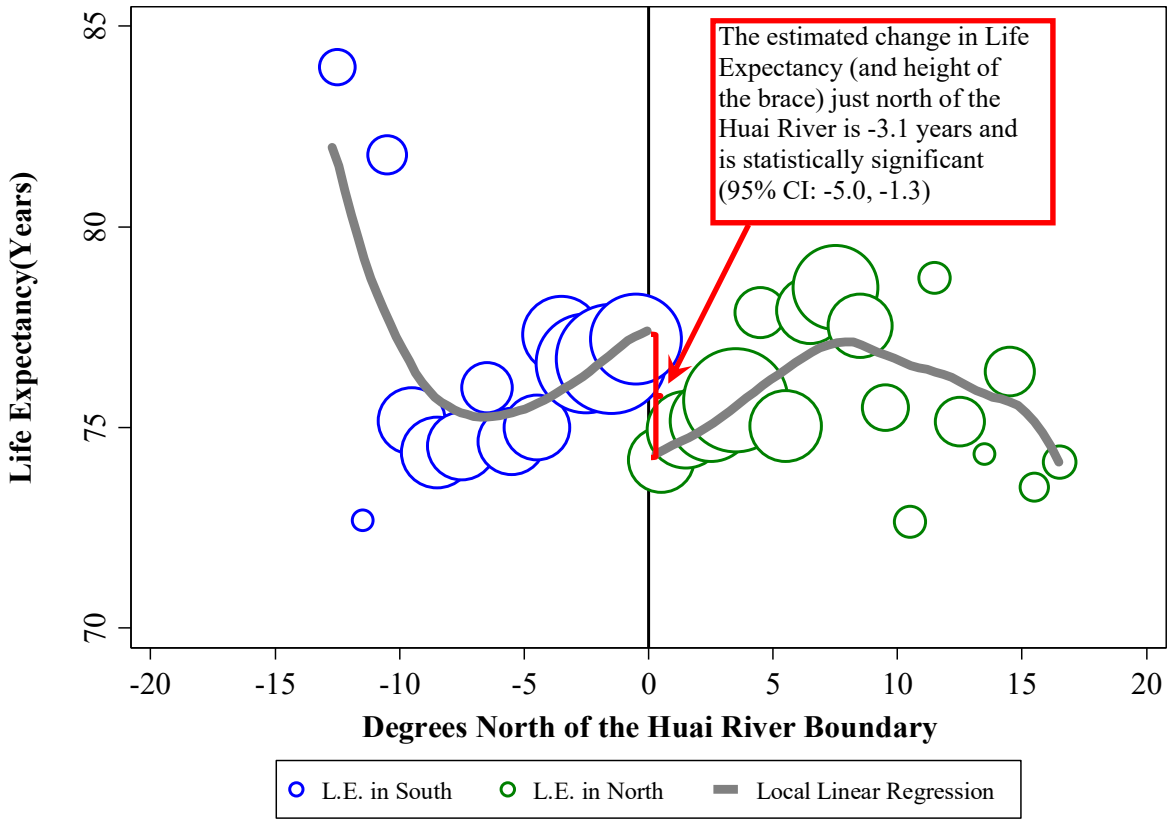
Particulate Matter Levels (PM_{10}) South and North of the Huai River Boundary



Notes : Each observation (circle) is generated by averaging PM_{10} across the Disease Surveillance Point locations within a 1 degree latitude range, weighted by the population at each location. The size of the circle is in proportion to the total population at DSP locations within the 1 degree latitude range. The plotted line reports a local linear regression plot estimated separately on on each side of the Huai River.

Figure 3

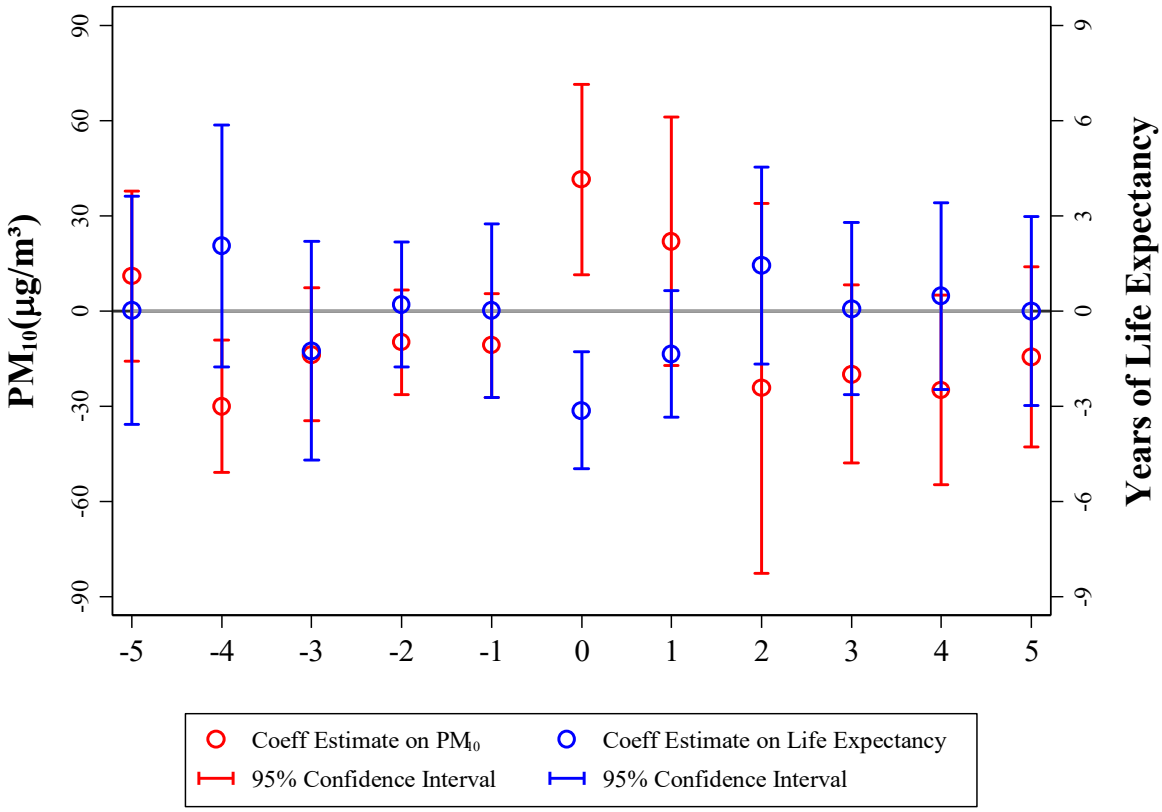
Life Expectancy South and North of the Huai River Boundary



Notes : Each observation (circle) is generated by averaging life expectancy across the Disease Surveillance Point locations within a 1 degree latitude range, weighted by the population at each location. The size of the circle is in proportion to the total population at DSP locations within the 1 degree latitude range. The plotted line reports a local linear regression plot estimated separately on on each side of the Huai River.

Figure 4

Placebo Testing: Estimated Discontinuity in Pollution and Life Expectancy at Displaced Huai River Boundaries



Notes : The figure reports the change in PM_{10} and life expectancy at the Huai River and at discontinuities estimated at 1 degree latitude displacements from the actual Huai River. Each discontinuity is estimated using a triangular kernel and with bandwidth chosen according to the method proposed by Imbens and Kaylanaraman (2012).

SUPPORTING INFORMATION

The Impact of Sustained Exposure to Particulate Matter on Life Expectancy: New Evidence from China's Huai River Policy

PART 1: Description of Data and Additional Summary Statistics

I. Mortality Rate and Life Expectancy Data

Our sample of mortality in China is taken from the Disease Surveillance Points (DSP) system, which forms a nationally-representative sample of mortality for 2004-2012. This sample represents a significant improvement in data quality relative to data collected by the DSP during the 1990s. The earlier DSP sample, collected between 1991 and 2000, provided coverage for roughly 10 million residents across 145 locations and have been relied on by earlier scholarship examining health in China (e.g. Chen et al. (2013)).[1] However, as a result of the SARS outbreak in 2003, the Chinese CDC obtained funds to dramatically expand the DSP sample to cover *all* inhabitants of 161 counties throughout China. This represented a roughly 8-fold increase in the sample size and provides a unique opportunity to examine mortality in a developing country with using reliable data. In this study, we examine a data extract of the DSP that was made available to the research team for this project; a full listing of the causes of death in our data are reported in Table S1. For each category, we report the cause of death description, the Chinese CDC code, and the analogous International Classification of Disease Revision 9 (ICD-9) and ICD-10 codes. It is worth noting that our sample means match the official death rates recorded by the Chinese Disease Surveillance Points system.

Table S2 reports our basic results of the change in mortality observed at the Huai River separately for different illness categories. Our primary division is between cardiorespiratory illnesses and non-cardiorespiratory illnesses, which we identify using the Chinese coding scheme. Cardiorespiratory illnesses include respiratory infections (U038), lung cancer (U067), heart diseases (U104-U107, U109-U110) and stroke (U108). Non-cardiorespiratory illnesses include cancers other than lung (U060-U065, U078), injuries (U148), and all other causes.[2] The cardiorespiratory causes are presumed to be affected by pollution exposure (and expected to rise

[1] The DSP was collected between 2001 and 2003 but the data were of limited quality and never released publicly.
[2] These codes are provided by the Chinese CDC. The mapping from these categories to ICD-9 and ICD-10 is reported in Table S1.

at the Huai River), whereas the others act as placebo outcomes. This prediction is borne out by the data: a statistically significant increase in cardiorespiratory mortality rates is found at the Huai River using our polynomial approach and using local linear regressions in all but one specification. This increase is driven largely by increases in mortality from heart disease and stroke, though lung cancers and respiratory illnesses are also higher just north of the Huai River. In contrast, the change in mortality for non-cardiorespiratory mortality at the Huai River is much more modest; for example, cancers other than lung-cancer actually decrease at the Huai River line, though the decrease is not statistically significant.

Since the paper's goal is to determine the relationship between mortality rates and long-term exposure to ambient air pollution, it is critical that the DSP accurately reports mortality among all permanent residents. There are two possible threats to the accurate calculation of such mortality rates. First, although China's *hukou* (household registration) system restricted migration, it did not prevent all migration, and indeed migration became more common in China by the end of our sample period. Second, sick rural residents are referred to urban hospitals with an unknown frequency; if their deaths are registered where they die instead of where they reside, our estimates could be biased. There are two primary reasons that we believe that mismeasurement of location-specific mortality rates are unlikely to explain the paper's results. First, the DSP's approach to assigning mortality rates to locations mitigates these potential problems because, like the *hukou* system, it assigns a death to an individuals' *hukou* or place of birth. At the DSP locations, the covered population is tracked from a registry designed at the outset of the program, and deaths are usually recorded by personnel at local hospitals. In the circumstance that a death occurs away from home, the family is instructed to report on the death wherever it occurs, and a site official is expected to follow up on and accurately record each death at the site of original registration. Further, the *hukou* system requires that all death certificates report the *hukou* or place of birth, and this information is checked by a local government official who provides an official stamp of verification to the death certificate. Second, the regression discontinuity design is likely to provide a solution to any residual confounding. Specifically, in the context of this design these two behaviors (migration and excess urban deaths) will not be a source of confounding unless their frequency changes discretely at the Huai River boundary. We are unaware of a reason that this would be the case.

Our analysis is conducted with cross-sectional data on mortality rates and their

determinants. We converted the panel DSP data into a cross-section by taking averages of the cause-specific mortality rates, age-specific mortality rates, and life expectancy, from all annual observations for 2004-2012 for each of the 161 DSP locations.[3]

II. Air Pollution and Weather Data

The creation of the air pollution and weather exposure data sets involved several steps. Our pollution data are formed by combining several sources of Chinese air pollution, and is to our knowledge the most comprehensive data ever assembled for the period 1981-2012. These data were compiled through hand entry from Chinese language publications; most readings were taken from China's *Environmental Yearbook*s 1990-2012 and China's *Environmental Quality Annual Reports* 2001-2012, and verified against each other. In more recent years, the Chinese EPA provided us electronic copies of the data sources, which we used to verify our readings taken from the hard-copy publications.

Our data contain measures of several pollutants. The main pollutant we focus on is particulate matter (PM_{10}), since this is considered the most harmful form of air pollution regularly measured by the Chinese monitoring system. The Chinese monitoring system began tracking PM_{10} in 2003 and our analysis relies on the readings taken between 2003 and 2012.

SO_2 concentrations have been reported consistently during our sample period (1990-2012). Concentrations of NO_X were reported for selected cities before 2001 but since 2001, all reporting is in terms of NO_2.

A. Creating a Panel Data Set of City-Level Pollution

The air pollution data were collected from several sources. Our compilation of readings for China between 1981-2012 are, to our knowledge, the most comprehensive collection ever assembled. These data were compiled in three steps.

1. A collection of electronic files were combined to form the main source of our data. Air pollution readings for roughly 90 cities from 1981-1995 were taken from the World Bank online archive of air quality readings, and supplemented by electronic resources provided by China's

[3] The DSP was expanded to 158 locations in 2004 and to 3 additional locations in 2006. Our analysis is conducted on the 161 locations.

Environmental Protection Agency as well as electronic copies of China's *Environmental Yearbooks*.

2. A print collection of China's *Environmental Yearbooks* and *Environmental Annual Quality Reports* at several universities were used to produce hand-entered data by two different researchers on our team, with nearly perfect data agreement.

3. Steps 1 and 2 created a data set of validated measurement of five distinct pollutants: TSP, PM_{10}, SO_2, NO_X, and NO_2. We have made the full collection of these data available online for the period 1981-2012 for scholarship. For our pollution measure of primary interest in this study, PM_{10}, we were able to obtain 1,971 validated measurements from 2003-2012. This is composed of 106 observations for 2003, 113 observations for 2004-2008, 318 for 2009, 333 for 2010, 324 for 2011, and 325 for 2012.

We supplement our direct PM_{10} readings with PM_{10} readings imputed from available TSP readings for 2003-2012, where we assume PM_{10} is a certain percentage of measured TSP. This is advantageous for our study by providing us wider coverage in early years of our sample before all stations began monitoring PM_{10}. The imputation was performed in either one of two ways: (1) in instances where a monitoring station tracked both pollutants for at least one year (which is 44 stations), we impute missing PM_{10} from TSP assuming it represented the same proportion as in the years in which both measures were available for the station; (2) in instances where a monitoring station never tracked both pollutants, we use the PM_{10}/TSP ratio of the nearest station with both PM_{10} and TSP readings. This imputation increases our sample of PM_{10} by 38 readings in 2003, 63 readings in 2004, 2 readings in 2009, and 1 reading in 2011.[4]

B. Creating a Panel Data Set of DSP location-Level Pollution Data from the City-Level Pollution Panel

The next step was to use the city-level panel data to create a DSP location-level panel of pollution data. This was done in three steps:

1. We first calculated the distance between each of these monitoring stations and our mortality sample taken from China's Disease Surveillance Points (DSP). The distance between

[4] Results where we exclude these imputed PM_{10} readings are presented in Part 3 of this appendix and are available in Table S10.

each of these 325 stations and the 161 DSP sites or locations yielded a full matrix of 161 X 325 calculated distances. The location of the stations and DSP sites are reported in Figure S1.[5]

2. Our measure of air pollution for a DSP location in a year was calculated as follows. If a DSP location was within 50 kilometers of a valid station reading, the nearest station's reading was used. If a DSP location was not within 150 kilometers of any of the stations, the DSP location was excluded from the sample. This resulted in the exclusion of 7 DSP locations. If a DSP location was within 150 kilometers of a station but not within 50 kilometers, the pollution was calculated as the weighted average of air pollution at each station with a valid reading (in that year) within 150 kilometers, with the weights determined by the inverse of the distance between the two points.[6] If a station had no valid PM_{10} reading for a particular year, it was assigned a zero weight for that year and did not enter into the calculation. These readings are then supplemented by interpolation, where readings are imputed in missing years with valid readings (when available).

3. At each DSP location, the air pollution exposure was measured as the average PM_{10} reading in all previous years. For example, the reading for a DSP location in 2007 is the average of PM_{10} readings from 2003-2006 at the location. This creates a panel data set of year by DSP location mortality readings with validated air pollution data averaged across all previous years.

C. Creating a Cross-Sectional Data Set of DSP-Level Pollution Data from the Panel DSP Data File.

The panel data set was then averaged across all observations for each DSP location to create a single observation for each of the 154 DSP within 150 kilometers of a reading that represent the main sample for our analysis.

A similar method was used to calculate weather exposure at each DSP location. Daily temperature readings for 1981-2012 were obtained from the World Meteorological Association (3). Our analysis is limited to the 339 weather stations with nearly-complete weather data for 1981-2012.[7] At each of these stations, we calculated the total heating and cooling degree days for

[5] The centroid of the county containing the DSP location and the centroid of the city containing the monitoring station were used to calculate an exact distance between the two.
[6] The results are robust to different choices for the functional form. This is discussed further in Part 3.
[7] Only stations with at least 360 days of data for every year in the period were included in the sample of weather stations. If a station had fewer than 360 readings in a particular year, the station was assigned a zero weight when estimating the weather for the DSP location for that year.

each year. The total heating and cooling degree exposure for a DSP location in a given year was calculated as the weighted average of the degree days at the 339 stations, with the weights determined by the distance between the DSP location and the station. The weather exposure for a DSP location in a year is the average of heating and cooling degrees in all previous years during the sample period assigned in a manner similar to our method for assigning air pollution exposure.

III. Health-Relevant Behavioral Data

The Chinese CDC closely tracks a range of health-relevant behaviors among the coverage population of each DSP location. The data used here were collected in 2010 across residents of all 161 counties that were covered by the DSP. Survey participants were asked to self-report whether they (a) smoke (b) drink heavily (c) eat excessive amounts of red meat or (d) exercise an insufficient amount. The health behavior measures are useful for two reasons. First, they enable us to more precisely measure the relationship between air pollution and health. Second, they enable us to directly test our empirical strategy by providing information on whether the residents south and north of the Huai River are similar along health-relevant dimensions other than pollution exposure. As reported in Table S3, the results indicate that while there are some observable differences in these behaviors on either side of the Huai River, they do not seem to change discontinuously at the river. In particular, after controlling for a polynomial in latitude or using the local linear regression techniques, no observable change in these behaviors is found near the Huai River.

IV. Other Covariates

As a complement to the DSP, we also use the China 2010 county census data to control for confounders that might vary across locations in China. We assigned county-level information to each of the DSP locations by spatially merging the DSP locations with ArcGIS shape files of the 2010 census. For example, in the main regressions, we include controls for average years of education and share of minority population. These additional covariates are taken from the Harvard Geospatial Library collection of China census data, and linked to the DSP locations by county. We also include a dummy for whether the DSP location is within 100km of the coast, which is calculated using ArcGIS. Note also that the 2010 DSP survey which provides us with

information on health-relevant behavior also solicited information on income per capita. This variable's mean is reported in Table 1 and used as a control variable in the analysis and in the calculation of predicted life expectancy.

PART 2: Heterogeneity by Sex and Age

I. Estimates by Sex

As shown in Table S4, the results when estimated separately for men and women fail to reveal important differences by sex in the effect of PM_{10} on mortality. For example, in our preferred specification using local linear regression (columns 3 and 6), we estimate that the decline in life expectancy at the Huai River is a 3.5/2.6 years for men and women respectively. Similarly, the increase in cardiorespiratory mortality at the Huai River is 40/33 percent for men and women respectively. The results are qualitatively similar across sex, which is consistent with an interpretation that shared exposure to air pollution is driving the results, rather than a spurious correlation with an omitted variable. In addition, the somewhat larger effects for men could potentially indicate that smoking and PM_{10} are particularly harmful in combination, as men in China have much higher smoking rates than women.

II. Estimates by Age

In Table S5, we examine the relationship between PM_{10} exposure and mortality rates separately for different age groups. While previous studies generally focus on the impact of temporary increases in pollution and its impact on short-term mortality, they generally focus on infant mortality, or on the overall increase in mortality (which will be over-represented by the elderly). Our analysis of long-term mortality enables us to consider the effect of air pollution throughout the life-cycle. As reported in Table S5, our results using both parametric and non-parametric methods indicate that a significant increase in cardiorespiratory mortality at the Huai River throughout the life-cycle. Figure S2 plots the change in cardiorespiratory mortality rates at the Huai River throughout the adult lifecycle using the local linear regression estimates of the magnitude of the discontinuity, including a 2 standard error width band. The results indicate that the increase in cardiorespiratory mortality is statistically significant for a large window of the adult life cycle. These striking findings make help explain the paper's central finding of a large impact of PM_{10} on life expectancy; since PM_{10} increases mortality risk from cardiorespiratory

mortality even at relatively young ages (when people have many years of remaining life expectancy), PM_{10} exposure imposes a large cumulative decline in life expectancy.

PART 3: Robustness of Results to Alternative Specifications and Samples

In this section, we examine the robustness of our results according to several different tests to confirm whether our results were affected qualitatively by the decisions made in our paper along several dimensions, such as data assignment, sample selection, functional forms of our models, and bandwidth selection methods.

I. Alternative Approaches to Implementation of the Parametric and Non-Parametric Regression Discontinuity Design

In Table S6, we explore the robustness of the results to alternative approaches to implementing the parametric RD design by varying the choice of functional form, starting with a linear polynomial in column interacted with a North dummy (1) and then progressively adding a higher order term to each subsequent specification as one moves from left to right. We consider separately the overall sample (154 DSP locations) in columns (1)-(5) and the restricted sample of locations within 5 degrees latitude of the Huai River in columns (6)-(10). For example, columns (3) and (8) includs latitude, its square, and its cube, and each term's interaction with a North dummy in the overall sample. Below the point estimate and its standard errors, we report the Akaike Information Criterion (AIC) statistic for that specification. The minimum element is listed in bold for the overall sample and the restricted sample. This exercise also provides our choice of functional form for our parametric results, and we focus on the functional form dictated by the AIC for our main outcome: life expectancy.

In the overall sample, the AIC favors a cubic polynomial in latitude (interacted with a North dummy) and in the restricted sample, the AIC favors a linear polynomial in latitude (interacted with a North dummy). While the AIC favors the choice of a linear polynomial for all our key outcome variables in the restricted sample, the AIC points to different optimal functional forms for different outcome variables in the parametric specifications. For example, for PM_{10}, the AIC statistic is minimized with a quartic polynomial, although the cubic is the next favored model. It is worth noting that the AIC is not entirely conclusive as a method for choosing the functional form for our running variable, which supports our use of non-parametric estimation

methods, which are less sensitive to these decisions.

Table S7 explores the sensitivity of the results to changing our bandwidth selection method for our non-parametric results. In the main paper, we use the method proposed by Imbens and Kalyanaraman (2012). In this table, we re-estimate our local linear regression results using alternative bandwidth choice criteria, such as the method proposed by Ludwig and Miller (2007), referred to as the cross validation method, or the method proposed by Calonico et al. (2014a). For each bandwidth selection method, we use three different kernel types (triangle, Epachenikov, uniform). The results are qualitatively similar across these different bandwidth selection methods and choice of kernel type, suggesting that our findings are not sensitive to how we generate our local linear regression estimates.

In Table S8, we re-estimate our models using a method proposed by Calonico et al. (2014a) in which local linear regression estimates can be "bias-corrected" for bias that can result from choice of bandwidth. Calonico et al. (2014a) also suggest an alternative method for calculating standard errors that is more conservative than conventional standard errors. Using their proposed methods, we generate results which are again qualitatively similar to the results featured in our main analysis. It is worth noting that the results are somewhat less statistically significant; while the main results were generally significant at the 1% level, the results reported in Table S7 are significant at the 5% level in most specifications. Note also that this stability of the results in Tables S7 and S8 are supportive of our reliance in this paper on non-parametric estimation methods for our baseline results, rather than parametric specifications, which are more sensitive to choice of functional form (as reported in Table S6).

In Figure S3, we re-run our placebo test of estimating the discontinuity in life expectancy and PM_{10} at 1 degree displacements of the Huai River (similar to Figure 4 in the main paper). In this version, instead of using the Imbens and Kalyanaraman (2012) method, we use bandwidth selected by the method proposed by Calonico et al. (2014a) and a triangular kernel. Note that only at the actual Huai River is a discontinuity in either life expectancy or PM_{10} observed, providing supportive evidence of our overall empirical strategy.

II. Assignment of PM_{10} Concentrations to DSP Locations

Table S9 examines the sensitivity of the results to alternative approaches to calculating distance weighted averages of PM_{10} and weather for each DSP location, and to our choice of

whether to use the closest monitoring station versus a distance-weighted set of monitoring stations. One potential concern is that using distance-weighted averages between a DSP location and PM_{10} readings from monitors will cause the calculated PM_{10} change to vary smoothly with latitude, and this may incorrectly attenuate the estimated effect of the Huai River policy on PM_{10} concentrations. We chose the distance-weighted method as it represents a conservative estimate of the impact of the policy on PM_{10}, though we acknowledge that other choices could have been made. In Table S8, each cell represents a separate regression, with a dummy for "North" the reported independent variable. All models include a cubic polynomial in latitude interacted with a "North" dummy. In columns (1)-(4), the PM_{10} for a given DSP location are calculated as the weighted average of the nearby monitoring stations, with the weight given by the inverse of the distance, the square of the distance, the cube of the distance, and the quartic of the distance respectively. The results are not very sensitive to using alternative methods for assigning PM_{10} to DSP locations for either the parametric (Panel A), non-parametric (Panel B) with the bandwidth estimated separately by column, and non-parametric results with fixed bandwidth (where we set the optimal bandwidth to the choice in column 1, as in the paper). The results across the various specifications are qualitatively similar and remain significant at the 5% level.

In columns (5)-(11), we examine the sensitivity of the results to alterations in the threshold for the cases where we only use the nearest monitor (rather than a weighted average). The analysis in the paper assigns PM_{10} based only on the nearest station when the nearest station is within 50 km of the DSP location and uses weighted averages across stations in cases where the nearest station is further than 25 km but less than 150 km away. In these columns, we re-estimate the discontinuity in pollution varying the decision rule where we use the closest station's reading (rather than the distance-weighted reading) only if the closest station is within 10km, 25km, 50km, or a larger distance away from DSP location. As shown in the three panels, the results are somewhat weakened by using weighted averages across many stations even when a single stations is relatively close. This supports our decision to use the closest station's reading when a reading is available within 50 km. More generally there is little evidence that the results are affected in any meaningful way by the rule to assign PM_{10} to DSP locations.

III. Restricting the Sample to DSP Locations Near Monitoring Stations

The main analysis is restricted to all DSP locations within 150 kilometers of TSP

monitoring station. Table S10 examines the robustness of the results to using stricter criteria and only including DSP locations within 150, 125, 100, 75, 50, or 25 kilometers of a monitoring station. While the results fluctuate to some extent, the discontinuity observed in both PM10 and life expectancy is robust to further restricting the sample to DSP locations near pollution monitoring stations, until the sample is limited to stations within 75 or 50 kilometers, and then the sample shrinks significantly, reducing the significance of the results. Note also that the results for life expectancy are more robust to this restriction. However, on the whole, these estimates are qualitatively similar to those in our main analysis and suggest our core finding is not sensitive to our choice of acceptable distance from a monitoring station.

IV. Expanding the Sample to Include All DSP Locations

Table S11 examines the relationship between health outcomes and living north of the Huai River among *all* DSP locations, even those that are not within 150 kilometers of an air quality monitoring station. For this larger sample of locations without complete PM_{10} data, we simply report the reduced form relationship between health outcomes and an indicator variable for "North", after controlling for our polynomial in latitude interacted with the North dummy. The sample includes all 161 DSP locations, relative to the sample of 154 locations in the main analysis. Among DSP locations within 5 degrees latitude of the Huai River, the expanded sample includes 82 locations, relative to the 79 locations with valid TSP data in the primary sample. In this larger sample, the estimated effect of living north of the Huai River is qualitatively similar to the results reported in the main analysis. Specifically, cardiorespiratory mortality rates are estimated to be 21-41% higher whereas no significant increase is found among non-cardiorespiratory illnesses. The range of estimates for the decline in life expectancies at the Huai River is also similar to those in the overall sample, ranging from 1.8 years to 3.2 years.

V. Restricting the Sample Period to Post-2009

Since the monitoring system in China expanded significantly in 2009, a natural test of the robustness of our results is to restrict the sample of mortality and pollution to the period after the expansion. Presumably, this sample will have greater precision in terms of accurate assignment of pollution to DSP locations, but will have the drawback that a shorter sample of mortality is being analyzed. The results in Table S12 are again qualitatively similar to the results generated

from the full sample period, with a range of estimated declines in life expectancy at the Huai River (2.4-3.8 years). Our IV estimates are somewhat larger than our results using the complete sample, with an additional exposure 10 $\mu g/m^3$ of PM_{10} reducing life expectancy by 0.99-1.5 years. This supports our claim that the health consequences of PM_{10} continue to affect mortality in China, even in recent years, and underscores the urgency of this public health issue.

VI. Dropping any Measurements of PM_{10} that were imputed from TSP

Our data are primarily composed of direct measurements, but in a limited set of cases, we use imputed PM_{10} measurements which are generated from available TSP readings. This is useful for providing better coverage of pollution exposure in the early years, when most but not all monitoring stations had begun to monitor PM_{10}. In Table S13, we examine the sensitivity of our results to restricting our sample to only direct measurements of PM_{10}. The results indicate that our findings are qualitatively unchanged by dropping any imputed measures of PM_{10}; the estimate of the increase in PM_{10} at the Huai River ranges from 31-49 $\mu g/m^3$ and is statistically significant across all specifications. The IV estimates are also in a relatively narrow range, with a 10 $\mu g/m^3$ increase associated with a .77-.99 decrease in years of life expectancy. This suggests that our inclusion of imputed PM_{10} does not significantly change our studies' conclusions.

PART 4: Do other government policies change at the Huai River?

A natural concern related to the research design is that the government used the Huai River as the demarcation line for changes in other government policies related to public health, and this would confound the estimates of PM_{10} on health. This possibility is mitigated by the fact that the Huai River is not a border used for administrative purposes. The Huai River follows the January zero degree average temperature line (Celsius), and this was in fact the basis of its choice as a method to divide the country for free heating. Further, local policies generally hew to administrative boundaries associated with cities and provinces; indeed, the Huai River cuts through several provinces. Nevertheless, we identified and compiled variables on policies that are plausibly related to health from the China City Statistical Yearbooks for 134 of our 154 DSP locations. Table S14 reports on the results of fitting regressions for these policy variables, where the parameter of interest is associated with an indicator for North. The outcome variables in Panel A are meant to capture health-relevant policies: counts of hospitals (per 10,000 residents), number of physicians

(per 10,000 residents), social security and employment spending (per capita). The outcome variables in Panel B are meant to capture measures of water pollution and wastewater treatment policies, since water pollution is another major environmental challenge in China. The variables in Panel B are water grade (1=clean, 6=polluted), share of wastewater that is treated, and share of solid waste that is treated. These variables are calculated as averages for the years 2004-2012, our sample period. Column (1) reports the difference between north and south after adjusting the estimate for a cubic in latitude interacted with the North dummy, while column (2) restricts the sample to cities that are roughly 5° latitude north or south of the Huai River (i.e., cities between 27.5° and 37.5° latitude) and adjusts for a linear function of latitude interacted with the North dummy. In columns (3)-(5), we estimate the discontinuity using our local linear regressions methods. The results fail to suggest any substantial differences in policies north and south of the Huai River. While some of our measures reveal small (and in some instances, statistically significant) differences between the two regions, the magnitudes are relatively small, and no statistically-significant discontinuity is observed using the local linear methods at the 5% level. Though the results cannot confirm that no differences are found in policies on either side of the river, they are broadly reassuring that the Huai River was not used as a boundary for policies other than those related to home heating.

References

1. *China's Environmental Yearbook*. 1981-2012. China Environmental Protection Agency and National Bureau of Statistics.
2. *World Meteorological Association.* Data taken from the World Weather Watch Program according to WMO Resolution 40 (Cg-XII). http://www.ncdc.noaa.gov/
3. Akaike, Hirotugu (1974). "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* 19 (6): 716–723.
4. Ludwig, J., and D. L. Miller. 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics* 122(1): 159-208.

Table S1

Registry of Causes of Death in China's Disease Surveillance Points System

Code	Description	ICD-9 code	ICD-10 code
U000	All causes		
U038	Respiratory infections	460–466, 480–487, 381–382	J00–J06, J10–J18, J20–J22, H65–H66
U039	Lower respiratory infections	466, 480–487	J10–J18, J20–J22
U040	Upper respiratory infections	460–465	J00–J06
U041	Otitis media	381–382	H65–H66
U060	Malignant neoplasms	140–208	C00–C97
U061	Mouth and oropharynx cancers	140–149	C00–C14
U062	Esophageal cancer	150	C15
U063	Stomach cancer	151	C16
U064	Colon and rectal cancers	153–154	C18–C21
U065	Liver cancer	155	C22
U067	Trachea, bronchus, and lung cancers	162	C33–C34
U078	Other neoplasms	210–239	D00–D48
U104	Cardiovascular diseases	390–459	I00–I99
U105	Rheumatic heart disease	390–398	I01–I09
U106	Hypertensive heart disease	401–405	I10–I13
U107	Ischemic heart disease	410–414	I20–I25
U108	Cerebrovascular disease	430–438	I60–I69
U109	Inflammatory heart diseases	420, 421, 422, 425	I30–I33, I38, I40, I42
U110	Other cardiovascular diseases	415–417, 423–424, 426–429, 440–448, 451–459	I00, I26–I28, I34–I37, I44–I51, I70–I99
U111	Respiratory diseases	470–478, 490–519	J30–J98
U112	Chronic obstructive pulmonary disease	490–492, 495–496	J40–J44
U113	Asthma	493	J45–J46
U114	Other respiratory diseases	470–478, 494, 500–508, 510–519	J30–J39, J47–J98
U148	Injuries	E800–999	V01–Y89
U156	Intentional injuries	E950–978, 990–999	X60–Y09, Y35–Y36, Y870, Y871

Notes : The list above is a catalogue of the disease classification scheme recorded by China's Disease Surveillance Points System (DSPS). The first column lists the code assigned by Chinese CDC. The second column is a description of the category of disease. The third column is the list of categories included under the International Classification of Diseases - 9th Revision (ICD-9). The fourth column lists the categories included under ICD-10. Note that this list of causes is what was made available to the research team by the Chinese CDC for the purpose of this project, and is not an exhaustive list of the data maintained by the DSPS. In the empirical analysis, cardiorespiratory mortality is composed of U038, U067, and U104. Non-cardiorespiratory mortality is all other causes (U000-U038-U067-U104).

Table S2

Regression Discontinuity Estimates of the Impact of the Huai River Policy by Cause of Death

	Polynomials				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
All Cause Mortality	0.26***	0.20**	0.27***	0.15	0.26***	0.26***	0.22**
(per 100,000, log)	(0.09)	(0.10)	(0.07)	(0.10)	(0.08)	(0.09)	(0.10)
Cardiorespiratory	0.42***	0.30**	0.39***	0.22	0.37***	0.39***	0.40***
(per 100,000, log)	(0.12)	(0.14)	(0.11)	(0.13)	(0.11)	(0.11)	(0.14)
Non-Cardiorespiratory	0.07	0.06	0.12*	0.07	0.13	0.13	0.16
(per 100,000, log)	(0.09)	(0.09)	(0.07)	(0.08)	(0.08)	(0.08)	(0.10)
<u>Sub-Categories of Cardiorespiratory Mortality</u>							
Heart	0.58***	0.43***	0.58***	0.43***	0.33*	0.32*	0.33*
	(0.19)	(0.15)	(0.19)	(0.15)	(0.17)	(0.17)	(0.17)
Stroke	0.49***	0.36**	0.46***	0.23	0.32**	0.32**	0.33*
	(0.15)	(0.18)	(0.14)	(0.17)	(0.15)	(0.16)	(0.17)
Lung Cancers	-0.17	0.08	-0.15	0.15	-0.32	-0.36	-0.43
	(0.17)	(0.12)	(0.15)	(0.11)	(0.28)	(0.29)	(0.31)
Respiratory Illnesses	0.44*	0.21	0.56**	0.38	0.51	0.49	0.44
	(0.23)	(0.27)	(0.24)	(0.27)	(0.33)	(0.33)	(0.36)
<u>Sub-Categories of Non-Cardiorespiratory Mortality</u>							
Cancers Other	-0.15	-0.08	-0.13	-0.14	-0.06	-0.06	-0.08
than Lung	(0.13)	(0.15)	(0.11)	(0.12)	(0.14)	(0.14)	(0.16)
Injuries	0.31	0.20	0.38**	0.29	0.32**	0.31**	0.40**
	(0.19)	(0.28)	(0.15)	(0.26)	(0.14)	(0.14)	(0.17)
Other Causes	0.19	0.15	0.26	0.21	0.33*	0.37*	0.44*
	(0.16)	(0.18)	(0.16)	(0.17)	(0.20)	(0.21)	(0.24)
Observations	154	154	79	79			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Demographic Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes: Each cell in the table represents a separate regression. In columns (1)-(4), we report OLS estimates of the coefficient on a "North of the Huai River" dummy after controlling for a polynomial in distance from the Huai river interacted with a North dummy. The results in columns 2 and 4 include the demographic covariates reported in Table 1. In columns (5)-(7), we report the estimated discontinuity at the Huai River using local linear regression and bandwidth selected by the method proposed by Imbens and Kalyanaraman (2012) for different kernel weighting methods. Heteroskedastic-consistent standard errors are reported below the coefficients in columns (1)-(4) and conventional local linear regression discontinuity standard errors are reported in columns (5)-(7). * significant at 10% ** significant at 5%. *** significant at 1%.

Table S3

Health-related Behavioral Patterns by Sex, South and North of the Huai River

	North	South	Difference in Means	Adjusted Difference (polynomial)	Adjusted Difference (local linear)
	(1)	(2)	(3)	(4)	(5)
Panel A: Men Only					
Smoking Rate (%)	62.4	66.1	-3.8*	-1.2	-4.3
	(10.9)	(11.5)	(2.0)	(7.5)	(7.2)
Heavy Drinker (%)	15.9	14.1	1.70	-5.6	-9.7
	(7.2)	(5.6)	(1.2)	(5.8)	(7.1)
Excessive Red Meat (%)	23.0	41.2	-18.1***	0.97	-8.6*
	(14.4)	(18.2)	(2.8)	(6.18)	(4.5)
Insufficient Exercise (%)	26.3	24.4	1.9	-4.2	-3.7
	(13.5)	(9.8)	(2.1)	(7.8)	(5.5)
Panel B: Women Only					
Smoking Rate (%)	6.5	4.6	2.0**	0.16	-2.0
	(5.2)	(3.6)	(0.8)	(2.07)	(1.7)
Heavy Drinker (%)	0.6	0.7	-0.05	-0.12	-0.1
	(0.9)	(0.7)	(0.14)	(0.40)	(0.3)
Excessive Red Meat (%)	13.6	27.6	-14.0***	0.76	-2.8
	(11.1)	(16.8)	(2.4)	(4.96)	(3.4)
Insufficient Exercise (%)	22.4	15.6	6.9***	0.61	1.6
	(15.1)	(9.3)	(2.2)	(8.56)	(7.4)

Note : Responses are reported for the sub-sample of 154 DSP locations which are used for the results reported in Tables 1-3. The results in column (4) are adjusted for a cubic in degrees latitude from the Huai River boundary (interacted with a North dummy). In column (5), we report the estimated (bias-corrected) discontinuity at the Huai River using local linear regression and bandwidth selected by the method proposed by Imbens and Kalyanaraman (2012) using a triangular kernel. Heteroskedastic-consistent standard errors are reported below the coefficients in columns (1)-(4) and conventional local linear regression discontinuity standard errors are reported in columns (5). * significant at 10% ** significant at 5%. *** significant at 1%. Source: China Health Behavioral Survey (2010).

Table S4

Regression Discontinuity Estimates of the Huai River Policy and the Impact of 10 $\mu g/m^3$ of PM_{10} on Health Outcomes by Sex

	Men Only			Women Only		
	Polynomials		Local Linear	Polynomials		Local Linear
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Discontinuity at the Boundary						
Life Expectancy at Birth (years)	-2.5**	-2.1*	-3.5***	-2.1**	-1.6	-2.6***
	(1.0)	(1.1)	(0.9)	(1.0)	(1.0)	(1.0)
Cardiorespiratory (per 100,000, log)	0.30**	0.21*	0.40***	0.34**	0.22*	0.33***
	(0.12)	(0.11)	(0.10)	(0.16)	(0.13)	(0.11)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}						
Life Expectancy at Birth (years)	-0.72*	-0.69**	-0.82***	-0.91	-1.0	-0.65***
	(0.42)	(0.35)	(0.30)	(0.69)	(0.8)	(0.23)
Cardiorespiratory (per 100,000, log)	0.09*	0.07*	0.11**	0.15	0.14	0.08**
	(0.05)	(0.04)	(0.05)	(0.09)	(0.10)	(0.04)
Observations	154	79		154	79	
Polynomial Function	3rd	Linear		3rd	Linear	
Sample	All	5 Degree		All	5 Degree	
Demographic Controls	Yes	Yes		Yes	Yes	
Kernel			Triangle			Triangle

Notes: Each cell in the table represents a separate regression. The results in Panel A report the discontinuity in life expectancy and cardiorespiratory mortality by sex in the same manner as those reported in Table 2. The results in Panel B report the impact of 10 $\mu g/m^3$ of PM_{10} on these outcomes, using the Huai River to generate IV estimates in the same manner as those reported in Table 3. * significant at 10% ** significant at 5%. *** significant at 1%.

Table S5

The Huai River Policy and the Impact of PM_{10} on Cardiorespiratory Mortality Throughout the Life Cycle

Age	Polynomial Estimates (2SLS)	Local Linear Estimates (Fuzzy RD)
20	0.12	0.12***
	(0.08)	(0.05)
25	0.15	0.13***
	(0.10)	(0.05)
30	0.20*	0.18**
	(0.11)	(0.07)
35	0.17*	0.12**
	(0.09)	(0.05)
40	0.22**	0.14**
	(0.10)	(0.06)
45	0.22**	0.15***
	(0.10)	(0.06)
50	0.23**	0.16***
	(0.11)	(0.06)
55	0.22*	0.15***
	(0.11)	(0.05)
60	0.19*	0.14***
	(0.10)	(0.05)
65	0.15*	0.14**
	(0.08)	(0.06)
70	0.13*	0.12**
	(0.08)	(0.05)
75	0.10	0.10**
	(0.06)	(0.04)
80	0.07	0.07*
	(0.06)	(0.04)
85+	0.03	0.03
	(0.05)	(0.04)

Note : In column (1), we report the 2SLS IV estimate of PM_{10} on cardiorespiratory mortality using "North of Huai River" as the instrunmental variable, after controlling for a cubic polynomial in degrees latitude from the Huai River interacted with a North dummy variable. In column (2), we estimate the impact of PM_{10} on cardiorespiratory mortality with the Huai River representing a "fuzzy" discontinuity in the level of PM_{10} with a triangular kernel and the bandwidth selection method recommended by Imbens and Kalyanaraman (2012). * significant at 10% ** significant at 5%. *** significant at 1%.

Table S6

Robustness Checks of Choice of Functional Form for Latitude

	Full Sample					Restricted Sample (within 5^0 latitude)				
	Linear	Quadratic	Cubic	Quartic	Quintic	Linear	Quadratic	Cubic	Quartic	Quintic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Life Expectancy	-0.4	-3.5***	-2.3**	-2.3*	-3.8***	-2.1*	-3.3**	-3.4*	-4.6	3.4
at Birth (years)	(0.9)	(0.9)	(1.0)	(1.2)	(1.3)	(1.2)	(1.5)	(1.7)	(3.1)	(4.1)
Akaike Info. Criterion	724.8	701.7	**700.0**	701.4	703.0	**333.6**	335.9	339.6	337.6	334.4
Particulate Matter (PM_{10})	14.6**	16.3*	27.6***	30.5***	14.6	21.7**	18.1*	26.0*	27.1	-10.9
	(6.3)	(8.6)	(9.2)	(10.2)	(13.6)	(9.0)	(10.2)	(15.4)	(27.0)	(43.1)
Akaike Info. Criterion	**1411.6**	1414.4	1415.4	1418.6	1419.6	**719.9**	721.5	724.1	727.2	729.8
Cardiorespiratory	0.08	0.32***	0.30**	0.39**	0.41**	0.22	0.37**	0.37*	0.23	-0.48
Mortality (per 100,000)	(0.10)	(0.11)	(0.14)	(0.16)	(0.19)	(0.13)	(0.19)	(0.22)	(0.30)	(0.45)
Akaike Info. Criterion	5.3	-7.5	-4.3	**-8.0**	-4.3	**-20.0**	-18.7	-15.6	-14.5	-15.3

Note : The overall sample includes 154 DSP locations and the restricted sample includes 74 DSP locations within 5 degrees latitude of the Huai River. Each cell in the table represents the coefficient from a separate regression where we report the magnitude of a "North of Huai River" dummy after controlling for the polynomial in latitude of degree listed in the column heading. Heteroskedastic-consistent standard errors are reported in parentheses. The value of the Akaike's Information Criterion (AIC) statistic is reported below the standard error, with the minimum element for the full and restricted sample in bold.* significant at 10% ** significant at 5%. *** significant at 1%.

Table S7

Regression Discontinuity Estimates of the Impact of the Huai River Policy by Alternative Bandwidth Selection Methods

	Calonico-Cattaneo-Titiunik			Cross-Validation		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Discontinuity at the Boundary						
Particulate Matter (PM_{10})	33.5***	38.1***	44.8***	41.6***	42.9***	43.3**
	(11.6)	(13.7)	(15.2)	(15.3)	(16.3)	(18.2)
Life Expectancy at Birth (years)	-3.4***	-3.6***	-3.7***	-3.1***	-3.2***	-3.3***
	(0.8)	(0.8)	(0.9)	(0.9)	(1.0)	(1.2)
Cardiorespiratory (per 100,000, log)	0.30***	0.32***	0.34***	0.37***	0.39***	0.40***
	(0.08)	(0.09)	(0.09)	(0.11)	(0.11)	(0.14)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}						
Life Expectancy at Birth (years)	-0.92***	-0.95***	-1.08***	-0.75***	-0.74***	-0.77**
	(0.29)	(0.31)	(0.36)	(0.27)	(0.26)	(0.30)
Cardiorespiratory (per 100,000, log)	0.09***	0.09***	0.10***	0.10**	0.10**	0.10**
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Kernel	Triangle	Epanech.	Uniform	Triangle	Epanech.	Uniform

Notes: Each cell in the table represents a separate regression. These results can be compared to the results presented in columns (4)-(6) of Table 2, but are estimated using alternative bandwidth selection methods. The Calonico-Cattaneo-Titunik method (columns 1-3) is the method proposed by Calonico et al. (2014a) and the Cross-Validation method (columns 4-6) is the method proposed by Ludwig and Miller (2007). * significant at 10% ** significant at 5%. *** significant at 1%.

Table S8

Regression Discontinuity Estimates of the Impact of the Huai River Policy using Bias-Corrected Coefficient Estimates and Robust Standard Errors Proposed By Calonico-Catteneo-Titiunik (2014a)

	Imbens-Kalyanaraman			Calonico-Cattaneo-Titiunik			Cross-Validation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Discontinuity at the Boundary									
Particulate Matter (PM_{10})	44.9**	46.3**	54.1**	47.1***	45.8**	38.2**	34.2**	36.2**	44.8**
	(17.7)	(19.1)	(21.9)	(16.5)	(18.0)	(18.9)	(15.8)	(17.2)	(20.9)
Life Expectancy	-3.4***	-3.7***	-4.1***	-3.2***	-2.9**	-2.7**	-2.8**	-2.9**	-3.1**
at Birth (years)	(1.2)	(1.3)	(1.4)	(1.1)	(1.2)	(1.3)	(1.2)	(1.3)	(1.5)
Cardiorespiratory	0.43***	0.45***	0.49***	0.39***	0.34***	0.30**	0.42***	0.44***	0.44**
(per 100,000, log)	(0.13)	(0.13)	(0.14)	(0.11)	(0.12)	(0.14)	(0.13)	(0.14)	(0.17)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}									
Life Expectancy	-1.1***	-1.3**	-1.6**	-0.7*	-0.6	-0.4	-0.7**	-0.6*	-0.6*
at Birth (years)	(0.4)	(0.5)	(0.7)	(0.4)	(0.4)	(0.6)	(0.3)	(0.3)	(0.4)
Cardiorespiratory	0.13**	0.14**	0.19**	0.08*	0.07	0.06	0.12**	0.13**	0.11*
(per 100,000, log)	(0.05)	(0.06)	(0.09)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Kernel	Triangle	Epanech.	Uniform	Triangle	Epanech.	Uniform	Triangle	Epanech.	Uniform

Notes : These results can be compared to the results presented in columns (4)-(6) of Table 2 and rows (1) and (2) of Table 3, but are estimated using the bias-correction method proposed by Calonico et al. (2014a) and the robust standard errors they propose. We present these results for three bandwidth-selection methods: the method proposed by Imbens and Kalyanaraman (2012) in columns (1)-(3), the method proposed by Calonico et al. (2014a) in columns (4)-(6), and the method proposed by Ludwig and Miller (2007), or the Cross-Validation method, in columns (7)-(9). * significant at 10% ** significant at 5%. *** significant at 1%.

Table S9

Robustness Checks using Weighted Averages of PM_{10} Across Monitoring Stations

	Weighted by:				Only use Nearest Station for Sites within (x) of a Monitor:					
	Distance	Distance^2	Distance^3	Distance^4	10 km	25 km	50 km	75 km	100 km	125 km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Polynomial Approach										
Particulate Matter	36.3***	34.0***	31.9***	30.1***	31.2***	29.9***	36.3***	28.8**	25.3**	25.4**
	(9.6)	(9.9)	(10.3)	(10.6)	(8.4)	(8.7)	(9.6)	(12.2)	(12.3)	(12.2)
Panel B: Local Linear Regression										
Particulate Matter	41.6***	39.8***	38.2**	36.3**	33.3***	35.0***	41.6***	31.6*	26.7	26.5
	(15.3)	(15.5)	(15.6)	(15.9)	(12.2)	(13.3)	(15.3)	(17.2)	(17.4)	(17.4)
Panel C: Local Linear Regression with a Fixed Bandwidth										
Particulate Matter	41.6***	40.1***	38.7**	37.6**	38.1**	37.9**	41.6***	34.8**	33.3**	33.4**
	(15.3)	(15.4)	(15.5)	(15.6)	(15.0)	(15.1)	(15.3)	(16.5)	(16.3)	(16.3)

Note : Each cell in the table represents the coefficient from a separate regression. In Panel A, we report the coefficient of a dummy for "North" with controls included for a cubic in latitude interacted with North. In Panel B, we report the magnitude of the discontinuity at the Huai River as estimated by local linear regression with the optimal bandwidth chosen by the method of Imbens and Kalyaranaman (2012) separately for each column. In the first four columns, PM_{10} for a given DSP site is calculated as the weighted average of the nearby monitoring stations, with the weight given by the inverse, the square, the cube, and the quartic of the distance' respectively. In columns 5-10, we assign to DSP sites sufficiently close to a monitoring station the value for the station, and assign all others using weighted averages where the weight is given by the inverse of the distance. For example, in column 5, any DSP location within 10 kilometers of a station is assigned the value at the closest station instead of a weighted average of the value at multiple stations. In Panel C, we reproduce the analysis of Panel B, but with a fixed bandwidth. Columns 1-4 all use the bandwidth chosen as optimal for the esimation of column 1, and columns 5-10 use the bandwidth chosen as optimal for estimation of column 7. * significant at 10% ** significant at 5%. *** significant at 1%.

Table S10

Robustness Checks of Choice of Acceptable Distance from DSP Locations to Monitoring Stations

	<150KM (4)	<125KM (5)	<100KM (6)	<75KM (7)	<50KM (8)
Panel A: Polynomial Approach					
Particulate Matter (PM_{10})	27.6***	28.5***	20.6**	-3.4	11.4
	(9.2)	(10.3)	(10.3)	(10.3)	(18.5)
Life Expectancy at Birth	-2.2**	-2.4**	-2.1*	-1.8	-1.9*
	(1.0)	(1.0)	(1.1)	(1.5)	(1.0)
Panel B: Local Linear Regression					
Particulate Matter (PM_{10})	41.6***	41.3***	29.0*	6.2	50.4*
	(15.3)	(15.0)	(15.5)	(19.0)	(28.3)
Life Expectancy at Birth	-3.1***	-3.2***	-3.0***	-4.8***	-2.5***
	(0.9)	(0.9)	(0.9)	(1.8)	(0.8)
Panel C: Local Linear Regression with a Fixed Bandwidth					
Particulate Matter (PM_{10})	41.6***	41.5***	36.9**	27.1	50.2*
	(15.3)	(15.0)	(15.6)	(17.5)	(25.9)
Life Expectancy at Birth	-3.1***	-3.1***	-3.1***	-3.7**	-2.4***
	(0.9)	(0.9)	(0.9)	(1.5)	(0.8)
Observations	154	153	149	133	106

Note: These results are comparable to the results in Table 2, but for different sample selection criteria. In the main results, all DSP locations within 150 kilometers are included in the sample (column 4). In these specifications, we evaluate how the results change when we change the cutoff for which DSP locations are included in our sample. In Panel A, we report the coefficient of a dummy for "North" with controls included for a cubic in latitude interacted with North. In Panel B, we report the magnitude of the discontinuity at the Huai River as estimated by local linear regression with the optimal bandwidth chosen for each sample. In Panel C, we report the magnitude of the discontinuity at the Huai River as estimated by local linear regression with the same bandwidth that is used in column 1, which is chosen by the method of Imbens and Kalayanaraman (2012) . * significant at 10% ** significant at 5%. *** significant at 1%.

Table S11

Regression Discontinuity Estimates of the Impact of the Huai River Policy using the Full Sample

	Polynomials				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Discontinuity at the Boundary							
Life Expectancy at Birth (years)	-3.1***	-2.1**	-3.2***	-1.8	-3.1***	-3.1***	-2.6**
	(1.0)	(1.1)	(0.9)	(1.1)	(1.0)	(1.1)	(1.2)
Cardiorespiratory (per 100,000, log)	0.41***	0.29**	0.38***	0.21*	0.37***	0.40***	0.41***
	(0.12)	(0.14)	(0.11)	(0.13)	(0.11)	(0.11)	(0.13)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}							
Life Expectancy at Birth (years)	-1.0*	-0.85	-1.1**	-0.98*	-0.75***	-0.74***	-0.77**
	(0.6)	(0.54)	(0.5)	(0.55)	(0.27)	(0.26)	(0.30)
Cardiorespiratory (per 100,000, log)	0.13**	0.11*	0.12**	0.10*	0.10**	0.10**	0.10**
	(0.06)	(0.06)	(0.05)	(0.06)	(0.04)	(0.05)	(0.05)
Observations	161	161	82	82			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Demographic Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes : These results were estimated in the same manner as those in Table 2, but estimated using the full sample of 161 DSP locations, which includes the additional 7 DSP locations which were not within 150 kilometers of a monitoring station. * significant at 10% ** significant at 5%. *** significant at 1%.

Table S12

Results Using Pollution and Mortality Data from 2009-2012

	Polynomials				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Discontinuity at the Boundary							
Particulate Matter (PM_{10})	25.7**	20.6**	26.3**	17.9**	33.6**	34.6**	34.3**
	(10.2)	(8.9)	(10.8)	(7.3)	(13.6)	(14.6)	(16.1)
Life Expectancy	-3.8***	-2.4*	-4.1***	-2.4*	-3.4***	-3.4***	-3.4***
at Birth (years)	(1.2)	(1.3)	(1.0)	(1.3)	(1.0)	(1.0)	(1.1)
Cardiorespiratory	0.52***	0.36**	0.50***	0.27*	0.42***	0.43***	0.40***
(per 100,000, log)	(0.14)	(0.16)	(0.12)	(0.14)	(0.11)	(0.12)	(0.13)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}							
Life Expectancy	-1.5*	-1.1	-1.5**	-1.3*	-1.01***	-0.98***	-0.99**
at Birth (years)	(0.8)	(0.8)	(0.7)	(0.7)	(0.38)	(0.38)	(0.43)
Cardiorespiratory	0.20*	0.17	0.19**	0.15*	0.14**	0.14**	0.13*
(per 100,000, log)	(0.10)	(0.11)	(0.08)	(0.08)	(0.06)	(0.06)	(0.07)
Observations	161	161	82	82			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Demographic Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes: These results were estimated in the same manner as those in Table 2, but estimated using only pollution and mortality data from 2009-2012, after a major expansion in the Chinese pollution monitoring system which occured in 2008. * significant at 10% ** significant at 5%. *** significant at 1%.

Table S13

Regression Discontinuity Estimates of the Huai River Policy and the Impact of 10 μg/m^3 of PM_{10} on Health Outcomes Using only Direct PM_{10} Measurements

	Polynomials				Local Linear Regressions		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Discontinuity at the Boundary							
Particulate Matter (PM_{10})	49.2***	31.4***	42.9***	32.0***	33.5**	31.6**	32.8*
	(11.8)	(11.9)	(10.6)	(8.5)	(15.3)	(15.8)	(16.9)
Life Expectancy	-4.1***	-2.8**	-3.6***	-2.5**	-3.2***	-3.1***	-2.8**
at Birth (years)	(1.1)	(1.2)	(0.9)	(1.2)	(0.9)	(1.0)	(1.2)
Cardiorespiratory	0.56***	0.40**	0.47***	0.28*	0.57***	0.59***	0.57***
(per 100,000, log)	(0.13)	(0.16)	(0.12)	(0.15)	(0.12)	(0.14)	(0.15)
Panel B: Instrumental Variables Estimates of 10 μ g/m^3 of PM $_{10}$							
Life Expectancy	-0.83**	-0.88*	-0.84***	-0.77**	-0.99**	-0.92**	-0.87*
at Birth (years)	(0.34)	(0.50)	(0.26)	(0.31)	(0.41)	(0.41)	(0.45)
Cardiorespiratory	0.11***	0.13**	0.11***	0.09***	0.15**	0.16**	0.17*
(per 100,000, log)	(0.04)	(0.06)	(0.03)	(0.03)	(0.07)	(0.08)	(0.10)
Observations	124	124	68	68			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Demographic Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes: The results in Panel A report the discontinuity in life expectancy and cardiorespiratory mortality in the same manner as those reported in Table 2, but without imputed values of PM_{10} included in the sample. The results in Panel B report the impact of 10 μg/m^3 of PM_{10} on these outcomes, using the Huai River to generate IV estimates in the same manner as those reported in Table 3, but without imputed values of PM_{10} from TSP included in the sample. * significant at 10% ** significant at 5%. *** significant at 1%.

Table S14

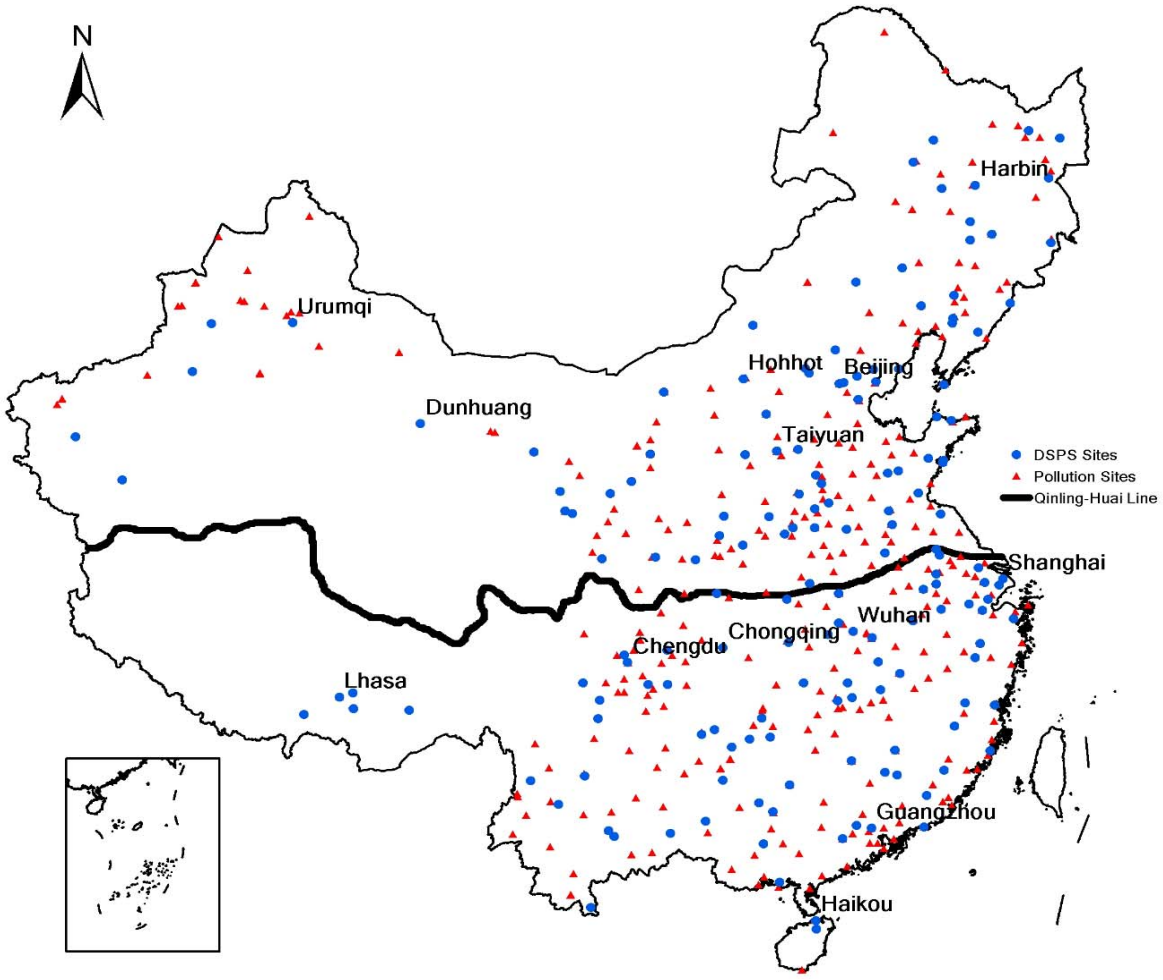
Patterns in Other Government Policies South and North of the Huai River Among DSP Locations

	Adjusted Difference (polynomial)		Adjusted Difference (local linear)		
	(1)	(2)	(3)	(4)	(5)
Panel A: Health Policy Variables					
Number of Hospitals	-0.03*	-0.01	0.0	0.0	0.0
(per 10,000 residents)	(0.02)	(0.01)	(0.0)	(0.0)	(0.0)
Number of Physicians	-10.5***	-7.6**	-5.1	-5.3	-6.9
(per 10,000 residents)	(3.50)	(3.10)	(4.9)	(5.4)	(5.8)
Social Security and Empl.	-0.15	-0.11	-0.1*	-0.1	-0.1
(per capita, yuan)	(0.11)	(0.09)	(0.1)	(0.1)	(0.1)
Panel B: Measures of Water Pollution and Wastewater Treatment Policies					
Water Pollution Grade (1-6)	0.89	1.17	-1.0	-1.0	-0.8
	(1.1)	(0.9)	(1.1)	(1.2)	(1.2)
Wastewater Treatment Rate (%)	-7.8*	-5.9	-5.0	-4.9	-4.6
	(4.7)	(4.8)	(5.4)	(5.5)	(5.8)
Solid Waste Treatment Rate (%)	2.49	4.79	0.2	0.6	-0.1
	(11.16)	(9.73)	(7.5)	(7.8)	(8.2)
Observations	134	68			
Polynomial Function	Cubic	Cubic			
Sample	All	5 Degree	All	All	All
Kernel			Triangle	Epanech.	Uniform

Notes: These statistics are taken from China's *Prefectural Statistical Yearbooks* for each DSP location and averaged between 2004 and 2012. In Panel A, we report the estimate of the discontinuity in hospitals, physicians, and social security spending at the Huai River. In Panel B, we report average water grade (1=clean, 6=dirty), the estimate of the proportion of wastewater that is treated, and the proportion of solid waste that is treated.

Figure S1

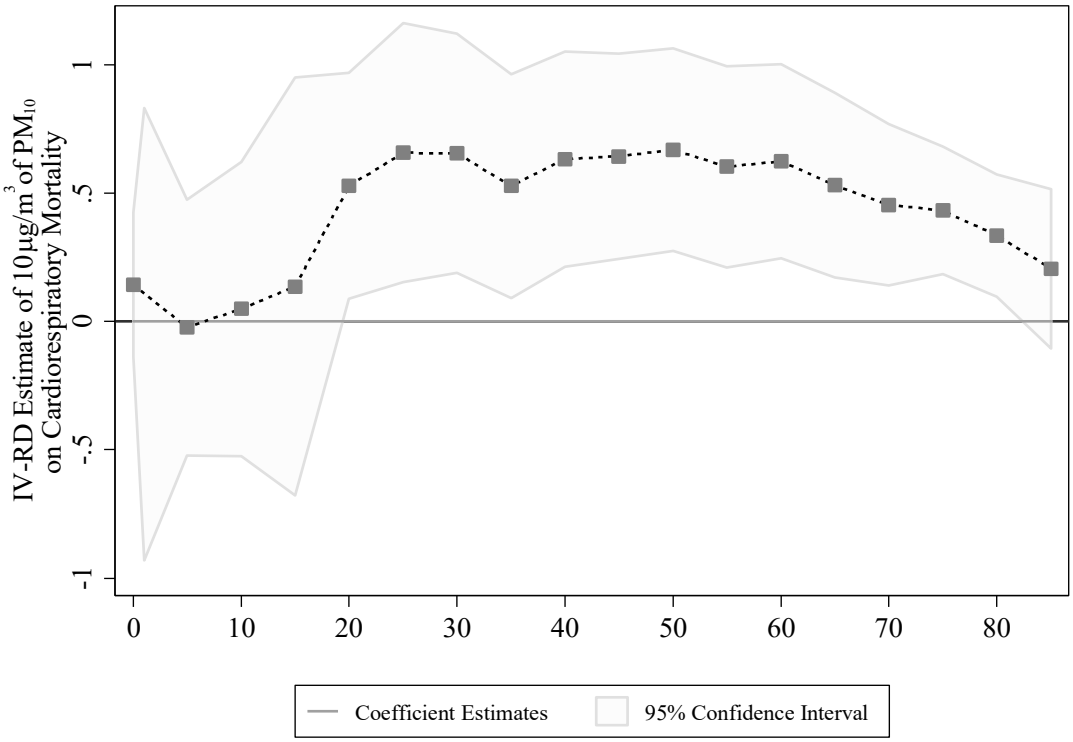
Location of DSP Sites and Pollution Monitoring Stations



Note : The figure plots the locations of the 161 Disease Surveillance Points and 325 pollution monitoring stations. Cities north of the solid line were covered by the home heating policy.

Figure S2

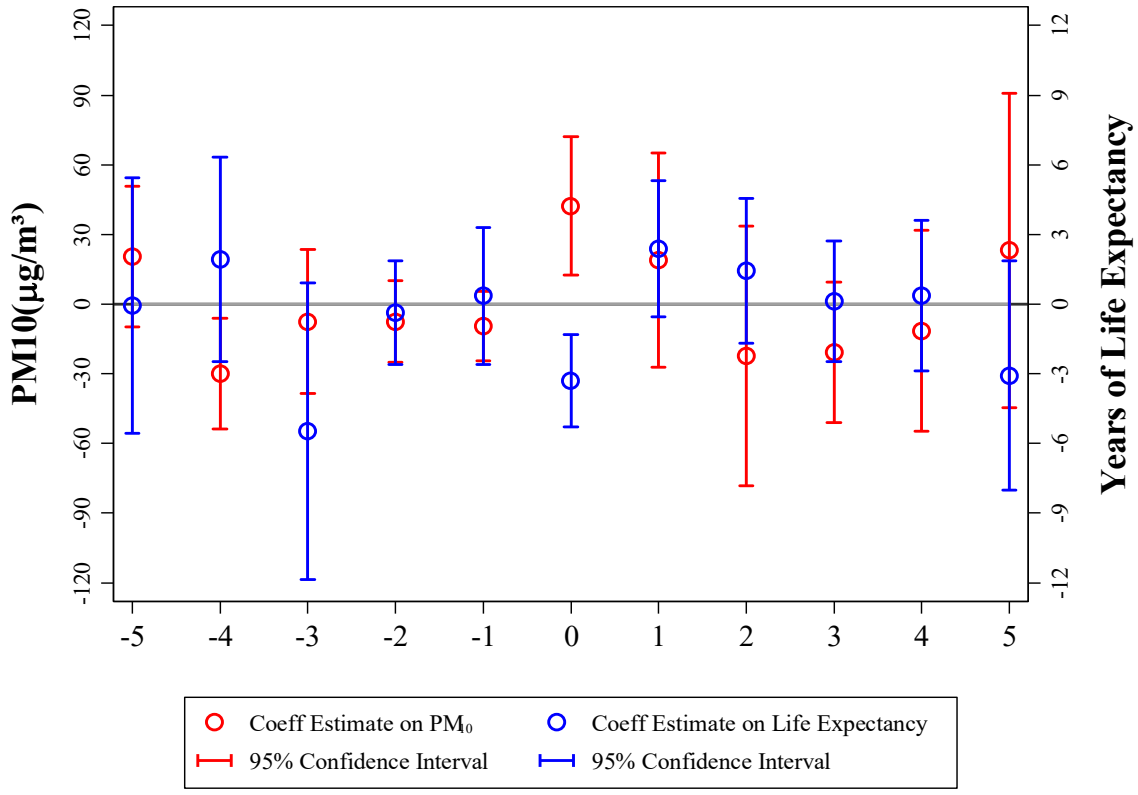
Instrumental Variables Estimates of the Impact of an Additional 10 μg/m^3 Exposure to PM_{10} on Cardiorespiratory Mortality by Age Group



Note : The figure plots the IV estimate of the impact of an additional 10 μg/m^3 exposure to PM_{10} on the log of cardiorespiratory mortality (per 100,000) at 5-year age intervals. The IV point estimate is estimated as a "fuzzy" RD, where it is estimated as the Wald ratio of the discontinuity in PM_{10} to the discontinuity in life expectancy, using the optimal bandwidth chosen for life expectancy to estimate both (Calonico et al. 2014). The discontinuities are estimated using the optimal bandwidth chosen by the method recommended by Imbens and Kalyanaraman (2012).

Figure S3

Placebo Testing: Estimated Discontinuity in Pollution and Life Expectancy at Displaced Huai River Boundaries using the Bandwidth Selection Method Proposed by Calonico et al. (2014a)



Notes : The figure reports the change in PM_{10} and life expectancy at the Huai River and at discontinuities estimated at 1 degree latitude displacements from the actual Huai River. Each discontinuity is estimated using a triangular kernel and with bandwidth chosen according to the method proposed by Calonico et al. (2014a).