New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy

Avraham Ebenstein[a,1], Maoyong Fan[b,1], Michael Greenstone[c,d,1,2], Guojun He[e,f,g,1], and Maigeng Zhou[h,1]

[a]Department of Environmental Economics and Management, Hebrew University of Jerusalem, Rehovot 76100, Israel; [b]Department of Economics, Ball State University, Muncie, IN 47304; [c]Department of Economics, University of Chicago, Chicago, IL 60637; [d]National Bureau of Economic Research, Cambridge, MA 02138; [e]Division of Social Science, The Hong Kong University of Science and Technology, Hong Kong; [f]Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong; [g]Department of Economics, The Hong Kong University of Science and Technology, Hong Kong; and [h]National Center for Chronic and Non-Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 100050, China

Edited by William C. Clark, Harvard University, Cambridge, MA, and approved July 3, 2017 (received for review October 27, 2016)

This paper finds that a 10-μg/m^3 increase in airborne particulate matter [particulate matter smaller than 10 μm (PM$_{10}$)] reduces life expectancy by 0.64 years (95% confidence interval = 0.21–1.07). This estimate is derived from quasiexperimental 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 those to the south. The findings are derived from a regression discontinuity design based on distance from the Huai River, and they are robust to using parametric and nonparametric 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 caused by 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 3.7 billion life-years.

airborne particulate matter | China | life expectancy | Huai River | regression discontinuity

Is airborne particulate matter (PM) today's greatest environmental risk to human health? Among the 5.9 billion people who live in countries where PM readings are available, 4.5 billion are currently exposed to PM concentrations that are at least twice the concentration that the WHO considers safe (1). The health effects of inhaling PM have been widely studied and found to be robustly associated with elevated risk of heart disease, stroke, and lung cancer (2–4). However, questions regarding the health effects of PM 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 also influence the probability that the world will face disruptive climate change, because the combustion of fossil fuels that causes PM also causes 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 composed 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.), such as has been emphasized in a recent *Science* article (5). Second, the available evidence is largely based on examinations of populations exposed to the modest levels of PM that are commonly observed in developed countries, where reliable pollution and health data are more readily available (6). PM concentrations in many developing countries (e.g., India and China) are 5–10 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. [Notable exceptions include the works by Dockery et al. (3) and Pope et al. (6, 7). However, these studies have at least one of the following limitations: they (*i*) exploit observational variation in PM, (*ii*) use small samples, (*iii*) assume no selective migration, or (*iv*) focus on low levels of pollution found in the United States.] As a consequence, the existing literature focuses on shorter-run variation in PM exposure and often examines outcomes (e.g., hospitalization, infant outcomes) that are only indirectly related to longer-run outcomes, like life expectancy.

This paper estimates the effect of sustained exposure to particulate matter smaller than 10 μm (PM$_{10}$) on life expectancy with recent data from China and, in so doing, addresses each of the previous literature's limitations. First, the quasiexperimental 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 is in cities to the south; the legacy of that policy is evident even today, with very different rates of indoor heating north and south of the Huai River. Consequently, the findings are derived from a regression discontinuity (RD) design based on distance from the Huai River.

Significance

An estimated 4.5 billion people are currently exposed to particulate matter (PM) levels at least twice the concentration that the WHO considers safe. Existing evidence linking health to air pollution is largely based on populations exposed to only modest levels of PM and almost entirely composed of observational studies, which are likely to confound air pollution with other unobserved determinants of health. This study uses quasiexperimental variation in particulate matter smaller than 10 μm (PM$_{10}$) generated by an arbitrary Chinese policy to find that a 10-μg/m^3 increase in PM$_{10}$ reduces life expectancy by 0.64 years. The estimates imply that bringing all of China into compliance with its Class I standards for PM$_{10}$ would save 3.7 billion life-years.

Author contributions: A.E., M.F., M.G., G.H., and M.Z. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The measures of PM$_{10}$ for each DSP location and the program files that produce the tables and figures have been posted as Dataset S1. Mortality measurements are available upon request and at the discretion of the Chinese Center for Disease Control and Prevention.

[1]A.E., M.F., M.G., G.H., and M.Z. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: mgreenst@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1616784114/-/DCSupplemental.

Second, the average PM_{10} concentration in China during the examined period is 103 $\mu g/m^3$ (i.e., more than five times the WHO standard), and, therefore, the results are informative for the 1.36 billion people currently living in China as well as the several billion other people who are exposed to high PM_{10} concentrations around the world. Furthermore, the analysis is conducted with the most comprehensive data file ever compiled on health and pollution in a developing country (Fig. 1).

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 PM in the early 1980s (and very likely earlier) and persist even today. Importantly, the mortality data cover the years 2004–2012 when restrictions on migration had been loosened, but Disease Surveillance Points (DSP) mortality was still generally recorded at an individual's birthplace or hukou (additional details are in *SI Appendix*). The possibility of individuals moving away from their hukou, as a compensatory response to high levels of air pollution, but still having their death recorded at their birthplace means that this study provides an answer to a central question: what are the effects of PM_{10} concentrations at a person's birthplace on their life expectancy? Because individuals can migrate and undertake other compensatory responses to air pollution, the resulting estimates of the effect of PM_{10} on life expectancy are more likely to be externally valid to other countries than estimates from earlier periods when Chinese migration was greatly restricted (8).

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 41.7 $\mu g/m^3$ [95% confidence interval (95% CI) = 16.4–67.0] and a decline of 3.1 years (95% CI = 1.3–4.9) in life expectancy just to the north of the river. Furthermore, the elevated mortality rates are concentrated among cardiorespiratory causes of death, whereas 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 0.64-year (95% CI = 0.21–1.07) decline in life expectancy.

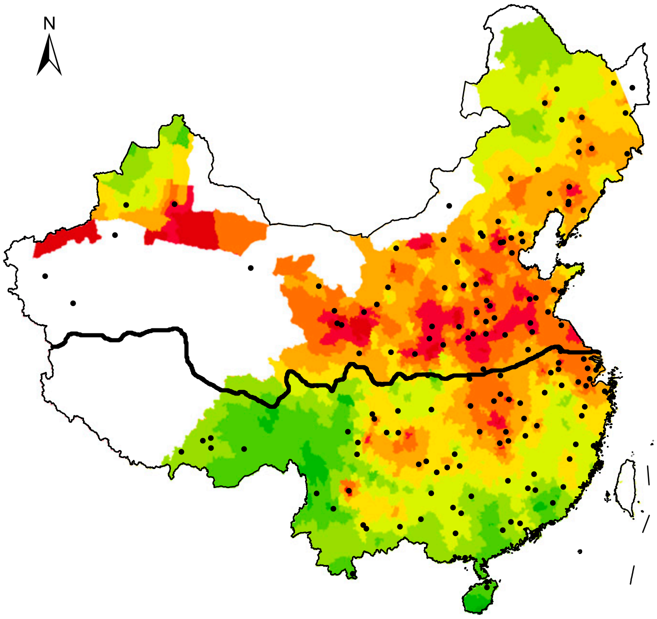

Fig. 1. China's Huai River/Qinling Mountain Range winter heating policy line and PM_{10} concentrations. Black dots indicate the DSP locations. Coloring corresponds to interpolated PM_{10} levels at the 12 nearest monitoring stations, where green, yellow, and red indicate areas with relatively low, moderate, and high levels of PM_{10}, respectively. Areas left in white are not within an acceptable range of any station.

This paper improves on the work by Chen et al. (8), which also exploits the Huai River Policy to measure the health impact of total suspended particulates (TSPs), by examining the health consequences of smaller particles that comprise PM_{10} and using significantly more accurate measures of mortality from a more recent time period (2004–2012) with a coverage population roughly eight times larger. We also probed the robustness of the results in several new and important ways, including but not limited to adjustment for important health behaviors (e.g., smoking), the choice of parametric vs. nonparametric estimation methods (9), a placebo exercise that finds that the discontinuity in pollution and life expectancy is only evident at the river itself, and a bounding exercise that allows for the possibility that pollution is correlated with an unobservable determinant of health (10).

The rest of the paper proceeds as follows. *Data Sources* describes the data sources, whereas *Econometric Model* outlines the empirical strategy. *Empirical Results* presents the results, and *Conclusion* provides some conclusions.

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 Control and Prevention (Chinese CDC) 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 73 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 noncardiorespiratory causes (i.e., cancers other than lung and all other causes). 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. 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 dataset; the panel data are collapsed in this manner, because the Huai River RD design is fundamentally a cross-sectional design. These collapsed measures of PM_{10} for each DSP are available in *SI Appendix* and are based on what we believe is the most comprehensive archive of Chinese air pollution ever assembled.

Three other datasets 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 Chinese CDC survey conducted in 2010. This survey is almost ideal for this study's purposes, because it was designed to capture patterns in behavior specifically for the coverage population at each DSP location. We use the 2005 census as a secondary dataset, which enables us to adjust the estimates for demographic covariates included in the census and not available from the Chinese CDC survey (e.g., share minority, average years of education). Finally, we consider other local factors that are plausibly related to health in China, including surface water pollution grade from China's Environmental Yearbooks (2004–2012). *SI Appendix* describes the data in greater detail.

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 (OLS) to fit the following equation to the cross-sectional data file:

$$Y_j = \beta_0 + \beta_1 PM_j + X_j \Gamma + \varepsilon_j, \tag{1}$$

where j references a DSP location in China. PM_j is the PM_{10} concentration in city j, X_j is a vector of the observable characteristics of the location 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 DSP location j or its residents' life expectancy, which is a simple function of age-specific mortality rates.

Table 1. Differences in pollution and other determinants of health at the Huai River

Outcome	North [1]	South [2]	Difference in means [3]	Adjusted difference (polynomial) [4]	Adjusted difference (local linear) [5]
PM_{10}	119.5 (31.5)	90.8 (25.3)	28.8*** (5.0)	48.3*** (12.2)	41.7*** (12.9)
Predicted life expectancy, y	76.2 (1.6)	76.2 (1.8)	−0.0 (0.3)	−1.3 (1.0)	−1.2 (1.0)
P value from joint test of equality	—	—	<0.01***	<0.01***	0.23

The sample is restricted to DSP locations ($n = 154$) within 150 km of an air quality monitoring station. 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 with a triangular kernel and bandwidth selected by the method proposed by Imbens and Kalyanaraman (14) chosen separately for each variable. Differences in predicted life expectancy are calculated by OLS using all of the covariates in *SI Appendix*, Table S1. The local linear joint test of equality uses the same set of covariates and bandwidth selection method proposed by Imbens and Kalyanaraman (14) with a uniform kernel.
***Significant at 1%.

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 (5, 11). Furthermore, pollution concentrations are measured with error, and it is well-known that classical measurement error will attenuate the coefficient associated with PM_{10}.

The second approach leverages the RD design implicit in the Huai River Policy to measure its impact on PM_{10} concentrations and life expectancy. 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 (12, 13).

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 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:

$$PM_j = \alpha_0 + \alpha_1 N_j + f(L_j) + N_j f(L_j) + X_j \gamma + u_j \qquad \textbf{[2a]}$$

$$Y_j = \delta_0 + \delta_1 N_j + f(L_j) + N_j f(L_j) + X_j \varphi + \varepsilon_j, \qquad \textbf{[2b]}$$

where j references a DSP location in China. PM_j is the average annual ambient concentration of PM_{10} in location j over the period 2004–2012, and Y_j is a measure of location j's mortality rate or life expectancy at birth. N_j is an indicator variable equal to one for locations that are north of the Huai River line, $f(L_j)$ is a polynomial in degrees north of the Huai River that is 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 nonparametric identification of **[2a]** and **[2b]**. For example, consider the following setup for estimation by local linear regression:

$$PM_j = \alpha_0 + \alpha_1 N_j + \alpha_2 L_j + \alpha_3 N_j L_j + u_j \qquad \textbf{[3a]}$$

$$Y_j = \delta_0 + \delta_1 N_j + \delta_2 L_j + \delta_3 N_j L_j + \varepsilon_j, \qquad \textbf{[3b]}$$

such that L_j is within h latitude degrees of the Huai River. We rely on a choice rule, which determines the optimal h as a function of the data (14, 15). Additionally, we report results using three separate kernels or weighting functions—triangle, uniform, and Epanechnikov.

Finally, we also report results from the parametric RD approach described in Eqs. **2a** and **2b** that restricts the sample to locations within 5° latitude of the Huai River; this sample restriction is an informal way of implementing the local linear methods that use bandwidths and kernels to focus comparisons near the discontinuity.

Empirical Results reports on estimation of Eqs. **2a**, **2b**, **3a**, and **3b** for PM_{10}, life expectancy at birth, and several other outcomes of interest, such as cardiorespiratory mortality. The parameters of interest are α_1 and 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. If the RD assumptions hold, estimates of δ_1 will provide an unbiased estimate of the life expectancy consequences of birth in a location just to the north of the Huai River. Importantly, this parameter is not a laboratory-style estimate of the consequences of exposure to air pollution where all other factors are held constant, because it reflects individuals' actions to protect themselves from the resulting health problems of pollution. Although the laboratory-style estimate might be of interest for researchers interested in how PM affects the human body, its relevance for understanding the real world consequences of air pollution is unclear. In fact, an appealing feature of the estimates of δ_1 is that they reflect all of the compensatory behavior that individuals undertake to protect themselves from air pollution, including migration to less polluted locations and other defensive measures, such as purchasing indoor air purifiers (16).

Importantly, the results in Eqs. **2a**, **2b**, **3a**, and **3b** 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

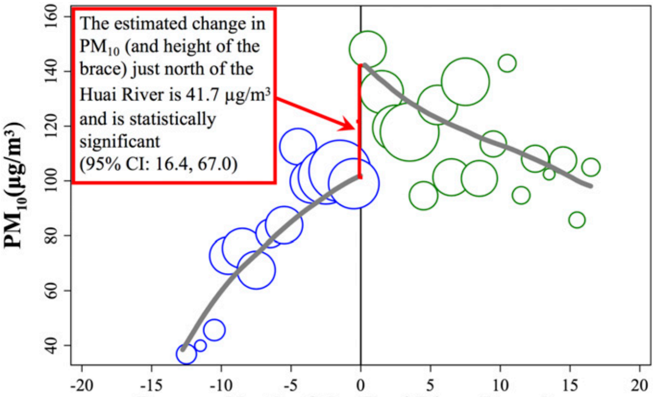

Fig. 2. Fitted values from a local linear regression of PM_{10} exposure on distance from the Huai River estimated separately on each side of the river.

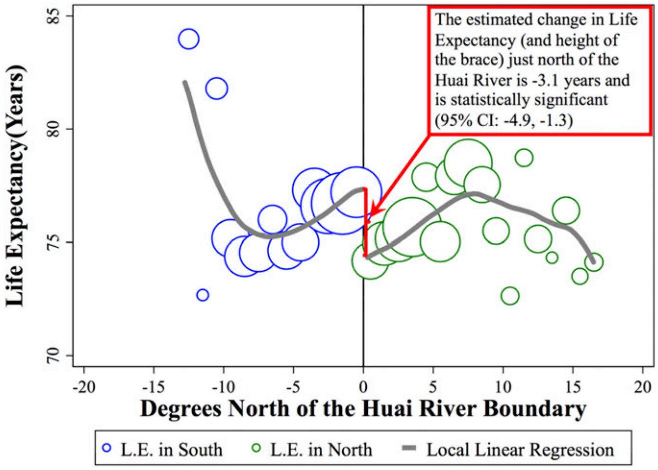

Fig. 3. Fitted values from a local linear regression of life expectancy (L.E.) on distance from the Huai River estimated in the same manner as in Fig. 2.

mortality through its impact on PM_{10}, then it is valid to treat Eq. **2a** as the first stage in a two-stage least squares (2SLS) system of equations. The appeal of the 2SLS approach is that it produces estimates of the impact of units of PM_{10} on life expectancy, and therefore, the results are applicable to other settings (e.g., India and other developing countries with comparable PM_{10} concentrations). The second-stage equation is

$$Y_j = \beta_0 + \beta_1 \widehat{PM}_j + f\left(L_j\right) + N_j f\left(L_j\right) + X_j \varphi + \varepsilon_j, \qquad \textbf{[2c]}$$

where $\widehat{PM}_j$ represents the fitted values from estimating **[2a]**; 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}.

A nonparametric analog 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 variable (IV) Wald estimate of the impact of PM_{10} on life expectancy that is analogous to the 2SLS estimates produced in **[2c]**, and it is based on the recommendations of Calonico et al. (17) 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, where the ratio is estimated as the ratio of two "sharp" discontinuities; in practice, we use the optimal bandwidth for life expectancy as the bandwidth for both life expectancy and PM_{10}. The work by Calonico et al. (15) has additional details.

Empirical Results

Assessing the Validity of the Huai River RD Design. Table 1 reports the summary statistics for PM_{10} exposure at DSP locations and provides evidence on the validity of the RD design. Columns [1] and [2] report the means along with the SDs in locations north and south of the Huai River line. Column [3] reports the mean differences between the north and the south along with the associated SEs. Column [4] also reports the differences (and SEs), but here, they are adjusted for a 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 (14).

There are large differences in PM_{10} exposure among southern and northern Chinese residents. In contrast, *SI Appendix*, Table S1 shows that concentrations of nitrogen dioxide and sulfur dioxide are statistically equivalent on both sides of the river after implementation of either the parametric or nonparametric RD approach. A potential explanation is that both sulfur dioxide and nitrogen dioxide are gaseous air pollutants that are lighter and travel farther than PM_{10}. Therefore, PM_{10} exposure can be isolated from other air pollutants as a potential health risk caused by living north of the Huai River.

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.) *SI Appendix*, Table S1 reports on the full set of individual covariates, whereas Table 1 provides two approaches to summarize differences in the available covariates and test for a discontinuity to the north of the river in these variables. The second row reports predicted life expectancy based on all of the potential covariates that collectively explain a substantial portion of the variation in life expectancy ($R^2 =$ 0.35). However, the null hypothesis of equal predicted life expectancies on the north and south sides of the river is not rejected

Table 2. **RD estimates of the impact of the Huai River Policy**

Outcome	[1]	[2]	[3]
Pollution and life expectancy			
PM_{10}	27.4*** (9.5)	31.8*** (9.1)	41.7*** (12.9)
Life expectancy at birth, y	−2.4** (1.0)	−2.2* (1.1)	−3.1*** (0.9)
Cause-specific mortality (per 100,000, log)			
Cardiorespiratory	0.30** (0.14)	0.22* (0.13)	0.37*** (0.11)
Noncardiorespiratory	0.06 (0.10)	0.08 (0.09)	0.13 (0.08)
RD type	Polynomial	Polynomial	LLR
Polynomial function	Third	Linear	
Sample	All	5°	

Column [1] reports 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 using the full sample (n = 154) and the control variables from *SI Appendix*, Table S1. Column [2] reports this estimate for the restricted sample (n = 79) of DSP locations within 5° of the Huai River. Column [3] presents estimates from local linear regression (LLR), with triangular kernel and bandwidth selected by the method proposed by Imbens and Kalyanaraman (14).

*Significant at 10%.

**Significant at 5%.

***Significant at 1%.

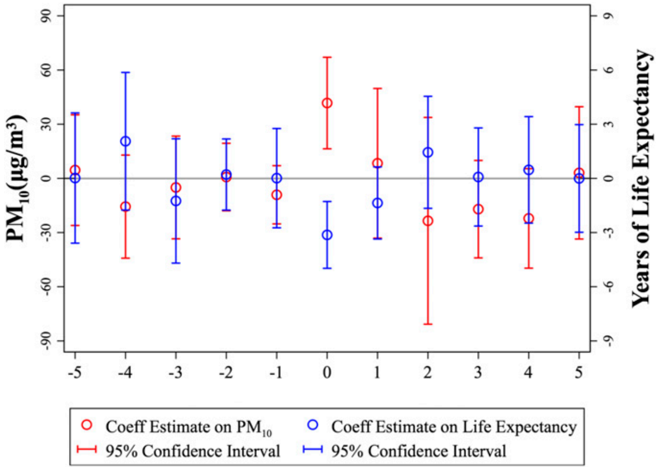

Fig. 4. RD estimates of the change in PM$_{10}$ and life expectancy at the Huai River and discontinuities estimated at 1°-latitude displacements from the actual Huai River.

with either application of the RD design. The third row reports P values from a joint test that the covariates are equal on the two sides of the river for all of the available covariates. The results of this test are less conclusive; they indicate that the null hypothesis of no difference in the covariates can be rejected using the polynomial application of the RD design but cannot be rejected using the local linear regression approach. These findings lend greater support to emphasizing the results from the local linear regression application of the RD design, but as an additional check, we probe the robustness of the results in several ways below.

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. Fig. 2 plots PM$_{10}$ at DSP locations against their degrees north of the Huai River line. The circles in Fig. 2 represent the average PM$_{10}$ concentration across locations within 1°-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 in Fig. 2 is generated by using a kernel-weighted local linear regression on either side of the river, which is similar to a nonparametric RD approach. This is estimated with a triangular kernel and bandwidth chosen by the method prescribed by Imbens and Kalyanaraman (14). Fig. 2 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 μg/m^3.

The plot in Fig. 3 is almost a mirror image of Fig. 2. It reveals a striking discrete decline in life expectancy at the border of roughly 3 years. Together, Figs. 2 and 3 reveal a sharp increase in PM$_{10}$ and a sharp decline in life expectancy at precisely the location where the Huai River Policy went into effect.

Table 2 statistically summarizes the graphical findings in Figs. 2 and 3 by reporting estimates (and SEs) associated with the north indicator variable from fitting equations for several variables of interest using a variety of RD approaches. Columns [1] and [2] apply the parametric RD approach from Eqs. **2a** and **2b** using the full sample and a subsample of DSP locations within 5° latitude of the Huai River, respectively. The estimates are adjusted for the full set of available covariates. We focus on specifications where latitude is interacted with the north dummy, so that latitude is allowed to affect outcomes differently north and south of the Huai River. In both samples, we use the Akaike Information Criteria of goodness of fit for life expectancy to choose the functional form for the polynomial in latitude, which recommends a cubic polynomial and linear term in columns [1] and [2] respectively. Column [3] reports on the estimated discontinuity at the Huai River using local linear regression with a triangular kernel and bandwidth selected by the method proposed by Imbens and Kalyanaraman (14). This nonparametric approach has the benefit that no functional form needs be imposed on the data. Furthermore, it places greater weight on DSP locations near the Huai River.

Table 2 presents evidence of significant increases in PM$_{10}$ and decreases in life expectancy at the Huai River. At the boundary, PM$_{10}$ rises by 27/32 μg/m^3 and life expectancy declines by 2.4/2.2 years in the full and restricted samples, respectively. The decline in life expectancy is driven by a statistically significant increase in cardiorespiratory mortality rates of 30/22% at the boundary in the full and restricted samples, respectively (Table 2, Cardiorespiratory). In contrast, there is little systematic evidence of a meaningful discontinuity in noncardiorespiratory mortality rates north of the river. The estimates from the nonparametric approach tend to be of similar magnitude: the estimated increase in PM$_{10}$ north of the river is 42 μg/m^3, whereas the decline in life expectancy is 3.1 years and again, driven by elevated cardiorespiratory mortality rates. *SI Appendix*, Table S7 shows a fuller set of results, all of which are qualitatively similar.

A powerful placebo test to assess the significance of these findings is to explore whether discontinuities are observed in other regions of China. Fig. 4 reports on the estimated discontinuities in PM$_{10}$ and life expectancy at 1°-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° displacement). The discontinuities are estimated from versions of Eqs. **3a** and **3b** using a triangular kernel and the bandwidth selection approach of Imbens and Kalyanaraman (14). Fig. 4 shows 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% CI. These results provide additional evidence that the effects in Figs. 2 and 3 and Table 2 are because of the Huai River Policy rather than an artifact of this application of the RD approach.

Estimates of the Effect of PM$_{10}$ on Life Expectancy. Table 3 reports on the estimated effect of 10 μg/m^3 PM$_{10}$ on life expectancy and cardiorespiratory mortality rates from alternative estimation approaches. Column [1] reports on the conventional OLS approach detailed in Eq. **1** and provides a basis for comparison with the RD IV or 2SLS approaches reported in columns [2] and [3].

The RD IV and 2SLS approaches suggest a substantially larger estimate of the health effects of PM$_{10}$. Specifically, the OLS estimate suggests that an additional 10 μg/m^3 sustained exposure to PM$_{10}$ is associated with a statistically significant decline in life expectancy of 0.27 years. The column [2] estimate from the parametric

Table 3. Comparing OLS and RD estimates of PM's impact on health outcomes

Outcome	[1]	[2]	[3]
Life expectancy at birth, y	−0.27*** (0.09)	−0.86* (0.51)	−0.64*** (0.22)
Cardiorespiratory (per 100,000, log)	0.02*** (0.01)	0.11* (0.06)	0.08*** (0.03)
Estimation method	OLS	IV	IV
RD type		Polynomial	LLR

In column [1], we report OLS estimates of the association between PM$_{10}$ and the listed outcome. In column [2], we report the 2SLS IV estimates using north of the Huai River as the IV and a cubic polynomial in degrees latitude from the Huai River interacted with a north dummy variable. In column [3], we estimate the impact of PM$_{10}$ on the listed outcomes using local linear regression (LLR), 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 PM$_{10}$ exposure. Results are reported in terms of the impact of an additional 10 μg/m^3 long-term PM$_{10}$ exposure. Results in columns [1] and [2] are based on the full sample (n = 154) and include the covariates listed in *SI Appendix*, Table S1. Column [3] is based on the bandwidth selection method proposed by Imbens and Kalyanaraman (14) with a triangular kernel.
*Significant at 10%.
***Significant at 1%.

RD IV approach indicates that an additional 10 $\mu g/m^3$ PM_{10} reduces life expectancy by 0.86 years, which is significant at the 10% level. The estimated effect from the favored nonparametric approach is –0.64 years and would be judged highly statistically significant by conventional criteria. The larger magnitude of these IV estimates suggests that some combination of omitted variables (e.g., more polluted areas are richer) and measurement error reduces the magnitude of the OLS estimates relative to the true effect of PM_{10} on life expectancy. *SI Appendix*, Table S8 shows a fuller set of results.

Robustness and Interpretation. *SI Appendix* explores heterogeneity in the results across different populations as well as the results' sensitivity to a rich set of robustness checks. Among the wide set of these results, we find that estimates do not differ significantly between sexes (*SI Appendix*, Tables S4 and S5), the impacts on cardiorespiratory mortality rates are generally evident over the entire course of the lifecycle (*SI Appendix*, Fig. S2 and Table S6), and different parametric and nonparametric applications of the RD design support the choice of polynomials in latitude in Tables 2 and 3 (*SI Appendix*, Table S9) and the robustness of the local linear regression results to different bandwidth selection methods and alternative choices of kernels (*SI Appendix*, Tables S10 and S11). Furthermore, the results are not very sensitive to our pollution assignment method or choice of acceptable distance from a monitoring station (*SI Appendix*, Tables S12 and S13), expanding the sample to include sites that are not near any pollution monitor (*SI Appendix*, Table S14), limiting the sample to data with more recent pollution data (*SI Appendix*, Table S15), or removing observations where PM_{10} is imputed from TSP (*SI Appendix*, Table S16). We also examined how the results are affected by using residual life expectancy as the outcome, which enables an examination of how local linear regression results are affected by the inclusion of covariates (*SI Appendix*, Table S17), and the results again remain qualitatively similar to the main results. *SI Appendix*, Tables S18 and S19 verify that the results are robust to inclusion of a control for distance of each DSP site from the coast and that the Huai River did not serve as a demarcation line for changes in other government policies that could confound the estimates of PM_{10} on health.

SI Appendix, Tables S20 and S21 explore the potential impact of migration on the estimates using information in the 2005 census on respondents' place of origin, place of current residence, and timing of migration. Although these results are from a different dataset, they suggest that the degree of compensatory migration in response to the higher PM_{10} levels induced by the Huai River Policy was limited during this period and that pollution concentrations at an individual's birthplace or hukou are likely to be a reliable measure of their lifetime exposure to pollution. It is worth noting that, because roughly 1.5% of deaths are registered outside an individual's birth hukou, we cannot entirely rule out the possibility that selective migration could influence our point estimates. Nevertheless, the results of these tests suggest that compensatory migration is unlikely to significantly bias the estimates of the effect of the Huai River Policy on lifetime PM_{10} exposure or in turn, the IV estimates of the effect of PM_{10} on life expectancy.

SI Appendix, Table S22 presents the results of bounding tests to consider how the point estimates would change if there are unobserved differences among the populations south and north of the river following the proposed method by Oster (10). The results fail to contradict the paper's qualitative findings and are presented in detail in *SI Appendix*.

Finally, *SI Appendix* explores the relationship between the estimates presented here and those in the work by Chen et al. (8). It is noteworthy that the application of the Huai River RD design produces estimates of the relationship between airborne PM and life expectancy that are qualitatively identical, despite the fact that they are derived from two different decades.

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 41.7 $\mu g/m^3$ (95% CI = 16.4–67.0) or 46% higher in the north and reductions in life expectancies of 3.1 years (95% CI = 1.3–4.9) in the north caused by elevated rates of cardiorespiratory mortality. More broadly, the results suggest that sustained exposure to an additional 10 $\mu g/m^3$ PM_{10} is associated with a 0.64-year (95% CI = 0.21–1.07) decline in life expectancy.

The implications of these results for human wellbeing are potentially enormous. The application of the paper's estimates suggests that bringing all of China into compliance with their Class I PM_{10} standard of 40 $\mu g/m^3$ would lead to a gain of 3.7 billion life-years for their current population. Furthermore, a growing body of evidence finds that individuals devote substantial resources to protecting themselves from air pollution, and these defensive expenditures represent very real costs of air pollution that are in addition to the direct mortality and morbidity effects (16, 18).

The risks to life expectancy from PM exposure are not confined to China. 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, for most people in the world, there is currently no greater environmental risk to health than airborne PM.

ACKNOWLEDGMENTS. We thank our seminar participants and colleagues for their insightful comments, including Songnian Chen, Rema Hanna, Douglas Miller, Paulina Oliva, and Nicholas Sanders. Faraz Hayat, Harshil Sahai, Patrick Schwarz, and Dan Stuart provided outstanding research assistance. This study is supported by China National Science and Technology Pillar Program 2013 Grant 2013BAI04B02 from the Ministry of Science and Technology and a School-Based Initiatives Research Grant from the Hong Kong University of Science and Technology.

1. World Health Organization (2014) Data from "Ambient Air Pollution Database." Available at www.who.int/phe/health_topics/outdoorair/databases/cities-2014/en/. Accessed October 26, 2016.
2. Pope CA, 3rd (1989) Respiratory disease associated with community air pollution and a steel mill, Utah Valley. *Am J Public Health* 79:623–628.
3. Dockery DW, et al. (1993) An association between air pollution and mortality in six U.S. cities. *N Engl J Med* 329:1753–1759.
4. Wellenius GA, Schwartz J, Mittleman MA (2005) Air pollution and hospital admissions for ischemic and hemorrhagic stroke among Medicare beneficiaries. *Stroke* 36:2549–2553.
5. Dominici F, Greenstone M, Sunstein CR (2014) Science and regulation. Particulate matter matters. *Science* 344:257–259.
6. Pope CA, 3rd, Ezzati M, Dockery DW (2009) Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med* 360:376–386.
7. Pope CA, 3rd, et al. (2002) Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 287:1132–1141.
8. Chen Y, Ebenstein A, Greenstone M, Li H (2013) Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proc Natl Acad Sci USA* 110:12936–12941.
9. Gelman A, Imbens G (2014) Why high-order polynomials should not be used in regression discontinuity designs. Working paper 20405 (NBER, Cambridge, MA). Available at http://www.nber.org/papers/w20405. Accessed July 24, 2017.
10. Oster E (2016) Unobservable selection and coefficient stability: Theory and evidence. *J Bus Econ Stat*, 10.1080/07350015.2016.1227711.
11. Chay K, Dobkin C, Greenstone M (2003) The Clean Air Act of 1970 and adult mortality. *J Risk Uncertain* 27:279–300.
12. Cook TD, Campbell DT (1979) *Quasi-Experimentation: Design and Analysis for Field Settings* (Rand McNally, Chicago).
13. Imbens G, Lemieux T (2008) Regression discontinuity designs: A guide to practice. *J Econom* 142:615–635.
14. Imbens G, Kalyanaraman K (2012) Optimal bandwidth choice for the regression discontinuity estimator. *Rev Econ Stud* 79:933–959.
15. Calonico S, Cattaneo MD, Titiunik R (2014a) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82:2295–2326.
16. Ito K, Zhang S (2016) Willingness to pay for clean air: Evidence from air purifier markets in China. Working paper 22367 (NBER). Available at http://www.nber.org/papers/w22367. Accessed July 24, 2017.
17. Calonico S, Cattaneo MD, Titiunik R (2014b) Robust data-driven inference in the regression-discontinuity design. *Stata J* 14:909–946.
18. Deschenes O, Greenstone M, Shapiro JS (2017) Defensive investments and the demand for air quality: Evidence from the NOx budget program. *Am Econ Rev*, in press.