SUPPORTING INFORMATION

New Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China's Huai River Policy

PART 1: Description of Data and Summary Statistics

I. Summary Statistics

Table S1 is an extended version of Table 1 and reports the summary statistics for several of the key determinants of mortality rates north and south of the Huai River, as well as our rich set of covariates. These data are described in detail in the sub-sections below, including the data sources and the construction of the key variables. Two key points emerge from this table, which are not immediately apparent in Table 1. First, the large differences in PM_{10} exposure among Southern and Northern Chinese residents are not accompanied by differences in nitrogen dioxide and sulfur dioxide levels, after implementation of either RD approach as reported in columns (4) and (5). Second, other observable determinants of life expectancy are generally well-balanced just to the north of the Huai River, relative to just to 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. Further, there are fewer instances of imbalance with the local linear RD approach. In contrast when one compares all of the North to all of the South as in column (3), there are a greater number of statistically significant differences, underscoring the value of the RD design in this setting.

II. Mortality Rate and Life Expectancy Data

Our sample of mortality in China is taken from the Disease Surveillance Points (DSP) system (1), 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 has been relied on by earlier scholarship examining health in China (e.g. (2)).[1] However, as a result of the SARS outbreak in

[1] A DSP sample was collected between 2001 and 2003 but the data were of limited quality and never released publicly.

2003, the Chinese CDC obtained funds to dramatically expand the DSP sample to cover *all* inhabitants of 161 urban districts/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 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 S2. 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 S3 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 diseases (U111), lung cancer (U067), heart diseases (U104-U107, U109-U110) and stroke (U108). Non-cardiorespiratory illnesses include cancers other than lung (U060-U065) and all other causes.[2] The cardiorespiratory causes are presumed to be affected by pollution exposure (and expected to rise 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 specifications. This increase is driven largely by increases in mortality from heart disease and stroke, though 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.

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

[2] These codes are provided by the Chinese CDC. The mapping from these categories to ICD-9 and ICD-10 codes is reported in Table S2.

observations for 2004-2012 for each of the 161 DSP locations.[3]

III. 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 (3) and China's *Environmental Quality Annual Reports* 2001-2012 (4), 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 and 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 *Environmental Protection Agency* as well as electronic copies of China's *Environmental Yearbooks*.

[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.

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 38 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[4]; (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.[5] These readings are then supplemented by interpolation, where readings are imputed in missing years with valid readings (when available).

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:

[4] In instances where a station had multiple years in which readings of TSP and PM_{10} were available, the average ratio of PM_{10} to TSP was used for imputation. The ratio averages 0.469 in 2003, 0.41 in 2004, and 0.37 in 2009 and 2010.

[5] Results where we exclude these imputed PM_{10} readings are presented in Part 3 of this appendix and are available in Table S13.

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 each of these 386 stations and the 161 DSP sites or locations yielded a full matrix of 161 X 386 calculated distances. The location of the stations and DSP sites are reported in Figure S1.[6]

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.[7] 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.[8]

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 (5). Our analysis is limited to the 304 weather stations with nearly-complete weather data for

[6] 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.
[7] The results are robust to different choices for the functional form. This is discussed further in Part 3.
[8] We also performed a separate robustness check where DSP locations are restricted to only being assigned pollution readings from the same side of the Huai (e.g. Northern DSP locations only receive readings from Northern monitoring stations). The results are very similar to our baseline results.

1981-2013.[9] At each of these stations, we calculated the total heating and cooling degree days for 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 304 stations, with the weights determined by the distance between the DSP location and the station. All stations within a 200 kilometer range of the DSP location were used to estimate the heating and cooling degrees in each year. The weather exposure for a DSP location in a year is the average of heating and cooling degrees in all previous years assigned in a manner similar to our method for assigning air pollution exposure.[10]

IV. 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 S4, 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 in columns 4 and 5 respectively, no observable change in these behaviors is found near the Huai River.

V. Other Covariates

As a complement to the DSP, we also use the China 2005 county census data to control for confounders that might vary across locations in China. We assigned county-level information

---

[9] Only stations with at least 350 days of data for every year in the period were included in the sample of weather stations. If a station had fewer than 350 readings in a particular year, the station was assigned a zero weight when estimating the weather for the DSP location for that year.
[10] The roll back for weather variables starts from the year 1990. For example, precipitation value for a DSP in 2007 is the average of the readings from 1990 to 2006.

to each of the DSP locations by matching the two data sets using China's county/district coding scheme. For example, in the main regressions, we include controls for average years of education, share of minority population, and the share of population employed as health professionals (e.g. doctors, nurses, etc.). 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. China's Environmental Yearbooks (2004 – 2012) provides us with the surface water pollution grade of each DSP. The 2010 DSP survey that provides us with information on health-relevant behavior also solicited information on income per capita. This variable's mean is reported in Table S1 and is 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 S5, 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 S6, we examine the relationship between PM_{10} exposure and mortality rates separately for different age groups. As previous studies generally examine the impact of temporary increases in pollution on short-term mortality, they tend to 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 S6, our results using both parametric and non-parametric methods reveal 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 the 95% confidence interval. The results indicate that the increase in cardiorespiratory mortality is statistically significant for a large range of the adult life cycle. These striking findings 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.

Note that while it may be surprising the estimated relationship between infant mortality and air pollution exposure is not stronger in Figure S2, and at odds with other influential studies examining this connection, there are two reasons that may explain our low point estimates for infants. First, surveys conducted by the Chinese CDC indicate that the completeness of deaths registration for infants and young children was lower than that of adult deaths (6). Second, for this population, there is also reason to be skeptical of the breakdown by causes. In particular, even though the DSP is an excellent resource for Chinese mortality data, it is still the case that the collection of rural mortality data (in which deaths occur at home rather than at a hospital) may be particularly unreliable in assigning causes to infant deaths, so the results for infants should be interpreted with caution.

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. Expanded Versions of Tables 2 and 3

In Table S7, we present an expanded version of Table 2 that includes a full set of parametric estimates with and without adjusting for the covariates listed in Panels B-D of Table S1, and local linear estimates using a variety of kernel weighting methods. The entries in columns (2) and (4) present parametric estimates identical to those in Table 2, which are adjusted using the full set of covariates and a cubic or linear polynomial in distance to the Huai River. Columns (1) and (3) report the estimates from these regressions without including covariates. The entries in column (5) match the local linear results in Table 2, which uses a triangular kernel, while columns (6) and (7) use an Epanechnikov and uniform kernel, respectively.

The parametric estimates of the change in PM_{10}, life expectancy and cardiorespiratory mortality at the Huai River in columns (1) – (4) indicate that the estimates with and without adjusting for the full set of covariates are qualitatively similar to one another and are all statistically significant. Similarly, there is little evidence of a discontinuous change in non-cardiorespiratory mortality at the boundary both with and without adjusting for covariates, indicating that the estimated decline in life expectancy is primarily driven by elevated cardiorespiratory mortality rates. The local linear estimates in columns (5) – (7) are statistically unchanged by the choice of kernel weighting method across all the listed outcome variables, and lie within a tighter range than the parametric regressions.

In Table S8, we similarly expand the set of results reported in Table 3 to include OLS and 2SLS estimates with and without adjusting for the full set of covariates as well as versions of the parametric regressions using the full and restricted sample. We also report local linear "fuzzy" discontinuity estimates using a variety of kernel weighting methods. Columns (1) – (4) report on the conventional OLS approach detailed in equation (1) of the manuscript; the full sample is used in the first two columns, while the sample in the last two columns is restricted to the 79 DSP sites within 5 degrees latitude of the Huai River. In each of these regressions, the full sample incorporates a cubic polynomial in distance from the Huai River, whereas the restricted sample uses a linear polynomial. 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 entries in columns (2), (6), and (9) match the entries in columns (1) – (3) of Table 3.

The traditional OLS estimates indicate substantially smaller estimates of the health effects of PM_{10}. In particular, 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 0.19 to 0.33 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 μg/m^3 of PM_{10} on life expectancy that range from -0.68 to -0.86 years in columns (5) through (8); three of the four estimates are statistically significant at the 5% level. The estimated effect from the non-parametric approach ranges from -0.64 to -0.68 years in columns (9) through (11) and would all be judged to be statistically significant by conventional criteria.[11]

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

In Table S9, 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) include 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) (8) 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 full sample. For example, the AIC statistic is minimized with a quadratic polynomial for PM_{10} but with a quartic polynomial for cardiorespiratory mortality. It is worth noting that the AIC is not entirely conclusive as a method

[11] Recall, we follow the method of Calonico et al. (7), which uses the same bandwidth determined when 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.

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 S10 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 (9). 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 (10), referred to as the cross validation method, or the method proposed by Calonico et al. (11). 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 S11, we re-estimate our models using a method proposed by Calonico et al. (11) in which local linear regression estimates can be "bias-corrected" for bias that can result from choice of bandwidth. Calonico et al. (11) 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. With more conservative standard errors, the results reported in Table S8 are still statistically significant at the 5% level in most specifications. Note also that this stability of the results in Tables S10 and S11 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 S9).

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 2 in the main paper). In this version, instead of using the Imbens and Kalyanaraman (9) method, we use bandwidth selected by the method proposed by Calonico et al. (11) 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.

III. Alternative Method of Assignment of PM_{10} Concentrations to DSP Locations

Table S12 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, though we acknowledge that other choices could have been made, which we investigate here. In Table S12, 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 inverse of the distance, the cubic inverse of the distance, and the quartic inverse 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 50 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 quite stable across different specifications and there is little evidence that the results are affected in any meaningful way by the rule to assign PM_{10} to DSP locations.

IV. Restricting the Sample to DSP Locations Near Monitoring Stations

The main analysis is restricted to all DSP locations within 150 kilometers of a PM_{10} monitoring station. Table S13 examines the robustness of the results to using stricter criteria and only including DSP locations within 200, 150, 125, or 100 kilometers of a monitoring station.

While the results fluctuate to some extent, the discontinuity observed using local linear regression in both PM_{10} and life expectancy is robust to expanding the sample to include stations within 200 kilometers or to further restricting the sample to DSP locations very near pollution monitoring stations, until the sample is limited to stations within 100 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, and the results are more robust in our non-parametric estimations. The stability of the results is most pronounced in Panel C, where we use non-parametric estimation but use a fixed bandwidth over all specifications, so that the discontinuity is estimated using a fixed set of observations. 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.

V. Expanding the Sample to Include All DSP Locations

Table S14 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 PM_{10} 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 22-41% higher. 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 1.7 years to 3.2 years.

VI. 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 S15 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 (2.3-4.1 years). Our IV estimates are somewhat larger than our results using the complete sample, with an additional exposure 10 μg/m^3 of PM_{10} reducing life expectancy by 0.89-1.3 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.

VII. Dropping any Estimates 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 that 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 S16, 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 μg/m^3 and is statistically significant across all specifications. The IV estimates are also in a relatively narrow range, with a 10 μg/m^3 increase associated with a 0.8-1.0 decrease in years of life expectancy. This suggests that our inclusion of imputed PM_{10} does not significantly change our studies' conclusions.

VIII. Examining the Discontinuity in Predicted Life Expectancy versus Residual Life Expectancy

A test of the validity of the RD design is the smoothness in the values for the covariates around the discontinuity. Our composite measure of the covariates, predicted life expectancy, is examined in Table S17. As mentioned, this variable is generated using all the demographic, health and environmental covariates presented in Table S1. In Table S17, we also examine residual life expectancy, which is the difference between actual and predicted life expectancy.[12] It is worth noting that there is a debate over whether actual or residual life expectancy should be the featured outcome in a regression discontinuity analysis (12). This issue is immaterial in 2SLS

[12] Note that since these two variables are generated as a function of other variables, the standard errors would generally be calculated accounting for this fact. Since this variable is only being used for expository purposes, we have not adjusted our standard errors for this and simply treat the variable as a standard variable emerging from a data-generating-process. See Lee and Lemieux (12) for a discussion of under what conditions that conventional standard errors can be used and it can be ignored that these variables were estimated.

when covariates can be easily included in a regression, but in non-parametric estimation, the inclusion of covariates can only be accomplished by using residual life expectancy as an alternative outcome variable. Lee and Lemieux (12) propose that in theory, if there is no violation of the RD assumption that unobservables are similar on both sides of the discontinuity, using a residualized outcome variable is desirable because it will improve the precision of estimates without a violation of the identification assumption. While we favor using actual life expectancy as the outcome in the in main analysis for reasons of transparency and faithfulness to the research design, this table examines the results' robustness to this alternative outcome variable.

The results in columns (1)-(6) demonstrate that across a range of specification choices, predicted life expectancy exhibits no discontinuity at the Huai River whereas residual life expectancy declines significantly at the Huai River. In columns (1)-(3), we estimate the discontinuity varying the bandwidth choice across the method proposed by Imbens and Kalyanaraman (9), the method proposed by Calonico et al. (11), and the method proposed by Ludwig and Miller (10). In columns (4)-(6), we alternatively estimate the discontinuity in the two variables using bias-adjusted estimates from the method proposed by Calonico et al. (11), as well as the robust standard errors they recommend. Note that in all specifications, the estimated discontinuity in predicted life expectancy is not significant, whereas the change in residual life expectancy is statistically significant. This supports the paper's identifying assumption that covariates change smoothly around the discontinuity and that the decline in life expectancy at the Huai River cannot be explained by the demographic and weather covariates included in our calculation of predicted life expectancy.

IX. Controlling for each DSP Location's Distance from the Coast

In our main empirical results using parametric estimation methods, we include a dummy variable for whether the DSP location was within 100 kilometers of China's coast. As a more flexible way to account for how inland cities may differ from cities closer to the coast, we present in Table S18 results in which we include either a linear or quadratic term in meters from the coast. The results are qualitatively similar to the regression results in Table 2 and Table 3 of the main paper (and by extension Table S7 and S8), suggesting that distance from the coast is not a significant factor in explaining our results.

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 these policies might 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 for 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 125 of the sample's 154 DSP locations. Table S19 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) and number of physicians (per 10,000 residents). 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 include 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. Columns (1)-(2) report sample averages north and south of the river, weighted by the population at each DSP location. Column (3) reports the simple difference in means between the north and south, and column (4) reports this difference after adjusting the estimate for a cubic in latitude. In column (5), we estimate the discontinuity using our local linear regressions methods in the same way as Table 1.

The results generally fail to lead to the rejection of the null hypothesis of no differences in policies north and south of the Huai River for the individual variables, but as in Table 1, we find tests of joint significance are inconclusive; using the polynomial approach, we reject the null of no difference between north and south, whereas with local linear regression, we fail to reject the null hypothesis. Overall, these findings are broadly reassuring that the Huai River was not used as a boundary for policies other than those related to home heating but also spur us to consider allowing for unobserved differences between north and south, which is addressed in Part 6 of the Supplementary Materials.

Note also these extra policy variables are not included in our main analysis because they are not available for all DSP locations and would thus reduce the sample size. However, adding these variables does not significantly alter the results we present in Table 2 and 3 of the paper.[13]

PART 5: The Consequences of Migration for Interpreting the Results

In China, migration has historically been tightly regulated and many social benefits are only available within one's origin or birthplace *hukou* (registration area). Specifically, China has a household registration system where people are permanently connected to their birthplace or *hukou*. However, in recent years, restrictions have been relaxed, and migration within China has become more common. This creates a potential challenge for our analysis, since some individuals reside outside their *hukou* for extended periods, and may die outside their *hukou*. The official DSP protocol is to consider the reference population at each site as the 'permanent population', defined as individuals who have resided in the location for more than six months, even those individuals who do not hold local *hukou*. However, since data collection on migrant populations is notoriously difficult, we investigated this issue more directly to examine where mortality is recorded for those without local *hukou*. We obtained DSP individual mortality records from 2008-2012 that recorded whether the deceased had local *hukou* and found that the vast majority of recorded deaths (98.5%) come from the local *hukou* population.[14] This finding is consistent with a known practice that seriously ill migrants return home for medical care and to be with their families in their final days, and thus have mortality registered at their *hukou*. Consequently, it seems reasonable to conclude that the measured mortality rate at a location is a close approximation to the mortality rate for individuals born at that location, regardless of whether they had previously been a migrant.

---

[13] For example, our reduced form estimate of the decline in life expectancy at the Huai River using a polynomial specification on all 154 locations versus the reduced sample of 125 locations is only changed from 2.4 years to 2.2 years, and remains statistically significant.

[14] During the 1990s through 2003, the DSP assigned a death to an individual's *hukou*. This practice was altered beginning in 2004, and deaths were instead intended to be recorded at an individual's place of residence if he or she had lived there for more than 6 months (regardless of *hukou* location). In 2008, the DSP began separately tabulating the number of deaths for those with and without local *hukou*. This extract of mortality at the DSP locations was provided to us and indicates that 98.5% of deaths occur among those with local *hukou*.

As such, while mortality will be properly recorded at the *hukou*, migration poses a potential challenge for our pollution exposure measure, which is based on assuming that lifetime exposure to pollution is at the levels observed at their *hukou*. Further, extensive migration also poses a potential challenge for interpreting the IV estimates of the effect of PM10 on life expectancy.

We explore the potential impact of migration on the results in several ways. First, as proposed by McCrary (13), we test for a discontinuity in population to the north of the Huai River. Table S20 implements several versions of this test using all counties in China's 2010 Census (columns 1, 3, and 5-7) and for the subset of counties included in the DSP that are used in our main analysis (columns 2, 4 and 8-10).[15] The point estimates tend to indicate that counties to the north have smaller populations but none of these tests would be judged statistically significant by conventional criteria. Further, the imprecision of the estimates from the counties containing DSP locations makes it difficult to draw strong conclusions regarding the subset of counties that are included in our paper's main analysis. Overall, this table fails to contradict the null hypothesis of equal population density near the Huai River but further investigation seems warranted.

Second, we examine directly whether there is a discontinuity in migration rates to the north of the Huai River. The results are presented in Table S21, which uses microdata from the 2005 Census to assess overall trends in migration; using the Census questions, we define a migrant as a respondent who is residing in a *hukou* that differs from their origin *hukou,* restricting the sample to the 154 Census counties that contain DSP locations used in our main analysis. The Census also provides information on how long the respondent has lived in their current location. Its limitation is that it only reports on the respondents' origin *hukou* and their current location so periods living in other locations, or even previous periods at the current location, cannot be observed. An additional limitation is that 24% of migrants report having lived outside their *hukou* for more than 6 years, which we impute as equal to ten years for all respondents.[16]

[15] Note that we focus on the 154 DSP counties with valid pollution measures used in our main analysis, rather than the entire DSP sample of 161 counties.
[16] We examined how the results change by assuming that the category of ">6" is equal to 15 or 20 years and find our main findings robust to this decision.

The column (1)–(2) entries indicate that 9.0% of this population lives outside their origin *hukou* and this fraction varies over the life cycle. For example, 18.5% of 20-29 year olds qualify as migrants by this criterion, while migration rates are much lower for the young and old (i.e., 4.7% for those 60 and older and 5.7% for ages 0-9). Although these migration rates are not trivial, columns (3)–(5) fail to find evidence of a statistically significant discontinuous change in migration rates at the river for the full sample or any subcategory. Thus, aggregate migration rates appear unrelated to the Huai River Policy, however this does not rule out migration as a source of mismeasurement in lifetime PM_{10} exposure.

The remaining columns exploit the availability of information in the Census on how long migrants have lived away from their origin hukou. Column (6) reports that the average Census respondent has spent 97% of their life in their birthplace hukou. Among migrants, this share is, of course, lower but is still quite high at 81.7% (see column (7)). These shares are also reported by age category.

Columns (8) – (10) use an alternative measure of lifetime PM_{10} exposure for individuals born at each of 154 DSP locations that accounts for migration. Specifically, we calculate respondents' lifetime PM_{10} exposure as the weighted-average of pollution in one's *hukou* and current residence, where the weights are the estimates of the share of their life spent in each location. Each DSP is then assigned the average lifetime PM_{10} exposure of all individuals whose origin *hukou* is that DSP location.

The Panel B entries report local linear RD estimates of lifetime PM_{10} exposure at the Huai River by age group. We continue to define the North indicator as equal to one for DSPs which are North of the Huai River. This means that the coefficient associated with the North indicator measures the discrete difference in lifetime PM_{10} exposure for individuals born in that DSP, accounting for any effect of that population's migration patterns on exposure. The results indicate that migration does not have a differential effect on lifetime PM_{10} exposure to the North of the Huai River. Specifically, the estimated discontinuity in Panel B is not materially different from the estimates that assume no migration as in Table 2 and Table S7 (reported again in Panel A of this Table).

While we emphasize that the 2005 Census only provides limited and incomplete information on migration patterns, these results, in conjunction with those in columns (1)-(5), fail to provide evidence that compensatory migration meaningfully biases 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. Overall, it appears that the paper's main findings would not be significantly altered if we were able to more accurately capture migration in the DSP data.

PART 6: Are the Results Robust to Oster's (14) Test for Selection Based on Unobserved Differences between DSP Locations North and South of the River?

As proposed by Oster (14), we perform a bounding exercise to consider the paper's findings sensitivity to the presence of unobserved selection of a particular variety.[17] The Oster (14) test examines the sensitivity of estimates when balancing exercises (e.g., as reported in Table S1) reveal differences between the control and treated populations. The spirit of this test is to assume that the magnitude of unobserved selection can be inferred by the amount of observed selection, which in turn is approximated by the observable differences between the control and treated populations – or in our context, the observable differences in our control variables near the Huai River. The intuition is that as controls are added, the magnitude of the change in the R squared is informative for the level of observable selection in the sample, and by assumption, for the level of unobserved selection potentially operating as well.

Table S22 reports the results of this bounding exercise, where we consider how the R squared changes in response to adding all available control variables from Panels B-D of Table S1. The results of this exercise fail to contradict the paper's primary conclusions, because this type of unobserved selection would not change the sign of the estimated effect. Furthermore, we find that selection on unobservables would have to be on average 2-3 times as powerful as selection on observables in order to zero-out the estimated effects.[18] This is particularly encouraging, as many of the observed controls are important determinants of the outcomes of interest, and so it is unlikely that the results are entirely explained by unobserved selection. While it is, of course, impossible to rule out the possibility that the results are affected by the presence of unobserved selection, the results of the exercise proposed by Oster (14) generally lend additional credibility to the paper's findings.

[17] Oster's (14) method for bounding estimates by specifying the form of unobserved selection was originally proposed in spirt by Altonji, Elder, and Taber (15).
[18] Oster (14) and Altonji, Elder, and Taber (15) suggest that results are heuristically robust when selection on unobservables need to be at least as important as selection on observables to zero-out the treatment effect.

PART 7: Ambiguity of the Location of the Huai River Line and Examining Municipal Compliance with the Huai River Policy

I. Ambiguity of the Location of the Huai River Line

The Qin Mountain and Huai River line is historically regarded as the geographical dividing line between northern and southern China. The line approximately follows the 0° January isotherm (Celsius) and the 800 mm isohyet in China. In the past few centuries, the geographical conditions in the Huai River basin have changed substantially, which introduced some ambiguity on China's north-south divide.

Before the Northern Song Dynasty, the Huai River directly entered the Yellow Sea (northern part of East China Sea) at Yuntiguan. However, starting from Southern Song Dynasty (1200s), because the Yellow River in northern China repeatedly changed its course southwards and ran into the Huai River, the geography of the Huai River basin was significantly changed. A variety of new geographical features, such as new high lands, lakes, and the built-up silt of the Yellow River's historical southern course, were created after the Yellow River changed back to its northerly course.[19] These changes prevented water in the midsection of the Huai River from flowing to the lower section, while water in the lower section could not find an estuary to the sea. Gradually the downstream water from pooled up into Lake Hongze and eventually entered Yangtze River. Some water enters the Yellow Sea through the North Jiangsu Irrigation Canal, and other water flows into Huaishu River which heads north to Lianyungang City and eventually enters Haizhou Bay.

The changes in waterways of the Huai River create ambiguity in drawing an "accurate" north-south dividing line, particularly for the downstream segment. Previous studies have used slightly different versions of the Huai River line since there is no official documentation specifying its location.

In this paper, we draw the Huai River line based on its major waterway, which originates in Tongbai Mountain in Henan province, flows through southern Henan, northern Anhui, and northern Jiangsu, and then finally enters the Yangtze River at Yangzhou in Jiangsu Province. This Huai River line is also used in several other studies, such as Almond et al. (17), Chen et al. (2) and Makinen (18). Slightly different version of the Huai River line can be found in Talhelm

[19] Please refer to (16) for more details.

et al. (19), Xiao et al. (20), and Ito and Zhang (21). The major difference between these two sets of the Huai River line is whether one uses Yangtze River or the North Jiangsu Irrigation Canal as the Huai River's estuary to the sea.[20]

The north-south assignment of the DSP locations in this paper is not affected by slight adjustments to the Huai River line even though the exact position of the Huai River line is still debatable. This is because the surveillance points located in the Huai River's downstream basin (Jiangsu province) are relatively far from the Huai River line.

II. The Winter Heating Policy and Non-Compliance among Municipalities

China established the winter heating system during the 1950s with assistance from the Soviet Union. Facing resource constraints, Chinese leaders proposed to provide free winter heating only for northern China using the Qin-Huai line as the cutoff. This divide roughly traces China's Qin Mountains and Huai River near the 33° latitude line, which also corresponds to China's January temperature 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. In a carry-over from China's highly centralized economic planning, the central heating policy was kept in place and the government continues to provide free or heavily subsidized winter heat to residents in northern China (22). However, the actual free winter heating provision does not completely follow the Qin-Huai line because of ambiguity of the Huai River Line. Further, a few cities in provinces that the Huai River line cuts through are non-compliers. For example, in Jiangsu province, Xuzhou is located north to the Huai River and has central heating, but its adjacent city, Suqian, does not provide central heating even though it is also located to the north of the Huai River. In contrast, we are unaware of any evidence of non-compliance in provinces that are further away from the Huai River, either to the north or south.

Due to this possibility of non-compliance in provinces that are divided by the Huai River, we carefully checked whether each municipality that has at least one DSP location in these provinces and made sure of the implementation of the free central winter heating policy. The results and sources are reported in Table S23. We only identified one non-complier in the sample: Yancheng City in Jiangsu Province is north of the Huai River but has no central heating.

[20] Talhelm et al. (19) uses coarser north-south divide that is based on provincial boundaries. For example, the entire Jiangsu province is treated as south in Talhelm et al. (19).

We checked the robustness of the main results in three different ways: (1) assign Yancheng City as a treated city despite the fact that there is no winter heating, then estimate a sharp regression discontinuity; (2) drop Yancheng City and estimate a sharp regression discontinuity; and (3) treat Yancheng City as a non-complier and estimate a fuzzy regression discontinuity for PM_{10} and life expectancy. The results are not sensitive to our treatment of Yancheng and the findings remain quantitatively unchanged across these three different specifications.

PART 8: Comparison of Results to Chen et al. (2)

Chen et al. (2) 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 (2) 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 45.4% of TSP. Thus, the Chen et al. (2) estimates translate into a prediction that 10 $\mu g/m^3$of PM_{10} reduce life expectancies by 0.66 years (3/4.54) that is well within sampling error of this paper's estimate of 0.64 years. Note also that if TSPs are composed of roughly the same historical proportion of PM_{10}, then 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 helping to explain why the life expectancy penalty for living north of the river has declined from 5.5 years to 3.1 years. 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.

PART 9: Estimating Total Years of Life Lost using the Coefficient Estimates

The estimate of total years of life lost from current levels of PM_{10} is generated by applying the estimated impact of PM_{10} on life expectancy (0.64 years per 10 $\mu g/m^3$) with population exposure estimates for PM_{10} taken from An et al. (23). They estimated exposure in each of 11 regions in China, and find that over 85% of China's population resides in locations where PM_{10} levels exceed the county's own Class II air quality standard of 70 $\mu g/m^3$. Additionally, PM_{10} levels in all 11 regions of China were found to exceed the county's Class I air quality standard of 40 $\mu g/m^3$. We calculate the number of life years that would be saved by reducing PM_{10} to the national standards levels (40 $\mu g/m^3$ or 70 $\mu g/m^3$) 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.

The estimates suggest that reducing PM_{10} to meet China's Class II air quality standard would save roughly 1.2 billion life years. Further reducing PM_{10} to the county's Class I standard would save an additional 2.5 billion life years, although there is greater uncertainty about the validity of applying our estimates at these lower PM_{10} concentrations.

The risks to life expectancy from particulate matter exposure are not confined to China. Some notable examples of countries with high concentrations 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 $\mu g/m^3$. This paper's results suggest that for most people in the world there is currently no greater environmental risk to health than airborne particulate matter.

References

1. *China's Disease Surveillance Points: 2004-2012*. Chinese Center for Disease Control and Prevention.
2. Chen Y, Ebenstein A, Greenstone M, and Li H (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.
3. *China's Environmental Yearbook: 1981-2012*. China Environmental Protection Agency and National Bureau of Statistics.
4. *Environmental Quality Annual Reports*: *2001-2012*. Ministry of Environmental Protection.
5. *World Meteorological Association.* Data from the World Weather Watch Program according to WMO Resolution 40 (Cg-XII). Available at http://www.ncdc.noaa.gov/. Accessed October 26, 2016.
6. Yu L, Lin X, Liu H, Shi J, Nong Q, Tang H, Mao Z (2015) Sex and Age Differences in Mortality in Southern China, 2004-2010. *International Journal of Environmental Research and Public Health* 12: 7886-7898.
7. Calonico S, Cattaneo MD, Titiunik R (2014b) Robust Data-Driven Inference in the Regression-Discontinuity Design. *Stata Journal* 14(4): 909-946.
8. Akaike, H (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19(6): 716–723.
9. Imbens G, Kalyanaraman K (2012) Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies* 79(3): 933-959.
10. Ludwig J, Miller DL (2007) Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics* 122(1): 159-208.
11. Calonico S, Cattaneo MD, Titiunik R (2014a) Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82: 2295–2326.
12. Lee DS, Lemieux T (2010) Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48(2): 281-355.
13. McCrary, J (2008) Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics* 142(2): 698-714.
14. Oster, E (2016) Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*.
15. Altonji JG, Elder TE, Taber CR (2005) Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113(1): 151-184.
16. The Editorial Board of Encyclopedia of Rivers and Lakes in China (2010) *Encyclopedia of Rivers and Lakes in China - Huai River Issue* (China Water & Power Press).
17. Almond D, Chen Y, Greenstone M, Li H (2009) Winter Heating or Clean Air? Unintended Impacts of China's Huai River Policy. *American Economic Review Papers and Proceedings* 99(2):184-190.
18. Makinen, J (2014) For Central Heat, China has a North-South Divide at Qin-Huai line. *Los Angeles Times*, Nov. 15. Available at http://www.latimes.com/world/asia/la-fg-china-heat-20141115-story.html. Accessed October 26, 2016.

19. Talhelm T, Zhang X, Oishi S, Shimin C, Duan D, Lan X, Kitayama S (2014) Large-Scale Psychological Differences within China Explained by Rice Versus Wheat Agriculture. *Science* 344(6184): 603-608.
20. Xiao Q, Ma Z, Li S, Liu Y (2015) The Impact of Winter Heating on Air Pollution in China. *PLOS ONE* 10(1): e0117311.
21. Ito K, Zhang S (2016) Willingness to Pay for Clean Air: Evidence from Air Purifier Markets in China. NBER working paper 22367. Available at http://www.nber.org/papers/w22367. Accessed October 26, 2016.
22. Century A (2013) Northern Chinese Brace for Smog Spike as Winter Heating Kicks In. *The New York Times*, Nov. 14. Available at http://sinosphere.blogs.nytimes.com/2013/11/14/northern-chinese-brace-for-smog-spike-as-winter-heating-kicks-in/. Accessed October 26, 2016.
23. An X, Hou Q, Li N, and Zhai S. 2013. Assessment of Human Exposure Level to PM10 in China. *Atmospheric Environment* 70: 376-386.

Table S1

Summary Statistics, Means and (Standard Deviations)

Variables	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})	119.5	90.8	28.8***	48.3***	41.7***
	(31.5)	(25.3)	(5.0)	(12.2)	(12.9)
Sulphur Dioxide (SO_2)	58.5	46.1	12.4***	16.1	6.4
	(25.4)	(21.7)	(4.0)	(10.6)	(12.5)
Nitrogen Oxide (NO_X)	37.7	34.0	3.8	-3.6	-4.1
	(12.0)	(13.8)	(2.4)	(6.8)	(7.0)
Panel B: Climate at the Disease Surveillance Points					
Heating Degrees	6.0	2.7	3.3***	1.0**	1.8
(1,000 degrees)	(1.8)	(1.5)	(0.3)	(0.5)	(1.3)
Cooling Degrees	1.2	2.1	-0.9***	-0.3	-0.6
(1,000 degrees)	(0.4)	(0.86)	(0.1)	(0.2)	(0.4)
Annual Rainfall	25.3	53.5	-28.2***	-3.0	-9.9**
	(7.8)	(10.6)	(1.7)	(3.4)	(4.1)
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	0.13	-0.06	-0.33
	(0.6)	(0.8)	(0.13)	(0.45)	(0.49)
Share of Minority (%)	7.6	8.8	-1.2	8.4	15.4*
	(16.6)	(20.8)	(2.7)	(5.4)	(8.5)
Coastal City Dummy (100km)	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	-0.93	-0.58	-2.9
	(6.4)	(6.6)	(1.17)	(4.06)	(4.1)
Heavy Drinker (%)	8.3	7.5	0.85	-2.8	-5.0
	(3.7)	(2.9)	(0.62)	(3.1)	(3.8)
Excessive Red Meat (%)	18.4	34.5	-16.0***	0.85	-5.9
	(12.5)	(17.2)	(2.6)	(5.40)	(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: Supply of Health Care and Water Pollution Measures					
Health Profession Employ. Rate	0.02	0.01	0.00	-0.01	0.00
	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)
Water Pollution Grade (1-6)	4.1	3.1	1.0***	0.83	-1.0
	(1.5)	(1.3)	(0.3)	(1.08)	(1.2)
Panel E: Summary of Observable Determinants of Life Expectancy in Panels B-D					
Predicted Life	76.2	76.2	-0.0	-1.3	-1.2
Expectancy (years)	(1.6)	(1.8)	(0.3)	(1.0)	(1.0)
P-value from Joint Test of Equality	-	-	<0.01***	<0.01***	0.23

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^o 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-D. 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. The optimal bandwidth is chosen separately for each variable. All results in columns (1) - (4) 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). Panel E reports differences in predicted life expectancy after controlling for the covariates listed in Panels B-D, and also reports the p-values from a joint test of equality between the north and south for the covariates in Panels B-D. The local linear joint test of equality uses the bandwidth selection method proposed by Imbens and Kalyanaraman (2012) and a uniform kernel. * 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 S2

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 S3

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***	0.16*	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*	0.08	0.13	0.13	0.16
(per 100,000, log)	(0.09)	(0.10)	(0.07)	(0.09)	(0.08)	(0.08)	(0.10)
Sub-Categories of Cardiorespiratory Mortality							
Heart	0.58***	0.45***	0.48***	0.29*	0.33*	0.32*	0.33*
	(0.19)	(0.15)	(0.18)	(0.15)	(0.17)	(0.17)	(0.17)
Stroke	0.49***	0.37**	0.46***	0.26	0.32**	0.32**	0.33*
	(0.15)	(0.18)	(0.14)	(0.16)	(0.15)	(0.16)	(0.17)
Lung Cancers	-0.17	0.05	-0.15	0.14	-0.32	-0.36	-0.43
	(0.17)	(0.13)	(0.15)	(0.12)	(0.28)	(0.29)	(0.31)
Respiratory Illnesses	0.44*	0.15	0.56**	0.26	0.51	0.49	0.44
	(0.23)	(0.27)	(0.24)	(0.25)	(0.33)	(0.33)	(0.36)
Sub-Categories of Non-Cardiorespiratory Mortality							
Cancers Other	-0.15	-0.10	-0.13	-0.15	-0.06	-0.06	-0.08
than Lung	(0.13)	(0.15)	(0.11)	(0.13)	(0.14)	(0.14)	(0.16)
Other Causes	0.19	0.16	0.26	0.23	0.33*	0.37*	0.44*
	(0.16)	(0.17)	(0.16)	(0.15)	(0.20)	(0.21)	(0.24)
Observations	154	154	79	79			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
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 covariates reported in Panels B-D of Table S1. 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 S4

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	1.7	-5.6	-9.7
	(7.2)	(5.6)	(1.2)	(5.8)	(7.1)
Excessive Red Meat (%)	23.0	41.2	-18.1***	1.0	-8.6*
	(14.4)	(18.2)	(2.8)	(6.2)	(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 (%)	0.60	0.65	-0.05	-0.12	-0.13
	(0.95)	(0.73)	(0.14)	(0.40)	(0.34)
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)

Notes : 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 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 column (5). * significant at 10% ** significant at 5% *** significant at 1%. Source: China Health Behavioral Survey (2010).

Table S5

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**	-2.2*	-3.5***	-2.2**	-1.8*	-2.6***
	(1.1)	(1.1)	(0.9)	(1.0)	(1.0)	(1.0)
Cardiorespiratory (per 100,000, log)	0.30**	0.23*	0.40***	0.34**	0.23*	0.33***
	(0.12)	(0.12)	(0.10)	(0.16)	(0.12)	(0.11)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}						
Life Expectancy at Birth (years)	-0.95	-0.65**	-0.71***	-0.79	-0.62*	-0.53***
	(0.58)	(0.32)	(0.25)	(0.50)	(0.33)	(0.19)
Cardiorespiratory (per 100,000, log)	0.12*	0.07**	0.08***	0.12**	0.08**	0.07***
	(0.06)	(0.03)	(0.03)	(0.06)	(0.04)	(0.03)
Observations	154	79		154	79	
Polynomial Function	3rd	Linear		3rd	Linear	
Sample	All	5 Degree		All	5 Degree	
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 S6

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	0.11	0.10***
	(0.08)	(0.04)
25	0.15	0.11***
	(0.10)	(0.04)
30	0.20*	0.14***
	(0.10)	(0.05)
35	0.17*	0.10**
	(0.09)	(0.04)
40	0.22**	0.12***
	(0.10)	(0.04)
45	0.21**	0.13***
	(0.10)	(0.04)
50	0.23**	0.13***
	(0.10)	(0.04)
55	0.22**	0.12***
	(0.10)	(0.04)
60	0.19**	0.12***
	(0.09)	(0.04)
65	0.15*	0.11***
	(0.08)	(0.04)
70	0.13*	0.09***
	(0.07)	(0.03)
75	0.10*	0.07***
	(0.06)	(0.03)
80	0.07	0.05**
	(0.05)	(0.02)
85+	0.04	0.03
	(0.05)	(0.02)

Notes : In column (1), we report the 2SLS IV estimate of PM_{10} on cardiorespiratory mortality using "North of Huai River" as the instrumental 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 S7

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})	48.3***	27.4***	49.9***	31.8***	41.7***	41.0***	40.2***
	(12.2)	(9.5)	(12.3)	(9.1)	(12.9)	(13.5)	(13.8)
Life Expectancy at Birth (years)	-3.3***	-2.4**	-3.5***	-2.2*	-3.1***	-3.2***	-3.3***
	(1.0)	(1.0)	(0.8)	(1.1)	(0.9)	(1.0)	(1.2)
Panel B: Cause-specific Mortality							
Cardiorespiratory (per 100,000, log)	0.42***	0.30**	0.39***	0.22*	0.37***	0.39***	0.40***
	(0.12)	(0.14)	(0.11)	(0.13)	(0.11)	(0.11)	(0.14)
Non-Cardiorespiratory (per 100,000, log)	0.07	0.06	0.12*	0.08	0.13	0.13	0.16
	(0.09)	(0.10)	(0.07)	(0.09)	(0.08)	(0.08)	(0.10)
Observations	154	154	79	79			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
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 covariates reported in Panels B-D of Table S1. 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. The number of observations (bandwidth) in columns (5)-(7) respectively are 132(9.3), 127(8.7), 110(7.3) for PM10, 93(5.9), 89(5.5), 78(4.6) for life expectancy, 77(4.3), 72(4.0), 62(3.4) for cardiorespiratory mortality and 96(6.3), 93(5.9), 79(5.0) for non-cardiorespiratory mortality. 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 S8

Comparing OLS and Regression Discontinuity Estimates of Particulate Matter's Impact 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)	-0.19**	-0.27***	-0.33***	-0.33**	-0.69**	-0.86*	-0.71***	-0.68**	-0.64***	-0.64***	-0.68**
	(0.08)	(0.09)	(0.08)	(0.13)	(0.32)	(0.51)	(0.25)	(0.31)	(0.22)	(0.22)	(0.27)
Cardiorespiratory (per 100,000, log)	0.03***	0.02***	0.04***	0.03***	0.09***	0.11*	0.08***	0.07**	0.08***	0.08***	0.08**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
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. 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 of Huai River" as the instrumental variable. The results in columns (5)-(8) also include a polynomial in degrees latitude from the Huai River (cubic in columns 5-6 and linear in columns 7-8) 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 the listed outcome, using the optimal bandwidth chosen for the listed outcome 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. The number of observations (bandwidth) in columns (9)-(11) are the same as those reported in Table S7 for the listed outcome. Controls include all the covariates listed in Panels B-D of Table S1. 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%.

Table S9

Robustness Checks of Choice of Functional Form for Latitude

	Full Sample					Restricted Sample (within 5^{o} latitude)				
	Linear	Quadratic	Cubic	Quartic	Quintic	Linear	Quadratic	Cubic	Quartic	Quintic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Life Expectancy	-1.15	-3.59***	-2.35**	-2.29*	-3.56***	-2.16*	-3.18**	-3.09*	-4.30	3.94
at Birth (years)	(0.87)	(0.91)	(1.05)	(1.22)	(1.35)	(1.11)	(1.40)	(1.60)	(3.15)	(4.27)
Akaike Info. Criterion	722.6	704.7	**701.8**	703.3	705.4	**336.1**	338.9	342.5	340.3	336.3
Particulate Matter (PM_{10})	16.63**	20.07**	27.36***	42.49***	30.90**	31.75***	27.39**	36.48*	18.19	0.63
	(7.17)	(8.26)	(9.53)	(11.15)	(13.79)	(9.09)	(12.02)	(19.67)	(27.27)	(45.11)
Akaike Info. Criterion	1395.1	**1395.0**	1398.1	1398.5	1400.9	**707.2**	710.8	714.1	717.2	720.8
Cardiorespiratory	0.15	0.33***	0.30**	0.38**	0.38**	0.22*	0.35*	0.31	0.18	-0.51
(per 100,000, log)	(0.10)	(0.11)	(0.14)	(0.16)	(0.19)	(0.13)	(0.18)	(0.20)	(0.29)	(0.45)
Akaike Info. Criterion	3.0	-5.9	-3.3	**-6.5**	-2.7	**-18.7**	-16.9	-14.1	-12.8	-13.3

Notes : The overall sample includes 154 DSP locations and the restricted sample includes 79 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 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 S10

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})	49.8***	50.3***	45.1**	38.1***	36.6***	48.1***
	(14.5)	(16.2)	(18.4)	(11.7)	(11.5)	(15.1)
Life Expectancy at Birth (years)	-3.3***	-3.5***	-4.0***	-3.4***	-3.6***	-3.7***
	(1.0)	(1.1)	(1.2)	(0.8)	(0.8)	(0.9)
Cardiorespiratory (per 100,000, log)	0.38***	0.41***	0.45***	0.30***	0.32***	0.34***
	(0.11)	(0.11)	(0.12)	(0.08)	(0.09)	(0.09)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}						
Life Expectancy at Birth (years)	-0.65***	-0.73***	-0.88**	-0.83***	-0.88***	-1.02***
	(0.24)	(0.27)	(0.37)	(0.25)	(0.28)	(0.34)
Cardiorespiratory (per 100,000, log)	0.08***	0.08***	0.10**	0.08***	0.08***	0.09***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
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 S7, 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 S11

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})	46.7***	44.9***	47.0*	55.1***	54.1***	53.4**	51.5***	50.9***	49.7***
	(15.3)	(16.8)	(24.1)	(15.9)	(18.0)	(20.8)	(15.8)	(16.4)	(17.7)
Life Expectancy at Birth (years)	-2.8**	-2.9**	-3.1**	-3.4***	-3.7***	-4.1***	-3.2***	-2.9**	-2.7**
	(1.2)	(1.3)	(1.5)	(1.2)	(1.3)	(1.4)	(1.1)	(1.2)	(1.3)
Cardiorespiratory (per 100,000, log)	0.42***	0.44***	0.44**	0.43***	0.45***	0.49***	0.39***	0.34***	0.30**
	(0.13)	(0.14)	(0.17)	(0.13)	(0.13)	(0.14)	(0.11)	(0.12)	(0.14)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}									
Life Expectancy at Birth (years)	-0.50*	-0.50*	-0.54	-0.67**	-0.78**	-0.96**	-0.55	-0.40	-0.23
	(0.27)	(0.28)	(0.33)	(0.29)	(0.33)	(0.44)	(0.35)	(0.39)	(0.51)
Cardiorespiratory (per 100,000, log)	0.08***	0.09**	0.08**	0.08***	0.09***	0.12***	0.07**	0.06	0.03
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Kernel	Triangle	Epanech.	Uniform	Triangle	Epanech.	Uniform	Triangle	Epanech.	Uniform

Notes : These results can be compared to the results presented in columns (5)-(7) of Table S7 and columns (9)-(11) of Table S8, 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 S12

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 (PM_{10})	27.4***	27.4***	27.2***	26.9***	28.0***	26.6***	27.4***	27.2***	27.0***	26.9***
	(9.5)	(9.5)	(9.5)	(9.5)	(9.6)	(9.6)	(9.5)	(10.1)	(10.1)	(10.1)
Panel B: Local Linear Regression										
Particulate Matter (PM_{10})	41.7***	42.2***	42.6***	42.6***	43.2***	39.5***	41.7***	45.2***	43.2***	43.3***
	(12.9)	(12.9)	(13.0)	(13.0)	(13.6)	(12.8)	(12.9)	(13.8)	(13.4)	(13.4)
Panel C: Local Linear Regression with a Fixed Bandwidth										
Particulate Matter (PM_{10})	41.7***	41.9***	42.0***	41.9***	39.3***	39.2***	41.7***	42.1***	42.8***	42.8***
	(12.9)	(12.8)	(12.8)	(12.8)	(12.7)	(12.7)	(12.9)	(13.2)	(13.2)	(13.2)

Notes : 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 of the distance, square, cube, and quartic, 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) (main specification), and columns (5)-(10) use the bandwidth chosen as optimal for estimation of column (7) (main specification). * significant at 10% ** significant at 5% *** significant at 1%.

Table S13

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

	<200KM (1)	<150KM (2)	<125KM (3)	<100KM (4)
Panel A: Polynomial Approach				
Particulate Matter (PM_{10})	16.5	27.4***	35.5***	-2.1
	(10.9)	(9.5)	(9.5)	(16.2)
Life Expectancy at Birth	-2.0*	-2.4**	-2.3**	-2.4**
	(1.1)	(1.0)	(1.1)	(1.1)
Panel B: Local Linear Regression				
Particulate Matter (PM_{10})	34.9***	41.7***	48.0***	16.5
	(12.4)	(12.9)	(14.0)	(20.3)
Life Expectancy at Birth	-3.1***	-3.1***	-3.1***	-3.2***
	(0.9)	(0.9)	(0.9)	(0.9)
Panel C: Local Linear Regression with a Fixed Bandwidth				
Particulate Matter (PM_{10})	35.1***	41.7***	44.9***	35.3**
	(13.1)	(12.9)	(12.7)	(14.5)
Life Expectancy at Birth	-3.0***	-3.1***	-3.1***	-3.1***
	(0.9)	(0.9)	(0.9)	(0.9)

Notes: These results are comparable to the results in Tables 2 and S7, but for different sample selection criteria. In the main results, all DSP locations within 150 kilometers are included in the sample (column 2). 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 for each sample, which is chosen by the method of Imbens and Kalayanaraman (2012). 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 (2) (main specification). * significant at 10% ** significant at 5% *** significant at 1%.

Table S14

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)
Life Expectancy at Birth (years)	-3.1***	-2.0*	-3.2***	-1.7	-3.1***	-3.1***	-2.6**
	(1.0)	(1.1)	(0.9)	(1.0)	(1.0)	(1.1)	(1.2)
Cardiorespiratory (per 100,000, log)	0.41***	0.29**	0.38***	0.22*	0.37***	0.40***	0.41***
	(0.12)	(0.14)	(0.11)	(0.12)	(0.11)	(0.11)	(0.13)
Observations	161	161	82	82			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes : These results were estimated in the same manner as those in Tables 2 and S7, 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 S15

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})	34.0***	17.6**	36.6***	24.2***	32.9***	33.6***	38.7***
	(9.7)	(6.8)	(10.0)	(6.8)	(11.7)	(12.3)	(12.9)
Life Expectancy at Birth (years)	-3.8***	-2.3*	-4.1***	-2.5**	-3.4***	-3.4***	-3.4***
	(1.2)	(1.3)	(1.0)	(1.2)	(1.0)	(1.0)	(1.1)
Cardiorespiratory (per 100,000, log)	0.52***	0.34**	0.50***	0.27*	0.42***	0.43***	0.40***
	(0.14)	(0.17)	(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 at Birth (years)	-1.12**	-1.29	-1.11***	-1.04**	-0.89***	-0.89***	-0.91**
	(0.52)	(0.91)	(0.40)	(0.45)	(0.31)	(0.32)	(0.38)
Cardiorespiratory (per 100,000, log)	0.15**	0.19	0.14***	0.11**	0.11***	0.12**	0.11**
	(0.06)	(0.12)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)
Observations	154	154	79	79			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
Controls	No	Yes	No	Yes			
Kernel					Triangle	Epanech.	Uniform

Notes : These results were estimated in the same manner as those in Tables 2-3 and S7-S8, 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 S16

Regression Discontinuity Estimates of the Huai River Policy and the Impact of 10 $\mu g/m^3$ 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***	30.7***	42.9***	30.9***	33.5**	31.6**	32.8*
	(11.8)	(10.7)	(10.6)	(8.3)	(15.3)	(15.8)	(16.9)
Life Expectancy	-4.1***	-2.8**	-3.6***	-2.7**	-3.2***	-3.1***	-2.8**
at Birth (years)	(1.1)	(1.2)	(0.9)	(1.1)	(0.9)	(1.0)	(1.2)
Cardiorespiratory	0.56***	0.41**	0.47***	0.34***	0.57***	0.59***	0.57***
(per 100,000, log)	(0.13)	(0.16)	(0.12)	(0.12)	(0.12)	(0.14)	(0.15)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}							
Life Expectancy	-0.83**	-0.90*	-0.84***	-0.89***	-0.99**	-0.92**	-0.87*
at Birth (years)	(0.34)	(0.51)	(0.26)	(0.34)	(0.41)	(0.41)	(0.45)
Cardiorespiratory	0.11***	0.13**	0.11***	0.11***	0.15**	0.16**	0.17*
(per 100,000, log)	(0.04)	(0.06)	(0.03)	(0.04)	(0.07)	(0.08)	(0.10)
Observations	124	124	68	68			
Polynomial Function	3rd	3rd	Linear	Linear			
Sample	All	All	5 Degree	5 Degree			
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 Tables 2 and S7, but without imputed values of PM_{10} included in the sample. 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 Tables 3 and S8, but without imputed values of PM_{10} from TSP included in the sample. * significant at 10% ** significant at 5% *** significant at 1%.

Table S17

Regression Discontinuity Estimates of the Change in Predicted and Residual Life Expectancy at the Huai River

	Local Linear Regressions (conventional)			Local Linear Regressions (bias-corrected)		
	IK	CCT	CV	IK	CCT	CV
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Life Expectancy at Birth (years)	-1.2	-0.8	-1.3	-0.7	-0.7	-1.1
	(1.0)	(1.3)	(0.9)	(1.5)	(1.6)	(1.4)
Residual Life Expectancy at Birth (years)	-2.0**	-2.1**	-2.0***	-2.4**	-2.5**	-2.1**
	(0.9)	(0.9)	(0.6)	(1.0)	(1.0)	(0.9)
Kernel	Triangle	Triangle	Triangle	Triangle	Triangle	Triangle

Notes : Each cell in the table represents a separate regression. Predicted life expectancy is generated from an OLS regression of life expectancy on the variables in Panels B-D of Table S1. Residual life expectancy is calculated as the differences between actual life expectancy and predicted life expectancy. The bandwidth selection methods in columns (1)-(3) are the Imbens-Kalyanaranam method (2012), the method proposed by Calonico et al. (2014a), and the cross-validation method (Ludwig and Miller 2007). In columns (4)-(6), we repeat this calculation using the bias-correction method proposed by Calonico et al. (2014), with the standard errors they recommend below the coefficient estimates. * significant at 10% ** significant at 5% *** significant at 1%.

Table S18

Robustness Checks of Choice of Functional Form for Distance from the Coast

	Adding a Linear Distance Term				Adding a Quadratic Distance Term			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Discontinuity at the Boundary								
Particulate Matter (PM_{10})	37.6***	25.0***	39.5***	34.8***	38.6***	26.5***	41.3***	35.7***
	(12.3)	(9.5)	(11.9)	(8.4)	(13.1)	(9.4)	(12.5)	(7.6)
Life Expectancy at Birth (years)	-2.8***	-2.4**	-2.8***	-2.3**	-2.3**	-2.2**	-2.7***	-2.3**
	(1.1)	(1.1)	(0.8)	(1.1)	(1.0)	(1.1)	(0.8)	(1.0)
Cardiorespiratory (per 100,000, log)	0.38***	0.29**	0.32***	0.25**	0.32***	0.27*	0.30***	0.25**
	(0.13)	(0.14)	(0.10)	(0.11)	(0.11)	(0.14)	(0.09)	(0.11)
Non-Cardiorespiratory (per 100,000, log)	0.06	0.07	0.10	0.08	0.02	0.05	0.09	0.08
	(0.09)	(0.10)	(0.07)	(0.09)	(0.09)	(0.10)	(0.07)	(0.09)
Panel B: Instrumental Variables Estimates of 10 $\mu g/m^3$ of PM_{10}								
Life Expectancy at Birth (years)	-0.76*	-0.96*	-0.71**	-0.67**	-0.60	-0.83*	-0.65**	-0.64***
	(0.42)	(0.57)	(0.32)	(0.28)	(0.37)	(0.49)	(0.28)	(0.24)
Cardiorespiratory (per 100,000, log)	0.10**	0.12*	0.08**	0.07**	0.08**	0.10*	0.07**	0.07***
	(0.05)	(0.07)	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.03)
Observations	154	154	79	79	154	154	79	79
Polynomial Function	3rd	3rd	Linear	Linear	3rd	3rd	Linear	Linear
Sample	All	All	5 Degree	5 Degree	All	All	5 Degree	5 Degree
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes : In this table, we report results where we add a linear (columns 1-4) or quadratic (columns 5-8) control in distance (in meters) from the coast to our standard set of controls for the parametric analysis. The results in Panel A report the discontinuity in life expectancy and cardiorespiratory mortality in the same manner as those reported in columns (1)-(4) of Table S7, but with the added control variable. 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 S8, but with the added control variable. * significant at 10% ** significant at 5% *** significant at 1%.

Table S19

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

Variables	North	South	Difference in Means	Adjusted Difference (polynomial)	Adjusted Difference (local linear)
	(1)	(2)	(3)	(4)	(5)
Panel A: Health Policy Variables					
Number of Hospitals	0.06	0.05	0.01	-0.04*	-0.02
(per 10,000 residents)	(0.1)	(0.02)	(0.01)	(0.02)	(0.02)
Number of Physicians	16.7	15.1	1.6	-10.7***	-5.4
(per 10,000 residents)	(7.5)	(6.5)	(1.4)	(3.5)	(5.3)
Observations			125	125	
Panel B: Measures of Water Pollution and Wastewater Treatment Policies					
Wastewater Treatment Rate (%)	71.0	60.6	10.5***	-12.9**	-5.9
	(11.4)	(14.7)	(2.6)	(5.6)	(8.0)
Solid Waste Treatment Rate (%)	80.9	79.7	1.2	-5.6	-4.6
	(15.0)	(21.4)	(3.9)	(9.1)	(7.2)
Observations			125	125	
Panel C: Summary of Observed Differences Using All Available Covariates in Table 1 and Table S16					
Predicted Life Expectancy (years)	76.1	76.5	-0.4	-1.2*	-0.3
	(1.4)	(1.4)	(0.3)	(0.6)	(0.6)
P-value from Joint Test of Equality	-	-	<0.01***	<0.01***	0.28

Notes: The covariates listed in Panels A and B are collected from multiple sources. When DSP county-level statistics are available, we use county-level statistics; otherwise, DSP prefectural-level statistics are used. In Panel A, number of hospitals and number of physicians in hospitals are taken from China's Prefectural Statistical Yearbook in 2005 at the DSP prefectural-level. In Panel B, industrial wastewater treatment rate and solid waste treatment rate at the DSP prefectural-level are collected from the CEIC China Database (http://www.ceicdata.com/en/countries/china) from 2004 to 2012. All results in columns (1)-(4) are weighted by the population at the DSP location. 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. Panel C reports differences in predicted life expectancy in the same manner as Tables 1 and S1, and p-values from joint tests of equality between the north and south using the covariates listed in Panels A and B in addition to the variables listed in Panels B-D of Table S1. These joint tests incorporate all 154 DSPs for the covariates listed in Table S1 and the subsample of 125 DSPs for the covariates listed in Panels A and B of this table. The local linear joint test of equality uses the bandwidth selection method proposed by Imbens and Kalyanaraman (2012) and a uniform kernel. Note that these variables are not included as covariates in Tables 1 and S1 since they are not available for all DSP locations.

Table S20

Regression Discontinuity Estimates of Change in Log Population at the Huai River

	Polynomials				Local Linear					
	Census Sample	DSP Sample	Census Sample	DSP Sample		Census Sample			DSP Sample	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Population (Log)	-0.08	-0.67	-0.04	-0.50	-0.01	-0.01	-0.00	0.04	-0.01	0.03
	(0.12)	(0.49)	(0.10)	(0.41)	(0.09)	(0.09)	(0.09)	(0.61)	(0.60)	(0.63)
Observations	2,852	154	1,516	79	1,955	1,865	1,630	69	66	52
Polynomial Function	3rd	3rd	Linear	Linear						
Sample	All	All	5 Degree	5 Degree	All	All	All	All	All	All
Kernel					Triangle	Epanech.	Uniform	Triangle	Epanech.	Uniform
Bandwidth					IK	IK	IK	IK	IK	IK

Notes : The results report the estimated discontintuity in log population using either all counties in China's 2010 census (columns 1, 3, 5-7) and the counties that include DSP locations (columns 2, 4, 8-10), respectively. The parametric and non-parametric specifications are estimated in a manner similar to those in the main paper (e.g. Table 2).

Table S21

Regression Discontinuity Estimates of Out-Migration and Migration-Weighted Particulate Matter (PM_{10}) Exposure at the Huai River

	Total Individuals	Total Migrants	Out-Migration Rates Local Linear Discontinuity Estimates			Average % Life Spent in Hukou		Migration-Weighted Particulate Matter (PM_{10}) Local Linear Discontinuity Estimates		
						All Individuals	Migrants			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Assuming Entire Life is Spent at Hukou (Table 2)										
Full Sample						100%	100%	41.7***	41.0***	40.2***
								(12.9)	(13.5)	(13.8)
Panel B: Allowing for Migration (2005 Census and Chinese CDC Mortality Extract)										
Full Sample	161,687	14,590	0.04	0.04	0.05	97.0%	81.7%	41.3***	40.6***	39.9***
	100.0%	9.0%	(0.05)	(0.05)	(0.05)			(12.9)	(13.4)	(13.7)
Ages 0 - 9	17,313	991	-0.01	-0.02	-0.03	91.8%	28.0%	41.1***	40.5***	40.0***
	100.0%	5.7%	(0.04)	(0.05)	(0.05)			(12.8)	(13.3)	(13.5)
Ages 10 - 19	26,569	2,098	0.02	0.03	0.03	96.3%	76.6%	41.5***	40.9***	40.1***
	100.0%	7.9%	(0.04)	(0.04)	(0.05)			(12.9)	(13.4)	(13.7)
Ages 20 - 29	21,172	3,916	0.04	0.04	0.05	95.8%	85.2%	40.9***	40.2***	39.6***
	100.0%	18.5%	(0.10)	(0.10)	(0.10)			(12.8)	(13.4)	(13.6)
Ages 30 - 39	29,834	3,466	0.08	0.08	0.07	97.1%	85.5%	41.3***	40.6***	40.0***
	100.0%	11.6%	(0.09)	(0.09)	(0.10)			(12.9)	(13.4)	(13.7)
Ages 40 - 49	26,418	1,999	0.04	0.04	0.02	98.2%	88.1%	41.5***	40.8***	40.0***
	100.0%	7.6%	(0.04)	(0.05)	(0.06)			(12.9)	(13.4)	(13.7)
Ages 50 - 59	19,778	1,149	-0.02	-0.01	-0.02	98.9%	90.3%	41.6***	40.9***	40.1***
	100.0%	5.8%	(0.02)	(0.03)	(0.03)			(12.9)	(13.4)	(13.7)
Ages 60 +	20,603	971	-0.01	-0.01	-0.01	99.3%	92.0%	41.6***	40.9***	40.1***
	100.0%	4.7%	(0.02)	(0.02)	(0.02)			(12.9)	(13.4)	(13.7)
Kernel			Triangle	Epanech.	Uniform			Triangle	Epanech.	Uniform

Notes: This table reports the results from an analysis of individuals in the 2005 Census microdata whose *hukou* is in the same county as one of the 154 DSP locations. Columns (1) and (2) report the total number of such respondents and the subset which are classified as migrants, respectively. A migrants is defined as someone who does not reside in their *hukou* county at the time of the Census. Percents are expressed as a percent of total individuals. Columns (3)-(5) report local linear discontinuity estimates in county-level out-migration rates at the Huai River. Column (6) reports the average percent of life spent living in the *hukou county*. This proportion is 100% for non-migrants and calculated for migrants as 100%*(1-(years lived outside hukou/age in 2005)). Column (7) reports the same estimate for migrants only. In columns (8) – (10) of Panel B, each DSP is assigned a PM_{10} value equal to the average lifetime exposure across all individuals whose hukou is a county associated with a DSP. Individual lifetime exposure is calculated as a weighted-average of measured PM_{10} concentrations at their hukou and 2005 residence PM_{10}, where weights are equal to (1-(years lived outside hukou/age in 2005)) and (years lived outside hukou/age in 2005), respectively. Local linear regressions are then conducted based on these adjusted PM_{10} averages for the DSP locations. The bandwidth used for each kernel type in columns (8)-(10) of Panel B is fixed to the corresponding bandwidth used in Panel A. Throughout this analysis, migrants who reported moving to their current residence "more than 6 years ago", which is the maximum number of years they could enter into the Census form, are assumed to have moved 10 years ago, and results are robust to varying this assumption to different periods of time (e.g. 15, 20, and 25 years).

Table S22

Coefficient Robustness to Unobservable Selection Based on Oster (2016)

Treatment Variable	Baseline Effect (Std. Error), [R^2]	Controlled Effect (Std. Error), [R^2]	R_{max}	δ for $\beta = 0$ Given R_{max}	Identified Set
	(1)	(2)	(3)	(4)	(5)
Panel A: Full Sample					
Particulate Matter (PM_{10})	48.3*** (12.2) [0.38]	27.4*** (9.5) [0.61]	0.79	1.31	[7.3, 27.4]
Life Expectancy at Birth (years)	-3.3*** (1) [0.21]	-2.4** (1) [0.47]	0.62	3.0	[-2.4, -1.7]
Cardiorespiratory Mortality (per 100,000, log)	0.42*** (0.12) [0.18]	0.3** (0.14) [0.45]	0.58	3.19	[0.2, 0.3]
Non-Cardiorespiratory Mortality (per 100,000, log)	0.07 (0.09) [0.28]	0.06 (0.1) [0.45]	0.59	2.83	[0, 0.1]
Panel B: 5 Degree Latitude Restricted Sample					
Particulate Matter (PM_{10})	49.9*** (12.3) [0.2]	31.8*** (9.1) [0.65]	0.84	2.7	[22.3, 31.8]
Life Expectancy at Birth (years)	-3.5*** (0.8) [0.12]	-2.2* (1.1) [0.55]	0.71	2.8	[-2.2, -1.5]
Cardiorespiratory Mortality (per 100,000, log)	0.39*** (0.11) [0.16]	0.22* (0.13) [0.57]	0.75	2.28	[0.1, 0.2]
Non-Cardiorespiratory Mortality (per 100,000, log)	0.12* (0.07) [0.03]	0.08 (0.09) [0.39]	0.50	4.8	[0.07, 0.08]

Notes: This table shows the results for the coefficient bounding exercise on unobservable selection presented in Oster (2016). Column (1) reports OLS estimates of the coefficient on a "North of the Huai River" dummy after controlling for a polynomial (cubic in Panel A, linear in Panel B) in distance from the Huai River interacted with a North dummy. This is considered the "uncontrolled" or parsimonious regression without any additional explanatory variables in our implementation of Oster's estimation method. Column (2) reports the same coefficent after adding in the covariates listed in Panels B-D of Table S1. These two columns match the output reported in Table S7 and column (2) matches the output reported in Table 2. Column (3) reports the estimated maximum R^2 from a hypothetical regression of the outcome on both observed and unobserved controls, which is calculated as 1.3 times the R^2 from the regression with observed controls. The 1.3 multiple is selected based on the analysis in Oster (2016), which identified it as the value which allows 90% of results to survive in a sample of randomized papers published in a collection of top economics journals between 2008 and 2013. Column (4) reports the relative importance of unobservables compared to observables necessary to zero out the effect listed in column (2), assuming an R^2 from a hypothetical regression including all observable and unobservable covariates equal to R_{max}. Column (5) reports the estimated bounds on the "North of Huai River" coefficient using the controlled effect from column (2) and a recalculated effect assuming $\delta = 1$ and R_{max} from column (3).

Table S23

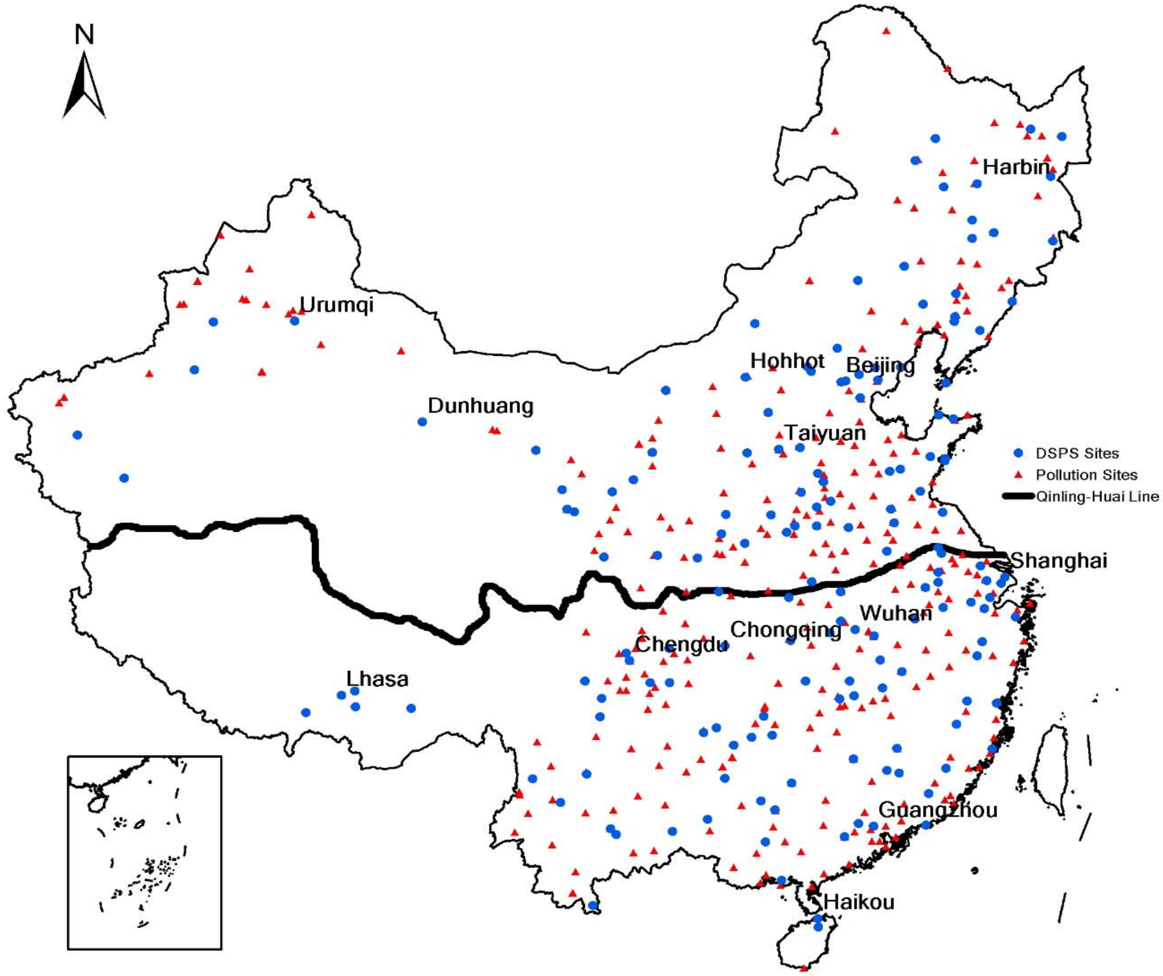
DSP Locations and their Home Heating Policies

Municipality	Name of DSP Locations	N/S	Winter Heating	Sources
Nanjing	Pukou District, Jiangsu	S	No	http://njcb.jschina.com.cn/mp3/html/2015-11/18/content_1337139.htm
Xuzhou	Yunlong District, Jiangsu	N	Yes	http://js.people.com.cn/html/2013/01/09/199449.html
Suzhou	Wuzhong District, Jiangsu	S	No	http://toutiao.com/i6212436723538182657/
Suzhou	Zhangjiagang City, Jiangsu	S	No	http://toutiao.com/i6212436723538182657/
Huai'an	Jinhu County, Jiangsu	S	No	http://www.js.xinhuanet.com/2014-12/12/c_1113620269.htm
Yancheng	**Xiangshui County, Jiangsu**	**N**	**No**	**http://www.js.xinhuanet.com/2014-12/12/c_1113620269.htm**
Ma'anshan	Yushan District, Anhui	S	No	http://www.masff.com/3g/newsandactive/content2.aspx?id=61061
Anqing	Daguan District, Anhui	S	No	http://wlwz.anqing.gov.cn/latter_view.php?latterid=93305&areaid=1
Chuzhou	Tianchang City, Anhui	S	No	http://ah.anhuinews.com/system/2011/01/17/003674548.shtml
Chaohu	Juchao District, Anhui	S	No	http://ah.anhuinews.com/system/2013/01/09/005398327.shtml
Bozhou	Mengcheng County, Anhui	N	Yes	http://www.bozhou.cn/2013/1225/123920.shtml
Xuancheng	Jing County, Anhui	S	No	http://www.xuanfang.org/news/374095.html
Zhengzhou	Zhongyuan District, Henan	N	Yes	http://news.shangdu.com/guanzhu/081110/
Luoyang	Jili District, Henan	N	Yes	http://news.lyd.com.cn/system/2014/10/10/010349104.shtml
Luoyang	Xin'an County, Henan	N	Yes	http://news.lyd.com.cn/system/2014/10/10/010349104.shtml
Anyang	Hua County, Henan	N	Yes	http://anyang.ljia.net/a/20141103/23371535.html
Xinxiang	Huixian City, Henan	N	Yes	http://www.henan.gov.cn/zwgk/system/2014/11/17/010508068.shtml
Nanyang	Tanghe County, Henan	N	Yes	http://henan.sina.com.cn/nanyang/m/2015-11-09/080640014.html
Shangqiu	Sui County, Henan	N	Yes	http://www.wutuojia.com/zixun/article/7026.html
Xinyang	Shihe District, Henan	S	No	http://www.ha.chinanews.com.cn/GNnews/1/2014/11/26/336182.shtml
Baoji	Mei County, Shaanxi	N	Yes	http://blog.sina.com.cn/s/blog_dfd97ccb0102vz0v.html
Weinan	Huayin City, Shaanxi	N	Yes	http://news.weinan.fang.com/2012-11-11/8947621.htm
Yan'an	Luochuan County, Shaanxi	N	Yes	http://news.cnwest.com/content/2015-11/02/content_13292829.htm
Ankang	Hanyin County, Shaanxi	S	No	http://news.sina.com.cn/c/2004-11-25/07394340460s.shtml

Notes : In this table, we report whether or not the municipality of DSP locations close to the Huai River line provides centralized winter heating. We examine compliance with the national policy for all the DSP locations in the four provinces (Jiangsu, Anhui, Henan, and Shaanxi) which the Huai River line cuts through. The only non-complier is Yancheng city, in bold. There is no non-compliance concern for DSP locations far away from the Huai River line: all of them are compliers.

Figure S1

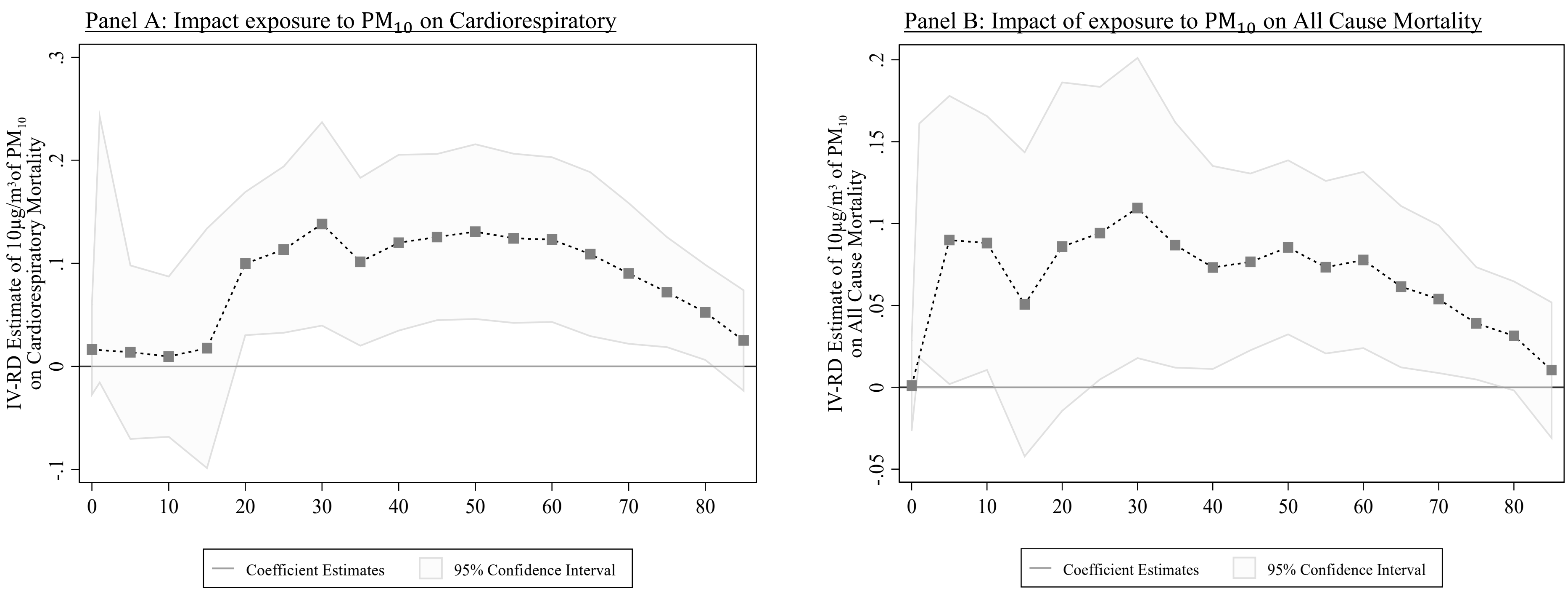
Location of DSP Sites and Pollution Monitoring Stations



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

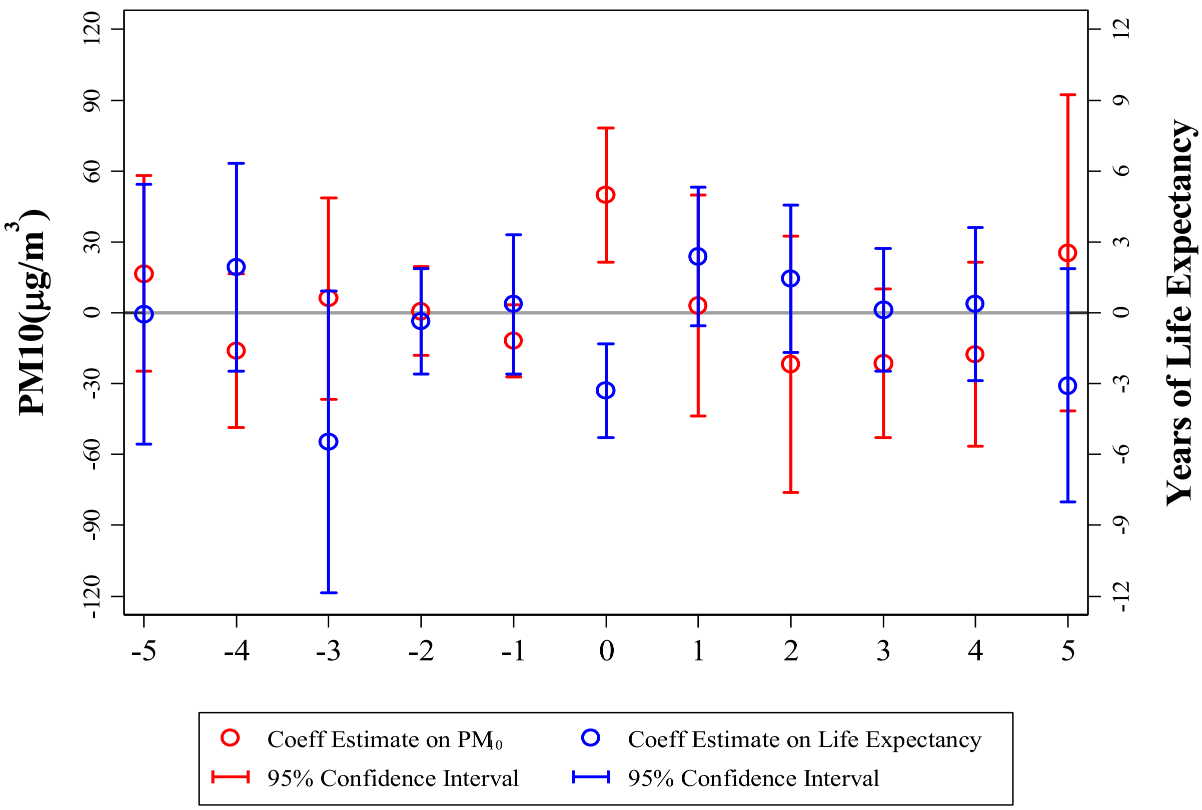
Instrumental Variables Estimates of the Impact of an Additional 10 μg/m^3 Exposure to PM_{10} on Cardiorespiratory and All Cause Mortality by Age Group



Notes : 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 (Panel A) and all cause mortality (Panel B) (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 using the optimal bandwidth chosen for mortality 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).