Supplementary Material for The Effect of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 U.S. Counties for the Period 2000 to 2007

Section A: Life Expectancy Calculation in Dataset 1

County–level life expectancies were calculated by applying a mixed-effects spatial Poisson model to mortality data from the National Center for Health Statistics (NCHS) and population data from the U.S. Census to obtain robust estimates of the number of deaths in each county.[1] These estimated counts are then used to calculate county life expectancies using standard life table techniques, which we discuss in more detail below. Specifically, the model is given by:

$$\log[E(y_{rjt})] = \beta_0 + \beta_1 income_{jt} + \beta_2 education_j + \beta_4 \sigma_j^{post} + \beta_4 race + \gamma_j t + \mu_j$$

"where y_{rjt} is the death count for race r within county j in year t; $income_{jt}$ is county per-capita income for year t; $education_j$ is the percent of adults within county j having completed high school in the 2000 census data; and *race* is a dummy variable for three race groups (white, black, and other). σ_j^{post} is a geospatial component, calculated as the average of the posterior mode of the county random intercept for counties adjacent to county j to account for residual spatial patterns, the values of which were derived from first running as a prior step the same model above without the geospatial component to derive the posterior values of the county random effect. Similarly, μ_j is the posterior value of the county random intercept. Lastly, γ_j is a random slope on time, t, for each county." These estimated counts, which are more robust due to the borrowing of information across space and time, are then used to calculate county life expectancies using standard life table techniques. Briefly, the estimated death counts within each age stratum, or interval, are divided by the mid-year population in that interval, providing us with an age-specific death rate for each age interval. This age-specific death rate is then used together with a term that estimates the average number of years lived by persons who die in each particular age interval – a term which, when expressed as a fraction, is often estimated as ½ – to estimate the probability of dying within each age interval. Using this age interval-specific probability of dying, one can then project the mortality experience of a hypothetical cohort that experiences the same age-specific probabilities of dying as our observed population. This is done for all counties. More details on calculating life expectancy in the life table setting are available in various texts.[2-4] Note then, that the hierarchical model only provides more robust estimates of the death counts within each age interval for each county, which often times can be small and unstable for smaller counties. The actual process of calculating life expectancy in a life table setting does not change.

Section B: Datasets 2 and 3

B1: Variables and Data Sources

The variables in Dataset 2 (211 counties, 1980 – 2000) were: life expectancy, $PM_{2.5}$, per capita income, population, and proportions of the population who were high school graduates, who had not lived in that county 5 years earlier (5-year in-migration), who had an urban residence, and who reported they were white, black, or Hispanic. Age-standardized death rates for lung cancer and chronic obstructive pulmonary disease (COPD) were included in the dataset to account for smoking prevalence in the population. Each variable had a value for both 1980 and 2000. This data and its sources are described in more detail elsewhere.[5]

The variables in Dataset 3 (211 counties, 2007) were the same as those in Dataset 2, and the data sources for these data were identical to those of the 545 county dataset with two exceptions: 1) proportion of the population that did not live in the county 5 years earlier, and 2) proportion of the population with an urban residence; these two variables are only available from the decennial Census, so we used year 2000 values as a proxy for 2007. Additionally, as in Dataset 1, due to the availability of NCHS data, 2005 death rates were used as a proxy for 2007.

Yearly average $PM_{2.5}$ for 2007 was calculated at the MSA level by averaging the yearly county-level $PM_{2.5}$ readings for all counties in a given MSA. We calculated both population-weighted and non-weighted averages. Combining Datasets 2 and 3 enabled us to extend the analysis in Pope et al[5] to the periods 1980 – 2007, and 2000 – 2007. When we exclusively analyzed the 211 counties in Pope et al,[5] regardless of the time period, we did so with $PM_{2.5}$ calculated at the MSA level for all counties, consistent with the original analysis. We also note that per capita income in Dataset 2 was obtained from the U.S. Census, while per capita income in Dataset 3 was obtained from the Bureau of Economic Analysis (BEA). BEA per capita income estimates were consistently higher than Census estimates, thus, for consistency we also obtained BEA per capita income estimates for 1980 and 2000, and results for the 211 counties from Pope et al[5] for the periods 1980 – 2007 and 2000 – 2007 were obtained using BEA per capita income estimates. When re-analyzing Dataset 2 (1980 – 2000), we obtained results using Census per capita income estimates as in Pope et al[5] and also using BEA per capita income estimates. When adjusting for changes in per capita income, the effect of $PM_{2.5}$ on life expectancy was not sensitive to the choice of the income variable.

We additionally note that the estimated counts used to calculate life expectancy in Dataset 3 for the year 2007 (described above in Section A) were calculated using a slightly different method than the one used to calculate the estimated death counts used to calculate life expectancy for the 211 counties in Pope et al[5] for the periods 1980 and 2000 (Dataset 2).[1,6] However, the two methods are only substantially different in locations with very small populations (pop < 7000),[1] which is not the case here as all of these counties are in metropolitan areas, and no counties had a year 2007 population less than 22,000. For the year 2000, where we have life expectancy estimates for the 211 counties using both methods, the correlation between the two was greater than 0.98.

B2: Results

eTables 1a and 1b report the summary statistics for the 211 counties for the periods 1980 – 2007 and 2000 – 2007, respectively. eTables 2a and 2b summarize the estimated regression coefficients for the association between changes in $PM_{2.5}$ and changes in life expectancy for those same counties for the periods 1980 –

2007 and 2000 – 2007, respectively. For the period from 1980 to 2007, decreases in $PM_{2.5}$ were significantly associated with increases in life expectancy in all models. The estimates were consistent with those reported previously. Across all models, estimates for the effect of a 10μg/m^3 decrease in $PM_{2.5}$ on life expectancy ranged from 0.56±0.19 years to 1.13±0.36 years, while for the period from 1980 to 2000,[5] estimates across models for the same effect ranged from 0.55±0.24 years to 1.01±0.25 years. In Model 5 of eTable 2a, a decrease of 10μg/m^3 of $PM_{2.5}$ was associated with an estimated 0.56±0.19 years of increased life expectancy.

For the period from 2000 to 2007 (eTable 2b), decreases in $PM_{2.5}$ were significantly associated with increases in life expectancies in all but the simplest model that includes only $PM_{2.5}$ as a predictor (Model 1) in the counties from Pope et al.[5] Here, however, our estimates were consistently much higher than those from the period 1980 – 2000, as the estimates for the effect of a 10μg/m^3 decrease in $PM_{2.5}$ on life expectancy ranged from an additional 0.66±0.34 years to an additional 1.60±0.53 years. Excluding the estimate from Model 1, the estimates were all greater than or equal to 1 (1.00±0.32 to 1.58±0.55). Similarly, when we restricted our county-level analysis to include only the 113 counties that were also in Pope et al,[5] the effect of a 10μg/m^3 decrease in $PM_{2.5}$ on life expectancy was associated with an average of 1.34±0.35 additional years of life expectancy for a model with the same covariate pattern as that of Model 3 in eTable 2b. For the 432 counties not in Pope et al,[5] the effect of a 10μg/m^3 decrease in $PM_{2.5}$ on life expectancy was only 0.07±0.16 additional years.

For the period from 1980 to 2000, while our reanalysis confirmed the original results, we additionally fitted separate models for males and females. As with our main analyses, we observed gender differences for the period 1980 to 2000. For the model corresponding to Model 4 in Pope et al,[5] a 10μg/m^3 decrease in $PM_{2.5}$ was associated with an increase of 0.29(±0.22) years for males, and an increase of 1.00(±0.27) years for females. This difference was statistically significant ($p = 0.012$), and again suggests that there may be different $PM_{2.5}$/mortality associations in males versus females.

Section C: Results from Cross-sectional Analyses

For Dataset 1 (545 counties, 2000 – 2007), simple models including only $PM_{2.5}$ as a predictor estimated an increase in life expectancy of 2.09±0.19 and 2.63±0.28 years for a $10\mu g/m^3$ decrease in $PM_{2.5}$ in 2000 and 2007, respectively ($p < 0.001$ for both). Models controlling for population, per capita income, proportion of the population that is black or Hispanic, and death rates for lung cancer and COPD showed markedly smaller associations, with $PM_{2.5}$ estimates of 0.33±0.11 ($p = 0.005$) and 0.39±0.17 ($p = 0.021$) years for 2000 and 2007, respectively.

Similarly, a cross-sectional analysis of the year 2007 of the 211 counties in Pope et al[5] gave a simple estimate of 2.80±0.64 ($p < 0.001$). $PM_{2.5}$ effects were attenuated when controlling for population, per capita income, proportion of the population that was black or Hispanic, and death rates for lung cancer and COPD (estimate = 0.30±0.38; $p = 0.44$). Cross-sectional analyses for the 211 counties for the years 1980 and 2000 were no different than originally reported.[5] In all datasets, however, additionally controlling for proportion of the population who are high school graduates shrank estimates of the effect of $PM_{2.5}$ towards zero, and yielded much higher p-values ($0.200 < p < 0.946$).

eTable 1a: Summary Characteristics of the 211 Counties Analyzed for the Period 1980 – 2007

Variable	Mean Value (±SD)
Life Expectancy (yr)	
1980	74.32±1.52
2007	78.12±1.86
Change	3.80±1.21
$PM_{2.5}$ ($\mu g/m^3$)	
1980	20.62±4.36
2007	12.44±2.17
Reduction	8.18±3.00
Per Capita Income (in thousands of $)	
1980	20.39±3.65
2007	33.64±8.58
Change	13.25±5.68
Population (in hundreds of thousands)	
1980	3.83±8.47
2007	5.17±10.49
Change	1.34±2.91
5-Year In-migration (prop. of population)	
1980	0.25±0.10
2007*	0.24±0.08
Change	-0.01±0.06
Urban residence (prop. of population)	
1980	0.58±0.33
2007*	0.78±0.22
Change	0.20±0.18
High School Graduates (prop. of population)	
1980	0.68±0.11
2007	0.87±0.05
Change	0.19±0.14
Black Population (prop. of population)	
1980	0.097±0.118
2007	0.116±0.128
Change	0.019±0.069
Hispanic Population (prop. of population)	
1980	0.035±0.072
2007	0.088±0.101
Change	0.053±0.053
Deaths from Lung Cancer (no./10,000 pop.)	
1980	14.38±2.95
2007†	15.25±3.37
Change	0.87±3.27
Deaths from COPD (no./10,000 pop.)	
1980	7.92±1.85
2007†	11.99±3.24
Change	4.07±3.13

* Values from the 2000 Census are used as a proxy for 2007.
† 2005 death rates were used a proxy for 2007 death rates.

eTable 1b: Summary Characteristics of the 211 Counties Analyzed for the Period 2000 – 2007

Variable	Mean Value (±SD)
Life Expectancy (yr)	
2000	77.04±1.82
2007	78.12±1.86
Change	1.08±0.64
$PM_{2.5}$ ($\mu g/m^3$)	
2000	14.10±2.86
2007	12.44±2.17
Reduction	1.67±1.25
Per Capita Income (in thousands of $)	
2000	31.69±8.01
2007	33.64±8.58
Change	1.95±2.70
Population (in hundreds of thousands)	
2000	4.82±10.13
2007	5.17±10.49
Change	0.35±0.79
High School Graduates (prop of population)	
2000	0.869±0.050
2007	0.875±0.046
Change	0.006±0.015
Black Population (prop of population)	
2000	0.115±0.130
2007	0.116±0.128
Change	0.001±0.028
Hispanic Population (prop of population)	
2000	0.068±0.093
2007	0.088±0.101
Change	0.019±0.016
Deaths from Lung Cancer (no./10,000 pop.)*	
2000	16.73±3.27
2007	15.25±3.37
Change	-1.48±1.96
Deaths from COPD (no./10,000 pop.)*	
2000	12.37±2.71
2007	11.99±3.24
Change	-0.38±2.15

* 2005 death rates were used as a proxy for 2007 death rates.

eTable 2a: Results of Selected Regression Models for Extended Analysis of 211 Counties, 1980 – 2007*

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6†	Model 7‡	Model 8‡
Intercept	3.01±0.32§	1.38±0.23§	2.58±0.34§	2.56±0.35§	2.58±0.35§	2.62±0.42§	1.69±0.70||	1.27±0.68
Reduction in $PM_{2.5}$ (10μg/m^3)	0.99±0.39||	0.83±0.20§	0.57±0.19§	0.64±0.18§	0.56±0.19§	0.60±0.22§	1.02±0.36§	1.13±0.36§
Change in income (in thousands of $)	--	0.10±0.02§	0.08±0.01§	0.09±0.01§	0.09±0.01§	0.08±0.01§	0.15±0.03§	0.13±0.02§
Change in population (in hundreds of thousands)	--	0.09±0.04||	0.06±0.03||	0.07±0.02§	0.06±0.02||	0.06±0.03||	0.04±0.03	0.03±0.03
Change in 5-yr In-migration (proportion of population)	--	3.27±1.13§	3.66±1.00§	--	3.57±0.89§	5.45±1.80§	3.95±2.35	4.97±2.27||
Change in high-school graduates (proportion of population)	--	1.25±0.83	0.19±0.65	--	--	--	-1.79±1.30	--
Change in urban residence (proportion of population)	--	-0.31±0.30	-0.10±0.29	--	--	--	-2.80±2.64	--
Change in black population (proportion of population)	--	-1.97±0.76||	-3.05±0.66§	-3.29±0.74§	-3.02±0.62§	-3.77±1.21§	-8.79±2.74§	-8.80±2.79§
Change in Hispanic population (proportion of population)	--	2.66±1.48	1.86±1.12	--	1.85±1.11	1.97±1.28	2.32±2.57	4.13±2.13
Change in lung cancer mortality rate (no./10,000 population)	--	--	-0.07±0.03||	-0.06±0.03	-0.07±0.03||	-0.11±0.03§	-0.09±0.05	-0.10±0.05||
Change in COPD mortality rate (no./10,000 population)	--	--	-0.10±0.04||	-0.10±0.03§	-0.10±0.04||	-0.06±0.04	-0.01±0.07	0.01±0.07
No. of county units	211	211	211	211	211	127	51	51

* Plus-minus values are regression coefficients ±SE. COPD denotes chronic obstructive pulmonary disease.
† This model included only counties with a 1986 population ≥ 100,000.
‡ This model included only counties with the largest 1986 population in the MSA.
§ Indicates $P < 0.01$.
|| Indicates $P < 0.05$.

eTable 2b: Results of Selected Regression Models for Restricted Analysis of 211 Counties, 2000 – 2007*

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6†	Model 7‡	Model 8‡
Intercept	0.98±0.05§	0.84±0.10§	0.83±0.09§	0.79±0.07§	0.83±0.05§	0.83±0.08§	0.94±0.19§	0.81±0.12§
Reduction in $PM_{2.5}$ (10μg/m^3)	0.66±0.34	1.28±0.36§	1.09±0.35§	1.05±0.36§	1.00±0.32§	1.58±0.48§	1.55±0.54§	1.60±0.53§
Change in income (in thousands of $)	--	0.02±0.02	0.02±0.02	0.02±0.02	--	--	-0.03±0.03	--
Change in population (in hundreds of thousands)	--	0.14±0.09	0.13±0.08	0.13±0.07	0.13±0.07	0.08±0.06	0.05±0.06	0.08±0.05
Change in high-school graduates (proportion of population)	--	-4.90±3.15	-3.47±3.03	--	--	--	-12.35±7.05	--
Change in black population (proportion of population)	--	-10.06±2.28§	-9.69±2.14§	-9.67±2.07§	-10.28±2.43§	-20.35±3.90§	-33.15±5.46§	-29.32±5.13§
Change in Hispanic population (proportion of population)	--	0.19±2.75	-1.26±3.03	--	--	--	-3.09±5.88	--
Change in lung cancer mortality rate (no./10,000 population)	--	--	-0.03±0.02	-0.03±0.02	-0.03±0.02	-0.00±0.05	0.04±0.05	0.03±0.05
Change in COPD mortality rate (no./10,000 population)	--	--	-0.05±0.02‖	-0.05±0.02‖	-0.05±0.02‖	-0.07±0.05	-0.03±0.05	-0.03±0.05
No. of county units	211	211	211	211	211	127	51	51

* Plus-minus values are regression coefficients ±SE. COPD denotes chronic obstructive pulmonary disease.
† This model included only counties with a 1986 population ≥ 100,000.
‡ This model included only counties with the largest 1986 population in the MSA.
§ Indicates P < 0.01.
|| Indicates P < 0.05.

eTable 3: Summary of selected regression analyses by baseline $PM_{2.5}$ levels for 545 counties (Dataset 1, 2000 – 2007)

Selected counties and analysis	# Counties	β (SE, p) for 10 μg/m^3 $PM_{2.5}$ (full model)*
2000 $PM_{2.5}$ < 10μg/m^3	100	-0.28(0.39, 0.482)
2000 $PM_{2.5}$ < 12μg/m^3	186	0.50(0.27, 0.065)
2000 $PM_{2.5}$ < 14μg/m^3	301	0.61(0.21, 0.004)
2000 $PM_{2.5}$ < 16μg/m^3	430	0.36(0.19, 0.064)
2000 $PM_{2.5}$ < 18μg/m^3	511	0.47(0.18, 0.009)
2000 $PM_{2.5}$ > 18μg/m^3	34	0.85(0.82, 0.314)
2000 $PM_{2.5}$ > 16μg/m^3	115	0.87(0.38, 0.023)
2000 $PM_{2.5}$ > 14μg/m^3	244	0.28(0.27, 0.305)
2000 $PM_{2.5}$ > 12μg/m^3	359	0.15(0.21, 0.462)
2000 $PM_{2.5}$ > 10μg/m^3	445	0.27(0.18, 0.126)

*Corresponds to the covariate pattern in Model 3 of Table 2 (main text) or Model 3 of eTable 2b. Covarites include change in income, change in population, change in high-school graduates, change in proportion of black population, change in proportion of Hispanic population, change in lung cancer mortality rate, change in COPD mortality rate. Analysis used: STATA 11.0, REGRESS clustered by MSA, using the “weight” statement.

eTable 4: Summary of selected stratified regression analyses for 545 counties (Dataset 1, 2000 – 2007)

Selected counties and analysis	# Counties	β (SE, p) for 10 $\mu g/m^3$ $PM_{2.5}$ (full model)*
2000 Pop. Den. >1000	96	0.86(0.45, 0.061)
2000 Pop. Den. >800	116	0.62(0.41, 0.139)
2000 Pop. Den. >600	145	0.81(0.32, 0.014)
2000 Pop. Den. >400	197	0.84(0.27, 0.003)
2000 Pop. Den. >200	307	0.72(0.22, 0.001)
2000 Pop. Den. < 200	238	-0.31(0.22, 0.165)
2000 urban rate >90%	169	0.95(0.31, 0.003)
2000 urban rate >95%	109	1.12(0.32, 0.001)
2000 Pop. Den. >200 & 2000 urban rate >90%	159	0.96(0.28, 0.001)
2000 urban rate <90%	376	-0.16(0.16, 0.299)
All counties, regression weighted by square root of 2000 Pop. Den.	545	0.74(0.24, 0.002)
All counties, regression weighted by inverse of county land area	545	0.96(0.27, 0.001)

*Covariates include change in income, change in population, change in high-school graduates, change in proportion of black population, change in proportion of Hispanic population, change in lung cancer mortality rate, change in COPD mortality rate. Analysis used: SAS 9.2, PROC SURVEYREG, clustered by MSA, using the "weight" statement, and STATA 11.0, REGRESS.

eTable 5: Comparison of results of select models for males vs. females (Dataset 1, 2000 – 2007)

Selected counties and analysis	**Males: β (SE, p) for 10 μg/m³ $PM_{2.5}$ (full model)***	**Females: β (SE, p) for 10 μg/m³ $PM_{2.5}$ (full model)***	**Signif. Different**
All counties	0.08(0.20, 0.681)	0.59(0.17, 0.001)	Yes
2000 Pop. Density > 200	0.44(0.25, 0.084)	0.85(0.24, 0.001)	No
2000 Pop. Density < 200	-0.55(0.27, 0.043)	-0.06(0.24, 0.805)	No
2000 urban rate > 90%	0.81(0.37, 0.033)	1.07(0.28, <0.001)	No
2000 urban rate < 90%	-0.44(0.20, 0.025)	0.08(0.19, 0.664)	No
All counties, regression weighted by square root of 2000 Pop. Den.	0.57(0.29, 0.047)	0.87(0.22, <0.001)	No
All counties, regression weighted by inverse of county land area	0.74(0.30, 0.013)	1.14(0.30, <0.001)	No

*Covariates include change in income, change in population, change in high-school graduates, change in proportion of black population, change in proportion of Hispanic population, change in lung cancer mortality rate, change in COPD mortality rate. Analysis used: SAS 9.2, PROC SURVEYREG, clustered by MSA, using the "weight" statement, and STATA 11.0, REGRESS.

References:

1. Kulkarni S, Lavin-Rector A, Ezzati M, Murray CJ. Falling behind: life expectancy in U.S. counties from 2000 to 2007 in an international context. *Population Health Metrics* 2011,9:16.
2. Chiang CL. Introduction to stochastic processes in biostatistics. Wiley, New York, 1968.
3. Keyfitz N and Flieger W. World population growth and aging: demographic trends in the late 20th century. U of Chicago, Chicago, 1990.
4. Preston S, Heuveline P, Guillot M. Demography: measuring and modeling population processes. Wiley-Blackwell, Malden (Massachusetts), 2001.
5. Pope CA III, Ezzati M, Dockery D. Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine* 2009;360(4):376-386.
6. Ezzati M, Friedman AB, Kulkarni S, Murray CJ. The reversal of fortunes: trends in county mortality and cross-country mortality disparities in the United States. *PLoS Med* 2008;5(4):e66.