Estimation of 1970 PM2.5 and Pollution Reductions in the US

Important Note

This year (2020), all calculations and graphs that have been generated (using the methodology described in the following text) pertain to a "county" level dataset.

Summary

We would like to evaluate the effects of the Clean Air Act of 1970 in the United States on PM2.5 concentrations and on consequent changes in life expectancy. Since the earliest year covered by the AQLI's satellite-derived PM2.5 data is 1998, we use the US Environmental Protection Agency's (EPA) monitor data dating back to the years around 1970. At that time, most monitors only measured total suspended particulates (TSP).

For the 236 counties for which we are able, we impute 1970 PM2.5 values by assuming constant PM2.5/PM10/TSP ratios over time in each county. Then, since monitor- and satellite-based pollution measurements for a given county may be systematically different, for each county, we scale the imputed 1970 PM2.5 values by the average ratio of satellite and monitor measurements post-1998 to be consistent with the AQLI's satellite data. So far, all data and calculations are at the county-level.

Finally, we combine the 1970 PM2.5 estimates with the AQLI's PM2.5 data for more recent years to calculate changes in PM2.5 concentrations and life expectancy gained due to the pollution reductions. We do this either at the county level or after aggregating to CBSA metropolitan areas.

Data Sources

From the US EPA, we obtained annual average concentrations of PM2.5, PM10, and TSP for all available monitor-years. In the EPA folder,

- *pm_annualsummary.csv* contains data up to 2014. This dataset was provided by Nick Mangus, Project Leader at the EPA's Air Quality System (National Air Group). Phone: (919) 541-5549. Email: mangus.nick@epa.gov.
- Individual CSV files for each year from 2015 onwards. Downloaded from the EPA site.

From the Census, we downloaded 2017 CBSA metropolitan area delineations: *Census/List1_17.xls*

From American FactFinder, we downloaded 2016 American Community Survey 5-Year Population estimates: *Census/ACS_16_5YR_B01003_with_ann.csv*

Finally, we use the AQLI's own US county-level data, *AQLI/us_county_satellite.csv*, for the latest year's PM2.5 concentrations.

Assumptions

1. PM2.5 particulates are a sub-group of PM10 particulates, which in turn are a subgroup of TSP. We assume that in each county, the ratios PM2.5/PM10, PM10/TSP, and PM2.5/TSP are constant over time, and are equal to the average of observed annual ratios.
2. When aggregating to CBSA metropolitan areas, we assume that within each metro area, the PM2.5 concentration is the same. In other words, if we have PM2.5 data for some but not all counties within a metropolitan area, we assume that the average PM2.5 concentration across available counties is the PM2.5 concentration across the entire metro area.

Procedure

Impute monitor-based 1970 PM2.5

For each county,

1. Use all years when multiple PM types are measured to calculate county-specific average ratios $\frac{\text{PM10}}{\text{PM2.5}}$, $\frac{\text{PM2.5}}{\text{TSP}}$, $\frac{\text{PM10}}{\text{TSP}}$. Make the simplifying assumption that in all years, these ratios hold.

2. Since only TSP was monitored in the years around 1970, the 1970 PM2.5 that we would like to calculate from the monitor data is

$$\text{PM2.5}_{1970}^{\text{EPA}} = \text{TSP}_{1970} \cdot \frac{\text{PM2.5}}{\text{TSP}} \tag{1}$$

where, for TSP_{1970}, we use the average TSP over the years 1968-1972.

However, since PM2.5 and TSP rarely overlap in terms of years when they are measured, $\frac{\text{PM2.5}}{\text{TSP}}$ is rarely observed. The solution is, since we already assume constant PM2.5/PM10/TSP ratios over time, it follows that

$$\frac{\text{PM2.5}}{\text{TSP}} = \frac{\text{PM2.5}}{\text{PM10}} \cdot \frac{\text{PM10}}{\text{TSP}},$$

which gives

$$\text{PM2.5}_{1970}^{\text{EPA}} = \text{TSP}_{1970} \cdot \frac{\text{PM2.5}}{\text{PM10}} \cdot \frac{\text{PM10}}{\text{TSP}}. \tag{2}$$

For each county, we will use imputation (1) if possible, and imputation (2) if imputation (1) is not possible but (2) is. This gives us imputed 1970 PM2.5 values for 236 counties. Of those, 174 use imputation (2) to get $\text{PM2.5}_{1970}^{\text{EPA}}$, while 62 directly impute $\text{PM2.5}_{1970}^{\text{EPA}}$ from TSP as in imputation (1).

Scaling monitor-based imputed PM2.5 to satellite data

On average, AQLI's satellite data records lower PM2.5 levels in the US than monitors record, due to the satellite data's exclusion of dust and sea salt and potential measurement error arising from cloud cover and calibration methodology. In order to make an apples-to-apples comparison between the latest year's satellite-derived PM2.5 data and the 1970 estimates, we need to scale the 1970 estimates to be consistent with the satellite data.

For each county and each year starting in 1998 when EPA monitors PM2.5, we take the ratio of AQLI's PM2.5 and EPA's PM2.5. We then average the ratios across years.

In other words, let $\{Y\}$ be the set of years within 1998-2018 in which the EPA monitored PM2.5 in a given county. For each of those years Y, let $\text{PM2.5}_Y^{\text{AQLI}}$ be the PM2.5 in that county-year according to the AQLI's satellite data, and let $\text{PM2.5}_Y^{\text{EPA}}$ be the PM2.5 in that county-year as measured by the EPA's monitors. The average ratio we want is

$$R = \frac{1}{|\{Y\}|} \sum_{\{Y\}} \frac{\text{PM2.5}_Y^{\text{AQLI}}}{\text{PM2.5}_Y^{\text{EPA}}} \tag{3}$$

We can calculate R for all 236 counties for which we were able to calculate $\text{PM2.5}_Y^{\text{EPA}}$. AQLI's satellite data has PM2.5 data for all counties and all years from 1998-2018. Since both imputation methods (1) and (2) require that PM2.5 was directly measured in at least one year, for all 236 counties for which we have $\text{PM2.5}_{1970}^{\text{EPA}}$, we also have directly measured PM2.5 in at least one year from 1998-2018. For our 236 counties, R has mean 0.96 and standard deviation 0.16.

Now, the satellite-adjusted imputed 1970 PM2.5 is $\text{PM2.5}_{1970}^{\text{EPA}}$ from imputations (1) and (2) scaled by (i.e. multiplied by) R from (3):

$$\text{PM2.5}_{1970}^{\text{AQLI}} = R \cdot \text{PM2.5}_{1970}^{\text{EPA}}$$

Life expectancy

Per Ebenstein et al. (2017),
in 1970, if PM2.5 were permanently reduced to the WHO guideline, life expectancy gained is

$$(\text{PM2.5}_{1970}^{\text{AQLI}} - 10) * 0.098 \text{ years}/(\mu g/m^2).$$

Gain in life expectancy from 1970 to 2018 due to PM reduction is:

$$(\text{PM2.5}_{1970}^{\text{AQLI}} - \text{PM2.5}_{2018}^{\text{AQLI}}) * 0.098 \text{ years}/(\mu g/m^2).$$

CBSA Metropolitan Area Aggregations

On the AQLI website, the Clean Air Act Policy Impacts page and the US Country Fact Sheet report PM2.5 changes since 1970 by metropolitan areas. To aggregate from county-level to metro-area-level, we use metro area delineations obtained from the Census Bureau to assign each of the 236 counties to its associated metropolitan or micropolitan area, if any. We then calculate population-weighted averages of 1970 and 2018 PM2.5 and life expectancy values within each metro and micro area. By Assumption 2, when 1970 PM2.5 estimates are not available for some counties within a metro area, we assume that the population-weighted average of values for available counties hold throughout the metro area.

The 236 counties for which we estimated 1970 PM2.5 have combined present population 157 million. When we extrapolate to CBSA metropolitan and micropolitan areas, then we have data for 163 metropolitan areas, 18 micropolitan areas, and 8 counties that do not belong to any CBSA area. The present total population of these CBSA areas and counties is 214 million.