eAppendices: Accountability assessment of health improvements in the United States associated with reduced coal emissions between 2005 and 2012

Lucas RF Henneman

Christine Choirat

Corwin Zigler

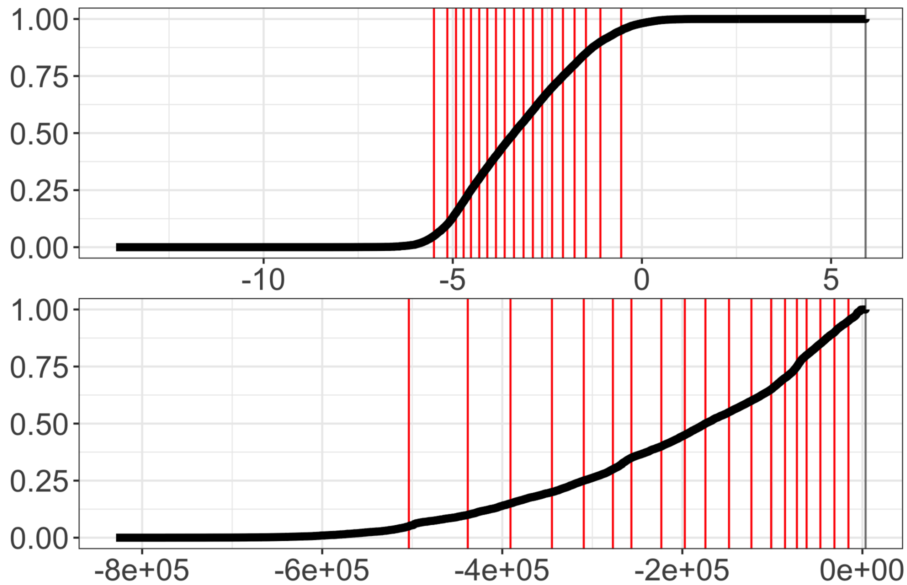
1 Coal exposure: exposure to air influenced by coal emissions

Coal exposure was modeled using a new approach described and evaluated in detail by Henneman et al. (2019).[28] The method uses unit level emissions for 2005 and 2012 measured using continuous emissions monitors, downloaded from the EPA's Air Markets Program Database[25] and linked with stack height data included in the National Emissions Inventory.[26] Units for which stack height was unavailable were assigned the average stack height. This yielded a database of 1,036 units in 2005 and 1,009 in 2012.

The method employs HYSPLIT, an air parcel trajectory and dispersion model.[27] We simulated the transport and dispersion of 100 parcels beginning at four instances per day—12:00 a.m., 6:00 a.m., 12:00 p.m., and 6:00 p.m.—from each coal unit. Parcel locations were tracked for 10 days after each emissions event. HYSPLIT was driven by NOAA reanalysis wind fields.[31]

Raw hourly parcel locations output from HYSPLIT were post-processed in three ways before linking to ZIP codes. First, parcels in hours 0 and 1 were removed to approximate plume transfer away from the source before emissions are vertically transported to ground-level. Second, parcels were trimmed after they reached a height of zero (i.e., no resuspension). Third, parcels were removed at points in time if they were above the planetary boundary layer, thus limiting the analysis to parcels that were in the mixed layer. Monthly mean planetary boundary layer heights were taken from 20th Century Reanalysis Data.[52]

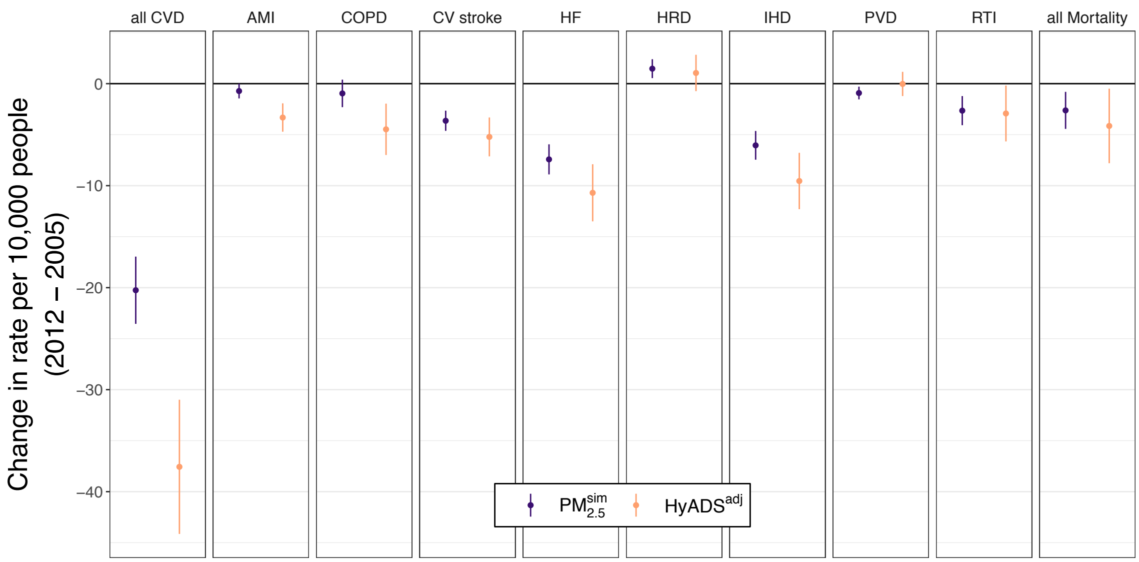
Hourly positions of trimmed parcel locations were combined by month for each unit and subsequently linked to ZIP codes using a multi-step process. First, parcel locations were aggregated by month and assigned to a fine ($0.19° \times 0.19°$) grid covering the United States. A concentration of parcels in each grid cell was taken as the number of parcels per volume, where the height was the planetary boundary layer height. ZIP code parcel concentrations were taken as the spatially-weighted mean of the grid cells over each ZIP code using the `over` command in the `sp` `R` package.[53] The resulting monthly measures of how emissions from each unit impact each ZIP code were then multiplied by the unit's monthly SO_2 emissions, yielding unit-less ZIP code-specific metrics of exposure to air parcels affected by coal SO_2 emissions. We summed monthly impacts to estimate total annual exposure to coal emissions.



eFigure 1: CDFs of exposure metric changes between 2005 and 2012 (negatives are a decrease). Vertical lines denote maximum values for each ventiles, and dark gray vertical lines denote ventiles with positive maximums. Top: $\mathrm{PM}_{2.5}^{sim}$, bottom: `HyADS`. Note that the horizontal axes are aligned based on each metric's range.

2 Important covariates in the mortality models

In both the $\mathrm{PM}_{2.5}^{sim}$ and `HyADS` models, the same covariates have coefficients different from zero at a statistical significance of 0.05: the change in median age of the population, the gender makeup and change over time, the racial makeup and change over time (in particular the change in America Indian fraction and change over time, the Black fraction and change over time, and the multiple race fraction), and the change in specific humidity.



eFigure 2: Changes in hospitalization rate per 10,000 person-years (beneficiaries for all Mortality) per unit decrease in each exposure metric ($-\beta_E^{\Delta}$) for the model in Eqn. 1 excluding the covariate terms ($\sum_{c=1}^{C} \beta_c X_{c,2005}$ and $\sum_{c=1}^{C-1} \beta_c^{\Delta} X_{c,2012-2005}^{\Delta}$). Negative values on the vertical axis denote decreases in health outcome rates associated with single unit decreases in exposure metrics. Error bars denote the 95% confidence intervals. Abbreviations: CVD - cardiovascular disease; AMI - acute myocardial infarction; COPD - chronic obstructive pulmonary disorder; CV - cardiovascular; HF - heart failure; HRD - heart rhythm disorders; IHD - Ischemic heart disease; PVD - peripheral vascular disease; RTI - respiratory tract infection.

	2005	2012	Difference
*Median age (years)	4.0e+1 (5.7e+0)	4.1e+1 (6.9e+0)	1.6e+0 (2.8e+0)
*Median housing income ($)	4.9e+4 (1.9e+4)	4.7e+4 (1.7e+4)	-1.7e+3 (7.0e+3)
*Per capita income ($)	2.4e+4 (8.2e+3)	2.4e+4 (8.2e+3)	-1.6e+2 (3.7e+3)
*Diversity index	2.6e+1 (3.6e+1)	2.7e+1 (3.7e+1)	2.3e+0 (5.4e+0)
*Female fraction	5.0e-1 (2.2e-2)	5.0e-1 (2.2e-2)	-1.9e-3 (1.3e-2)
*American Indian fraction	1.2e-2 (4.3e-3)	1.3e-2 (4.6e-3)	5.4e-4 (2.4e-3)
*Asian fraction	1.4e-2 (9.5e-3)	1.5e-2 (1.0e-2)	1.8e-3 (3.8e-3)
*Black fraction	8.9e-2 (8.0e-2)	8.5e-2 (7.7e-2)	-2.9e-3 (9.1e-3)
*Other race fraction	2.4e-2 (1.6e-2)	2.4e-2 (2.0e-2)	7.4e-4 (6.9e-3)
*Pacific Islander fraction	3.9e-4 (4.8e-4)	3.9e-4 (4.6e-4)	2.1e-5 (2.4e-4)
*Multiple race fraction	1.5e-2 (1.2e-2)	1.8e-2 (1.3e-2)	3.1e-3 (7.1e-3)
Smoking rate	2.6e-1 (5.4e-2)	2.3e-1 (5.9e-2)	-2.9e-2 (1.9e-2)
Temperature (K)	2.9e+2 (6.6e+0)	2.9e+2 (6.5e+0)	1.2e+0 (7.1e-1)
Specific humidity (kg/kg)	8.4e-3 (2.4e-3)	8.4e-3 (2.6e-3)	-2.5e-6 (3.4e-4)

eTable 1: ZIP code averages across the United States east of -110° longitude in 2005, 2012, and Δ means (IQR). A negative difference denotes a decrease from 2005 to 2012. Business Analysis census variables (*) were available starting in 2007, which is used as a proxy for 2005.

Response	Independant	N	β_E^{Δ}	95% CI
all CVD	$PM_{2.5}^{sim}$	24371	-8.4	(-12.67, -4.14)
	`HyADS`	24458	-26.11	(-33.89, -18.33)
AMI	$PM_{2.5}^{sim}$	18467	-0.01	(-0.93, 0.91)
	`HyADS`	18531	-2.29	(-3.91, -0.66)
COPD	$PM_{2.5}^{sim}$	17304	-0.47	(-2.14, 1.19)
	`HyADS`	17359	-5.19	(-8.13, -2.25)
CV stroke	$PM_{2.5}^{sim}$	20038	-1.95	(-3.2, -0.7)
	`HyADS`	20105	-4.24	(-6.48, -2.01)
HF	$PM_{2.5}^{sim}$	19762	-4.26	(-6.09, -2.43)
	`HyADS`	19831	-9.66	(-12.91, -6.41)
HRD	$PM_{2.5}^{sim}$	19668	0.96	(-0.21, 2.12)
	`HyADS`	19736	0.95	(-1.14, 3.05)
IHD	$PM_{2.5}^{sim}$	21631	-3.87	(-5.67, -2.08)
	`HyADS`	21703	-7.42	(-10.66, -4.18)
PVD	$PM_{2.5}^{sim}$	15145	-0.64	(-1.42, 0.15)
	`HyADS`	15200	-0.43	(-1.82, 0.96)
RTI	$PM_{2.5}^{sim}$	20227	-4.33	(-6.1, -2.55)
	`HyADS`	20297	-5.14	(-8.34, -1.95)
all Mortality	$PM_{2.5}^{sim}$	25143	-0.38	(-2.76, 2.01)
	`HyADS`	25234	-0.26	(-4.63, 4.11)

eTable 2: Regression coefficients and 95% confidence intervals plotted in Figure 2. Abbreviations: CVD - cardiovascular disease; AMI - acute myocardial infarction; COPD - chronic obstructive pulmonary disorder; CV - cardiovascular; HF - heart failure; HRD - heart rhythm disorders; IHD - Ischemic heart disease; PVD - peripheral vascular disease; RTI - respiratory tract infection.

Response	Independant	$\beta_1 - \beta_5$	$\beta_5 - \beta_{10}$	$\beta_{10} - \beta_{15}$	$\beta_{15} - \beta_{20}$
all CVD	$PM_{2.5}^{sim}$	-17.7	-52.8**	22.7	-7.1
	HyADS	22.3	1.9	-54.6***	-45.7**
AMI	$PM_{2.5}^{sim}$	-2.1	0.3	-1.2	3.9
	HyADS	3.7	0.4	0.2	-6.5
COPD	$PM_{2.5}^{sim}$	-2.5	-2.3	2.3	-6.5
	HyADS	0.0	16.9**	-20.8***	9.3
CV stroke	$PM_{2.5}^{sim}$	-0.6	-15.0**	7.7	-2.5
	HyADS	2.1	-9.0	-0.9	-4.8
HF	$PM_{2.5}^{sim}$	-0.5	-17.5*	5.5	-7.5
	HyADS	8.7	-13.2	-1.1	-20.8**
HRD	$PM_{2.5}^{sim}$	2.3	3.4	2.8	-5.5
	HyADS	-7.3	17.6***	-5.1	-5.5
IHD	$PM_{2.5}^{sim}$	-30.3***	-23.9***	6.7	8.1
	HyADS	-0.4	-4.6	-10.8	-11.0
PVD	$PM_{2.5}^{sim}$	-1.4	-2.8	-3.5	1.0
	HyADS	6.7	-4.4	4.9	-4.4
RTI	$PM_{2.5}^{sim}$	-9.0	-15.9*	-4.6	-4.1
	HyADS	-21.4*	-10.1	-5.2	5.2
all Mortality	$PM_{2.5}^{sim}$	-28.5**	-0.3	10.1	1.9
	HyADS	4.5	-19.4*	13.6	1.2

eTable 3: Differences between select regression parameters (β_q) from the models in Eqn. 2. Statistical significance labels follow the following convention: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. Abbreviations: CVD - cardiovascular disease; AMI - acute myocardial infarction; COPD - chronic obstructive pulmonary disorder; CV - cardiovascular; HF - heart failure; HRD - heart rhythm disorders; IHD - Ischemic heart disease; PVD - peripheral vascular disease; RTI - respiratory tract infection.

Response	Independent	β_q^* (95% CI)	$p_{\beta_q^*}$	p_{lin}
all CVD	$PM_{2.5}^{sim}$	-2.22 (-3.31, -1.13)	0.00	0.12
	HyADS	-2.77 (-3.91, -1.62)	0.00	0.00
AMI	$PM_{2.5}^{sim}$	-0.01 (-0.23, 0.22)	0.96	0.27
	HyADS	-0.22 (-0.45, 0.01)	0.06	0.00
COPD	$PM_{2.5}^{sim}$	-0.09 (-0.48, 0.30)	0.64	0.06
	HyADS	-0.63 (-1.04, -0.22)	0.00	0.00
CV stroke	$PM_{2.5}^{sim}$	-0.46 (-0.77, -0.16)	0.00	0.58
	HyADS	-0.64 (-0.96, -0.32)	0.00	0.03
HF	$PM_{2.5}^{sim}$	-1.02 (-1.47, -0.58)	0.00	0.00
	HyADS	-1.13 (-1.59, -0.67)	0.00	0.00
HRD	$PM_{2.5}^{sim}$	0.23 (-0.06, 0.51)	0.12	0.51
	HyADS	0.34 (0.04, 0.64)	0.03	0.01
IHD	$PM_{2.5}^{sim}$	-0.81 (-1.26, -0.36)	0.00	0.00
	HyADS	-1.10 (-1.57, -0.64)	0.00	0.00
PVD	$PM_{2.5}^{sim}$	-0.15 (-0.34, 0.04)	0.12	0.13
	HyADS	-0.03 (-0.22, 0.17)	0.81	0.00
RTI	$PM_{2.5}^{sim}$	-1.01 (-1.44, -0.58)	0.00	0.05
	HyADS	-0.84 (-1.30, -0.39)	0.00	0.00
all Mortality	$PM_{2.5}^{sim}$	0.04 (-0.58, 0.66)	0.90	0.23
	HyADS	-0.07 (-0.71, 0.58)	0.84	0.06

eTable 4: Regression parameters ($-\beta_q^*$) from the model in Eqn. 3 and two hypothesis tests related to the secondary analysis. $p_{\beta_q^*}$ summarizes the test with the null hypothesis that the slope of β_q's is not zero. p_{lin} is the result of the likelihood ratio test with the null hypothesis that the linear model in q (Eq. 3) is not different from a model that is nonlinear in q (Eq. 2). Abbreviations: CVD - cardiovascular disease; AMI - acute myocardial infarction; COPD - chronic obstructive pulmonary disorder; CV - cardiovascular; HF - heart failure; HRD - heart rhythm disorders; IHD - Ischemic heart disease; PVD - peripheral vascular disease; RTI - respiratory tract infection.