Annals of Epidemiology 25 (2015) 162–173

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States?

Louis Anthony (Tony) Cox Jr. PhD [a,b,*], Douglas A. Popken PhD [a,b]

[a] Cox Associates and University of Colorado, Denver, CO
[b] NextHealth Technologies, Englewood, CO

ARTICLE INFO

Article history:
Received 7 November 2014
Accepted 9 November 2014
Available online 20 November 2014

Keywords:
Air pollution
Health effects
Accountability research
Causality
Fine particulate matter
Ozone
PM2.5
PM10
Cardiovascular mortality risks
Respiratory mortality risk

ABSTRACT

Purpose: Between 2000 and 2010, air pollutant levels in counties throughout the United States changed significantly, with fine particulate matter (PM2.5) declining over 30% in some counties and ozone (O_3) exhibiting large variations from year to year. This history provides an opportunity to compare county-level changes in average annual ambient pollutant levels to corresponding changes in all-cause (AC) and cardiovascular disease (CVD) mortality rates over the course of a decade. Past studies have demonstrated associations and subsequently either interpreted associations causally or relied on subjective judgments to infer causation. This article applies more quantitative methods to assess causality.
Methods: This article examines data from these "natural experiments" of changing pollutant levels for 483 counties in the 15 most populated US states using quantitative methods for causal hypothesis testing, such as conditional independence and Granger causality tests. We assessed whether changes in historical pollution levels helped to predict and explain changes in CVD and AC mortality rates.
Results: A causal relation between pollutant concentrations and AC or CVD mortality rates cannot be inferred from these historical data, although a statistical association between them is well supported. There were no significant positive associations between changes in PM2.5 or O_3 levels and corresponding changes in disease mortality rates between 2000 and 2010, nor for shorter time intervals of 1 to 3 years.
Conclusions: These findings suggest that predicted substantial human longevity benefits resulting from reducing PM2.5 and O_3 may not occur or may be smaller than previously estimated. Our results highlight the potential for heterogeneity in air pollution health effects across regions, and the high potential value of accountability research comparing model-based predictions of health benefits from reducing air pollutants to historical records of what actually occurred.

Introduction: using data from natural experiments to understand causality

An aim of applied science in general and of epidemiology in particular is to draw sound causal inferences from observations. Students are taught to develop hypotheses about causal relations, devise testable implications of these causal hypotheses, carry out the tests, and objectively report and learn from the results to refute or refine the initial hypotheses. For at least the past two decades, however, epidemiologists and commentators on scientific methods and results have raised concerns that current practices too often lead to false-positive findings and to mistaken attributions of causality to mere statistical associations [1–4]. Formal training in epidemiology may be a mixed blessing in addressing these concerns, as concepts such as "attributable risk," "population attributable fraction," "burden of disease," "etiologic fraction," and even "probability of causation" are based on relative risks and related measures of statistical association and do not necessarily reveal anything about causation [5,6]. Limitations of human judgment and inference, such as confirmation bias (finding what we expect to find), motivated reasoning (concluding what it pays us to conclude), and overconfidence (mistakenly believing that our own beliefs are more accurate than they really are), do not spare health effects investigators. Experts in the health effects of particular compounds are not always experts in causal analysis, and published causal conclusions are often unwarranted, with a pronounced bias toward finding "significant" effects where none actually exists (false positives) [1,2,7,8]. This article considers ways to do better, borrowing ideas from econometrics and causal analysis. It illustrates them in the important practical domain of assessing public health risks from air pollution and estimating public health benefits from reducing it.

* Corresponding author. Cox Associates, 503 Franklin Street, Denver, CO 80218.
E-mail address: tcoxdenver@aol.com (L.A.(Tony) Cox).

http://dx.doi.org/10.1016/j.annepidem.2014.11.006
1047-2797/

Dominici et al. [9] recently noted that "[A]nalyses of observational data have had a large impact on air-quality regulations and on the supporting analyses of their accompanying benefits, [but] associational approaches to inferring causal relations can be highly sensitive to the choice of the statistical model and set of available covariates that are used to adjust for confounding. … There is a growing consensus … that the associational or regression approach to inferring causal relations—on the basis of adjustment with observable confounders—is unreliable in many settings." The authors demonstrate via example that the choice of regression model can result in either statistically significant positive or statistically significant negative associations between air pollutant levels and mortality rates. This implies that implicit modeling choices can greatly affect—or even determine—the results presented to decision makers and the public. Table 1 provides some examples of important policy-relevant conclusions and doubts about their validity from the recent air pollution health effects literature.

To overcome this difficulty, Dominici et al. [9] proposed the use of quasi-experiments (QEs), or natural experiments, in which outcomes are compared between a treatment and control group, but without random assignment or other determination of the treatment status by the researcher. As an example, they cite a study reporting significantly lower mortality rates in the 6 years after a ban on coal burning in Dublin County, Ireland compared with the 6 years before the ban [22]. Their proposal to use QEs to better assess causal relations between pollution levels and health effects has been hailed by some [27] as "a paradigm-shifting solution." Yet, ever since QEs were first introduced in social statistics over half a century ago, expert practitioners [28] have recognized that "in many QEs, one is most often left with the question: 'Are there alternative explanations for the apparent causal association?' Such alternative explanations constitute threats to the internal validity of causal inferences for the studied populations that must be refuted before valid causal inferences can be drawn from QEs [29]. A long tradition of refutationist approaches to causal inference in epidemiologic methodology makes a similar point [30,31].

For example, to be valid, the conclusion that a ban on coal burning caused an immediate reduction in all-cause (AC) and cardiovascular mortality [23] would have had to refute alternative explanations. A study design including a relevant historical or contemporaneous control group (using a pretest—posttest design or a nonequivalent control group design, respectively, in QE terminology) would have allowed the elimination of noncausal explanations, such as that (a) mortality rates were already declining before the ban and continued to do so without significant change during and afterward for reasons unrelated to the ban (the "History" threat to internal validity, in QE terminology); or (b) mortality rates declined at the same rate in areas not affected by the ban as in areas affected by it. For the Dublin study, both possibilities (a) and (b) proved to be true, so that no valid conclusions about the impact of the ban on AC or cardiovascular mortality rates can be drawn [24,25]. Indeed, on reanalysis using relevant control groups, no effect of the ban on these outcomes could be detected [26]. Yet, as Dominici et al., rightly note, natural experiments occur frequently and, if properly analyzed, they can provide crucial policy-relevant insights into causality (or lack thereof) in observed exposure-response relations. In the United States, for example, geographic heterogeneity in the rates at which pollutant levels have declined in different regions has created many natural experiments for assessing the effects of these changes on public health over time.

To take advantage of these natural experiments, this article compares changes in PM2.5 and O_3 levels from 2000 to 2010 to corresponding changes in AC and cardiovascular disease (CVD) age-specific mortality rates over the same interval, for hundreds of counties in the 15 largest states in the United States. Treating county as the unit of observation, as in the Dublin study and many

Table 1
Some conflicting claims about health effects known to be caused by air pollution

Pro (causal interpretation or claim)	Con (counter interpretation or claim)
"Epidemiological evidence is used to quantitatively relate PM2.5 exposure to risk of early death. We find that UK combustion emissions cause ~13,000 premature deaths in the UK per year, while an additional ~6000 deaths in the UK are caused by non-UK European Union (EU) combustion emissions" [10].	"[A]lthough this sort of study can provide useful projections, its results are only estimates. In particular, although particulate matter has been associated with premature mortality in other studies, a definitive cause-and-effect link has not yet been demonstrated" [11].
"[A]bout 80,000 premature mortalities [per year] would be avoided by lowering PM2.5 levels to 5 $\mu g/m^3$ nationwide" in the U.S. 2005 levels of PM2.5 caused about 130,000 premature mortalities per year among people over age 29, with a simulation-based 95% CI of 51,000 to 200,000 [12].	"Analysis assumes a causal relationship between PM exposure and premature mortality based on strong epidemiological evidence… However, epidemiological evidence alone cannot establish this causal link." [13] Significant negative associations have also been reported between PM2.5 and short-term mortality and morbidity rates [14], as well as between levels of some other pollutants [15,16] (e.g., NO_2 and ozone) and short-term mortality and morbidity rates.
"Some of the data on the impact of improved air quality on children's health are provided, including … the reduction in the rates of childhood asthma events during the 1996 Summer Olympics in Atlanta, Georgia, due to a reduction in local motor vehicle traffic" [17]. "During the Olympic Games, the number of asthma acute care events decreased 41.6% (4.23 vs. 2.47 daily events) in the Georgia Medicaid claims file," coincident with significant reductions in ozone and other pollutants [18].	"In their primary analyses, which were adjusted for seasonal trends in air pollutant concentrations and health outcomes during the years before and after the Olympic Games, the investigators did not find significant reductions in the number of emergency department visits for respiratory or cardiovascular health outcomes in adults or children." In fact, "relative risk estimates for the longer time series were actually suggestive of increased ED [emergency department] visits during the Olympic Games" [19].
"An association between elevated PM10 levels and hospital admissions for pneumonia, pleurisy, bronchitis, and asthma was observed. During months when 24-hour PM10 levels exceeded 150 micrograms/m3, average admissions for children nearly tripled; in adults, the increase in admissions was 44 per cent." [20].	"Respiratory syncytial virus (RSV) activity was the single explanatory factor that consistently accounted for a statistically significant portion of the observed variations of pediatric respiratory hospitalizations. No coherent evidence of residual statistical associations between PM_{10} levels and hospitalizations was found for any age group or respiratory illness." [21].
"Reductions in respiratory and cardiovascular death rates in Dublin suggest that control of particulate air pollution could substantially diminish daily death … Our findings suggest that control of particulate air pollution in Dublin led to an immediate reduction in cardiovascular and respiratory deaths." [22].	Mortality rates were already declining long before the ban, and occurred in areas not affected by it. "Serious epidemics and pronounced trends feign excess mortality previously attributed to heavy black-smoke exposure" [24]. "Thus, a causal link between the decline in mortality and the ban of coal sales cannot be established" [25]. "In contrast to the earlier study, there appeared to be no reductions in total mortality or in mortality from other causes, including cardiovascular disease, that could be attributed to any of the bans. That is, after correcting for background trends, similar reductions were seen in ban and non-ban areas." [26].
"The results could not be more clear, reducing particulate air pollution reduces the number of respiratory and cardiovascular related deaths immediately" [23].	

Adapted from a study by Cox [5].

others where individual-level exposure data are not available, invites application of longitudinal designs and methods in which each county's history of pollution levels and mortality rates serves as its own control group for purposes of determining how subsequent changes in pollution are associated with subsequent changes in mortality rates [29]. Using repeated observations on the same counties over time also allows the effects of unmeasured (and possibly unknown) confounders to be largely controlled for as changes in pollutant levels and mortality rates are calculated—the basic strategy of panel data analysis [32]. The goal of our analysis is to understand the extent to which historical associations between pollutant levels and mortality rates reflect a clear causal relation, rather than merely coincident trends, or the effect of confounders, or modeling choices.

Table 2 lists several quantitative methods for causal hypothesis testing, modeling, and analysis that have been extensively developed and applied over the past six decades [5]. Various advantages of these techniques, compared with qualitative causal criteria [31] such as the traditional Hill considerations and other weight-of-evidence and associational methods, are well explained and illustrated [6] in the references for Table 2, along with their limitations [33]. Prominent among these advantages is the development of empirically testable implications of causal hypotheses, such as conditional independence implications, timing implications, information-theoretic implications, and exogeneity implications, with conditional probability distributions of some variables being determined by the values of others. These testable implications capture the asymmetry inherent in the notion of causation, unlike correlations or other symmetric measures of association. They can be tested statistically using publically available standard computer codes, such as those in R (www.r-project.org/) and Python/NumPy (www.numpy.org/). This enables different investigators, perhaps with very different prior beliefs, to reach the same conclusions from the same data. This points the way toward greater objectivity and definitiveness in determining via such tests the extent to which data do or do not support causal hypotheses, based on their testable implications.

Other reasons why modern methods of quantitative causal analysis should be (and increasingly are) included among current approaches in the epidemiologist's tool kit are discussed in modern epidemiology textbooks [31] and monographs [42] and in the references to Table 2. The purpose of this article is not to further review these methods but to apply those that are most useful to the air pollution and mortality rate records in the United States.

Data and methods

Cause-specific mortality rates, by county and age group, were downloaded from the Centers for Disease Control and Prevention Wonder "Compressed Mortality, 1999 to 2010" database [43]. To

Table 2
Some formal methods for modeling and testing causal hypotheses

Method and references	Basic idea	Appropriate study design
Quasi-experimental design and analysis [29]	Can control group comparisons refute alternative (noncausal) explanations for observed associations between hypothesized causes and effects, e.g., coincident trends and regression to the mean? If so, this strengthens causal interpretation.	Observational data on subjects exposed and not exposed to interventions that change the hypothesized cause(s) of effects.
Conditional independence tests [33,34]	Is hypothesized effect (e.g., CVD mortality rate) statistically independent of hypothesized cause (e.g., PM2.5 concentration), given (i.e., conditioned on) the values of other variables, such as education and income? If so, this undermines causal interpretation.	Cross-sectional data; can also be applied to multiperiod data (e.g., in dynamic Bayesian networks)
Panel data analysis [32,35]	Are changes in exposures followed by changes in the effects that they are hypothesized to help cause? If not, this undermines causal interpretation; if so, this strengthens causal interpretation. Example: are reductions in PM2.5 levels followed (but not preceded) by corresponding changes in CVD mortality rates?	Panel data study: collect a sequence of observations on same subjects or units of observation (e.g., counties) over time
Granger causality test [36]	Does the history of the hypothesized cause improve ability to predict the future of the hypothesized effect? If so, this strengthens causal interpretation; otherwise, it undermines causal interpretation. Example: can CVD mortality rates be predicted better from time series histories of PM2.5 levels and mortality rates than from the time series history of mortality rates alone?	Time series data on hypothesized causes and effects
Intervention analysis and change point analysis [37,38]	Does the best-fitting model of the observed data change significantly at or following the time of an intervention? If so, this strengthens causal interpretation. Do the quantitative changes in hypothesized causes predict and explain the subsequently observed quantitative changes in hypothesized effects? If so, this strengthens causal interpretation. Example: Do mortality rates fall faster in counties where pollutant levels fall faster than in other counties?	Time series observations on hypothesized effects, and knowledge of timing of intervention(s) Quantitative time series data for hypothesized causes and effects
Counterfactual and potential outcome models [39]	Do exposed individuals have significantly different response probabilities than they would have had if they had not been exposed? Example: do people have lower mortality risk after historical exposure reductions than they would have had otherwise?	Cross-sectional and/or longitudinal data, with selection biases and feedback among variables allowed
Causal network, path analysis and structural equations models of change propagation [40]	Do changes in exposures (or other causes) create a cascade of changes through a network of causal mechanisms (represented by equations), resulting in changes in the effect variables? Example: do relatively large variations in daily levels of fine particulate matter (PM2.5) air pollution create corresponding variations in markers of oxidative stress in the lungs?	Observations of variables in a dynamic system out of equilibrium
Negative controls (for exposures or for effects) [41]	Do exposures predict health effects better than they predict effects that cannot be caused by exposures? Example: do pollutant levels predict cardiovascular mortality rates better than they explain car accident mortality rates? If not, this weakens causal interpretation of the CVD associations.	Observational studies

Adapted from a study by Cox [5].

create a geographically diverse sample, mortality rates were extracted at the county level for the 15 largest states in the United States (California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, Georgia, Michigan, North Carolina, New Jersey, Virginia, Washington, Massachusetts, and Arizona) representing approximately 65% of the total US population. We extracted mortality rates (per 100,000 person-years) for all causes of death and then created three disease subcategories: (1) diseases of the circulatory system (*International Classification of Diseases, 10th revision codes* I00–I99), (2) all external causes of death (*International Classification of Diseases, 10th revision* codes V01–Y89), and (3) total disease-related mortalities (all causes of death excluding external causes). The dependent variables shown in subsequent tables thus included the following:

- CVRatePer100K—mortality rate (per 100,000 people per year) because of all heart and/or circulatory diseases
- ExtRatePer100K—mortality rate because of external causes (used as a negative control). (To investigate whether the methods used can detect causal known relationships, we also used a positive control in which a known causal effect was simulated, as discussed later.)
- ACRatePer100K—mortality rate because of all disease-related (nonexternal) causes.

Most of our analyses were restricted to ages 65+ years as they have the highest CVD mortality rates. Age was categorized as 65 to 74 years, 75 to 84 years, and 85+ years.

County-level air quality data for PM2.5 (daily 24-hour mean) and O_3 (daily maximum 8-hour moving average) were downloaded from the US Environmental Protection Agency Air Quality System (AQS) for all monitors located in each county ($n = 483$) of the 15 states listed previously [44]. Data were obtained for the years 2000 to 2010. The two pollutant measures were summarized as county-level annual averages in our analyses.

The mortality and air quality data were merged by state/county and year. The resulting merged data file contained data for 483 distinct counties from 2000 to 2010, although not all counties collected both ozone (O_3) and PM2.5 data for all years. These merged data files are freely available from the authors on request.

Statistical analysis methods

The methods in Table 2 that are most useful for the air pollution and mortality rate data sets just described include conditional independence tests, longitudinal comparisons of changes in death rates and changes in pollution levels, Granger causality tests, and negative controls comparing presumably noncausal associations between longitudinal changes in accident and other "external" (nondisease) death rates and changes in pollutant levels to associations between changes in disease mortality rates and changes in pollutant levels. These are described in the following paragraphs. All statistical computations were carried out using the Statistica 12.5 (https://support.software.dell.com/download-install-detail/5503316) statistical computing environment, with the exception of the Granger causality tests, described in the following paragraphs. Other methods in Table 2, such as change-point analysis and intervention analysis for an intervention that occurs at a single point in time (e.g., closing a steel mill or banning coal burning in Dublin) are less relevant for these data because both changes in PM2.5 and changes in mortality rates occurred gradually over a decade, rather than abruptly from before to after some intervention.

Association-based methods: correlation and regression

Although they are not methods of causal analysis, association-based methods such as correlation and regression analysis are widely used in air pollution health effects research [9]. We used these methods also to test whether applying them in this data set produced similar results to past studies. Intuitively, the absence of any association might be interpreted to suggest that causation is unlikely [45,46]. We used Pearson product-moment linear correlation coefficients and linear regression coefficients as measures of linear association because past research suggests an approximately linear association of PM2.5 and O_3 with mortality [47].

Conditional independence tests

If a statistically significant association between exposure and response variables is found, for example, based on linear correlation and regression tests, then an important screening test for potential causation is the conditional independence test: does a significant association remain even after conditioning on potential confounders, such as age or year? For example, if a significant association between PM2.5 and CVD mortality were hypothesized to be created by confounding by year (because both PM2.5 and CVD mortality rates declined with time, even if one did not cause the other), then one could condition on year (i.e., holding it fixed at a given value, such as 2010) and test whether the conditional association vanishes within the subset of records with that value (e.g., with year = 2010).

To avoid biasing results by manual selection of variables to condition on, we relied on automated backward stepwise variable selection in our multiple regression models. This is a standard—but deservedly controversial—technique. We do not advocate it for general use, as it over-fits models to data, producing excess false positives in simple settings. We therefore have used it only as a readily available automated approach that may be more familiar and easily available than alternatives such as Bayesian Model Averaging; but we have also verified the main conclusions using multiple disjoint random samples of the data (20% cross-validation) to guard against the defects of backward stepwise selection. The backward stepwise selection procedure uses successive F tests to determine whether dropping individual variables (e.g., O_3 concentration) from the set of potential explanatory variables significantly decreases the ability of the model to predict values of the dependent variable (e.g., CVD mortality risk). If not, that is, if the F test indicates that the dependent variable is conditionally independent of a potential explanatory variable (such as O_3), given the values of other variables in the model, then that variable is automatically dropped from the final set of explanatory variables. Despite its flaws, use of this technique reduces subjectivity in choosing explanatory variables. We used the default settings in *Statistica* (e.g., P values of .05 to define significant associations).

Correlations among changes over time

Perhaps the most important screening test we use for potential causality is examining whether changes in an exposure help to predict and explain changes in a response. A frequent confusion in epidemiology is to interpret the slope of a concentration-response relation as indicating the future change in response (e.g., mortality rates) that would be caused by a unit change in future exposure concentration. This is incorrect because many concentration-response associations are not entirely causal (e.g., because of confounders or modeling biases). Rather than using slopes of cross-sectional regression lines as proxies for causal impacts, we directly tested whether there were significant positive correlations and regression coefficients between longitudinal changes in county-specific PM2.5 and O_3 levels from 2000 to 2010 and corresponding longitudinal changes in county-specific and age-specific

mortality rates; and whether counties with more rapid declines in PM2.5 and O_3 had more rapid declines in mortality than those with slower declines, or where concentrations increased.

Granger causality and negative and positive controls

A more general approach than studying associations between changes in exposure concentrations and changes in mortality rates over a single time interval is to use time series analysis to test whether past values of exposure help to predict present and future mortality rates more accurately than they can be predicted from past mortality rates alone. This is the basic idea of the Granger causality test [36]. If the future of a mortality rate time series is conditionally independent of the past and present exposure time series, given the past and present mortality rate series, so that knowing exposure does not improve ability to predict future mortality rates, then exposure is not a Granger cause of mortality. The Granger causality test produces a *P* value for the null hypothesis that one time series does not improve prediction of another compared with using lagged values of the dependent variable itself.

We performed the Granger tests, using the *grangercausalitytests* function in the Python *statsmodels* module, for each county and age category combination described previously, with the restriction that the combination must have at least 10 consecutive annual values available for analysis. We tested lags of 1 to 3 years, as many previous studies suggest that reductions in PM2.5 and other pollutants lead to almost immediate reductions in mortality rates, for example, within as little as a few days, and certainly well within a year or two [18,22,47,48]. The Python Granger function *grangercausalitytests* provides *P* values for each of four separate test statistics (two based on the *F* distribution and two on the χ^2 distribution), all of which yield closely similar results. We evaluated the proportion of counties for which these tests produced a *P* value of 0.05 or less; random variation alone could explain this occurring in about 5% of counties. Significantly higher levels would be suggestive of a Granger causality effect.

In addition to formal test statistics, we also compared the statistical association between changes in exposures and changes in disease-related mortality rates, on the one hand, to the association between changes in exposures and changes in nondisease-related (external-cause) mortality rates, on the other. The external-cause mortality rates include deaths due to accidents and assaults, in which the observed temporal trends are presumably not caused by changes in pollution levels. Such negative controls test whether hypothesized causal associations are stronger than those presumed to be noncausal [41]. As discussed further later, we also simulated the effects of a positive causal relation between changes in pollution levels and changes in mortality rates. This simulation-based analysis served as a type of positive control to test whether sample sizes are large enough and whether the statistical methods we applied are powerful enough to detect such genuine causal effects if they are present. Finally, we briefly examined the geographic pattern of results to determine whether findings appeared to hold consistently in different parts of the United States.

Results

Descriptive statistics

Figure 1 shows trends in average pollution levels, population, and mortality rates for all counties from 2000 to 2010. For each time series, values are normalized by dividing by the value in 2000, so that all time series values in 2000 are defined as 1.0. PM2.5 and CVD mortality rates declined most steeply over this interval (two lowest curves), whereas population levels and external-cause mortality rates (e.g., from accidents) increased, perhaps reflecting a longer-lived aging population.

Figure 2 shows how the age-specific mortality curve, plotting annual deaths per capita versus age, has shifted downward over time. (The horizontal positions for the rates have been spread out to allow easy visualization of trends. Vertical bars indicate 95% confidence intervals (CIs) for the mean mortality rates but are very narrow because of the large sample sizes.) Clearly, age-specific mortality rates have declined for all age groups, but most for the older age groups.

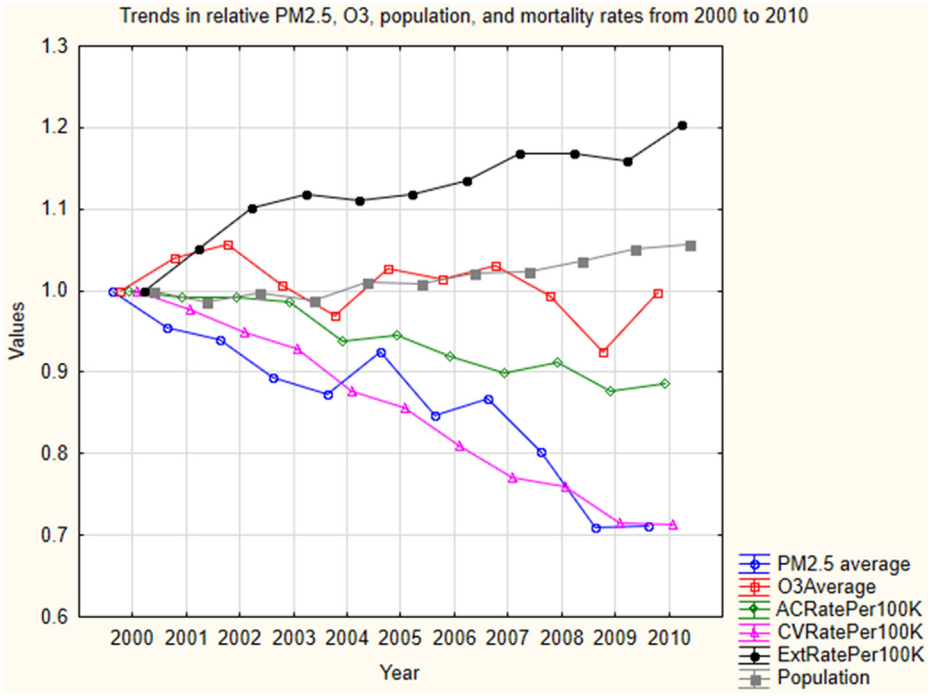

Fig. 1. Trends in relative values of pollutants, mortality rates, and population, 2000 to 2010.

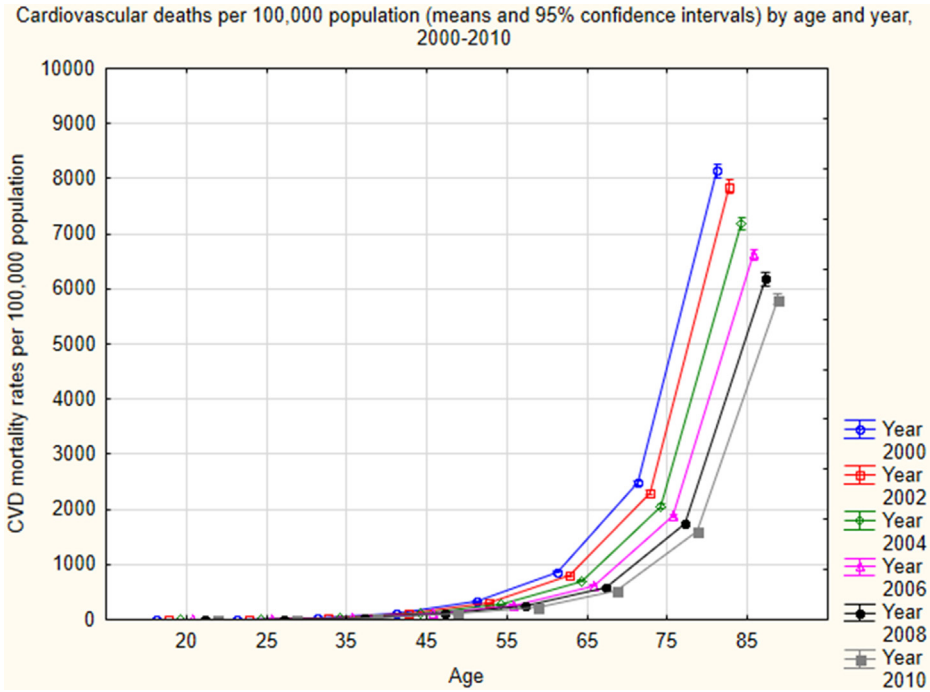

Fig. 2. Declines of age-specific CVD mortality rates over time (top curve is for year 2000, bottom curve is for year 2010).

Figure 3 shows analogous curves for age groups 55 to 64, 65 to 74, 75 to 84, and 85 years or older, abbreviated 55, 65, 75, and 85 years, respectively, for different average PM2.5 levels in 2000 (left) and 2010 (right). At all PM2.5 levels, age-specific mortality rates declined conspicuously from 2000 to 2010. In both years, mortality rates in the oldest age categories were higher at PM2.5 levels of 12 $\mu g/m^3$ than at 3 $\mu g/m^3$, suggesting a possible persistent positive association between PM2.5 concentrations and elderly mortality rates.

There was substantial geographic heterogeneity in both PM2.5 values and CVD mortality rates among the counties in this study, allowing the relation between them to be studied with considerable statistical power despite the smoothing effects of using county-level data [49]. PM2.5 average levels ranged from below 2 to above 20 $\mu g/m^3$, and cardiovascular deaths per 100,000 people per year ranged from close to zero (for younger age groups) to more than 10,000 deaths per 100,000 person-years (for the oldest age group in early years). Even for a single age group (e.g., 75–84-

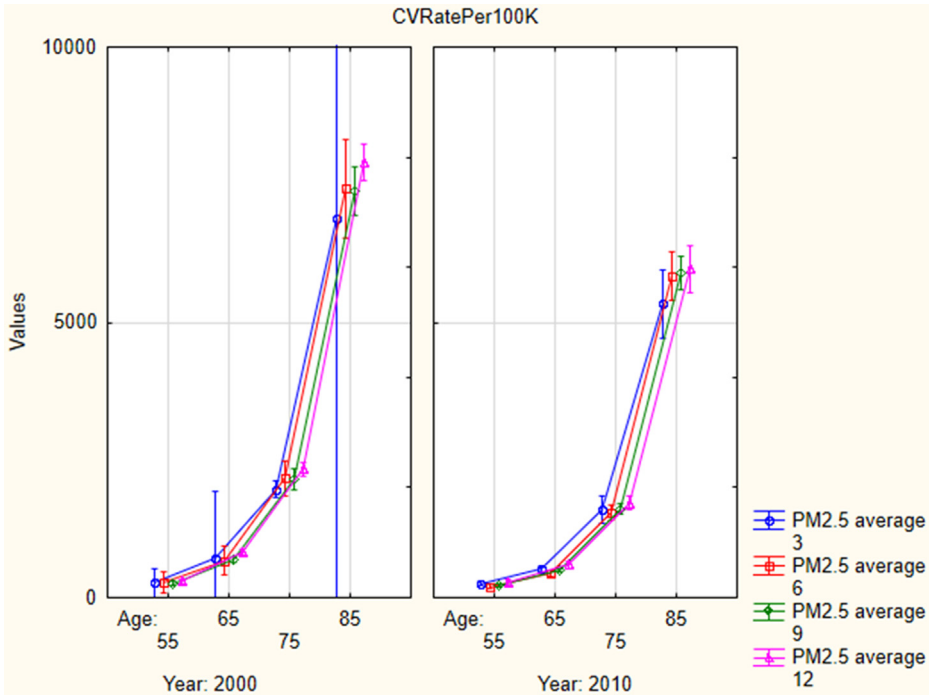

Fig. 3. Decline of older age-specific mortality rates over time (left panel is for year 2000, right panel is for year 2010) for counties with different average PM2.5 levels (microgram per cubic meter).

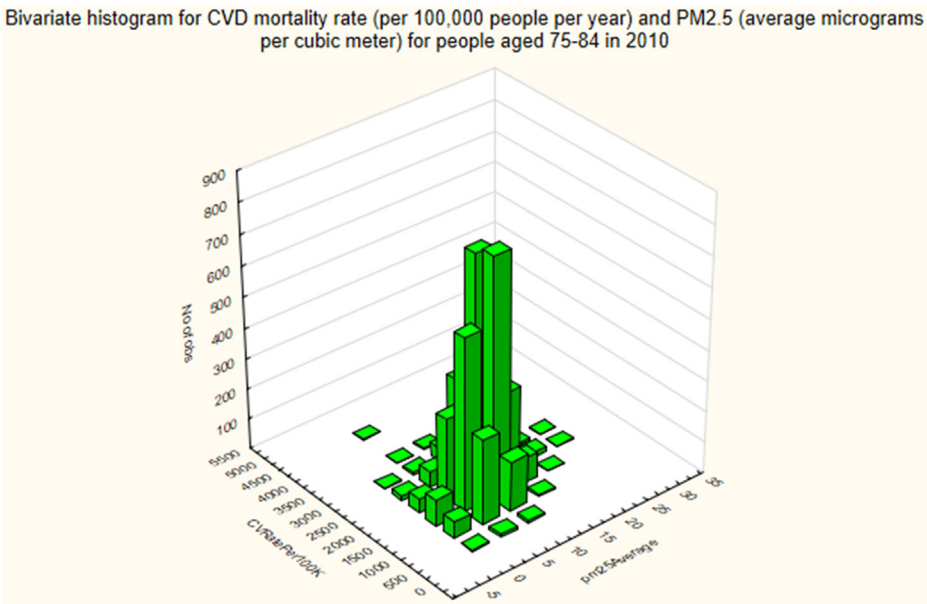

Fig. 4. There is substantial geographic heterogeneity in PM2.5 levels and CVD mortality rates even within a single age group and year (here, 75–84 year olds in 2010).

year-olds) and a single year (2010), there is a greater than fivefold variation in CVD mortality rates and a more than eightfold variation in average PM2.5 levels among counties as shown in Figure 4.

Results on statistical associations between pollutant levels and mortality rates

Table 3 shows the Pearson correlation coefficients between PM2.5 and O_3 levels, county population sizes, and AC, cardiovascular, and external-cause (nondisease) mortality rates, holding year and age fixed at 2010 and 75 to 84 years, respectively. Similar correlations hold for other years. All off-diagonal correlation coefficients in Table 3 are statistically significant from zero ($P < .05$) except for the −.09 correlation between PM2.5 levels and nondisease mortality rates (ExtRatePer100k). Specifically, Table 3 shows the following significant associations:

- PM2.5 and O_3 concentrations are positively associated with each other (correlation $r = 0.28$)
- Both PM2.5 and O_3 concentrations are positively associated with both AC and cardiovascular mortality rates.
- O_3 is also positively associated with nondisease mortality rates but PM2.5 is not. (All positive correlations in Table 3 are significant, but the −0.09 numbers are not.)
- Population size of a county is positively associated with PM2.5 and is negatively associated with O_3 and with all mortality rates.
- All mortality rates (disease related and nondisease related) are positively associated with each other, but negatively associated with population size.

The associations in Table 3 may or may not be causal, but they are not explained by coincident historical trends (since the year is held fixed at 2010) nor by confounding by age category because the age category is also held fixed at 75 to 84 years. Whether confounding by education, income, temperature, or other variables might account for some of these associations—for example, if mortality rates and PM2.5 are both elevated on cold days or in colder regions; or if lower-income families tend to live in more polluted areas and also to have higher age-specific mortality rates irrespective of location—cannot be determined from the exposure and mortality rate data alone.

In multiple linear regression modeling of the association between explanatory variables and elderly (75 to 84-years-old) CVD mortality rate using automated backward stepwise variable selection via *F* tests, only the regression coefficient between PM2.5 and CVD mortality rate, but not O_3 and CVD mortality rate, remains significant. Thus, there is a positive association between PM2.5 levels and CVD mortality rates among the elderly that is not explained by coincident historical trends nor by confounding by age or population or O_3; but the correlations between O_3 and CVD mortality rates and between O_3 and all-disease mortality rate, vanish after conditioning (via multiple linear regression) on PM2.5 and population size for all disease-related mortalities. In short, PM2.5, but not O3, passes this conditional independence test for

Table 3
Pearson correlations between pairs of exposure and response variables for elderly (aged 75–84 years) people in 2010

Variable	Correlations between county-specific average PM2.5 and O_3 concentrations and mortality rates for 75–84-year-olds in 2010						
	Means	PM2.5 average	O_3 average	Population	ACRatePer100K	CVRatePer100K	ExtRatePer100K
PM2.5 average	9.16	1.00	**0.28**	**0.14**	**0.17**	**0.22**	−0.09
O_3 average	0.04	**0.28**	1.00	**−0.20**	**0.30**	**0.014**	**0.20**
Population	15,783.16	**0.14**	**−0.20**	1.00	**−0.33**	**−0.15**	**−0.34**
ACRatePer100K	4855.06	**0.17**	**0.30**	**−0.33**	1.00	**0.72**	**0.38**
CVRatePer100K	1614.13	**0.22**	**0.14**	**−0.15**	**0.72**	1.00	**0.19**
ExtRatePer 100K	137.28	−0.09	**0.20**	**−0.34**	**0.38**	**0.19**	1.00

Off-diagonal correlations in bold differ from 0 at the conventional 5% significance level ($P < .05$).

Table 4
County-specific average PM2.5 concentration is significantly positively associated with county-specific CVD mortality rates across all age categories and years

$N = 21,613$	Regression summary for dependent variable: CVRatePer100K ($R = 0.78$, $R^2 = 0.605$, adjusted $R^2 = 0.605$)					
	b^*	Standard error of b^*	b	Standard error of b	$t(21,608)$	P value
Intercept	—	—	**114,927.7**	**7043.87**	**16.3160**	**.000000**
Year	**−0.08**	**0.0046**	**−60.3**	**3.51**	**−17.2049**	**.000000**
Age	**0.81**	**0.0048**	**120.0**	**0.70**	**170.4133**	**.000000**
PM2.5 average	**0.04**	**0.0047**	**33.6**	**3.54**	**9.4979**	**.000000**
Population	**0.08**	**0.0048**	**0.0**	**0.00**	**17.4577**	**.000000**

Lower and upper 95% confidence limits, −95% CL and +95 CL, approximated as $b \pm 1.96 \times$ standard error of b.
Off-diagonal correlations in bold differ from 0 at the conventional 5% significance level ($P < .05$).

being a potential causal driver of elderly mortality rates. Similarly, for all age categories and years, PM2.5 average levels, but not O_3 levels, help to predict CVD mortality rates.

Table 4 shows the results of a multiple linear regression with backward stepwise variable selection; results were also confirmed in multiple disjoint random samples (20% cross-validation samples). The b^* column contains standardized regression coefficients (scaling each variable in terms of standard deviations) and the b column contains the unstandardized regression coefficients. As expected, year is negatively associated with CVD mortality risk, and age is positively associated with CVD mortality risk. Age is quantitatively by far the most important predictor of risk. PM2.5 average concentration makes the smallest, but still highly statistically significant ($P < .000001$), contribution to predicting CVD values. Population (specific to each county and age group) is also a significant predictor of CVD risk. Results for all-disease–related mortality (AC) risks are similar, with the standardized regression coefficient for PM2.5 increasing to 0.06, with the exception that both O3 and population size are significantly negatively associated with AC mortality rates (standardized regression coefficients of −0.12 for population and −0.02 for O_3). Interpretively, the coefficient for PM2.5 in Table 4 ($b = 33.6$) indicates that CVD mortality risk increases by 33.6 deaths per 100,000 person-years for each microgram per cubic meter increase of PM2.5 in air, assuming other variables are held constant. The mean CVD mortality rate averaged over all age categories and years is 1931.6 deaths per 100,000 person-years, so a change in PM2.5 of 10 $\mu g/m^3$ corresponds to a change in CVD mortality rate of approximately (10 $\mu g/m^3$) × (33.6 deaths per 100,000 person-years per $\mu g/m^3$)/(1931.6 deaths per 100,000 person-years) = 336/1931.6 = 17.4%. This slope factor could be described as a 17.4% increase in mortality per 10 $\mu g/m^3$increase in PM2.5 concentration.

Results on correlations between changes in variables over time

Tables 5 and 6 show correlations between changes in AC mortality, CVD mortality, and nondisease mortality, respectively (the columns), and different possible predictors (the rows), for all counties included in the study. Table 5 presents results for the 75- to 84-year-old group, and Table 6 repeats the analysis for all age groups.

For the 75- to 85-year-old age category, changes in AC and CVD mortality rates are significantly positively correlated with each other, as expected, and with changes in external-cause mortality rates. They are significantly negatively correlated with increases in population. Neither is significantly correlated with changes in PM2.5 or changes in O_3. For all age groups, changes in PM2.5 are significantly but weakly positively correlated with changes in external-cause mortality rates. Changes in O_3 are significantly positively correlated both with changes in AC mortality rates and with changes in CVD mortality rates. Increases in population are significantly correlated with reductions in all mortality rates.

Table 5
Pearson correlations between changes in variables from 2000 to 2010 for elderly (aged 75–84 years) people

Variable	Correlates of changes in mortality rates from 2000 to 2010 for 75–84-year-olds		
	Delta AC mortality	Delta CVD mortality	Delta external rate
Delta PM2.5	−0.07	−0.08	0.04
Delta O_3	0.03	0.03	0.06
Delta population	**−0.59**	**−0.56**	**−0.44**
Delta AC mortality	1.00	**0.99**	**0.81**
Delta CVD mortality	**0.99**	1.00	**0.79**
Delta external rate	**0.81**	**0.79**	1.00
PM2.5 average	−0.10	−0.12	−0.10
O_3 average	−0.04	−0.04	**−0.18**
Population 2010	0.07	0.07	0.06
ACRatePer100K	−0.05	−0.07	−0.07
CVRatePer100K	0.06	0.06	−0.02
ExtRatePer100K	−0.16	**−0.17**	0.14

Off-diagonal correlations in bold differ from 0 at the conventional 5% significance level ($P < .05$).

In multivariate analysis using multiple linear regression, changes in both AC and CVD mortality rates are conditionally independent of changes in both PM2.5 and O_3, given changes in population size, changes in external-cause mortality rates, and age in 2010. These three explanatory variables are automatically selected by backward stepwise variable selection, whereas changes in PM2.5 and O_3 are dropped as they provide no additional information useful for predicting the AC or CVD mortality rates. Thus, by this criterion, changes in PM2.5 and O_3 levels do not help to predict or explain changes in CVD or AC mortality rates, undermining a causal interpretation of the positive associations between them in the cross-sectional analysis in Table 3.

Other, perhaps unexpected, correlations between changes in variables in Table 6 include a strong positive correlation (0.59) between changes in external-cause mortality rates and changes in CVD mortality rates; and positive correlations between baseline levels of mortality rates and changes in their levels. Thus, relatively high-risk areas in 2000 tended to become more risky by 2010. As expected, older age categories saw relatively large reductions in disease mortality rates (but increases in nondisease mortality rates).

Table 6
Pearson correlations between changes in variables from 2000 to 2010 for all age groups

Variable	Correlates of changes in mortality rates from 2000 to 2010		
	Delta AC mortality	Delta CVD mortality	Delta external rate
Delta PM2.5	−0.00	−0.01	**0.06**
Delta O_3	**0.06**	**0.08**	0.02
Delta population	**−0.15**	**−0.15**	**−0.17**
Delta AC mortality	1.00	**0.95**	**0.69**
Delta CVD mortality	**0.95**	1.00	**0.59**
Delta external rate	**0.69**	**0.59**	1.00
Age 2000	−0.03	**−0.14**	**0.24**
PM2.5 average 2000	−0.02	−0.04	−0.04
O_3 average 2000	**−0.05**	**−0.07**	−0.05
ACRatePer100K 2000	**0.147**	−0.01	**0.04**
CVRatePer100K 2000	**0.23**	0.02	**0.04**
ExtRatePer100K 2000	**0.21**	−0.03	**0.56**

Off-diagonal correlations in bold differ from 0 at the conventional 5% significance level ($P < .05$).

Granger causality test and control results

Granger tests using standard time series regression models with maximum lags of 1, 2, or 3 years show that, for all age categories tested (65–74, 75–84, and 85 years or older) and for all mortality outcomes considered (CVD, all-disease, and external-cause mortality rates), both PM2.5 and O_3 histories are not useful for predicting mortality rates in most (over 90%) of the counties. PM2.5 and O_3 have predictive coefficients for CVD and all-disease mortality rates that are significantly different from zero in only a small minority of counties (7% for AC mortality, 6% for CVD mortality, and 7% for external-cause mortality, which was used as a negative control), roughly consistent with, although slightly higher than, the 5% false-positive error rate that might occur by chance due to the 5% significance level used in the tests. (For 483 counties and a true false-positive rate of 5%, there is about a 26% probability that the sample proportion of false positives would exceed 6% or be less than 4% by chance.) Perhaps more importantly, the negative control (external-cause mortalities) also shows that O_3 and PM2.5 histories on time scales of several years are not Granger-causes of CVD or all disease-related deaths any more than they are of external-cause deaths. For example, the age group and lag with the highest fraction of Granger-positive associations between PM2.5 and CVD rate is the 85+ years age group with a lag of 1 year: this fraction is 11%. But the corresponding fraction for Granger-positive associations between PM2.5 and external-cause mortalities is greater, at 14%. Thus, the Granger tests do not support a conclusion of a genuine causal effect, that is, positive results clearly above what might occur by chance and what is found for the negative controls.

Given the well-known limitations of *P* values and significance testing, it may also be useful to consider that, if pollutant levels were detectable causal drivers of increased mortality rates at recent historical levels, then this causal relation should have been visible in a large majority of counties. The fractions in Table 7 might all be expected to exceed 50% in the presence of clear Granger causality, that is, most counties should have shown evidence of a Granger-positive association between PM2.5 and mortality rates caused by them. Intuitively, as suggested by Figure 1, although pollutant levels declined substantially in most counties from 2000 to 2010, declines in CVD and AC mortality rates did not appear to proceed more quickly when PM2.5 declined quickly than when it did not or than when it increased. The Granger test results confirm this suggestion at the level of individual counties and for time lags of 1 to 3 years.

Positive controls: does absence of evidence constitute evidence of absence?

Might the absence of a significant association between county-specific changes in PM2.5 levels and changes in mortality rates between 2000 and 2010, shown in Tables 5 and 6 and in corresponding multiple linear regression models, be due to limited statistical power to detect changes in the presence of substantial heterogeneity and variability in the data? To check the statistical power of these methods, we modified the observed data by adding a known "signal"—a 2.6% decrease in CVD mortality rate per microgram per cubic meter decrease in PM2.5 concentration based on the slope estimate of Lepeule [47]. We then tested whether this known signal is detectable through the noise in the data using the methods we have applied.

Table 8 shows the results of multiple linear regression applied to the artificial data set with a simulated known causal impact of exposure. The simulated effect of changes in PM2.5 on changes in CVD mortality rates based on the 2.6% slope coefficient for change in mortality rate per microgram per cubic meter change in PM2.5 was successfully detected. (All predictors remain significant using backward stepwise variable selection.) This suggests that an effect of this size would probably have been detected in the real data if it had been present. This type of positive control gives some reassurance that the substantial variability and heterogeneity in county-level time series data would not hide causal effects of the sizes that have sometimes been estimated from standard associational (regression-based) models by assuming that slope coefficients are causal, if such causal effects were actually present.

Finally, we briefly examined the geographic distribution of associations. Previous investigators have reported that chronic exposure to PM2.5 is associated with mortality in the eastern and central regions of the United States, but not in the Western region [50]. In our data set, for the main elderly population (75–84-year-olds) in 2010, PM2.5 was statistically significantly positively correlated with CVD mortality in Florida and overall in pooled data from counties in all states. It was statistically negatively correlated with all-disease (AC) mortality rate in Arizona and statistically positively correlated with AC mortality rate in Florida and overall. Otherwise, state-specific correlations in 2010 were not individually statistically significant at the conventional 0.05 significance level and were a mix of nonsignificant positive and negative correlations with no obvious geographic distribution.

Discussion and conclusions: caveats for causal interpretations of regression coefficients

The epidemiologic and risk assessment literature on human health effects of air pollution contains dozens of studies that attribute reductions in mortality risks to reductions in air pollution levels and that estimate the slope of the concentration-response association between exposures to pollutants and corresponding mortality rates [12,20,22,47,51,52]. The work reported here

Table 7
Fractions of counties with positive Granger causality tests for PM2.5 and AC, CVD, and external-cause mortality rates, for different age groups and lags (1–3 years)

Age/lag	AC mortality rate	CVD rate	External rate
65	**0.06**	**0.06**	**0.07**
1	0.09	0.08	0.10
2	0.04	0.05	0.10
3	0.05	0.05	0.02
75	**0.08**	**0.06**	**0.06**
1	0.10	0.08	0.05
2	0.08	0.06	0.08
3	0.04	0.05	0.06
85	**0.08**	**0.06**	**0.08**
1	0.15	0.11	0.14
2	0.06	0.03	0.09
3	0.04	0.03	0.01
Overall	0.07	0.06	0.07

Results averaged over all three lags are shown in bold.

Table 8
Multiple linear regression detects simulated PM2.5 effects on mortality rates of the sizes predicted from previously published regression slope coefficients [47]

$N = 1425$	Regression of CVD mortality rate with simulated effect of PM2.5 $R = 0.76465$, $R^2 = 0.58469$, adjusted $R^2 = 0.5838$ $F(3, 1421) = 06966.84$, $P < .0000$, standard error of estimate: 1178.0					
	b^*	Standard error of b^*	b	Standard error of b	$t(1421)$	P value
Intercept	—	—	**541.5**	**74.8**	**7.2**	**.000**
Delta PM2.5	**0.04**	**0.017**	**37.9**	**16.5**	**2.3**	**.022**
CVRatePer199K 2000	**−0.75**	**0.017**	**−0.5**	**0.0**	**−43.9**	**.000**
Delta population	**−0.16**	**0.017**	**−0.0**	**0.0**	**−9.2**	**.000**

Lower and upper 95% confidence limits, −95% CL and +95 CL, approximated as $b \pm 1.96 \times$ standard error of b.

contributes a new data set to this literature. It supports previous findings of positive PM2.5-mortality associations based on PM2.5 (and O_3) and age-specific mortality data, in county-level data from the 15 largest US states over the years from 2000 to 2010. Confirming earlier studies such as Lepeule et al. [47], we found a statistically significant positive association between PM2.5 (and also O_3) concentrations and both all-disease related and CVD mortality rates, as well as a significant positive association between O_3 and external-cause mortalities, which we used as a negative control (Tables 3 and 4).

However, such associations between historical levels of exposure and response variables do not necessarily describe causal relations. In our examination of historical changes in pollutant levels and mortality rates (Tables 5 and 6 and multiple regression models and Granger causality tests), actual changes in PM2.5 and O_3 levels over time did not significantly help to predict or explain corresponding observed changes in all-disease or CVD mortality rates over time. This argues against facile causal interpretations of the significant statistical associations between pollution levels and mortality rates. Such causal interpretations of slope coefficients are commonly made in air pollution health effects (and other) epidemiology. For example, the study of Lepeule et al. [47], updating the influential Harvard Six Cities Study, offers the important causal interpretation that "These results [i.e., that each 10 $\mu g/m^3$ increase in PM2.5 was associated with a 26% increase in cardiovascular mortality risk] suggest that further public policy efforts that reduce fine particulate matter air pollution are likely to have continuing public health benefits." But such policy-relevant causal conclusions are unwarranted if the exposure-response association discussed is not a causal relation, and if the changes referred to are only the hypothetical ones implied by a slope coefficient, rather than actual changes in the levels of exposure and mortality time series.

Study limitations

Our study and conclusions have several limitations. Although our analysis of county-level data does not provide evidence that the roughly 30% reduction in PM2.5 levels from 2000 to 2010 (Fig. 1) caused any detectable effect on disease-related mortality rates, it remains possible that such an effect was present that is too small to detect. For example, if each 10 $\mu g/m^3$ change in PM2.5 concentration causes only a 1.03% change in CVD mortality rate, as estimated by Dai et al. [52], then the power of our data set would not be great enough to distinguish this from zero. In addition, like many other studies, our analysis lacked individual-level exposure data. Our basic units of observation are death counts, by cause, within age categories, years, and counties; finer resolution would require a different data set. Age and death are available at the individual level, making this a semi-individual design [53], rather than a purely ecological design; but other individual covariates are not available. On the other hand, the fact that we follow the same counties over multiple years contributes one of the strengths of a panel study design: the effects of fixed (or slowly changing) possible confounders or effect modifiers, such as differences in income or education or regional climate, cancel out when changes (deltas) in mortality rates are calculated for the same locations in successive years. In addition, our study substantially meets several criteria [49] for useful ecological studies: marked variation across geographic units (counties); unlikely confounding (due to the longitudinal panel design, in which counties serve as their own controls for purposes of subtracting out fixed effects of confounders when computing changes over time); opportunities to include negative controls (external-cause mortalities); and simulated positive controls (via simulation of postulated causal impacts).

A remaining question is, if the significant associations between PM2.5 and O_3 on the one hand and CVD and AC mortality on the other are not due to a causal relationship between pollutant exposure and disease, then what does explain them? Our analyses have ruled out coincident trends (because the associations hold even within single years) and chance (because the correlation and regression coefficients reported are statistically significant), as well as fixed confounders (because of the panel design) as plausible explanations. Possible confounders that might covary with exposure levels over time and thus offer explanations, range from co-pollutants to temperature (e.g., if very hot or very cold areas have higher levels of PM2.5, perhaps due in part to coal-fired power plants that power air conditioning or heating, and independently have higher mortality rates). Attaching more variables to the county-specific mortality rate and pollution level data, such as daily temperature (high and low), could potentially help to answer this question. But at present, the answer is unknown.

Finally, by focusing on changes in annual average pollutant levels and mortality rates at the individual county level, we have foregone opportunities to model or "adjust" for effects of seasonality, more granular spatial variations, and measured or latent confounders. Dominici et al. [9] suggest that it is not uncommon for different regression models based on different modeling choices and assumptions to produce very different answers. For example, regression coefficients that are significantly positive in one model may be significantly negative in another depending on which variables and interaction terms are included. By using several different approaches (conditional independence tests, Granger tests, positive and negative controls, and automated variable selection) and relatively simple measures of association (correlations and linear regression coefficients, fractions of counties with Granger-positive associations) computed using standard widely available software for all tests, we have sought conclusions that are more robust and objective, minimizing opportunities for manual intervention to shape the results.

Comparisons with conclusions from other studies

The coefficient for PM2.5 in Table 4 ($b = 33.6$), corresponding to a 17% increase in mortality per 10 $\mu g/m^3$ increase in PM2.5 concentration, is well within the range of other recent association-based estimates based on regression relations. For example Dai et al. [52], in a study of 75 US cities between 2000 and 2006, reported a 1.03% (95% CI: 0.65%–1.41%) increase in CVD mortality with each 10 $\mu g/m^3$ increase in PM2.5 averaged over a 2-day period. In their update of the Harvard Six Cities Study, Lepeule et al. [47] estimated a 26% (95% CI: 14%–40%) increase in CVD mortality for each 10 $\mu g/m^3$ increase in PM2.5, averaged over the three prior years. Thus, our value of 17.4% falls between these two estimates and is within the 95% CI of the Lepeule et al. [47] study. Some other recent studies have not detected clearly significant associations between PM2.5 levels and most CVD or AC mortality rates [54] or found no association between local trends in mortality and local trends in yearly average PM2.5 after adjusting for national trends and local differences [55]. For the US county data set we have analyzed, our main conclusions are that (a) there are statistically significant associations between PM2.5 and both all-disease and CVD mortality risks; but (b) there is no clear evidence of a causal relation between PM2.5 and O_3 concentration levels and mortality rates. These results differ both from studies that do not find clear associations and also from some authoritative opinions, including views in an Expert Elicitation Study [56], that statistically significant exposure-response associations between PM2.5 and CVD mortality are probably causal.

Although our results do not support some previous expert judgment-based assessments of causality, this is consistent with studies showing that firmly expressed opinions of key experts [23] about air pollution health effects associations being causal have later proved to be unwarranted [26]. The practice of applying human judgment using weight-of-evidence considerations to measures of association (such as relative risks, odds ratios, population attributable fractions, burden-of-disease estimates, and regression coefficients) to determine whether an inference of causality is supported has been widespread in epidemiology, although some methodologists [57] have argued that logically valid causal inferences cannot be derived from such associations in purely observational studies without interventions. This makes natural experiments, where interventions such as pollution reductions occur differently for different subpopulations, potentially valuable aids to understanding causation.

The calculations in this article illustrate that a significant positive association between historical levels of PM2.5 and historical mortality rates does not necessarily provide a sound basis for inferring a positive association between changes in levels of PM2.5 and changes in mortality rates. This methodological point confirms the importance of using QEs or other appropriate formal methods of causal study design and analysis (Table 2) to draw causal conclusions. Free publicly available data sets such as the U.S. Environmental Protection Agency and Centers for Disease Control and Prevention data sets used in this study and free publicly available software such as R and Python now make it relatively easy to test whether changes in PM2.5 and O_3 help to predict changes in disease mortality rates on time scales from days to over a decade. The data files and software scripts used in this study are freely available from the authors on request. We hope that this will encourage others to investigate further the relation between longitudinal changes in pollutant levels and changes in mortality rates and to clarify the crucial distinction between positive statistical associations and evidence of causality in air pollution health effects epidemiology.

Acknowledgments

The research described here was supported in part by the American Petroleum Institute and by Cox Associates, LLC. The research questions asked, methods used, and conclusions reached are solely those of those the authors. The authors thank reviewers for their comments and suggestions, which improved the clarity of the exposition.

References

[1] Lehrer J. Trials and errors: Why science is failing us. Wired. 2012. Available from: http://www.wired.co.uk/magazine/archive/2012/02/features/trials-and-errors?page=all. Accessed October 15 2014.
[2] Sarewitz D. Beware the creeping cracks of bias. Nature 2012;485:149.
[3] Ottenbacher KJ. Quantitative evaluation of multiplicity in epidemiology and public health research. Am J Epidemiol 1998;147:615–9.
[4] Imberger G, Vejlby AD, Hansen SB, Møller AM, Wetterslev J. Statistical multiplicity in systematic reviews of anesthesia interventions: a quantification and comparison between cochrane and non-cochrane reviews. PLoS One 2011;6(12):e28422.
[5] Cox Jr LA. Improving causal inference in risk analysis. Risk Anal 2013;33(10):1762–71.
[6] Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol 2002;31(5):1030–7.
[7] Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2(8):e124.
[8] The Economist (2013). Trouble at the lab: scientists like to think of science as self-correcting. To an alarming degree, it is not. 2013. Available from: www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble. Accessed October 15 2014.
[9] Dominici F, Greenstone M, Sunstein CR. Particulate matter matters. Science 2014;344:257–8.
[10] Yim SH, Barrett SR. Public health impacts of combustion emissions in the United Kingdom. Environsci Technol 2012;46(8):4291–6.
[11] NHS. Air pollution 'kills 13,000 a year' says study. 2012. Available from: www.nhs.uk/news/2012/04april/Pages/air-pollution-exhaust-death-estimates.aspx. Accessed October 15 2014.
[12] Fann N, Lamson AD, Anenberg SC, Wesson K, Risley D. Hubbell. Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. Risk Anal 2012;32(1):81–95.
[13] EPA (U.S. Environmental Protection Agency). The benefits and costs of the clean air act from 1990 to 2020. Final report – Rev. A. Washington D.C.: Office of Air and Radiation; 2011.
[14] Krstić G. Apparent temperature and air pollution vs. elderly population mortality in Metro Vancouver. PLoS One 2011;6(9):e25101.
[15] Kelly F, Armstrong B, Atkinson R, Anderson HR, Barratt B, Beevers S, et al, HEI Health Review Committee. The London low emission zone baseline study. Res Rep Health Eff Inst 2011;(163):3–79.
[16] Powell H, Lee D, Bowman A. Estimating constrained concentration–response functions between air pollution and health. Environmetrics 2012;23(3):228–37.
[17] Buka I, Koranteng S, Osornio-Vargas AR. The effects of air pollution on the health of children. Paediatr Child Health 2006;11(8):513–6.
[18] Friedman MS, Powell KE, Hutwagner L, Graham LM, Teague WG. Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma. JAMA 2001;285(7):897–905.
[19] Peel JL, Klein M, Flanders WD, Mulholland JA, Tolbert PE, Health Effects Institute (HEI). Impact of improved air quality during the 1996 summer olympic games in Atlanta on multiple cardiovascular and respiratory outcomes. HEI Research Report #148. Boston, MA: Health Effects Institute; 2010. Available from: http://pubs.healtheffects.org/getfile.php?u=564.
[20] Pope 3rd CA. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. Am J Public Health 1989;79(5):623–8.
[21] Lamm SH, Hall TA, Engel E, White LD, Ructer FH. PM 10 particulates: are they the major determinant in pediatric respiratory admissions in Utah County, Utah (1985-1989)? Ann Occup Hyg 1994;38:969–72.
[22] Clancy L, Goodman P, Sinclair H, Dockery DW. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. Lancet 2002;360(9341):1210–4.
[23] Harvard School of Public Health. Press release: "ban on coal burning in Dublin cleans the air and reduces death rates". 2002. Available from: www.hsph.harvard.edu/news/press-releases/archives/2002-releases/press10172002.html. Accessed October 15 2014.
[24] Wittmaack K. The big ban on bituminous coal sales revisited: serious epidemics and pronounced trends feign excess mortality previously attributed to heavy black-smoke exposure. InhalToxicol 2007;19(4):343–50.
[25] Pelucchi C, Negri E, Gallus S, Boffetta P, Tramacere I, La Vecchia C. Long-term particulate matter exposure and mortality: a review of European epidemiological studies. BMC Public Health 2009;9:453.
[26] Health Effects Institute (HEI). Did the Irish coal bans improve air quality and health?. Boston, MA, USA: Health Effects Institute; 2013. Available from: http://pubs.healtheffects.org/getfile.php?u=929. Accessed February 1 2014.
[27] Harvard Law Today. 2014. Available from: http://today.law.harvard.edu/improving-the-pollution-mortality-link/. Accessed October 15 2014.
[28] Harris AD, McGregor JC, Perencevich EN, Furuno JP, Zhu J, Peterson DE, et al. The use and interpretation of quasi-experimental studies in medical informatics. J Am Med Inform Assoc 2006;13(1):16–23.
[29] Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally; 1966.
[30] Maclure M. Taxonomic axes of epidemiologic study designs: a refutationist perspective. J Clin Epidemiol 1991;44(10):1045–53.
[31] Rothman KJ, Greenland S, Lash TL, editors. Chapter 2 of modern epidemiology. 3rd ed. Philadelphia: Lippincott-Wolters-Kluwer; 2008.
[32] Angrist JD, Pischke J-S. Mostly harmless econometrics: an empiricist's companion. Princeton, NJ: Princeton University Press; 2009.
[33] Freedman DA. Graphical models for causation, and the identification problem. Eval Rev 2004;28(4):267–93.
[34] Friedman N, Goldszmidt M. Learning Bayesian networks with local structure. In: Jordan MI, editor. Learning in Graphical Models. Cambridge, MA: MIT Press; 1998. p. 421–59.
[35] Stebbings Jr JH. Panel studies of acute health effects of air pollution. II. A methodologic study of linear regression analysis of asthma panel data. Environ Res 1978;17(1):10–32.
[36] Eichler M, Didelez V. On Granger causality and the effect of interventions in time series. Lifetime Data Anal 2010;16(1):3–32.
[37] Helfenstein U. The use of transfer function models, intervention analysis and related time series methods in epidemiology. Int J Epidemiol 1991;20(3):808–15.
[38] Gilmour S, Degenhardt L, Hall W, Day C. Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: a general method and example. BMC Med Res Methodol 2006;6:16.
[39] Moore KL, Neugebauer R, van der Laan MJ, Tager IB. Causal inference in epidemiological studies with strong confounding. Stat Med 2012;31(13):1380–404.
[40] Hack CE, Haber LT, Maier A, Shulte P, Fowler B, Lotz WG, et al. A Bayesian network model for biomarker-based dose response. Risk Anal 2010;30(7):1037–51.

[41] Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2010;21(3):383–8.
[42] Hernan MA, Robins JM. (2011, 2015). Causal inference. Chapman and Hall/VCFC Press (forthcoming). Available from: www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.
[43] Centers for Disease Control and Prevention (CDC). Wonder "Compressed Mortality, 1999–2010amprdquosemicolon database. 2014. Available from: http://wonder.cdc.gov/cmf-icd10.html. Accessed October 15 2014.
[44] U.S. Environmental Protection Agency (EPA), 2014. Available from: www.epa.gov/airdata/ad_data_daily.html. Accessed October 15 2014.
[45] Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58(5):295–300.
[46] Weed DL. Meta-analysis and causal inference: a case study of benzene and non-Hodgkin lymphoma. Ann Epidemiol 2010;20(5):347–55.
[47] Lepeule J, Laden F, Dockery D, Schwartz J. Chronic exposure to fine particles and mortality: an extended follow-up of the Harvard six cities study from 1974 to 2009. Environ Health Perspect 2012;120(7):965–70.
[48] Yang Y, Li R, Li W, Wang M, Cao Y, Wu Z, et al. The association between ambient air pollution and daily mortality in Beijing after the 2008 Olympics: a time series study. PLoS One 2013;8(10):e76759.
[49] Savitz DA. Commentary: a niche for ecologic studies in environmental epidemiology. Epidemiology 2012;23(1):53–4.
[50] Zeger SL, Dominici F, McDermott A, Samet JM. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000-2005). Environ Health Perspect 2008;116(12):1614–9.
[51] Cesaroni G, Badaloni C, Gariazzo C, Stafoggia M, Sozzi R, Davoli M, et al. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. Environ Health Perspect 2013;121(3):324–31.
[52] Dai L, Zanobetti A, Koutrakis P, Schwartz JD. Associations of fine particulate matter species with mortality in the United States: a multicity time-series analysis. Environ Health Perspect 2014;122(8):837–42.
[53] Künzli N, Tager IB. The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies. Environ Health Perspect 1997;105(10):1078–83.
[54] Beelen R, Stafoggia M, Raaschou-Nielsen O, Andersen ZJ, Xun WW, Katsouyanni K, et al. Long-term exposure to air pollution and cardiovascular mortality: an analysis of 22 European cohorts. Epidemiology 2014;25(3):368–78.
[55] Greven S, Dominici F, Zeger S. An approach to the estimation of chronic air pollution health effects using spatio-temporal information. J Am Stat Assoc 2011;106(494):396–406.
[56] EPA. Expanded expert judgment assessment of the concentration-response relationship between PM2.5 exposure and mortality. 2006. Available from: www.epa.gov/ttn/ecas/regdata/Uncertainty/pm_ee_report.pdf. Accessed October 15 2014.
[57] Ward AC. The role of causal criteria in causal inferences: Bradford Hill's "aspects of association". Epidemiol Perspect Innov 2009;6:2.