Chapter 1
Scientific Method for Health Risk Analysis: The Example of Fine Particulate Matter Air Pollution and COVID-19 Mortality Risk

Introduction: Scientific Method for Quantitative Risk Assessment

Applied science is largely about how to use observations to learn, express, and verify predictive generalizations—causal laws stating that if certain antecedent conditions hold, then certain consequences will follow. Non-deterministic or incompletely known causal laws may only determine conditional probabilities or occurrence rates for consequences from known conditions (Spirtes 2010). For example, different exposure concentrations of air pollution might cause different mortality incidence rates or age-specific hazard rates for people with different values of causally relevant covariates. A defining characteristic of sound science is that causal laws and their predictions are formulated and expressed unambiguously, using clear operational definitions, so that they can be independently tested and verified by others and empirically confirmed, refuted, or refined as needed using new data as it becomes available. Comparing unambiguous predictions to observations (using statistics if the predictions are probabilistic) determines the extent to which they are empirically supported. The authority of valid scientific conclusions rests on their testability, potential falsifiability, and empirically demonstrated predictive validity when tested. Using new data to constantly question, test, verify, and if necessary correct and refine previous predictive generalizations, and wider theories and networks of assumptions into which they may fit, is a hallmark of sound science. Its practical benefit in risk analysis is better understanding of what truly protects people, and what does not—for example, the unexpected discovery that administering retinol and beta carotene to subjects at risk of lung cancer increased risk instead of decreasing it (Omenn et al. 1996; Goodman et al. 2004).

Recent proposals to apply this concept of science at the United States Environmental Protection Agency (EPA) focused on the following proposed normative principles for better understanding and communicating human health
effects associated with air pollution (Clean Air Scientific Advisory Committee (CASAC) 2019).

1. Consider all of the relevant, high-quality data before forming conclusions. For example, if some studies report that reducing current ambient levels of particulate air pollution substantially reduces all-cause mortality rates (Clancy et al. 2002), but other, equally good (or better) studies report that large reductions in pollution levels have produced no detectable effect on all-cause mortality rates (Dockery et al. 2013), then consider both sets of studies before drawing conclusions about how changing air pollution levels affects public health. Acknowledge apparent discrepancies in available data (Burns et al. 2020). Seek to understand and resolve them (Zigler and Dominici 2014). Such openness to observations, even when—or especially when—they conflict with what has previously been believed, helps to overcome confirmation bias and to learn how the world works (Kahneman 2011).

2. Provide explicit, independently verifiable derivations of conclusions from data. Sound scientific conclusions are derived from observations by explicit reasoning that can be checked by others. This is often done by testing null hypotheses implied by proposed causal laws (or implied by theories comprised of networks of such laws). For example, if exposure to a substance is hypothesized to be a cause of increased mortality risk, then testing the null hypothesis that mortality risk is conditionally independent of exposure, given the values of other variables (such as income, daily high and low temperatures in the weeks prior to mortality, or co-exposures and co-morbidities) and showing that the data allow confident rejection of this null hypothesis would support the causal hypothesis. Failing to reject the null hypothesis would not support it. Explicitly stating the null hypotheses tested, the data and procedures (and, where necessary, any assumptions) used to test them, and the results of the tests provides a dispassionate basis for stating, verifying, and defending conclusions. There is neither room nor need in this procedure for authoritative holistic judgments, which are often not as readily independently testable or falsifiable, as a basis for conclusions. Rather, the validity and credibility of conclusions rests on the data and methods used to test them, independent of the credentials, expertise, authority, beliefs, or preferences of those making them.

3. Use terms with clear conceptual and operational definitions in stating conclusions. For example does a determination that an observed statistical association between exposure to fine particulate matter (PM2.5) and mortality risk is “causal” imply that reducing PM2.5 alone would be sufficient to reduce mortality rates, other conditions remaining fixed (“manipulative causation”); or does it simply mean that the association satisfies some or all of the Bradford Hill considerations (see Chap. 9) for associations (“attributive causation”)? (Neither of these implies the other, as discussed in more detail in Chaps. 9 and 15.) Must mortality risk be reduced by at least a certain minimal amount if exposure is removed for an association to be described as causal, or is no such minimum effect size intended by the description? More generally, those offering scientific
advice to policy makers and the public should define exactly what they mean by key terms, such as “causal” or “likely to be causal,” used to communicate conclusions to policy makers. They should specify how observations can be used to determine whether these terms hold.

4. Most importantly, show results from empirical tests of causal conclusions (predictive generalizations) implied by causal theories or models against observational data. That is, do not simply report causal conclusions at the end of a study, but test them against data and report the results. Causal determinations, conclusions, and interpretations of observations should be tested systematically for both external validity (can the specific causal conclusions from individual studies be synthesized and generalized to yield valid general causal conclusions that apply across studies and to new conditions or interventions?) and internal validity (are stated causal conclusions implied by, or at least consistent with, the data from the specific studies that generated the data? Are rival (non-causal) conclusions refuted by data?) (Campbell and Stanley 1963). The results of these tests should be reported wherever they are available. Technical methods for systematically testing internal validity have been well developed since the early 1960s for quasi-experiments and other observational studies, using relevant comparison groups and adjustment sets (Campbell and Stanley 1963; Textor et al. 2016). Methods for assessing external validity and identifying externally valid (“transportable”) causal conclusions from data have been developed more recently, based largely on conditional probabilities and empirical constraints that are invariant across settings and interventions (Schwartz et al. 2011; Bareinboim and Pearl 2013; Triantafillou and Tsamardinos 2015; Peters et al. 2016).

5. Correctly qualify causal conclusions and interpretations. Clearly identify any untested or unverified assumptions used in deriving stated causal conclusions, and state conclusions as being contingent on the validity of these assumptions. If a body of evidence is consistent with several alternative causal interpretations, perhaps under alternative assumptions, then present all of them; do not select just one. Where possible, use observations from multiple studies to rule out rival potential explanations for observations, narrowing the set of potential valid causal conclusions and interpretations that are consistent with the entire body of observations from multiple studies and populations (Campbell and Stanley 1963; Triantafillou and Tsamardinos 2015; Cox Jr. 2018). Do not characterize observations as supporting one interpretation, such as that exposure to air pollution causes increased mortality, if they support alternative interpretations at least as well, such as that both air pollution and mortality rates are correlated over time due to coincident trends, or that confounders (such as poverty, or daily temperature extremes in the week or two before death) explain the observed association between them.

This book describes and illustrates methods for putting these principles into practice. Chapters 2, 3, 4, 5, and 6 discuss methods for estimating and simulating exposure-related health risks, focusing primarily on events that take place within the human body because of exposure to air pollutants or chemicals in air—pharmacokinetics,
pharmacodynamics, and induction of disease processes. In addition to such “micro” analyses of changes within an individual, Chap. 2 also introduces methods for “macro” analyses that estimate changes in the health risks of exposed populations using epidemiological data. These methods are discussed and applied further in Parts 2 and 3 of the book. Chapter 2 compares statistical regression models, which quantify statistical associations between exposure and response in the presence of other variables, with simulation and causal Bayesian network methods for elucidating why and how much changing some variables—especially, exposure concentrations and durations—will change others, including health risks. Chapter 3 provides a relatively accessible introduction to the key concepts of simulation-based modeling of exposure-related health effects, seeking to build intuition for how and why different time patterns of exposure affect risk differently, even if they deliver the same cumulative exposure per unit time (e.g., per week or per year). Chapters 4 and 5 develop and illustrate the simulation-based approach in more detail and apply it to respirable crystalline silica and asbestos, respectively—two occupational hazards of great current interest. Chapter 6 emphasizes the importance of nonlinear responses in exposure-related diseases, due largely to the fact that induction of chronic inflammation mediates many such diseases. Chronic inflammation involves biological mechanisms with clear thresholds and nonlinearities, in which doubling concentration more than doubles risk. Chapter 6 closes Part 1 of the book by considering how regulations to protect occupational health might be made more effective by paying more attention to variability in exposure concentrations over time, arguing that regulating average exposure concentrations and frequencies of excursions above permitted exposure concentration limits is often inadequate for limiting risks that are primarily driven by repeated high exposure concentrations spaced closely together in time.

Part 2 critically examines the use of statistical regression models in public health risk assessment, building on the insights and methods discussed in Chap. 2. Chapters 7, 8, 9, 10, 11, and 12 argue that regression modeling is not a suitable replacement for quantitative risk assessment (QRA) methods that model causal processes. These chapters apply the association-causation distinction to past studies of health effects associated with—but not necessarily caused by—emissions from factory farms (Chap. 7), spatially distributed sources of natural asbestos (Chap. 8), low levels of the heavy metals lead and molybdenum (Chaps. 10 and 11), and low-level occupational exposures to benzene (Chap. 12). These applications are powered by advances in theory and techniques for causal modeling. Chapter 9 reviews the traditional Bradford-Hill considerations for trying to make judgments about whether observed associations are best explained by causation. It suggests that quantitative methods of causal analysis and machine learning developed over the past century can add considerable value to the Bradford-Hill considerations, helping to shift the scientific basis for causal inferences away from the psychology of judgment emphasized by Hill, and toward more objective and reliable data-driven considerations, such as whether health effects are found to depend on exposures when causal analysis methods are used to hold the values of other variables fixed. The principles and methods of causal analysis reviewed in Part 2 emphasizes robust data-driven inferences, based largely on information theory and non-parametric tests, to minimize
dependence on modeling assumptions. Other causal modeling techniques (such as regression-based and propensity score model-based methods) that rely on judgments about the plausibility of untested parametric modeling assumptions are not emphasized.

Part 3 turns to health effects associated with fine particulate matter (PM2.5) air pollution, consisting of atmospheric particulate matter (PM) with particles having a diameter of less than 2.5 μm. Decades of regression modeling, funded largely by the United States Environmental Protection Agency (US EPA), have produced a vast and still growing literature documenting statistical associations between PM2.5 levels and a wide variety of adverse effects, from autism to violent crime, in a wide variety of regression models (Chaps. 14, 15, and 16). Yet, the central paradox of this area of research is that there is very little clear evidence that reducing ambient levels of particulate air pollution actually causes reductions in the various adverse effects attributed to it (Chaps. 17, 18, and 19). Nonetheless, a vocal community of air pollution health effects investigators, activists, advocates, and reporters take it as axiomatic that PM2.5 air pollution kills, and that many deaths per year could be prevented by further reducing ambient concentrations of PM2.5 (Goldman and Dominici 2019). This firmly held and widely shared belief, constantly assumed in calculations of regulatory benefits (Chap. 14) and asserted for years as a well-established fact in countless news stories, op-eds, and scientific journals (Chap. 19), is not easily shaken by empirical observations and findings to the contrary (Henneman et al. 2017; Burns et al. 2020). Instead, expert “weight-of-evidence” (WoE) judgments and “causal determination” judgments (Chap. 18) are frequently advocated as a more dependable basis than causal analysis of data for choosing what should be believed or assumed and communicated to policy makers and the public (e.g., Goldman and Dominici 2019). Part 3 takes issue with this judgment-based approach. It notes that often-cited associations between air pollution and adverse health effects do not fully control for important confounders (Chaps. 13 and 15); that association-based studies of air pollution health effects have typically assumed causality without rigorously testing other explanations (Chaps. 14 and 15); that realistic errors in estimates of exposure concentrations distort the shapes of estimated concentration-response functions (Chap. 16); and that, in practice, reducing air pollution in many cities and areas has not caused the improvements in public health that were projected based on statistical associations (Chap. 17). Part 3 closes by arguing that both the essentially political practice of using consensus expert judgments to make “causal determinations” about health effects of exposures (Chap. 18) and the communication of such assumptions and judgments to the public and policy makers as if they were scientific facts can be substantially improved (Chap. 19). The key to better informing policy makers and the public is to better distinguish between trustworthy, empirically demonstrable and reproducible effects, and hypothesized effects based on unverified modeling assumptions and judgments. Headlines reporting model-based projections of deaths attributed to pollution do not imply that reducing pollution would reduce deaths, although the language used often obscures this fact. Chapter 19 concludes with recommendations for communicating more accurately about what is and is not implied by studies of air pollution.
pollution-health effects associations that might be useful in crafting effective, well-informed regulations and policies.

This book makes repeated use of current data science and machine learning techniques, such as partial dependence plots estimated from random forest model ensembles, and Bayesian networks. For readers unfamiliar with these techniques, Chap. 2 provides a brief nontechnical introduction, and Chap. 9 and its appendixes and references provide more technical detail. In an effort to make the chapters relatively self-contained, relevant techniques are briefly reviewed in less detail in other chapters as needed. Similarly, the early chapters introduce examples and analyses that are expanded on in later chapters. Readers primarily interested in specific substances (e.g., crystalline silica, lead, asbestos, benzene, or fine particulate matter) should find sufficient explanation of the technical methods used in each chapter to follow the analyses without reading all of the preceding chapters. Conversely, readers more interested in general principles may wish to focus on Chaps. 2, 9, 18 and 19, as the remaining chapters mainly deal with applications of the principles and methods discussed in these methodological chapters.

Scientific Method vs. Weight-of-Evidence Consensus Judgments as Paradigms for Regulatory Risk Analysis

Efforts to apply the foregoing principles 1–5 to reviews of air pollution health effects have provoked considerable backlash, enthusiastic *ad hominem* attacks, and much reporting describing them as dangerous attacks on the established science of air pollution health effects research and an attempt to subvert years of progress in environmental regulation (Tollefson 2019; Drugmond 2020). For example, a 29 April, 2019 Editorial in *Nature* stated the following.

- “Research linking fine particulate pollution and premature deaths is under attack.” A less dramatic account, and the point of view adopted in this book, is that decades of research associating exposure to fine particulate matter with increased mortality rates is well accepted, but has recently come under scrutiny to see what valid causal conclusions it can yield (Burns et al. 2019; Cox 2017; Henneman et al. 2017). In less contentious times, this reexamination might be viewed not as an “attack” on such research, but as an attempt to understand and use it correctly. Likewise, demands for clear definitions (principle 3 above)—especially whether the designation “causal” is intended to refer to manipulative causation or to Hill causation (or to something else)—are viewed in this book not as an attack on science, or an effort to create an impossible (or any) burden of proof (Goldman and Dominici 2019; Tollefson 2019), but rather as an effort to bring much-needed clarity to an area with a long history of using vague and ambiguous terms (such as pollution being “linked” to health effects) that have obscured the distinction between correlation and causality. Part 3 argues that such clarity is crucial for well-informed and causally effective policy-making.
and communication (Chap. 19); pursuing it advances rather than undermines sound science.

• “… Sceptics often argue that the epidemiological evidence cannot prove that air pollution causes premature deaths. But that is deliberately ignoring the weight of evidence from an array of rigorous epidemiological studies, aligned with other sources of evidence.” This misrepresents crucial points. As explained in Part III of this book, the key scientific issue is not that epidemiological evidence cannot prove that reducing current ambient air pollution levels causes reduced death rates, but rather that, to date, multiple studies in multiple countries do not do so (Burns et al. 2020; Cox Jr. 2017; Henneman et al. 2017). More precisely, existing epidemiological data sets enable powerful statistical tests of specific null hypotheses, including whether mortality rate is conditionally independent of air pollution, given levels of known potential confounders (e.g., income and daily high and low temperatures in the weeks before death (Zhang et al. 2017)); and whether mortality rates have declined no more quickly where air pollution has decreased substantially than where it has increased or remained unchanged (Chap. 17). The view espoused in this book is that such null hypotheses can and should be directly tested using available data (e.g., via conditional independence tests and rate of change comparisons) (Chaps. 13, 15, and 17), and that the evidence base considered by EPA should be expanded to include results of more such studies (Clean Air Scientific Advisory Committee (CASAC) 2019). In multiple studies in multiple countries, the null hypothesis of no causal effect of changes in exposure on changes in mortality rate is not clearly rejected when rigorous tests are applied to epidemiological data, although the different null hypothesis of no association between exposure and mortality is often decisively rejected (e.g., Burns et al. 2019; Henneman et al. 2017; Stenlund et al. 2009; Zigler and Dominici 2014, Chap. 17). Those who assert the contrary—that data from famous studies of association, such as the Harvard Six Cities study, have already amply shown that ambient levels of air pollution cause increased mortality—conflate association and causation, often using ambiguous words such as “links” and “linking” that needlessly obscure this crucial distinction (e.g., Tollefson 2019). They refer to “rigorous epidemiological studies” without noting that most of these studies only address associations and do not fully control for obvious potential confounders, such as poverty or humidity and daily high and low temperatures lagged by more than a day or two. We follow other investigators who rely on data rather than authoritative judgment for their conclusions in remaining open to the possibility that reducing air pollution does in fact reduce mortality rates (e.g., Burns et al. 2019), but that the causal effect is small enough to have escaped detection (Chap. 17).

• “But questioning the evidence won’t make it go away. …Now is not the time to undermine efforts to clean air—it is time to strengthen them.” Much of the backlash against the principles of sound science proposed above seems to start with policy convictions, such as that now is the time to strengthen efforts to clean air; to then assert that it is well known and long established that current ambient levels of fine particulate matter kill, but without presenting causal analyses or
empirical data demonstrating that this is likely to be true; and to refer to ambiguous and conflicting evidence of association in the presence of incompletely controlled confounders as if it supported these conclusions (Goldman and Dominici 2019). But sounder science and better informed policy-making both require a different procedure that first seeks to understand what valid causal conclusions are justified by observations, as a prelude to understanding what policies are needed to fully and effectively protect public health, including the health of sensitive subpopulations, with an adequate margin of safety. This approach, developed in the chapters that follow, might support policy recommendations in either direction.

This book argues that in air pollution health effects research, as in other areas of applied science, investigators can best pursue objective scientific truth by using dispassionate, sound analysis of data to inform policy, rather than letting policy preferences and judgments inform or constrain selection and interpretation of evidence. This reversal has been vociferously opposed by advocates of a “weight of evidence” (WoE) framework that makes judgments of selected experts—and especially, “causal determination” judgments (Chap. 18)—central to interpretation of evidence and conclusions communicated to policy makers (Goldman and Dominici 2019; Krimsky 2005; Linkov et al. 2015). The WoE framework does not require use of the foregoing principles of sound science. However, as argued in the following chapters, following these principles better serves the discovery and communication of valid causal conclusions, the needs of policy makers to be informed about the public health consequences caused by alternative policy choices, and hence the public interest.

A Recent Example: PM2.5 and COVID-19 Mortality

To illustrate why statistical modeling of exposure-response associations accompanied by judgments about causal interpretations of statistical associations and regression coefficients—the current weight-of-evidence (WoE) approach favored in much current regulatory risk analysis for air pollutants—is not a valid basis for determining whether or to what extent risk of harm to human health would be reduced by reducing exposure, and why the traditional scientific method based on testing predictive generalizations against data remains a more reliable paradigm for risk analysis and risk management, we consider the question of whether past exposure to fine particulate matter (PM2.5) increases risk of COVID-19 mortality. As COVID-19 mortalities mounted worldwide in the first two quarters of 2020, headlines and scientific articles warned that fine particulate matter air pollution (PM2.5) increases risk of COVID-19-related illness and death. For example, Jiang et al. (2020) used a Poisson regression model to conclude that PM2.5 and humidity increased the risk of daily COVID-19 incidence in three Chinese cities, while coarse particulate air pollution (PM10) and temperature decreased this risk. Bashir
et al. (2020) calculated significant ordinal correlations between PM2.5 and other air pollutants (PM10, SO2, NO2, and CO) and COVID-19 cases in California, and concluded that such correlations should encourage regulators to more tightly control pollution sources to prevent harm. Most famously, Wu et al. (2020) fit a negative binomial regression model to county-level data in the United States, and interpreted their finding of a significant positive regression coefficient for PM2.5 as implying that “A small increase in long-term exposure to PM2.5 leads to a large increase in the COVID-19 death rate.” This interpretation attracted national headlines and widespread political concern (Friedman 2020).

These examples follow a common technical approach, also widely applied in many other areas of current regulatory risk assessment and public health, with the following steps:

1. Collect data on estimated air pollution levels, one or more adverse health outcomes of interest (such as COVID-19 mortality), and any covariates of interest (e.g., humidity, temperature, population density, etc.)
2. Fit one or more regression model to the data, treating air pollution levels as predictors and adverse health outcomes as dependent variables. Include other variables as predictors at the modeler’s discretion.
3. If the regression coefficient for a pollutant as a predictor of an adverse health outcome is significantly positive in the one or more regression models, use judgment to interpret this as evidence that reducing levels of the pollutant would reduce risk of the adverse health outcome.
4. Communicate the results to policy makers and the press using the policy-relevant language of causation and change—that is, claim that a given reduction in pollution would create a corresponding reduction in adverse health outcomes—rather than in the (technically accurate) language of association and difference: that a given difference in estimated exposures is associated with a corresponding difference in the conditional expected value of a dependent variable predicted by the selected regression model.

Step 3 is based on a judgment that a positive regression coefficient in a modeler-selected regression model is evidence of a causal relationship: that it implies or suggests that reducing exposure would reduce risk, even if the experiment has not actually been made. In this respect, it incorporates the central principle of the WoE framework: that a well-informed expert scientist can make a useful judgment about whether the association indicated by a statistically significant positive regression coefficient is likely to be causal. This assumption is scrutinized next.

**Do Positive Regression Coefficients Provide Evidence of Causation?**

As noted by Dominici et al. (2014), either significant positive coefficients or significant negative regression coefficients (or no significant regression coefficient at all) for air pollution as a predictor of mortality risk can often be produced from the same
data, depending on the modeling choices made; thus “There is a growing consensus in economics, political science, statistics, and other fields that the associational or regression approach to inferring causal relations—on the basis of adjustment with observable confounders—is unreliable in many settings.” In the field of air pollution health effects research, however, investigators continue to rely on regression modeling in step 2 of the above approach. A skilled regression modeler can usually produce a model with a significant positive regression coefficient for exposure in step 2, allowing steps 3 and 4 to proceed. We illustrate how this can be done, using a data set on PM2.5 and COVID-19 mortality in the United States as an example. The data set, described and provided via a web link in Appendix, compiles county-level data on historical ambient PM2.5 concentration estimates, COVID-19 mortality rates and case rates (per 100,000 people) through April of 2020, along with numerous other county-level variables.

A key step in regression modeling is to select variables to include in the model. Figure 1.1 shows a random forest (nonparametric model ensemble) importance plot for county-level variables as predictors of COVID-19 mortality rates, where the “importance” of each variable is measured by the estimated percentage increase in mean squared prediction error if that variable is dropped as a predictor. The few most important predictors of COVID-19 mortality (DeathRate100k) are $PCT_{BLACK}$, the percentage of a county population that is Black; $PopDensity$, the average density of the population in the county (number of people per square mile) or its logarithm, $PopDensityLog$ (the log transform makes little difference to nonparametric methods such as random forest, but can be important for parametric regression models); $Longitude$, time since first case in the county ($FirstCaseDays$), average
A Recent Example: PM2.5 and COVID-19 Mortality

estimated PM2.5 concentration between 2000 and 2016 ($X_{2000.2016\text{AveragePM25}}$), average temperature during the winter months between 2000 and 2016 ($WINTERAVG\text{TMP}$), and Latitude. For the case rate (COVID-19 cases reported per 100,000 people), the most important predictors also include the average minimum temperature in February over the past two decades ($Feb\text{Min}\text{Tmp}_{2000.2019}$), percent Hispanic ($PCT\_HISP$), and percent of population with at least high school educations ($PCT\_ED\_HS$). These ten predictors alone explain about the same percentages of the variances in COVID-19 mortality and case rates across counties (48% and 40%, respectively) as the full set of over 60 variables, of which the most important are shown in Fig. 1.1.

Of course, predictors need not be statistically independent of each other. To visualize the statistical interdependencies among them, Fig. 1.2 shows a Bayesian network (BN) fit to the data (using the default hill-climbing (HC) algorithm in the bnlearn package in R), with Latitude and Longitude constrained to have only outward-pointing arrows and DeathRate100k constrained to have only inward-pointing arrows, to facilitate possible intuitive causal interpretations of the arrows leaving or entering these three nodes. (Presumably, latitude and longitude are not caused by anything else, and death does cause any of the other variables.) However, in general the arrows only signify statistical dependencies between variables, and not necessarily causal relationships. (Chapter 9 discusses principles of BN-learning and interpretation of BNs in more detail, and Chap. 2 provides a less technical introduction.)

For example, an arrow between PM2.5 and percent Hispanic ($X_{2000.2016\text{AveragePM25}}$ and $PCT\_HISP$) does not suggest that either causes the other: it simply reflects that counties with higher percentages of Hispanic populations tend

![Bayesian network for COVID-19 mortality (deaths per 100,000 people) showing statistical dependencies among variables. An arrow between two variables indicates that they are informative about each other (i.e., not statistically independent).](image-url)
to also have higher PM2.5 levels. However, if variables depend on their direct causes, then absence of an arrow between two variables corresponds to absence of empirical evidence in the BN that either directly causes the other. COVID-19 mortality in Fig. 1.2 is shown as depending directly on latitude and longitude (which are presumably surrogates for other biologically effective causes), as well as on time since first case in a county (FirstCaseDays), average winter temperature, and ethnic composition (PCT_BLACK and PCT_HISP). Figure 1.3 shows an analogous BN for COVID-19 case rate, which depends directly on latitude and longitude, ethnic composition (PCT_BLACK and PCT_HISP), time since first case in a county (FirstCaseDays), and education (PCT_ED_HS). Bayesian network learning is a relatively new technique for exploring and visualizing direct and indirect dependencies among variables. As an alternative, Fig. 1.4 shows a classification and regression tree (CART) tree for COVID-19 mortality. The CART algorithm (implemented in the rpart package in R) recursively partitions counties into clusters with significantly different COVID-19 mortality rates, based on the results of binary tests (“splits”), such as whether Longitude < -75.61 (yes = left branch, no = right branch). For example, the counties with Longitude < -75.61, PCT_BLACK < 0.2636, and time since first case < 44.5 days have an average COVID-19 mortality rate of less than 3 per 100,000 (2.436, although 3 significant digits is spurious precision), compared to a rate over 50 times greater (148.7 per 100,000) for counties further to the East with high population densities and longer times since first cases. Although CART trees are subject to residual confounding due their binary splits of continuous variables and are not very robust, in the sense that fitting them to multiple random samples from the same data set often produces different trees (e.g., with WINTERAVGTEMP in some and FebMinTemp2000.2019 in others), they provide a relatively simple, well-established nonparametric technique for exploring significant predictors of a selected dependent variable such as DeathRate100K. The
A Recent Example: PM2.5 and COVID-19 Mortality

Fig. 1.4 A classification and regression tree (CART) tree for COVID-19 mortality (DeathRate100K). The tree was generated by the rpart package in R

The predictors identified in Fig. 1.4 are Longitude, PCT_BLACK, WINTERAVGTMP, FirstCaseDays, and PopDensity.

Although we regard Figs. 1.2, 1.3, and 1.4 as only exploratory visualizations, they highlight the importance of confounders such as Longitude in understanding associations between PM2.5 and COVID-19 mortality. The usual way to control for measured confounders in regression modeling is to “adjust” for them by including them as predictors on the right side of a regression equation. For example, a multiple linear regression model that includes all of the variables identified in Figs. 1.2 and 1.4 on which COVID-19 mortality rate might directly depend, and that further hypothesizes a dependence on historical average ambient PM2.5 exposure concentration levels ($X_{2000.2016AveragePM25}$), would posit an equation of the form in Eq. (1.1).

$$E(\text{DeathRate100K}) = \text{Intercept} + b_1 \cdot X_{2000.2016AveragePM25}$$
$$+ b_2 \cdot \text{PCT\_BLACK} + b_3 \cdot \text{PCT\_HISP} + b_4 \cdot \text{Latitude} + b_5 \cdot \text{Longitude} + b_6 \cdot \text{FirstCaseDays}$$
$$+ b_7 \cdot \text{WINTERAVGTMP} + b_8 \cdot \text{PopDensity} \quad (1.1)$$

For simplicity, Eq. (1.1) follows Wu et al. (2020) in assuming that risk depends on a weighted sum of terms on the right side, ignoring interaction terms (e.g., that increasing PM2.5 should not increase death rates if population density = 0); consequences of this modeling choice are discussed later. Fitting Eq. (1.1) to the data set via ordinary least squares (OLS) regression yields Table 1.1.

The regression coefficient for past estimated average ambient PM2.5 exposure concentration (denoted by $X_{2000.2016AveragePM25}$ in Tables 1.1 and 1.2) is negative and not significantly different from 0 ($p = 0.87$), consistent with Fig. 1.2. However, regression modeling allows modelers to select variables to include in the model, which can drive the results that get published (Dominici et al. 2014). For example, dropping Longitude from the regression model yields Table 1.2. Now the regression coefficient for PM2.5 is positive instead of negative, and it is highly
significant \((p = 0.000053)\) instead of non-significant. In effect, PM2.5 acts as a partial surrogate for longitude for predicting COVID-19 mortality risk, so that omitting longitude induces PM2.5 to enter with a significant positive coefficient. Figure 1.5 suggests why: both PM2.5 and COVID-19 mortality rates tend to be higher in the East than in the West. Interpreting the positive regression coefficient for PM2.5 in Table 1.2 as evidence that an increase in PM2.5 increases PM2.5 mortality risk, as in Wu et al. (2020) would be mistaken: it is only evidence that the modelers omitted longitude from the model.

### Table 1.1 Multiple linear regression model for COVID-19 mortality rate

|                     | b*   | Std. err. of b* | b    | Std. err. of b | t(3000) | p-value |
|---------------------|------|-----------------|------|----------------|---------|---------|
| Intercept           | $-15.81$ | $9.28$ | $-1.7$ | $0.0885$ |
| 2000-2016AveragePM25 | $-0.00$ | $0.028$ | $-0.04$ | $0.23$ | $0.8718$ |
| PCT_HISP            | $0.09$ | $0.021$ | $13.17$ | $3.19$ | $4.1$ | $0.0000$ |
| PCT_BLACK           | $0.28$ | $0.021$ | $40.75$ | $2.99$ | $13.6$ | $0.0000$ |
| PopDensity          | $0.04$ | $0.018$ | $0.00$ | $0.00$ | $2.2$ | $0.0261$ |
| WINTERAVGTMP        | $0.09$ | $0.034$ | $0.17$ | $0.06$ | $2.7$ | $0.0663$ |
| FirstCaseDays       | $0.17$ | $0.018$ | $0.21$ | $0.02$ | $9.6$ | $0.0000$ |
| Latitude            | $0.18$ | $0.037$ | $0.77$ | $0.16$ | $4.8$ | $0.0000$ |
| Longitude           | $0.15$ | $0.025$ | $0.27$ | $0.05$ | $5.9$ | $0.0000$ |

The columns give standardized regression coefficients \((b^*)\) and their standard errors; unstandardized regression coefficients \((b)\) and their standard errors; and t-test values and significance levels \((p\text{-values})\) for each coefficient.

### Table 1.2 Multiple linear regression model for COVID-19 mortality rate with longitude omitted

|                     | b*   | Std. err. of b* | b    | Std. err. of b | t(3001) | p-value |
|---------------------|------|-----------------|------|----------------|---------|---------|
| Intercept           | $-38.08$ | $8.52$ | $-4.5$ | $0.0000$ |
| 2000-2016AveragePM25 | $0.09$ | $0.02$ | $0.77$ | $0.19$ | $4.0$ | $0.0001$ |
| PCT_HISP            | $0.06$ | $0.02$ | $9.05$ | $3.13$ | $2.9$ | $0.0038$ |
| PCT_BLACK           | $0.29$ | $0.02$ | $41.08$ | $3.00$ | $13.7$ | $0.0000$ |
| PopDensity          | $0.04$ | $0.02$ | $0.00$ | $0.00$ | $2.5$ | $0.0109$ |
| WINTERAVGTMP        | $0.04$ | $0.03$ | $0.08$ | $0.06$ | $1.3$ | $0.1876$ |
| FirstCaseDays       | $0.17$ | $0.02$ | $0.21$ | $0.02$ | $9.6$ | $0.0000$ |
| Latitude            | $0.15$ | $0.04$ | $0.63$ | $0.16$ | $4.0$ | $0.0001$ |

The regression summary for dependent variable: DeathRate100K, \(R = 0.39594357, R^2 = 0.16, \text{Adjusted } R^2 = 0.16\)
Positive Regression Coefficients Created by Model Specification Error and Other Causes

More generally, there are many reasons that PM2.5 might have a significant positive regression coefficient that do not imply that increasing PM2.5 would increase risk. As a simple conceptual example, suppose that \( \text{PopDensityLog} \) is a confounder of the PM2.5-Risk association, and that the structural equations describing the causal relationships among these variables are as follows:

\[
E(Risk) = \text{PopDensityLog}^2 \quad (1.2)
\]

\[
\text{PM2.5} = \text{PopDensityLog}^2 \quad (1.3)
\]

(In such structural equations, the values of the dependent variables on the left are causally determined by the values of the variables on the right: if the right-hand variables are exogenously changed, then the left-hand variables will change to make the equality hold.) Then fitting the following multiple linear regression model (1.4) to data generated by Eqs. (1.2) and (1.3) for a large number of counties with different population densities would yield the optimized parameter estimates \( b_0 = 0, b_1 = 0, b_2 = 1 \).

Fig. 1.5 Scatter plots of average estimated historical PM2.5 concentrations (in micrograms per cubic meter) (red squares) and COVID-19 deaths per 100,000 (blue circles) vs. longitude
\[ E(Risk) = b_0 + b_1 \times PopDensityLog + b_2 \times PM2.5 \] (1.4)

Thus, regression identifies a significant positive coefficient for PM2.5, and not for the confounder PopDensityLog, because doing so minimizes prediction error (MSE). But this coefficient has no relevance for determining how or whether changing PM2.5 would change Risk. A claim that such a regression analysis had “controlled for” potential confounding from PopDensityLog by including it in regression model (1.4), and yet still found that PM2.5 increased risk, would be deluded. A judgment that such an analysis provides evidence that increasing PM2.5 levels increases risk would be mistaken.

As a less hypothetical example, suppose we create a synthetic data set that is identical to the real one, except for the addition of a new Risk variable defined as 
\[ Risk = PopDensityLog^2 \]. In other words, we artificially create a variable that we know is determined only by population density, via the nonlinear formula \[ Risk = (\log\text{(population density)})^2 \]. (This example is suggested by Fig. 1.6, which shows a scatter plot of COVID-19 deaths per 100,000 against PopDensityLog.) Fitting a multiple linear regression model to the data with this artificial Risk variable as the dependent variable yields the results in Table 1.3. All but one of the predictors, including PM2.5 (2000-2016AveragePM25), have highly statistically significant positive regression coefficients, even though, by construction, Risk does not depend on anything other than PopDensityLog. The reason is that the multiple linear regression model’s assumption that risk depends only on a weighted sum of the predictors is false. As illustrated in Fig. 1.6 for the real risk variable (DeathRate100k), risk varies nonlinearly with PopDensityLog. The mistaken modeling assumption of linearity is sufficient to induce many other predictors to enter the regression model with significant positive coefficients, because including them helps to reduce the mean squared prediction error due to model misspecification. Again, interpreting such a positive regression coefficient for exposure as evidence that reducing exposure would reduce risk is mistaken, even though no confounders have been omitted in this example. Instead, positive regression coefficients are only evidence that the assumed regression model does not describe the data. Chapter 2 develops further the point that model specification error often accounts for positive regression coefficients.

Nonparametric methods help to avoid these difficulties. Figure 1.7 shows a CART tree for the same example as in Table 1.3. In this tree, as also in a nonparametric Bayesian network fit to the same data, the only predictor of PopDensityLog^2 is found to be PopDensityLog.

More generally, PM2.5 could have a significant positive regression coefficient as a predictor of COVID-19 mortality risk for any or all of the following reasons, as discussed further in Chap. 2.

- Model specification errors, e.g., if mortality rate is assumed to depend on a weighted sum of variables, but in fact its dependence is better described by a model with nonlinearities or interaction terms, as in Table 1.3.
Scatterplot of COVID-19 deaths per 100,000 (DeathRate100k) against PopDensityLog. A non-parametric (lowess) smoothing regression curve is superimposed to aid visual interpretation.

Table 1.3 A multiple linear regression model for the dependent variable PopDensityLog^2

| N = 3009 | b' | Std. err. of b' | b | Std. err. of b | t(3000) | p-value |
|----------|----|-----------------|----|----------------|---------|---------|
| Intercept| 47.30 | 4.19 | -11.3 | 0.000000 |
| 2000-2016AveragePM25 | 0.35 | 0.02 | 1.99 | 0.11 | 18.9 | 0.000000 |
| PCT_BLACK | -0.01 | 0.01 | 1.32 | 1.35 | 1.0 | 0.326789 |
| PCT_HISP | 0.16 | 0.01 | 17.09 | 1.44 | 11.9 | 0.000000 |
| Latitude | 0.37 | 0.02 | 1.09 | 0.07 | 15.0 | 0.000000 |
| Longitude | 0.17 | 0.02 | 0.21 | 0.02 | 10.1 | 0.000000 |
| WINTERAVGTEMP | 0.22 | 0.02 | 0.28 | 0.03 | 9.9 | 0.000000 |
| FirstCaseDays | 0.39 | 0.01 | 0.33 | 0.01 | 33.5 | 0.000000 |
| PopDensity | 0.38 | 0.01 | 0.00 | 0.00 | 33.1 | 0.000000 |
• Omitted confounders, as in the example of PM2.5 and COVID-19 mortality risk depending on latitude and longitude (independently of other factors, as shown in Fig. 1.2), if these factors are omitted;
• Measurement errors in explanatory variables, e.g., if PM2.5 is correlated with other variables that are measured or estimated with error, so that including PM2.5 in the regression reduces prediction error due to uncertainties about those variables;
• Residual confounding, e.g., if older people tend to live in more polluted areas and, independently, to have higher mortality rates, but age is only measured in wide categories such as “% of people aged 65 or older”;
• Use of surrogate variables, e.g., “Average winter temperature” since 2000, rather than more causally relevant variables such as low temperatures in the months of COVID-19 in 2020;
• Unmodeled interactions or dependencies among variables, e.g., if PM2.5 modifies or is modified by variables such as humidity and temperature that affect respiratory illnesses and COVID-19 mortality;

A positive regression coefficient explained by one or more of these sources does not provide evidence that reducing PM2.5 would reduce mortality risk.

**Conclusion: Regression Models and Judgment Should Complement Science, not Substitute for it**

We do not conclude from the foregoing considerations that PM2.5 does not increase risk of COVID-19 mortality; perhaps it does. Rather, we conclude that a positive regression coefficient *per se* does not provide useful evidence about the matter. Such regression coefficients are easily produced by modeling choices (Dominici et al. 2014), but lack clear causal interpretation. A judgment that such evidence provides reason to worry—that, in the words of a *New York Times* headline, “New Research Links Air Pollution to Higher Coronavirus Death Rates” (Friedman 2020)—is simply misleading: the “links” provided by positive regression coeffi-

---

**Fig. 1.7** A CART tree model for the dependent variable *PopDensityLog^2*
Coefficients are statistical links, not causal ones, and they may signify only that longitude was omitted, or that a linear form was assumed, or that different predictors are correlated with each other and estimated with errors, or that continuous variables have been categorized, and so forth. (All of these are true of the model of Wu et al. (2020) behind the headline.) This need not mean that the conclusion is false, but it does mean that the conclusion is not implied by the data and regression analyses from which it is said to be derived.

North (2020) wrote that “An established paradigm for interpreting epidemiological evidence causally, used by the US EPA, based on considerations proposed in 1965 by British epidemiologist Sir Austin Bradford Hill, is being challenged by another paradigm based on statistical procedures to distinguish between association and causation.” Perhaps the most fundamental prescription of the causal paradigm is to recognize that statistical “links” such as positive regression coefficients (or relative risks greater than 1, or positive attributable risks and burdens of disease, and so forth) are neither more nor less than indicators of statistical association, which do not necessarily or usually provide relevant evidence about causation (Pearl 2009). Modern methods such as the Bayesian network in Fig. 1.2 can help to discover what evidence a data set does provide that some variables depend, directly or indirectly, on others. For example, Fig. 1.2 suggests that latitude and longitude have direct effects on COVID-19 mortality risk (meaning, effects in addition to those mediated by the other variables in Fig. 1.2). This discovery might not have been anticipated intuitively by an investigator, leading to the omission of these confounders, as in Wu et al. (2020). Such computer-assisted discoveries from data may assist, but not replace, the scientific work of formulating testable predictions about whether and how much changes in some variables affect changes in others, and then testing these predictions against new data and reporting the results. If COVID-19 mortality risk appears to be conditionally independent of PM2.5 in non-parametric analyses with adequate power to detect even relatively small effects (Figs. 1.2 and 1.4), then parametric regression modeling that imposes assumptions on the data sufficient to create a positive regression coefficient for PM2.5 (Tables 1.2 and 1.3) should not be construed as evidence that changing PM2.5 would change COVID-19 mortality risk. But neither should it preclude a search for alternative hypotheses, backed by empirical testing, that better explain the observations. For example, for the same average PM2.5 concentration over the past 20 years, do counties with constant or increasing PM2.5 levels over time have significantly earlier first dates of COVID-19 mortalities than counties with PM2.5 levels that decreased over time? Figure 1.2 leaves open this possibility, and additional research might pursue it further.

The question of what scientific hypotheses are worth investigating further is surely a proper matter for expert judgment. Interpretation of regression coefficients as evidence that reducing exposure would reduce risk is not. Thus, we conclude that empirical testing of predictive generalizations against data should not be skipped in favor of applying judgment to regression coefficients to draw policy-relevant causal conclusions. Regression coefficients simply do not provide the information needed to determine—or to make sound judgments about—whether or to what extent they are likely to be causal (Table 1.3). Judgment that seeks to bridge the gap between
association and causation based on positive regression coefficients (e.g., Cromar and Ewart 2016) is akin to a Rorschach test: an expert may perceive evidence for causation in such coefficients, and may even use them to quantify health benefits to be expected from reducing exposure, but the perceived evidence and expectations of health benefits are solely in the mind of the expert, neither supported nor refuted by the regression coefficients themselves. The causal paradigm proposes that traditional scientific method, while often more time-consuming and difficult than applying judgment to regression models, is a far more reliable guide to determining whether and to what extent interventions that change exposure will cause risk to change. Verifying scientific models and predictions for PM2.5 and COVID-19 might take impractically long, and meanwhile decisions must be made and risks managed despite scientific uncertainties about causation. But we reiterate that unwarranted causal interpretation of statistical associations and regression coefficients in a WoE framework should not be substituted for sound science. A technically gullible press and policy makers should not be distracted by pre-scientific claims about health effects from PM2.5 based on judgment and regression modeling in the absence of traditional scientific method and careful evaluation of the predictive validity of such claims (Chap. 19). Risk analysis and the public interest can and should be better served by adhering to the principles of sound science articulated in the Introduction and to the principles of sound causal inference referred to by North (2020) and discussed in several chapters of this book, especially Chaps. 9 and 18.

The remainder of this book develops and illustrates these themes in more detail. Although regression modeling assisted by WoE judgments does not usually suffice to provide a sound or reliable basis for determining whether or to what extent reducing exposure would reduce risk, other techniques such as simulation modeling, Bayesian networks, and causal modeling using nonparametric machine-learning methods can do much to improve the risk analyst’s ability to predict the effectiveness of interventions, including reductions in exposure concentrations or changes in time patterns of exposure, in changing the health risks of exposed public or occupational populations. Chapter 2 introduces these methods.

Appendix 1: Data

We collected data from many sources, including most of those cited by Wu et al. (2020), but with alternate authoritative sources for temperature, humidity, and cases/deaths data. We used more recent data for PM2.5, demographics, temperatures, and cases/deaths, and added further sources or fields. For example, we collected USDA county level economic characterizations along with various county attributes compiled by the UC Berkeley Yu Group (2020). Table 1.4 summarizes data sources and variables. Data building was accomplished using python scripts. The full data set can be downloaded from http://cox-associates.com/CausalAnalytics/; it is the file “covidpm25.xlsx” (Table 1.5).
### Table 1.4  Data Sources and variable overview

| Data category     | Source                                                                 | Comments                                                                 |
|-------------------|------------------------------------------------------------------------|--------------------------------------------------------------------------|
| PM2.5             | Pm2.5 annual average data from the Atmospheric Composition Analysis Group ([http://fizz.phys.dal.ca/~atmos/martin](http://fizz.phys.dal.ca/~atmos/martin)). 0.01° x 0.01° grid resolution PM2.5 prediction in mcg/m³ | We averaged across grid cells in each county, and produced a 2000–2016 average, as well as separate values for each year 2000–2018 |
| County boundaries | U. S. Census [https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html](https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html) | Used to provide county boundaries for PM2.5 attribution, land area for popdensity, and centroids lat/lons |
| Demographic       | U.S. Census Bureau online API. 2018 ACS5 (American Community Survey 5-year data ending in 2018). County level data. [https://www.census.gov/data/developers/data-sets/acs-5year.html](https://www.census.gov/data/developers/data-sets/acs-5year.html) | List of variables in table below |
| Temperatures      | NOAA. County level annual data [ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/](https://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/) Average: climdiv-tmpccy-v1.0.0-20200504' Min: climdiv-tmincy-v1.0.0-20200504' Max: climdiv-tmaxcy-v1.0.0-20200504' Description: county-readme.txt | We averaged across years 2000–2019 for long term averages by month. We also extracted monthly averages for Jan–Apr 2020 |
| Humidity          | Humidity averages by U.S. weather station (city) through 2018. [https://www1.ncdc.noaa.gov/pub/data/ccd-data/relhum18.dat](https://www1.ncdc.noaa.gov/pub/data/ccd-data/relhum18.dat). City lat/lons from [https://simplemaps.com/data/us-cities](https://simplemaps.com/data/us-cities) For each county centroid, the humidity data from the closest (based on lat/lon coordinates) weather station is obtained |
| Hospital beds     | Hospital level data with county identifier from Homeland Infrastructure Foundation-Level Data (HIFLD) [https://hifld-geoplatform.opendata.arcgis.com/datasets/hospitals as of 10/7/2019](https://hifld-geoplatform.opendata.arcgis.com/datasets/hospitals) | Aggregated hospitals over counties. Converted to beds per 100K population. Log version also |
| Mitigation policies | State level governmental COVID-19 policies compiled by Raifman et al. Boston University School of Public Health, COVID-19 United States state policy database ([www.tinyurl.com/statepolicies](http://www.tinyurl.com/statepolicies)) | Used to compute days since stay-at-home order and days since closure of non-essential businesses (from 5/11/2020) |
| Behavioral        | County level data from Robert Wood Johnson (2020) [https://www.countyhealthrankings.org/](https://www.countyhealthrankings.org/) Smoking, obesity, and overall health |
| County population | [https://www2.census.gov/programs-surveys/popest/datasets/2010-2018/counties/totals/](https://www2.census.gov/programs-surveys/popest/datasets/2010-2018/counties/totals/) | Used to scale various variables |

(continued)
### Table 1.4 (continued)

| Data category       | Source                                                                 | Comments                                                                 |
|---------------------|------------------------------------------------------------------------|--------------------------------------------------------------------------|
| County attributes   | Selected variables from the COVID Severity Forecasting project. UC Berkeley Departments of Statistics, EECS led by Professor Bin Yu [https://github.com/Yu-Group/covid19-severity-prediction](https://github.com/Yu-Group/covid19-severity-prediction). See also [https://www.stat.berkeley.edu/~binyu/ps/papers2020/covid19_paper.pdf](https://www.stat.berkeley.edu/~binyu/ps/papers2020/covid19_paper.pdf) | List of variables in table below                                          |
| Economic characteristics | County level data from USDA—[https://www.ers.usda.gov/data-products](https://www.ers.usda.gov/data-products) | 3 county coding schemes described in table below                          |
| Outcomes (deaths, cases, days since first case) | Cumulative values by date downloaded from [https://github.com/nytimes/covid-19-data](https://github.com/nytimes/covid-19-data) and are as of 5/11/2020 | 5/11 values for cases and deaths extracted. Converted to per 100K. Days since first case computed by using first case date |

### Table 1.5 Additional variable details

| Variable               | Category            | Description                                                                 |
|------------------------|---------------------|-----------------------------------------------------------------------------|
| PCT_POVERTY            | Demographic         | % below poverty                                                             |
| PCT_OWNEDHOM           | Demographic         | % owning home                                                               |
| PCT_ED_HS              | Demographic         | % with high school education                                                |
| PCT_BLACK              | Demographic         | % black                                                                     |
| PCT_HISP               | Demographic         | % hispanic                                                                  |
| MED_INCOMERATIO        | Demographic         | Median income, converted to ratio relative to mean over counties            |
| MED_HOMERATIO          | Demographic         | Median home value, converted to ratio relative to mean over counties        |
| PCT_65PLUS             | Demographic         | % 65+ years                                                                 |
| PCT_45TO64             | Demographic         | % 45–64 years                                                               |
| PCT_15TO44             | Demographic         | % 15–44 years                                                               |
| Rural-urban_ContinuumCode_2013 | Economic characteristics | 1–9 code indicating county degree of urbanization. [https://www.ers.usda.gov/data-products/rural-urban-continuum-codes](https://www.ers.usda.gov/data-products/rural-urban-continuum-codes). Created binary column for each level |
| Urban_Influence_Code_2013 | Economic characteristics | 1–12 code indicating county degree of urban influence. [https://www.ers.usda.gov/data-products/urban-influence-codes](https://www.ers.usda.gov/data-products/urban-influence-codes). Created binary column for each level |
| Economic_typology_2015 | Economic characteristics | 1–6 code indicating county economic condition. [https://www.ers.usda.gov/data-products/county-typology-codes](https://www.ers.usda.gov/data-products/county-typology-codes). Created binary column for each level |

(continued)
Table 1.5 (continued)

| Variable        | Category         | Description                                                                                                                                 |
|-----------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| PopDensity[Log] | County population| 2018 population estimate divided by land area (square miles) from shape files. Log version also                                             |
| CensusRegionName| County attributes | Created binary column for each level                                                                                                           |
| CensusDivisionName| County attributes | Created binary column for each level                                                                                                          |
| StateName       | County attributes | Created binary column for each level                                                                                                          |
| dem_to_rep_ratio | County attributes | Ratio of registered democrats to republicans in county                                                                                       |
| #ICU_beds100K[Log] | County attributes | Number of ICU beds per 100K population. Log version also                                                                                     |

References

Bareinboim E, Pearl J (2013) Causal transportability with limited experiments. In: Proceedings of the 27th AAAI conference on artificial intelligence, pp 95–101

Bashir MF, Ma BJ, Bilal A, et al. Correlation between environmental pollution indicators and COVID-19 pandemic: a brief study in Californian context. Environ Res. 2020;187:109652. https://doi.org/10.1016/j.envres.2020.109652.

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess E. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst Rev. 2019;5:CD010919. https://doi.org/10.1002/14651858.CD010919.pub2.

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess E. Interventions to reduce ambient air pollution and their effects on health: an abridged Cochrane systematic review. Environ Int. 2020;135:105400. https://doi.org/10.1016/j.envint.2019.105400.

Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company; 1963.

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.

Clean Air Scientific Advisory Committee (CASAC) (2019) CASAC review of the EPA’s policy assessment for the review of the National Ambient Air Quality Standards for Particulate Matter (External Review Draft—September 2019). https://yosemite.epa.gov/sab%5Csubproduct.nsf/E2F6C71737201612852584D20069DFB1/$File/EPA-CASAC-20-001.pdf

Cox LA Jr. Do causal concentration-response functions exist? A critical review of association and causal relations between fine particulate matter and mortality. Crit Rev Toxicol. 2017;47(7):603–31. https://doi.org/10.1080/10408444.2017.1311838.

Cox LA Jr. Modernizing the Bradford Hill criteria for assessing causal relationships in observational data. Crit Rev Toxicol. 2018;15:1–31. https://doi.org/10.1080/10408444.2018.1518404.

Cromar K, Ewart G. Concentration-response associations used to estimate public health benefits of less pollution are not valid causal predictive models. Ann Am Thorac Soc. 2016;13(12):2281. https://doi.org/10.1513/AnnalsATS.201610-754LE.

Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, Kotlov T, HEI Health Review Committee. Effect of air pollution control on mortality and hospital admissions in Ireland. Res Rep Health Eff Inst. 2013;176:3–109.

Dominici F, Greenstone M, Sunstein CR. Science and regulation. Particulate matter matters. Science. 2014;344(6181):257–9. https://doi.org/10.1126/science.1247348.
Drugmond D (2020) EPA clean air panel chair dismisses his oil industry ties, slams Harvard study on air pollution and COVID risks. https://www.desmogblog.com/2020/05/18/epa-clean-air-committee-cox-harvard-air-pollution-covid
Friedman L (2020). https://www.nytimes.com/2020/04/07/climate/air-pollution-coronavirus-covid.html
Goldman GT, Dominici F (2019) Don’t abandon evidence and process on air pollution policy. Science. https://doi.org/10.1126/science.aaw9460
Goodman GE, Thornquist MD, Balmes J, Cullen MR, Meyskens FL Jr, Omenn GS, Valanis B, Williams JH Jr. The beta-carotene and retinol efficacy trial: incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. J Natl Cancer Inst. 2004;96(23):1743–50.
Henneman LR, Liu C, Mulholland JA, Russell AG. Evaluating the effectiveness of air quality regulations: a review of accountability studies and frameworks. J Air Waste Manag Assoc. 2017;67(2):144–72. https://doi.org/10.1080/10962247.2016.1242518.
Jiang Y, Wu XJ, Guan YJ. Effect of ambient air pollutants and meteorological variables on COVID-19 incidence. Infect Control Hosp Epidemiol. 2020;222:1–11. https://doi.org/10.1017/ice.2020.222.
Kahneman D. Thinking, fast and slow. New York: Farrar, Straus, and Giroux; 2011.
Krimsky S. The weight of scientific evidence in policy and law. Am J Public Health. 2005;95(S1):129–36.
Linkov I, Massey O, Keisler J, Rusyn I, Hartung T. From “weight of evidence” to quantitative data integration using multicriteria decision analysis and Bayesian methods. ALTEX. 2015;32(1):3–8. https://doi.org/10.14573/altex.1412231.
North DW. Commentary on “Should health risks of air pollution be studied scientifically?” by Louis Anthony Cox, Jr. Global Epidemiol. 2020;2:100021. https://doi.org/10.1016/j.gloepi.2020.100021.
Omenn GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL Jr, Valanis B, Williams JH Jr, Barnhart S, Cherniack MG, Brodkin CA, Hammar S. Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. J Natl Cancer Inst. 1996;88(21):1550–9.
Pearl J. Causal inference in statistics: an overview. Stat Surveys. 2009;3:96–146.
Peters J, Bühlmann P, Meinshausen N. Causal inference by using invariant prediction: identification and confidence intervals. J R Stat Soc Ser B. 2016;78(5):947–1012.
Schwartz S, Gatto NM, Campbell UB. Transportability and causal generalization. Epidemiology. 2011;22(5):745–6.
Spirtes P. Introduction to causal inference. J Mach Learn Res. 2010;11:1643–62.
Stenlund T, Lidén E, Andersson K, Garvill J, Nordin S. Annoyance and health symptoms and their influencing factors: a population-based air pollution intervention study. Public Health. 2009;123(4):339–45. https://doi.org/10.1016/j.puhe.2008.12.021.
Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package ‘dagitty’. Int J Epidemiol. 2016;45(6):1887–94.
Tollefson J (2019) Air pollution science under siege at US environment agency Top EPA adviser attacks agency decision-making ahead of major review of air pollution standards. Nature. https://www.nature.com/articles/d41586-019-00937-w
Triantafillou S, Tsamardinos I. Constraint-based causal discovery from multiple interventions over overlapping variable sets. J Mach Learn Res. 2015;16:2147–205.
Wu X, Nethery RC, Sabath BM, Braun D, Dominic F (2020) Exposure to air pollution and COVID-19 mortality in the United States: a nationwide cross-sectional study. medRxiv 2020. 04.05.20205402
Zhang L, Zhang Z, Wang C, et al. Int J Disaster Risk Sci. 2017;8:445. https://doi.org/10.1007/s13753-017-0149-2.
Zigler CM, Dominici F. Point: clarifying policy evidence with potential-outcomes thinking—beyond exposure-response estimation in air pollution epidemiology. Am J Epidemiol. 2014;180(12):1133–40. https://doi.org/10.1093/aje/kwu263.