

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON DC

REGULATORY STUDIES CENTER

Working Paper

August 2012

Improving Causal Inferences in Risk Analysis

Louis Anthony (Tony) Cox, Jr., Ph.D. ¹

tcoxdenver@aol.com

¹ Tony Cox is President of Cox Associates, LLC. This working paper was conducted under contract with the GW Regulatory Studies Center, which raises awareness of regulations' effects with the goal of improving regulatory policy through research, education, and outreach. It does not represent an official position of the GW Regulatory Studies Center or the George Washington University.

Perspective: Improving Causal Inferences in Risk Analysis

Louis Anthony (Tony) Cox, Jr., Ph.D.

tcoxdenver@aol.com

August 24, 2012

ABSTRACT

Recent headlines and scientific articles projecting significant human health benefits from changes in exposures too often depend on unvalidated subjective expert judgments and modeling assumptions, especially about the causal interpretation of statistical associations. Some of these assessments are demonstrably biased toward false positives and inflated effects estimates. More objective, data-driven methods of causal analysis are available to risk analysts. These can help to reduce bias and increase the credibility and realism of health effects risk assessments and causal claims. For example, quasi-experimental designs and analysis allow alternative (non-causal) explanations for associations to be tested, and refuted if appropriate. Panel data studies examine empirical relations between changes in hypothesized causes and effects. Intervention and change-point analyses identify effects (e.g., significant changes in health effects time series) and estimate their sizes. Granger causality tests, conditional independence tests, and counterfactual causality models test whether a hypothesized cause helps to predict its presumed effects, and quantify exposure-specific contributions to response rates in differently exposed groups, even in the presence of confounders. Causal graph models let causal mechanistic hypotheses be tested and refined using biomarker data. These methods can potentially revolutionize the study of exposure-induced health effects, helping to overcome pervasive false-positive biases and move the health risk assessment scientific community toward more accurate assessments of the impacts of exposures and interventions on public health.

KEY WORDS: Accountability research, causality, causal modeling, Granger tests, panel data, intervention analysis, change-point analysis, causal graphs, counterfactual models

The George Washington University Regulatory Studies Center
Cox Associates LLC.

www.RegulatoryStudies.gwu.edu
www.cox-associates.com

Introduction: The Challenge of Causal Inference in Risk Analysis

Public health risk managers and policy makers are frequently presented with conflicting accounts of how the world works, and are urged by various interest groups – often passionately – to take different prompt, decisive actions based on these rival causal theories. Members of Congress are implored by many climate scientists to do more to curb climate change, before it is too late. Simultaneously, other groups beseech them not to spend resources on expensive actions that might create no, or little, or uncertain, benefits. While many financial economists and risk analysts call for tighter regulation of complex financial instruments, or better-funded public safety nets for big banks, or quicker and larger stimulus expenditures, others warn that these efforts risk exacerbating the problems they are meant to solve. Experts in development economics are split between those who encourage increasing aid payments to poor countries to jump-start their economies, and those who say that such transfers merely cement the wealth and power, and contribute to the corruption, of existing power elites. In these and countless other disagreements, both sides usually have more-or-less plausible stories about how different actions will cause different consequences, but their stories do not agree. This puts risk managers and policy makers in the uncomfortable position of having to assess the credibility of different causal theories – a task for which compelling data, decisive expertise, and provably useful training are often in short supply.

Two natural reactions to the challenge of judging among rival causal theories are to trust one's common sense and intuition, deferring to gut feel when cognition must admit defeat; and to rely on trusted scientific experts, who specialize in the relevant technical disciplines, for candid advice about the probable consequences caused by different choices. But modern scholarship has diminished the luster and apparent trustworthiness of both intuitive and expert judgments in matters of causation. Psychologists have shown convincingly that all of us, including experts in science and statistics, are prone to over-confidence in our own judgments; misattribution of causes; excessive inclination to blame people instead of situations; affect bias (in which emotional responses color our beliefs about facts, inclining us toward causal theories that agree with our intuitive perceptions of what is good or bad); motivated reasoning (which prompts us to believe whatever seems most profitable for us to believe); and confirmation bias (which leads us to see only what we expect, and to seek and interpret information selectively to reinforce our

beliefs, rather than to learn from reality) ([Fugelsang, 2004](#); [Gardner, 2009](#); Sunstein, 2009). For over a decade, the peer-reviewed scientific literature on risks and causes has been found to reflect these very human biases, with a large excess of false-positive errors in published results and in confident public assertions about health effects of various interventions ([Sarewitz, 2012](#), [Ottenbacher, 1998](#); [Imberger et al., 2011](#)). Attempts that fail to replicate published results may carry little professional or academic reward, undermining incentives to try to independently replicate key claims ([Sarewitz, 2012](#); Yong, 2012). Scientists with deep subject matter expertise are not necessarily or usually also experts in causal analysis and valid causal interpretation of data, and their causal conclusions are often mistaken. This has led some commentators to worry that “science is failing us,” due largely to widely publicized but false beliefs about causation ([Lehrer, 2012](#)); and that, in recent times, “Most published research findings are wrong” ([Ioannadis, 2005](#)), with the most sensational and publicized claims being most likely to be wrong.

To feel the pull of rival causal theories, consider the contrasting accounts of public health effects caused by air pollution, shown in Table 1. On the left are quotes from studies usually interpreted as showing that exposure to air pollutants (mainly, fine particulate matter (PM_{2.5})) causes increased risks of adverse health effects (e.g., [Pope, 2010](#)), along with some quantitative risk estimates for these effects. On the right are caveats and results of studies suggesting that these associations may not be causal after all. Both seem more or less plausible at first glance.

**Table 1. Some examples of conflicting claims
about health effects known to be caused by air pollution**

Pro (causal interpretation or claim)	Con (counter-interpretation or claim)
<p>“Epidemiological evidence is used to quantitatively relate PM_{2.5} 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” (Yim and Barrett, 2012).</p>	<p>“[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” (NHS, 2012)</p>

<p>“[A]bout 80,000 premature mortalities [per year] would be avoided by lowering PM2.5 levels to 5 µg/m³ 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% confidence interval of 51,000 to 200,000 (Fann et al., 2012).</p>	<p>“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” (EPA, 2011, Table 5-11).</p> <p>Significant negative associations have also been reported between exposures to some pollutants (e.g., NO2 (Kelly et al., 2012), PM2.5 (Krstić 2010) and ozone (Powell et al., 2012)) and short-term mortality and morbidity rates.</p>
<p>“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” (Buka et al., 2006). “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 (Friedman et al., 2001).</p>	<p>“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” (Health Effects Institute, 2010)</p>
<p>“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.” (Pope, 1989)</p>	<p>“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 PM₁₀ levels and hospitalizations was found for any age group or respiratory illness.” (Lamm et al., 1996)</p>
<p>“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.” (Clancy et al., 2002) "The results could not be more clear, reducing particulate air pollution reduces the number of respiratory and cardiovascular related deaths immediately" (Harvard School of Public Health, 2002).</p>	<p>The same rate of reduction in death rates was already occurring long before the ban, and occurred in other parts of Europe and Ireland not affected by it. “Serious epidemics and pronounced trends feign excess mortality previously attributed to heavy black-smoke exposure” (Wittmaack, 2007). “Thus, a causal link between the decline in mortality and the ban of coal sales cannot be established” (Pelucchi et al., 2009).</p>

If one's own judgment, scientific expert opinion, and the authority of peer-reviewed publications are all suspect as guides to the truth about such basic questions as whether air pollution caused adverse health effects in these studies, then how might one more objectively determine what causal conclusions are warranted by available facts and data? A common approach in epidemiology is to use statistical tests to determine whether there is strong evidence for a non-random positive association between exposure and response, and then to check whether, in the judgment of knowledgeable experts, the association can correctly be described by adjectives such as "strong," "consistent," "specific," "temporal," and "biologically plausible." The problem with this very popular approach is that all of these (and other) laudatory adjectives can apply perfectly well to associations even when there is no causation. Such associations can be created by strong confounders with time delays; or by data- and model-selection biases; or by unmodeled errors in exposure estimates; or by regression to the mean, or contemporaneous historical trends, or a host of other well-known threats to valid causal inference ([Campbell and Stanley, 1966](#); [Cox, 2007](#)). Applying adjectives to associations, as proposed in the thoughtful and influential work of Sir Bradford Hill, and as subsequently implemented in many weight-of-evidence schemes, does not overcome the basic limitation that an association is still only an association. Even the best qualified association may not reveal anything about causation, including the correct sign (positive or negative) of the causal influence of exposure on risk, if there is one. For example, if elderly people consume more baby aspirin than younger people to reduce their risk of heart attacks, then level of aspirin consumption might be significantly positively *associated* with increased risk of heart attack, even if increasing aspirin consumption would *cause* reduced heart attack risk at every age.

More generally, causality in risk analysis is not mainly about statistical associations between *levels* of passively observed variables; but rather about how *changes*, if made, would propagate through systems ([Druzdzel and Simon, 1993](#); [Greenland and Brumback, 2002](#)). This distinction should be of critical importance to risk analysts advising policy makers on the probable consequences of proposed interventions, and also to policy makers considering how much weight to give such advice. As a real-world example of how much it matters, mortality rates among the elderly tend to be elevated where and when fine particulate pollutant concentrations are highest among 100 U.S. cities (namely, in cities and months with cold winter

days), and yet changes in these pollutant concentration levels from one year to the next are significantly *negatively* associated with corresponding changes in mortality rates, undermining any straightforward causal interpretation of the positive association between pollutant levels and mortality rates (Cox, 2012). Yet, this crucial distinction is often glossed over in the current language and presentation of health risk assessment results. For example, one recent article (Lepeue et al., 2012) announced that that, for six U.S. cities, “Using the Cox proportional hazards model, statistically significant associations between [fine particulate matter] PM2.5 exposure and all-cause, cardiovascular, and lung-cancer mortality were observed. ...Each 10- $\mu\text{g}/\text{m}^3$ increase in PM2.5 was associated with a 14% increased risk of all-cause death.” But the word “increase” here does *not* refer to any actual change (increase over time) in PM2.5 levels or risk over time. Instead, it refers to associations between higher *levels* of PM2.5 and higher levels of risk. The study then infers that “These results suggest that further public policy efforts that reduce fine particulate matter air pollution are likely to have continuing public health benefits.” But this causal conclusion about predicted effects of *changes* does not follow from the statistical association between *levels* of PM2.5, since the two may (and in fact, in the U.S., often do) have opposite signs (Cox, 2012). The contrasting statements on the left and right sides of Table 1 suggest that health effects researchers not infrequently leap from observations of associations to conclusion about causation, without carefully checking whether changes in inputs produce the changes in outputs that static associations between them suggest. This casual treatment of key causal questions must change, if risk analysis predictions are to become more accurate and trustworthy.

Risk management advice based on past statistical exposure-response associations (or other associations) may not be very useful for correctly predicting probable effects of future changes in exposures (or other variables) brought about by risk management interventions. Instead, an understanding of *causal mechanisms* – that is, of how changes in some variables change others – is usually necessary to correctly predict the effects of interventions (Greenland and Brumback, 2002; Freedman, 2004). This need not be difficult or mysterious. Simulation models (e.g., based on systems of differential and algebraic equations) describing flows of quantities among compartments over time, and the effects of interventions on flow rates, suffices to model the effects of interventions in many practical settings (Dash and Druzdzel, 2008;

[Druzdzel and Simon 1993](#); [Lu et al. 2000](#)). However, shifting the emphasis from making judgments about the causal interpretation or “weight-of-evidence” of statistical associations to rigorous formal testing of causal hypotheses, formulated in terms of propagation of changes along causal paths (or through more complex causal networks), requires a major change in commonly taught epidemiological practices.

How to Do Better: More Objective Tests for Causal Impacts

Happily, modern methods of causal analysis now enable risk analysts to address questions about causation by considering relatively objective evidence on how and whether changes in the inputs to a system propagate to cause changes in its outputs. This is a far more useful, and objective, approach than making judgments about statistical associations, for reasons given below. Well-developed methodological principles for drawing sound causal inferences from observational data include asking (and using data to answer) the following simple, systematically skeptical, questions about observed exposure-response associations, to test whether the observations are logically capable of providing evidence for a genuine causal relation.

- *Do the study design and data collected permit convincing refutation of non-causal explanations for observed associations* between levels of exposure and response (or between levels of other hypothesized cause and effect variables)? Potential non-causal explanations for associations include data selection and model selection biases, residual confounding by modeled confounders, unmodeled confounders, unmodeled errors in exposure estimates and covariates, unmodeled uncertainties in model form specification, regression to the mean, and so forth ([Cox, 2007](#)). These potential rival explanations can be ruled out by appropriate study designs, control group comparisons, and data analyses, if they indeed do not explain the observed associations ([Campbell and Stanley, 1966](#); [Maclure, 1990](#); [Cox, 2007](#)). Assuming that they have been ruled out, the next questions consider whether there is objective evidence that the observed relation might be causal.
- *Are significant positive associations also found between changes in exposures and changes in response rates?* If the answer is no, as revealed in some panel data studies of previously

reported positive associations between exposure and response levels ([Stebbing, 1978](#)), then this undermines causal interpretation of the positive associations.

- *Do changes in hypothesized causes precede changes in their hypothesized effects?* If not, e.g., if health effects are already declining before reductions in exposure, then this casts doubt on the latter being a cause of the former. Doubt is increased if, as in the Dublin study in Table 1, a steep reduction in exposure is not followed by any detectable corresponding change in the rate of decline in effects.
- *Are reductions in hypothesized effects significantly greater in times and places where exposure went down than where exposure remained the same or went up?* If not, as in the [HEI \(2010\)](#) analysis of the Atlanta Olympics data in Table 1, then this casts doubt on the hypothesis that reductions in exposure caused the reductions in effects.
- *Do changes in hypothesized causes (e.g., exposures) help to predict subsequent changes in their hypothesized effects?* If not, e.g., if changes in effects appear to be statistically independent of previous changes in the hypothesized causes, then this reduces the plausibility of a causal interpretation for a regression model, or other statistical model, relating them.

Such qualitative questions provide clear common-sense and logical foundations as screens for causal inference, and they are relatively easy to understand and ask.

Quantitative methods, although sometimes technically sophisticated, help to implement many of the same basic ideas, and to provide relatively objective answers using formal statistical tests. Among the most useful analytic methods for testing causal hypotheses and constructing valid causal models from data are the following.

- *Intervention analysis* ([Friede et al., 2006](#)), also called interrupted time series analysis, tests whether the best-fitting model of the data-generating process for an observed time series, such as daily mortality and morbidity counts, changed significantly at, or following, the time of an intervention. Intervention analysis provides methods to identify, test for, and estimate significant changes in time series that might have been caused by an intervention, and that cannot easily be explained by other (non-causal) hypotheses.
- *Change-point analysis* ([Helfenstein, 1991](#); [Gilmour et al., 2006](#)) searches for any significant changes in the data-generating process over an interval of time – for example, a change in the

slope of a long-term declining trend in cardiovascular mortality rates, or a change in the season-specific rate of hospital admissions for pediatric asthma. If a change point is detected at, or closely following, an intervention, such as an emissions ban that reduces pollution levels, then the intervention might have caused the change. If no change is detected, then there is no evidence that the intervention had a detectable effect.

- *Quasi-experimental designs* and methods ([Campbell and Stanley, 1966](#)) make use of control-group comparisons (including pre- and post-intervention observations on the same subjects) to try to systematically refute each of a list of identified methodological threats to valid causal inference, such as “History” (e.g., that the Dublin coal ban occurred during a long-term historical trend toward lower cardiovascular rates due to better prevention, diagnosis, and treatment), regression to the mean (unusual bursts of ill effects tend to be followed by lower levels even if any intervention that they may have triggered have no effect), aging of subjects, and so forth.
- *Panel data analysis* ([Angrist and Pischke, 2009](#)) examines how well *changes* in explanatory variables predict *changes* in responses, using repeated measures of the same observational units over time to control for unobserved confounders. In health risk assessment, comparing changes in exposures to changes in responses can give a very different understanding of the likely health consequences caused by changes in exposure than studying estimated (or assumed) statistical associations between exposure and response levels ([Cox, 2012](#), [Stebbings, 1976](#)).
- *Granger causality tests* ([Eichler and Didelez, 2010](#)). Changes in causes should help to predict subsequent changes in their effects, even if there is no intervention in the time series being observed. To formally test whether changes in exposure might be a contributing cause of changes in short-term daily mortality rates, for example, one could compare a simple predictive model, created by regressing future mortality rates against their own past (lagged) values, to a richer model that also regresses them against lagged values of exposure as possible predictors. If including exposure history does not improve predictions of mortality rates (e.g., producing smaller mean squared prediction errors or larger mutual information between predicted and observed values), then the time series data do not support the hypothesis that exposure causes mortality, in the sense of helping to predict it. This method

of testing causal hypotheses is incorporated in the *Granger causality test*. (It is now widely and freely available, for example, as the *granger.test* procedure in R.) In practice, Granger-causality testing may show that some significant correlates of short-term mortality rates (such as low temperature ([Mercer, 2003](#))) are also Granger-causes of the short-term mortality rates, while others (e.g., PM2.5) are not Granger-causes ([Cox, 2012](#)). Although Granger tests are subject to the usual limitations of parametric modeling assumptions, such as the use of a linear regression model, the lack of Granger-causation between exposure and response even when there is a clear, statistically significant positive regression relation between them, highlights the importance of distinguishing between positive regression relations and causal relations. (This distinction has not been prominent in the air pollution health effects accountability literature to date, but deserves to be in future.)

- *Conditional independence tests* ([Freedman, 2004](#), [Friedman and Goldszmidt, 1998](#)). In both cross-sectional and longitudinal data, a cause should provide some information about its effect that cannot be obtained from other sources. Conversely, if an effect is *conditionally independent* of a hypothesized cause, given the values of other explanatory variables (e.g., measured potential confounders and covariates), then the causal hypothesis is not supported by the data ([Freedman, 2004](#)). For example, if daily mortality rates are conditionally independent of pollution levels, given city and month and temperature, then there would be no evidence that pollution levels make a causal contribution to daily mortality rates. Conversely, if there is no way to eliminate the significant difference between mortality rates for very different pollutant levels, holding other covariate levels fixed, then pollutant levels would appear as direct causes (“parents”) of daily mortality rates in causal graph models ([Freedman, 2004](#), [Friedman and Goldszmidt, 1998](#)).
- *Counterfactual and potential outcome models*. One possible definition of the causal impact of exposure on mortality rates in a population or subpopulation is the difference between the average mortality rate with real exposures and the projected average mortality rate (typically derived from regression models) if all members had reduced or zero exposures. Although this requires considering counterfactual exposures and responses, since no individual can be both exposed and unexposed, there has much recent progress in technical methods for developing and fitting counterfactual regression models (“marginal structural models” and

their extensions) to predict what would have happened if exposure had been lower, or absent ([Robins et al. 2000](#); [Moore et al., 2012](#)). Such counterfactual causal models can yield insights and conclusions quite different from earlier regression models. For example, in one recent study, adverse effects of ozone exposure that are statistically significant in earlier regression models (which must make unverifiable modeling assumptions about what responses would be to combinations of predictors that do not occur in reality), are not significant when methods are applied that only use realistic exposure-response data ([Moore et al., 2012](#)).

- *Modeling causal mechanisms via propagation of changes through chains or networks of causal predecessors.* If exposure causes adverse health effects, it must do so via one or more causal pathways. Collecting biomarker data can allow specific causal hypotheses about the mechanisms of harm to be tested ([Hack et al., 2010](#)). For example, causal graph models ([Freedman, 2004](#)), although in general only providing a way to factor joint distributions into marginal and conditional distributions, can be constructed to preserve causal orderings discovered from structural equations or mechanistic models ([Dash and Druzdzel, 2008](#); [Druzdzel and Simon 1993](#); [Lu et al. 2000](#)). In this case, absence of changes in the intermediate variables that are hypothesized to mediate the transmission of causal impacts from exposure to health response would provide evidence against the hypothesized causal mechanism. Conversely, detecting and quantifying those changes (via the conditional probability relations at intermediate nodes in a causal graph model) allows prediction of the sizes of changes in health effects to be expected from changes in exposure, given the values of other variables in the causal model. For example, a recent study ([Chuang et al., 2007](#)) provided panel data to test the specific mechanistic hypotheses for air pollution health effects, including that PM_{2.5} causes adverse cardiovascular effects by increasing oxidative stress as measured by urinary 8-hydroxy-2'-deoxyguanosine (a marker for oxidative DNA damage). As summarized by [Kaufman \(2007\)](#), “blood and electrocardiographic markers were repeatedly collected over 3 months to examine multiple potential mechanistic pathways. They had the benefit of fairly large daily fluctuations in exposure, presumably dictated by meteorological conditions. While their inflammatory, oxidative stress, fibrinolysis, and coagulation health markers did not change consistently as hypothesized with fine particles,

they did detect associations with some PM components and credited these to traffic-related air pollution.... Their measure of “oxidative stress” (urinary 8-hydroxy-2'-deoxyguanosine, assessing oxidative DNA damage) was not associated with pollution exposures. Heart rate variability metrics, on the other hand, consistently demonstrated negative associations with all air pollutants examined, in a manner that appeared to be independent of inflammation.” This ability to refute expected causal hypotheses and to reveal unexpected time-ordered sequences of changes makes panel data especially valuable for learning from data by testing and improving mechanistic models.

Table 2 summarizes some of the best-developed quantitative methods for testing causal hypotheses and for quantifying the sizes of causal effects. These methods of causal analysis are relatively objective. Unlike expert judgments and opinions about the causal interpretation of statistical associations, they can be independently replicated by others, using standard statistical methods (such as *granger.test* in the R statistical computing environment). They focus on answering the following key factual questions.

1. *Can any effect (e.g., a significant change in a health effects time series following a change in exposures) be detected?* Methods for detecting such changes include change-point analysis, intervention analysis, and panel data analysis. If there is no apparent effect, as in the Dublin study data ([Wittmaack, 2007](#)), then there is nothing to explain, and proffered causal interpretations are superfluous.
2. *If so, how large is it?* This may be assessed via intervention analysis, change-point models, panel data, or quasi-experimental pre-post comparisons, with counterfactual causal models untangling the effects of confounders and estimating the remaining effect specifically caused by exposure ([Moore et al., 2012](#)). If this causal effect is only a fraction as large as the statistical “effect” estimated from a regression model, for example, then only a portion of the statistical association should be attributed to exposure, as opposed to confounding.
3. *Can changes in responses be explained or predicted as well without knowledge of a putative cause as with it?* This crucial screening question can be answered using Granger tests, conditional independence tests, and quasi-experimental analyses to refute other explanations. If knowledge of changes in a hypothesized cause does not improve ability to predict its

hypothesized effects, or, conversely, if the effects can be explained by other variables and are conditionally independent of the hypothesized cause, given the values of the other variables, then the causal hypothesis is not supported.

4. *Are the changes in causal predecessors implied by hypothesized causal mechanisms observed?* This can be addressed using causal graph models and panel data analysis applied to biomarker data

Using modern methods of causal analysis to address these factual questions can liberate risk analysts and policy makers from the need to rely on (potentially biased or unreliable) subjective judgments in addressing questions of causality. They provide an alternative to the traditional Hill-type criteria (such as strength, consistency, specificity, and temporality of associations).

Table 2. Some formal methods for modeling and testing causal hypotheses

Method and References	Basic Idea	Appropriate study design
Conditional independence tests (Freedman, 2004 , Friedman and Goldszmidt, 1998)	Is hypothesized effect statistically independent of other (“ancestor”) variables, given values of hypothesized direct causes (“parents” in causal graph model)? If so, this strengthens causal interpretation. Is hypothesized effect statistically independent of hypothesized cause, given the values of other variables? If so, this undermines causal interpretation.	Cross-sectional data Can also be applied to multi-period data (in dynamic Bayesian networks (DBNs))
Panel data analysis (Angrist and Pischke, 2009 , Stebbing, 1976)	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.	Panel data study: Collect a sequence of observations on same subjects or units over time
Granger causality test (Eichler and Didelez, 2010)	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.	Time series data on hypothesized causes and effects

<p>Quasi-experimental design and analysis (Campbell and Stanley, 1966)</p>	<p>Can control groups and other comparisons refute alternative (non-causal) explanations for observed associations between hypothesized causes and effects? For example, can coincident trends and regression to the mean be refuted as possible explanations? If so, this strengthens causal interpretation.</p>	<p>Longitudinal observational data on subjects exposed and not exposed to interventions that change the hypothesized cause(s) of effects.</p>
<p>Intervention analysis, change point analysis (Helfenstein, 1991; Gilmour et al., 2006; Friede et al., 2006)</p>	<p>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.</p> <p>Do the quantitative changes in hypothesized causes predict and explain the subsequently observed quantitative changes in hypothesized effects? If so, this strengthens causal interpretation.</p>	<p>Time series observations on hypothesized effects, and knowledge of timing of intervention(s)</p> <p>Quantitative time series data for hypothesized causes and effects</p>
<p>Counterfactual and potential outcome models (Robins et al. 2000; Moore et al., 2012)</p>	<p>Do exposed individuals have significantly different response probabilities than they would have had if they had not been exposed?</p>	<p>Cross-sectional and/or longitudinal data, with selection biases and feedback among variables allowed</p>
<p>Causal network models of change propagation (Hack et al., 2010, Dash and Druzdzal, 2008)</p>	<p>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?</p>	<p>Observations of variables in a dynamic system out of equilibrium</p>

Conclusions: Improving Causal Analysis of Health Effects

The principles and techniques sketched here for protecting against false conclusions and for drawing sound causal inference from data are ready for practical use in health effects research and risk assessment, both to reduce uncertainty about causality in concentration-response functions, and to more clearly delineate needed distinctions between causal and non-causal associations. Taking seriously the need to apply more objective methods to assess causality in health risk assessment suggests the following policy-relevant perspectives.

- *Expert judgment-based assessments of causality, and subjective causal interpretations of statistical associations, are unreliable and prone to error and bias.* This is illustrated in examples where confidently expressed expert conclusions and more formal causal analyses conflict (as for several studies in Table 1). The prevalence of confirmation bias ([Fugelsang et al., 2004](#); [Gardner 2009](#); Sunstein, 2009) makes it crucial for expert panels (or individuals) tasked with forming judgments about causation to seek out well-supported contrary views.
- *It is possible and practical to do better.* More objective methods for causal analysis are now readily available, and more informative designs and analyses (e.g., using panel data to study changes in exposure and response variables, instead of using regression models to study associations between their levels) can eliminate much of the speculation, controversy, and ambiguity surrounding causation in health effects research.
- *The credibility of conclusions about causation, and the credibility of risk assessments and health benefits projections based on them, should be assessed based on how well they provide sound, independently reproducible, answers to specific, factual, causal questions.* These include addressing whether observed changes in hypothesized causal predecessors do in fact precede and help to explain or predict observed changes in their hypothesized effects. Passionate or confident beliefs about causation expressed by subject matter experts who have not yet addressed these questions using data and independently reproducible analyses, should be regarded as expressions of personal belief, rather than as answers to scientific questions.

Following these recommendations could transform health effects accountability research ([Pope, 2010](#)), by promoting health benefits estimates for exposure reductions that are more realistic, and more solidly based on reproducible science and data, than those driving headlines today. This

would reduce needless controversies over the interpretation of ambiguous statistical associations; focus attention on the sizes of demonstrable real-world causal impacts; and shift the emphasis of health effects claims for emissions reductions toward more objective and independently verifiable risk analysis.

REFERENCES

- Angrist JD, Pischke J-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. 2009 Princeton University Press, Princeton, NJ.
- Buka I, Koranteng S, Osornio-Vargas AR. [The effects of air pollution on the health of children](#). *Paediatr Child Health*. 2006 Oct;11(8):513-6.
- California Air Resources Board. Effects of air pollution control on hospital admissions and mortality.” California Environmental Protection Agency. May 22, 2003. www.docstoc.com/docs/43067212/Effect-of-Air-Pollution-Control-on-Hospital-Admissions-and
- Campbell DT, Stanley JC. *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand McNally, 1966
- Chuang KJ, Chan CC, Su TC, Lee CT, Tang CS. [The effect of urban air pollution on inflammation, oxidative stress, coagulation, and autonomic dysfunction in young adults](#). *Am J Respir Crit Care Med*. 2007 Aug 15;176(4):370-6.
- Clancy L, Goodman P, Sinclair H, Dockery DW. [Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study](#). *Lancet*. 2002 Oct 19;360(9341):1210-4.
- Cox LA Jr. Warmer is healthier: Effects on mortality rates of changes in average fine particulate matter (PM2.5) concentrations and temperatures in 100 U.S. Cities. *Dose-Response*. 2012
- Cox LA Jr. [Hormesis for fine particulate matter](#) (PM2.5). *Dose-Response*. 2012
- Cox LA Jr. [Regulatory false positives: true, false, or uncertain?](#). *Risk Analysis* Risk Anal. 2007 Oct;27(5):1083-6
- Dash D, Druzdel MJ. A note on the correctness of the causal ordering algorithm. 2008. *Artificial Intelligence* 172:1800–1808. <http://www.pitt.edu/~druzdel/psfiles/aij08.pdf>
- Druzdel MJ, Simon HA. [Causality in Bayesian belief networks](#). In Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93), pages 3-11, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1993
- Elliott CT, Copes R. [Burden of mortality due to ambient fine particulate air pollution \(PM2.5\) in interior and Northern BC](#). *Can J Public Health*. 2011 Sep-Oct;102(5):390-3.
- EPA (U.S. Environmental Protection Agency). 2011. The Benefits and Costs of the Clean Air Act from 1990 to 2020. Final Report – Rev. A. Office of Air and Radiation, Washington D.C.
- EPA (U.S. Environmental Protection Agency). 2009. Integrated Science Assessment for Particulate Matter (Final Report) <http://cfpub.epa.gov/ncea/cfm/recorddisplay.cfm?deid=216546> .
-
- The George Washington University Regulatory Studies Center www.RegulatoryStudies.gwu.edu
Cox Associates LLC. www.cox-associates.com

Eichler M, Didelez V. [On Granger causality and the effect of interventions in time series.](#) Lifetime Data Anal. 2010 Jan;16(1):3-32.

Ezzati M, Hoorn SV, Lopez AD, Danaei G, Rodgers A, Mathers CD, Murray CJL. [Comparative Quantification of Mortality and Burden of Disease Attributable to Selected Risk Factors.](#) In: Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL, editors. Global Burden of Disease and Risk Factors. Washington (DC): World Bank; 2006. Chapter 4.

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 Analysis* Jan. 2012 32(1):81-95

Franklin M, Zeka A, Schwartz J. 2007. Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. *J Expo Sci Environ Epidemiol.* May;17(3):279-87.

Freedman DA. [Graphical models for causation, and the identification problem.](#) Eval Rev. 2004 Aug;28(4):267-93.

Friede T, Henderson R, Kao CF. [A note on testing for intervention effects on binary responses.](#) Methods Inf Med. 2006;45(4):435-40.

Friedman, N., & Goldszmidt, M. (1998) [Learning Bayesian networks with local structure.](#) In M.I. Jordan (Ed.), Learning in Graphical Models (pp 421-459). MIT Press. Cambridge, MA.

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 Feb 21;285(7):897-905.

Fugelsang JA, Stein CB, Green AE, Dunbar KN. [Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory.](#) Can J Exp Psychol. 2004 Jun;58(2):86-95.

Gardner, D. (2009). *The Science of Fear: How the Culture of Fear Manipulates Your Brain.* Penguin Group. New York, New York.

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 Apr 3;6:16.

Greenland S, Brumback B. [An overview of relations among causal modelling methods.](#) Int J Epidemiol. 2002 Oct;31(5):1030-7.

Hack CE, Haber LT, Maier A, Shulte P, Fowler B, Lotz WG, Savage RE Jr. [A Bayesian network model for biomarker-based dose response.](#) Risk Anal. 2010 Jul;30(7):1037-51.

Harvard School of Public Health, 2002. Press Release: "Ban On Coal Burning in Dublin Cleans the Air and Reduces Death Rates" www.hsph.harvard.edu/news/press-releases/archives/2002-releases/press10172002.html

Health Effects Institute (HEI) (2010). Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes. HEI Research Report #148. April, 2010. Authors: Jennifer L. Peel, Mitchell Klein, W. Dana Flanders, James A. Mulholland, and Paige E. Tolbert. Health Effects Institute. Boston, MA. <http://pubs.healtheffects.org/getfile.php?u=564>

Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. [Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study.](#) Lancet. 2002 Nov 23;360(9346):1646-52.

Helfenstein U. (1991) [The use of transfer function models, intervention analysis and related time series methods in epidemiology.](#) Int J Epidemiol. Sep;20(3):808-15.

Imberger G, Vejlbj AD, Hansen SB, Møller AM, Wetterslev J (2011) Statistical Multiplicity in Systematic Reviews of Anaesthesia Interventions: A quantification and comparison between Cochrane and non-Cochrane reviews. PLoS ONE 6(12): e28422. doi:10.1371/journal.pone.0028422

Ioannidis JPA. Why most published research findings are false. 2005. PLoS Med 2(8): e124. doi:10.1371/journal.pmed.0020124

Kaufman JD. [Air pollution and mortality: are we closer to understanding the how?](#) Am J Respir Crit Care Med. 2007 Aug 15;176(4):325-6.

Kelly F, Armstrong B, Atkinson R, Anderson HR, Barratt B, Beevers S, Cook D, Green D, Derwent D, Mudway I, Wilkinson P; HEI Health Review Committee. [The London low emission zone baseline study.](#) Res Rep Health Eff Inst. 2011 Nov;(163):3-79.

Krstić G. [Apparent temperature and air pollution vs. elderly population mortality in Metro Vancouver.](#) PLoS One. 2011;6(9):e25101

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., 38:969-972, 1994.

Lehrer J. Trials and errors: Why science is failing us. Wired. January 28, 2012. <http://www.wired.co.uk/magazine/archive/2012/02/features/trials-and-errors?page=all>

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 Jul;120(7):965-70.

Lu T-C, Druzdzel MJ, Leong T-Y. [Causal mechanism-based model construction](#). In Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-00), pages 353-362, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 2000

Maclure M. [Multivariate refutation of aetiological hypotheses in non-experimental epidemiology](#). Int J Epidemiol. 1990 Dec;19(4):782-7.

Mercer JB. [Cold--an underrated risk factor for health](#). Environ Res. 2003 May;92(1):8-13.

Moore KL, Neugebauer R, van der Laan MJ, Tager IB. [Causal inference in epidemiological studies with strong confounding](#). Stat Med. 2012 Feb 23. doi: 10.1002/sim.4469.

Moshammer H. (2010). Communicating Health Impact of Air Pollution, Air Pollution, Vanda Villanyi (Ed.), ISBN: 978-953-307-143-5, InTech, Available from:
<http://www.intechopen.com/books/air-pollution/communicating-health-impact-of-air-pollution>

NHS, 2012. Air pollution 'kills 13,000 a year' says study
www.nhs.uk/news/2012/04april/Pages/air-pollution-exhaust-death-estimates.aspx

Ottenbacher KJ (1998) Quantitative evaluation of multiplicity in epidemiology and public health research. Am J Epidemiol 147: 615–619.

Pope CA 3rd. [Respiratory disease associated with community air pollution and a steel mill, Utah Valley](#). Am J Public Health. 1989 May;79(5):623-8.

Pope CA 3rd. [Particulate pollution and health: a review of the Utah valley experience](#). J Expo Anal Environ Epidemiol. 1996 Jan-Mar;6(1):23-34.

Pope, CA (2010). Accountability Studies of Air Pollution and Human Health: Where are We Now, and Where Does the Research Need to Go Next?
<http://pubs.healtheffects.org/getfile.php?u=584>

Powell H, Lee D, Bowman A. Estimating constrained concentration–response functions between air pollution and health. Environmetrics. 2012 May;23(3): 228-237.

Roberts S. [Biologically plausible particulate air pollution mortality concentration-response functions](#). Environ Health Perspect. 2004 Mar;112(3):309-13.

Robins JM, Hernán MA, Brumback B. [Marginal structural models and causal inference in epidemiology](#). Epidemiology. 2000 Sep;11(5):550-60.

Röösli M, Künzli N, Braun-Fahrländer C, Egger M. [Years of life lost attributable to air pollution in Switzerland: dynamic exposure-response model](#). Int J Epidemiol. 2005 Oct;34(5):1029-35.

Sarewitz D. [Beware the creeping cracks of bias](#). Nature. 10 May 2012 485:149

Shindell D, Kuylenstierna JC, Vignati E, van Dingenen R, Amann M, Klimont Z, Anenberg SC, Muller N, Janssens-Maenhout G, Raes F, Schwartz J, Faluvegi G, Pozzoli L, Kupiainen K, Höglund-Isaksson L, Emberson L, Streets D, Ramanathan V, Hicks K, Oanh NT, Milly G, Williams M, Demkine V, Fowler D. [Simultaneously mitigating near-term climate change and improving human health and food security](#). Science. 2012 Jan 13;335(6065):183-9.

Stebbing JH Jr. [Panel studies of acute health effects of air pollution. II. A methodologic study of linear regression analysis of asthma panel data](#). Environ Res. 1978 Aug;17(1):10-32.

Sunstein, CR. 2009. [Going to Extremes: How Like Minds Unite and Divide](#). Oxford University Press. New York, New York.

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](#). Inhal Toxicol. 2007 Apr;19(4):343-50.

Wong CM, Atkinson RW, Anderson HR, Hedley AJ, Ma S, Chau PY, Lam TH. [A tale of two cities: effects of air pollution on hospital admissions in Hong Kong and London compared](#). Environ Health Perspect. 2002 Jan;110(1):67-77.

Wong CM, Ma S, Hedley AJ, Lam TH. [Effect of air pollution on daily mortality in Hong Kong](#). Environ Health Perspect. 2001 Apr;109(4):335-40.

Yim SH, Barrett SR. [Public health impacts of combustion emissions in the United Kingdom](#). Environ Sci Technol. 2012 Apr 17;46(8):4291-6.

Yong, E. [Replication studies: Bad copy](#). Nature 17 May, 2012. 485: 298–300