

---

# Quantifying and Reducing Uncertainty About Causality in Improving Public Health and Safety

# 43

Louis Anthony Cox, Jr.

---

## Abstract

Effectively managing uncertain health, safety, and environmental risks requires quantitative methods for quantifying uncertain risks, answering the following questions about them, and characterizing uncertainties about the answers:

- Event detection: What has changed recently in disease patterns or other adverse outcomes, by how much, when?
- Consequence prediction: What are the implications for what will probably happen next if different actions (or no new actions) are taken?
- Risk attribution: What is causing current undesirable outcomes? Does a specific exposure harm human health, and, if so, who is at greatest risk and under what conditions?
- Response modeling: What combinations of factors affect health outcomes, and how strongly? How would risks change if one or more of these factors were changed?
- Decision making: What actions or interventions will most effectively reduce uncertain health risks?
- Retrospective evaluation and accountability: How much difference have exposure reductions actually made in reducing adverse health outcomes?

These are all causal questions. They are about the uncertain causal relations between causes, such as exposures, and consequences, such as adverse health outcomes. This chapter reviews advances in quantitative methods for answering them. It recommends integrated application of these advances, which might collectively be called causal analytics, to better assess and manage uncertain risks. It discusses uncertainty quantification and reduction techniques for causal modeling that can help to predict the probable consequences of different policy

---

L.A. Cox, Jr., (✉)

Cox Associates and University of Colorado, Denver, CO, USA

e-mail: [tcoxdenver@aol.com](mailto:tcoxdenver@aol.com)

choices and how to optimize decisions. Methods of causal analytics, including change-point analysis, quasi-experimental studies, causal graph modeling, Bayesian Networks and influence diagrams, Granger causality and transfer entropy methods for time series, and adaptive learning algorithms provide a rich toolkit for using data to assess and improve the performance of risk management efforts by actively discovering what works well and what does not.

### Keywords

Adaptive learning • Change-point analysis (CPA) • Bayesian networks (BN) • Causal analytics • Causal graph • Causal laws • Counterfactual • Directed acyclic graph (DAG) • DAG model • Dynamic Bayesian networks (DBN) • Ensemble learning algorithms • Evaluation analytics • Granger causality • Influence diagram (ID) • Intervention analysis • Interrupted time series analysis • Learning analytics • Marginal structural model (MSM) • Model ensembles • Multi-agent influence diagram (MAID) • Path analysis • Predictive analytics • Prescriptive analytics • Propensity score • Quasi-experiments (QEs) • Simulation • Structural equations model (SEM) • Structure discovery • Transfer Entropy • Uncertainty analytics

### Contents

1	Introduction	1439
2	Some Limitations of Traditional Epidemiological Measures for Causal Inference: Uncertainty About Whether Associations Are Causal	1442
2.1	Example: Exposure-Response Relations Depend on Modeling Choices	1443
3	Event Detection and Consequence Prediction: What's New, and So What?	1448
3.1	Example: Finding Change Points in Surveillance Data	1449
4	Causal Analytics: Determining Whether a Specific Exposure Harms Human Health	1453
5	Causes and Effects Are Informative About Each Other: DAG Models, Conditional Independence Tests, and Classification and Regression Tree Algorithms	1454
6	Changes in Causes Should Precede, and Help to Predict and Explain, Changes in the Effects that They Cause	1455
6.1	Change-Point Analysis Can Be Used to Determine Temporal Order	1456
6.2	Intervention Analysis Estimates Effects of Changes Occurring at Known Times, Enabling Retrospective Evaluation of the Effectiveness of Interventions	1456
6.3	Granger Causality Tests Show Whether Changes in Hypothesized Causes Help to Predict Subsequent Changes in Hypothesized Effects	1459
7	Information Flows From Causes to Their Effects over Time: Transfer Entropy	1460
8	Changes in Causes Make Future Effects Different From What They Otherwise Would Have Been: Potential-Outcome and Counterfactual Analyses	1461
9	Valid Causal Relations Cannot Be Explained Away by Noncausal Explanations	1466
10	Changes in Causes Produce Changes in Effects via Networks of Causal Mechanisms	1466
10.1	Structural Equation and Path Analysis Models Model Linear Effects Among Variables	1467

10.2	Bayesian Networks Show How Multiple Interacting Factors Affect Outcome Probabilities . . . . .	1470
10.3	Quantifying Probabilistic Dependencies Among BN Variables . . . . .	1471
10.4	Causal vs. Noncausal BNs . . . . .	1473
10.5	Causal Mechanisms Are Lawlike, Yielding the Same Output Probabilities for the Same Inputs . . . . .	1474
10.6	Posterior Inference in BN Models . . . . .	1475
10.7	Causal Discovery of BNs from Data . . . . .	1477
10.8	Handling Uncertainty in Bayesian Network Models . . . . .	1479
10.9	Influence Diagrams Extend BNs to Support Optimal Risk Management Decision-Making . . . . .	1482
10.10	Value of Information (VOI), Dynamic Bayesian Networks (DBNs), and Sequential Experiments for Reducing Uncertainties Over Time . . . . .	1485
11	Causal Analytics . . . . .	1488
12	Summary and Conclusions: Applying Causal Graph Models to Better Manage Risks and Uncertainties . . . . .	1491
	Cross-References . . . . .	1493
	References . . . . .	1493

## 1 Introduction

Politically contentious issues often turn on what appear to be technical and scientific questions about cause and effect. Once a perceived undesired state of affairs reaches a regulatory or policy agenda, the question arises of what to do about it to change things for the better. Useful answers require understanding the probable consequences caused by alternative policy actions. For example,

- A century ago, policy makers might have asked whether a prohibition amendment would decrease or increase alcohol abuse.
- A decade ago, policy makers might have wondered whether an invigorated war on drugs would increase or decrease drug abuse.
- Do seatbelts reduce deaths from car accidents, even after accounting for “risk homeostasis” changes in driving behaviors?
- Does gun control reduce deaths due to shootings?
- Does the death penalty reduce violent crime?
- Has banning smoking in bars reduced mortality rates due to heart attacks?
- Do sex education and birth control programs in schools decrease teen pregnancy rates and prevalence of sexually transmitted diseases?
- Has the Clean Air Act reduced mortality rates, e.g., due to lung cancer or coronary heart disease (CHD) or to all causes?
- Will reformulations of childhood vaccines reduce autism?
- Would banning routine antibiotic use in farm animals reduce antibiotic-resistant infections in people?

Policy makers look to epidemiologists, scientists, and risk analysts to answer such questions. They want to know how policy actions will change (or already have changed) outcomes and by how much – how much improvement is caused how

quickly and how long does it last? They want to know what will work best and what has (and has not) worked well in reducing risks and undesirable outcomes without causing unintended adverse consequences. And they want to know how certain or uncertain the answers to these questions are.

Developing trustworthy answers to these questions and characterizing uncertainty about them requires special methods. It is notoriously difficult to quickly and accurately identify events or exposures that cause adverse human health outcomes, quantify uncertainties about causal relations and impacts, accurately predict the probable consequences of a proposed action such as a change in exposure or introduction of a regulation or intervention program, and quantify in retrospect what effects an action actually did cause, especially if other changing factors affected the observed outcomes. The following section explains some limitations of association-based epidemiological and regulatory risk assessment methods that are often used to try to answer these questions. These limitations suggest that association-based methods are not adequate for the task [21], contributing to an unnecessarily widespread prevalence of false positives in current epidemiology that undermines the credibility and value of scientific studies that should be providing trustworthy, crucial information to policy makers [22, 56, 70, 102]. New and better ideas and methods are needed, and are available, to provide better answers. The remaining sections review the current state of the art of methods for answering the following causal questions and quantifying uncertainties about their answers:

1. *Event detection: What has changed recently in disease patterns or other adverse outcomes, by how much, when, and why?* For example, have hospital or emergency room admissions or age-specific mortalities with similar symptoms recently jumped significantly, perhaps suggesting a disease outbreak (or a terrorist bio-attack)?
2. *Consequence prediction: What are the implications for what will probably happen next if different actions (or no new actions) are taken?* For example, how many new illnesses are likely to occur and when? How quickly can a confident answer be developed and how certain and accurate can answers be based on limited surveillance data?
3. *Risk attribution: What is causing current undesirable outcomes? Does a specific exposure harm human health? If so, who is at greatest risk (e.g., children, elderly, other vulnerable subpopulations) and under what conditions (e.g., for what exposure concentrations and durations or for what co-exposures)?* Answering this question is the subject of *hazard identification* in health risk assessment. For example, do ambient concentrations of fine particulate matter or ozone in air (possibly in combination with other pollutants) cause increased incidence rates of heart disease or lung cancer in one or more vulnerable populations? Here, “cause” is meant in the specific sense that reducing exposures would reduce the risks per person per year of the adverse health effects. (The following section contrasts this with other interpretations of “exposure causes disease *Y*,” such as “exposure *X* is strongly, consistently, specifically, temporally, and statistically

significantly associated with  $Y$ , and the association is biologically plausible and is stronger for greater exposures” or “the fraction of cases of  $Y$  attributable to  $X$ , based on relative risks or regression models, is significantly greater than zero.” These interpretations do *not* imply that reducing  $X$  will reduce  $Y$ , as positive associations and large attributable risks may reflect modeling choices or p-hacking, biases, or confounding rather than genuine causation.)

4. *Response modeling: What combinations of factors affect health outcomes and how strongly? How would risks change if one or more of these factors were changed?* For example, what is the quantitative causal relationship between exposure levels and probabilities or rates of adverse health outcomes for individuals and identifiable subpopulations? How well can these relationships be inferred from data, and how can uncertainties about the answers be characterized?
5. *Decision making: What actions or interventions will most effectively reduce uncertain health risks?* How well can the effects of possible future actions be predicted, such as reducing specific exposures, taking specific precautionary measures (e.g., flu shots for the elderly), or other interventions? This is the key information needed to inform risk management decisions before they are made.
6. *Retrospective evaluation and accountability: How much difference have exposure reductions actually made in reducing adverse health outcomes?* For example, has reducing particulate matter air pollution reduced cardiovascular mortality rates over the past decade, or would these reductions have occurred just as quickly without reductions in air pollution (i.e., are these coincident historical trends, or did one cause the other?)

These questions are fundamental in epidemiology and health and safety risk assessment. They are mainly about how changes in exposures affect changes in health outcomes and about how certain the answers are. They can be answered using current methods of causal analysis and uncertainty quantification (UQ) for causal models if sufficient data are available.

The following sections discuss methods for drawing valid causal inferences from epidemiological data and for quantifying uncertainties about causal impacts, taking into account model uncertainty as well as sampling errors and measurement, classification, or estimation errors in predictors. UQ methods based on model ensemble methods, such as Bayesian model averaging (BMA) and various forms of resampling, boosting, model cross validation, and simulation, can help to overcome over-fitting and other modeling biases, leading to wider confidence intervals for the estimated impacts of actions and reducing false-positive rates [50]. UQ has the potential to restore greater integrity and credibility to model-based risk estimates and causal predictions, to reveal the quantitative impacts of model and other uncertainties on risk estimates and recommended risk management actions, and to guide more productive-applied research to decrease key remaining uncertainties and to improve risk management decision-making via active exploration and discovery of valid causal conclusions and uncertainty characterizations.

## 2 Some Limitations of Traditional Epidemiological Measures for Causal Inference: Uncertainty About Whether Associations Are Causal

Epidemiology has a set of well-developed traditional methods and measures for quantifying associations between observed quantities. These include regression model coefficients and relative risk (RR) ratios (e.g., the ratio of disease rates for exposed and unexposed populations) as well as various quantities derived from them by algebraic rearrangements. Derived quantities include population attributable risks (PARs) and population attributable fractions (PAFs) for the fraction of disease or mortality cases attributable to a specific cause, global burden of disease estimates, etiologic fractions and probability-of-causation calculations, and estimated concentration-response slope factors for exposure-response relations [27, 98]. Although the details of calculations for these measures vary, the key idea for all of them is to observe whether more-exposed people suffer adverse consequences at a higher rate than less-exposed people and, if so, to attribute the excess risks in the more-exposed group to a causal impact of exposure. Conventional statistical methods for quantifying uncertainty about measures of association, such as confidence intervals and  $p$ -values for RR, PAF, and regression coefficients in logistic regression, Cox proportional hazards, or other parametric or semi-parametric regression models, are typically used to show how firmly the data, together with the assumptions embedded in these statistical models, can be used to reject the null hypothesis of independence (no association) between exposures and adverse health responses. In addition, model diagnostics (such as plots of residuals and formal tests of model assumptions) can reveal whether modeling assumptions appear to be satisfied; more commonly, less informative goodness-of-fit measures are reported to show that the models used do not give conspicuously poor descriptions of the data, at least as far as the goodness-of-fit test can determine. However, goodness-of-fit tests are typically very weak in detecting conspicuously poor fits to data. This is often illustrated by the notorious “Anscombe’s quartet” of qualitatively very different scatter plots giving identical least-squares regression lines and goodness-of-fit test values.

The main limitation of these techniques is that they only address associations, rather than causation. Hence, they typically do not actually quantify the fraction or number of illnesses or mortalities per year that would be prevented by reducing or eliminating specific exposures. Unfortunately, as many methodologists have warned, PAF and probability of causation, as well as regression coefficients, are widely misinterpreted as doing precisely this (e.g., [98]). Large epidemiological initiatives, such as the World Health Organization’s Global Burden of Disease studies, make heavy use of association-based methods that are mistakenly interpreted as if they indicated causal relations. This has become a very common mistake in contemporary epidemiological practice. It undermines the validity, credibility, and practical value of many (some have argued most) causal claims now being published using traditional epidemiological methods [70, 88, 102]. To what extent associations correspond to stable causal laws that can reliably predict future consequences of

policy actions is beyond the power of these traditional epidemiological measures to say [98] doing so requires different techniques.

## 2.1 Example: Exposure-Response Relations Depend on Modeling Choices

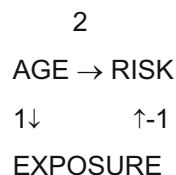
A 2014 article in *Science* [21] noted that “There is a growing consensus in economics, political science, statistics, and other fields that the associational or regression approach to inferring causal relations – on the basis of adjustment with observable confounders – is unreliable in many settings.” To illustrate this point, the authors cite estimates of the effects of total suspended particulates (TSPs) on mortality rates of adults over 50 years old, in which significantly positive associations (regression coefficients) are reported in some regression models that did not adjust for confounders such as age and sex, but significantly negative associations are reported in other regression models that did adjust for confounders by including them as explanatory variables. The authors note that the sign, as well as the magnitude, of reported exposure concentration-response (C-R) relations depends on details of modeling choices about which variables to include as explanatory variables in the regression models. Thus, the quantitative results of risk assessments presented to policy makers as showing the expected reductions in mortality risk per unit decrease in pollution concentrations actually reflect specific modeling choices, rather than reliable causal relations that accurately predict how (or whether) reductions in exposure concentrations would reduce risks.

A distinction from econometrics between structural equations and reduced-form equations [65] is helpful in understanding why different epidemiologists can estimate exposure concentration-response regression coefficients with opposite signs from the same data. The following highly simplified hypothetical example illustrates the key idea. Suppose that cumulative exposure to a chemical increases in direct proportion to age and that the risk of disease (e.g., the average number of illness episodes of a certain type per person per decade) also increases with age. Finally, suppose that the effect of exposure at any age is to decrease risk. These hypothesized causal relations are shown via the following two *structural equations*:

$$\begin{array}{ll} \text{EXPOSURE} = \text{AGE} & \text{SEM equations} \\ \text{RISK} = 2 * \text{AGE} - \text{EXPOSURE}. & \end{array}$$

These are equations with the explicit causal interpretation that a change in the variable on the right side causes a corresponding change in the variable on the left side to restore equality between the two sides (e.g., increasing age increases cumulative exposure and disease risk, but increasing exposure decreases risk at any age). These two structural equations together constitute a *structural equation model* (SEM) that can be diagrammed as in Fig. 43.1:

**Fig. 43.1** SEM causal graph model



In this diagram, each variable depends causally only on the variables that point into it, as revealed by the SEM equations. The weights on the arrows (the coefficients in the SEM equations) show how the average value of the variable at the arrow's head will change if the variable at its tail is increased by one unit, for example, increasing AGE by 1 decade (if that is the relevant unit) increases RISK directly by 2 units (e.g., 2 expected illnesses per decade, if that is the relevant unit), increases EXPOSURE by one unit, and thereby decreases RISK indirectly by 1 unit, via the path through EXPOSURE, for a net effect of a 1 unit increase in RISK per unit increase in AGE.

By contrast to such causal SEM models, what is called a *reduced-form model* is obtained by regressing RISK against EXPOSURE. Using the first SEM equation,  $EXPOSURE = 1 \cdot AGE$ , to substitute EXPOSURE for AGE in the second SEM equation,  $RISK = 2 \cdot AGE - EXPOSURE$ , yields the following reduced-form equation:

$$RISK = EXPOSURE$$

Reduced-form equation

This reduced-form model is a valid *descriptive statistical* model: it reveals that in communities with higher exposure levels, risk should be expected to be greater. But it is not a valid *causal* model: a prediction that reducing exposure would cause a reduction in risk would be mistaken, as the SEM equations make clear. The reduced-form equation is not a structural equation, so it cannot be used to predict correctly how changing the right side would cause the left side to change. The coefficient of EXPOSURE in the linear regression model relating exposure to risk is  $+1$  in the reduced-form model, but is  $-1$  in the SEM model, showing how different investigators might reach opposite conclusions about the sign of “the” exposure-response coefficient based on whether or not they condition on age (or, equivalently, on whether they use structural or reduced-form regression equations).

In current epidemiological practice, the distinction between structural and reduced-form equations is often not clearly drawn. Regression coefficients of various signs and magnitudes, as well as various measures of association based on relative risk ratios, are all presented to policy makers as if they had valid causal interpretations and therefore important implications for risk management policy-making. In air pollution health effects epidemiology, for example, it is standard practice to present regression coefficients as expected reductions in elderly mortality rates (or as expected increases in life span) per unit reduction in air pollution concentrations [24, 28], thereby conflating associations between historical levels



(e.g., pollutant levels and mortality rates both tend to be higher on cold winter days than during the rest of the year, and both have declined in recent decades) with a causal, predictive relation that implies that future reductions in pollution would cause further future reductions in elderly mortality rates. Since such association-based studies are often unreliable indicators of causality [21] or simply irrelevant for determining causality, as in the examples for Fig. 43.1, policy makers who wish to use reliable causal relations to inform policy decisions must seek elsewhere.

These limitations of association-based methods have been well discussed among methodological specialists for decades [98]. Key lessons, such as that the same data set can yield either a statistically significant positive exposure-response regression coefficient or a statistically significant negative exposure-response regression coefficient, depending on the modeling choices made by the investigators, are becoming increasingly appreciated by practitioners [21]. They illustrate an important type of uncertainty that arises in epidemiology, but that is less familiar in many other applied statistical settings: *uncertainty about the interpretation of regression coefficients* (or other association-based measures such as RR, PAF, etc.) as indicating causal relations *vs.* confounded associations or modeling biases *vs.* some of each. This type of uncertainty cannot be addressed by presenting conventional statistical uncertainty measures such as confidence intervals, p-values, regression diagnostics, sensitivity analyses, or goodness-of-fit statistics, since the uncertainty is not about how well a model fits data or about the estimated parameters of the model. Rather, it is about the extent to which the model is only descriptive of the past *vs.* predictive of different futures caused by different choices. Although this is not an uncertainty to which conventional statistical tests apply, it is crucial for the practical purpose of making model-informed risk management decisions. Policy interventions will successfully increase the probabilities of desired outcomes and decrease the frequencies of undesired ones only to the extent that they act causally on drivers of the outcomes and not necessarily to the extent that the models used describe past associations.

One way to try to bridge the gap between association and causation is to ask selected experts what they think about whether or to what extent associations might be causal. However, research on the performance of expert judgments has called into question the reliability of expert judgments, specifically including judgments about causation [61]. Such judgments typically reflect qualitative “weight of evidence” (WoE) considerations about the strength, consistency (e.g., do multiple independent researchers find the claimed associations?), specificity, coherence (e.g., are associations of exposure with multiple health endpoints mutually consistent with each other and with the hypothesis of causality?) temporality (do hypothesized causes precede their hypothesized effects?), gradient (are larger exposures associated with larger risks?), and biological plausibility of statistical associations and the quality of the data sources and studies supporting them. One difficulty is that a strong confounder (such as age in Fig. 43.1) with delayed effects can create strong, consistent, specific, coherent, temporal associations between exposure and risk of an adverse response, with a clear gradient associating larger risks with larger exposures, without providing any evidence that exposure actually causes increased risk.

Showing that an association is strong, for example, does not address whether it is causal, although many WoE systems simply assume that the former supports the latter without explicitly addressing whether the strong associations are instead explained by strong confounding, strong biases, or strong modeling assumptions. Similarly, showing that different investigators find the same or similar association does not necessarily show whether this consistency results from shared modeling assumptions, biases, or confounders. Conflating causal and associational concepts, such as evidence for the strength of an association and evidence for causality of the association, too often makes assessments of causality in epidemiology untrustworthy compared to methods used in other fields, discussed subsequently [51, 83]. Most epidemiologists are trained to treat various aspects of association as evidence for causation, even though they are not, and this undermines the trustworthiness of expert judgments about causation based on WOE considerations [83].

In addition, experts are sometimes asked to judge the *probability* that an association is causal (e.g., [25]). This makes little sense. It neglects the fact that an association may be partly causal and partly due to confounding or modeling biases or coincident historical trends. For example, if exposure does increase risk, but is also confounded by age, then asking for the probability that the regression coefficient relating exposure to risk is causal overlooks the realistic possibility that it reflects both a causal component and a confounding component, so that the probability that it is partly causal might be 1 and the probability that it is completely causal might be 0. A more useful question to pose to experts might be what *fraction* of the association is causal, but this is seldom asked. Common noncausal sources of statistical associations include model selection and multiple testing biases, model specification errors, unmodeled errors in explanatory variables in multivariate models, biases due to data selection and coding (e.g., dichotomizing or categorizing continuous variables such as age, which can lead to residual confounding), and coincident historical trends, which can induce statistically significant-appearing associations between statistically independent random walks – a phenomenon sometimes dubbed as spurious regression [17, 98].

Finally, qualitative subjective judgments and ratings used in many WoE systems are subject to well-documented psychological biases. These include confirmation bias (seeing what one expects to see and discounting or ignoring evidence that might challenge one's preconceptions), motivated reasoning (finding what it benefits one to find and believing what it pays one to believe), and overconfidence (not sufficiently doubting, questioning, or testing one's own beliefs and hence not seeking potentially disconfirming information that might require those beliefs to be revised) [61, 102].

That statistical associations do not in general convey information sufficient for making valid causal predictions has been well understood for decades by statisticians and epidemiologists specializing in technical methods for causal analysis (e.g., [31, 36]). This understanding is gradually percolating through the larger epidemiological and risk analysis communities. Peer-reviewed published papers and reports, including those relied on in many regulatory risk assessments, still too often make the fundamental mistake of reinterpreting empirical exposure-response (ER)

relations between historical levels of exposure and response as if they were causal relations useful for predicting how future changes in exposures would change future responses. Fortunately, this confusion is unnecessary today: appropriate technical methods for causal analysis and modeling are now well developed, widely available in free software such as R or Python, and readily applicable to the same kinds of cross-sectional and longitudinal data collected for association-based studies. Table 43.1 summarizes some of the most useful study designs and methods for valid causal analysis and modeling of causal exposure-response relations.

Despite the foregoing limitations, there is much of potential value in several WoE considerations, especially consistency, specificity, and temporality of associations, especially if they are used as part of a relatively objective, quantitative, data-

**Table 43.1** Some formal methods for modeling and testing causal hypotheses

Method and References	Basic Idea	Appropriate study design
Conditional independence tests [31, 32]	Is hypothesized effect (e.g., lung cancer) statistically independent of hypothesized cause (e.g., exposure to chemical), given values of other variables (e.g., education and income)? If so, this undermines causal interpretation.	Cross-sectional data Can also be applied to multi-period data (e.g., in dynamic Bayesian networks)
Panel data analysis [2, 109]	Are changes in exposures followed by changes in the effects that they are hypothesized to help cause? If not, this undermines causal interpretation; if so, this strengthens causal interpretation.  Example: Are changes in exposure levels followed (but not preceded) by corresponding changes in mortality rates?	Panel data study: Collect a sequence of observations on same subjects or units over time
Granger causality test [23], transfer entropy [81, 91, 99, 118]	Does the history of the hypothesized cause improve ability to predict the future of the hypothesized effect? If so, this strengthens causal interpretation; otherwise, it undermines causal interpretation.  Example: Can lung cancer mortality rates in different occupational groups be predicted better from time series histories of exposure levels and mortality rates than from the time series history of mortality rates alone?	Time series data on hypothesized causes and effects
Quasi-experimental design and analysis [12, 40, 41]	Can control groups and other comparisons refute alternative (noncausal) 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.	Longitudinal observational data on subjects exposed and not exposed to interventions that change the hypothesized cause(s) of effects.

(continued)

**Table 43.1** (continued)

Method and References	Basic Idea	Appropriate study design
Intervention analysis, change-point analysis [45], Gilmour et al. 2006	Does the best-fitting model of the observed data change significantly at or following the time of an intervention? If so, this strengthens causal interpretation. Do the quantitative changes in hypothesized causes predict and explain the subsequently observed quantitative changes in hypothesized effects? If so, this strengthens causal interpretation. Example: Did lung disease mortality rates fall significantly faster or sooner in workplaces that reduced exposures more or earlier than in workplaces that did not?	Time series observations on hypothesized effects and knowledge of timing of intervention(s)  Quantitative time series data for hypothesized causes and effects
Counterfactual and potential outcome models, including propensity scores and marginal structural models (MSMs) [82, 96]	Do exposed individuals have significantly different response probabilities than they would have had if they had not been exposed? Example: Do workers have lower mortality risk after historical exposure reductions than they would have had otherwise?	Cross-sectional and/or longitudinal data, with selection biases and feedback among variables allowed
Causal network models of change propagation [19, 39]	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?	Observations of variables in a dynamic system out of equilibrium
Negative controls (for exposures or for effects) [73]	Do exposures predict health effects better than they predict effects that cannot be caused by exposures more (e.g., reductions in traumatic injuries)?	Observational studies

driven approach to inferring probable causation. The following sections discuss this possibility and show how such traditional qualitative WoE considerations can be fit into more formal quantitative causal analyses.

### 3 Event Detection and Consequence Prediction: What's New, and So What?

In public health and epidemiology, surveillance data showing changes in hospital or emergency department admission rates for a specific disease or symptom category may provide the first indication that an event has occurred that has caused changes in health outcomes. Initially, the causes of the changes may be uncertain, but if the date of a change can be estimated fairly precisely and matches the date of an event that might have caused the observed effects, then the event might have caused the change

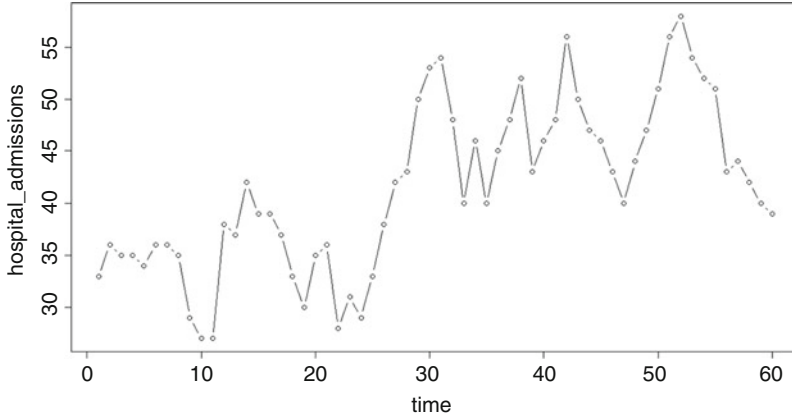
in admissions rates. This causal hypothesis is strengthened if occurrences of the same or similar event in multiple times and places are followed by similar changes in admission rates (consistency and temporality of association) and if these changes in admissions rates do not occur except when the event occurs first (specificity of association). To make this inference sound, the event occurrences must not be triggered by high levels of admissions rates, since otherwise interventions that respond to these high rates might be followed by significant reductions in admission rates due solely to *regression to the mean*, i.e., the fact that exceptionally high levels are likely to be followed by lower levels, even if the interventions have no impact [12].

The technical methods used to estimate when admission rates or other effect have changed significantly, such as counts of accidents or injuries or fatalities per person per week in a population, include several different statistical anomaly-detection and change-point analysis (CPA) algorithms (e.g., [108]). The key idea of these algorithms is to determine whether, for each point in time (e.g., for each week in a surveillance time series), the series is significantly different (e.g., in distribution or trend) before that time point than after it. If so – if a time series jumps at a certain time – that time is called a change point.

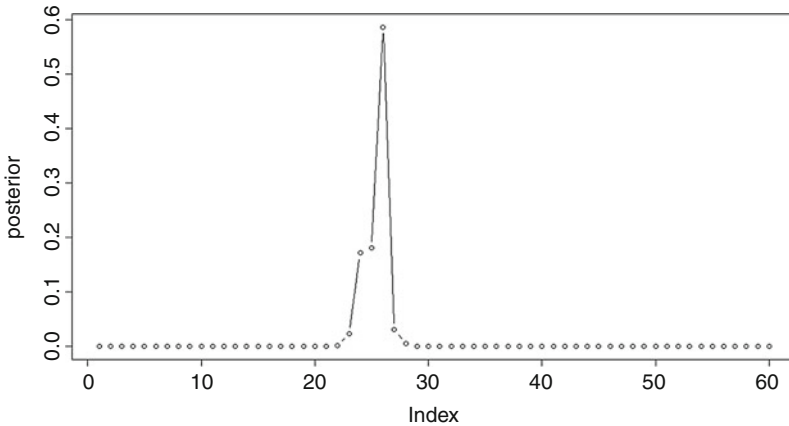
### 3.1 Example: Finding Change Points in Surveillance Data

As an example of change-point detection in surveillance data, consider the following example. Since 2001, when a letter containing anthrax led to 5 deaths and 17 infections from which the victims recovered, the US Environmental Protection Agency (EPA), the Centers for Disease Control and Prevention (CDC), and the Department of Health Services have invested over a billion dollars to develop surveillance methods and prevention and preparedness measures to help reduce or mitigate the consequences of bioterrorism attacks should they occur again [38]. Detecting a significant upsurge in hospital admissions with similar symptoms may indicate that a bioterrorism attack is in progress. The statistical challenge of detecting such changes against the background of normal variability in hospital admissions has motivated the development of computational intelligence methods that seek to reduce the time to detect attacks when they occur, while keeping the rates of false positives acceptably small [11, 106].

Well-developed, sophisticated techniques of statistical uncertainty quantification are currently available for settings in which the patterns for which one is searching are well understood (e.g., a jump in hospitalization rates for patients with similar symptoms that could be caused by a biological agent) and in which enough surveillance data are available to quantify background rates and to monitor changes over time. Figure 43.2 presents a hypothetical example showing weekly counts of hospital admissions with specified symptoms in a certain city. Given such surveillance data, the risk assessment inference task is to determine whether the hospitalization rate increased at some point on time (suggestive of an attack) and, if so, when and by how much. Intuitively, it appears that counts are greater on the



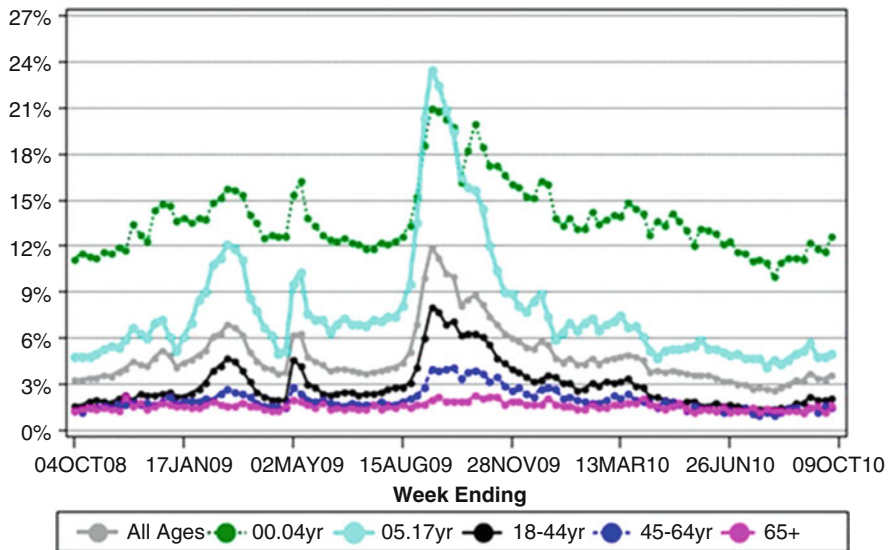
**Fig. 43.2** Surveillance time series showing a possible increase in hospitalization rates



**Fig. 43.3** Bayesian posterior distribution for the timing of the increase in Fig. 43.2 if one has occurred

right side of Fig. 43.2 than the left, but might this plausibly just be due to chance, or is it evidence for a real increase in hospitalization rates?

Figure 43.3 illustrates a typical result of current statistical technology (also used in computational intelligence, computational Bayesian, machine learning, pattern recognition, and data mining technologies) for solving such problems by using statistical evidence, together with risk models, to draw inferences about what is probably happening in the real world. The main idea is simple: the highest points indicate the times that are computed to be most likely for when a change in hospitalization rate occurred, based on the data in Fig. 43.2. (Technically, Fig. 43.3 plots the likelihood function of the data, assuming that at most one jump from one level to a different level has occurred for the hospitalization rate. The



**Fig. 43.4** Proportion of emergency department visits for influenza-like illness by age group for October 4, 2008–October 9, 2010, in a US Department of Health and Human Services region (Source: [62], <http://jamia.oxfordjournals.org/content/19/6/1075.long>)

likelihoods are rescaled so that their sum for all 60 weeks is 1, so that they can be interpreted as posterior probabilities if the prior is assumed to be uniform. More sophisticated algorithms are discussed next.) The maximum likelihood-based algorithm accurately identifies both the time of the change (week 25) and the magnitude of its effect to one significant decimal place (not shown in Fig. 43.3). The spread of the likelihood function (or posterior probability distribution) around the most likely value in Fig. 43.3 also shows how precise is the estimation of the change-point time.

Figure 43.4 shows a real-world example of a change point and its consequences for emergency department visits over time. Admissions for flu-like symptoms, especially among infants and children (0–4 and 5–17 year olds), increased sharply in August and declined gradually in each age group thereafter. Being able to identify the jump quickly and then applying a predictive model – such as a stochastic compartmental transition model with susceptible, infected, and recovered subpopulations (SIR model) for each age group – to predict the time course of the disease in the population can help forecast the care resources that will be needed over time for each age group.

More generally, detecting change points can be accomplished by testing the null hypothesis of no change for each time step and correcting for multiple testing bias (which would otherwise inflate false-positive rates, since testing the null hypothesis for each of the many different possible times at which a change might have occurred multiplies the occasions on which an apparently significant change occurs

by chance). Many CPA algorithms use likelihood-based Bayesian methods, as in Fig. 43.3, to identify when a change is most likely to have occurred and whether the hypothesis that it did provides a significantly better explanation (higher likelihood) for the observed data than the null hypothesis of no change. Likelihood-based techniques are fundamental for a wide variety of statistical detection and estimation algorithms. Practitioners can use free, high-quality algorithms available in the R statistical computing environment (e.g., <http://surveillance.r-forge.r-project.org/>; [57]), Python, and other statistics programs and packages to perform CPA analyses.

Algorithms for change-point detection have recently been extended to allow detection of multiple change points within multivariate time series, i.e., in time series containing observations of multiple variables instead of only one [57]. These new algorithms use nonparametric tests (e.g., permutation tests) to determine whether the distributions of the observations before and after the change point differ significantly, even if neither distribution is known, and hence no parametric statistical model can be specified [57]. The development of powerful nonparametric (“model-free”) methods for testing the null hypothesis of no change in the (unknown) distribution enables CPA that is much more robust to uncertainties in modeling assumptions than was possible previously. Assumptions that remain, such as that observations following a change point are drawn from a new distribution, independently of the observations preceding the change point, are statistically testable and weaker than the assumptions (such as approximately normally distributed observations) made in older CPA work.

The use of CPA to search for significant changes in surveillance time series, showing that the number of undesirable events per person per week in a population underwent significant changes at certain times, has allowed the probable causes of observed changes in health and safety to be identified in many applications, providing evidence for or against important causal relations between public policy measures and resulting health and safety effects. For example,

- Nakahara et al. [85] used CPA to assess the impact on vehicle crash fatalities of a program initiated in Japan in 2002 that severely penalized drunk driving. Fatality rates between 2002 and 2006 (the end of the available data series) were significantly lower than between 1995 and 2002. However, the CPA revealed that the change point occurred around the end of 1999, right after a high-profile vehicle fatality that was much discussed in the news. The authors concluded that changes in drunk-driving behavior occurred well before the new penalties were instituted.
- In Finland in 1981–1986, a nationwide oral poliovirus vaccine campaign was closely followed by, and partly overlapped with, a significant increase in the incidence of Guillain-Barré syndrome (GBS). This temporal association raised the important question of whether something about the vaccine might have caused some or all of the increase in GBS. Kinnunen et al. [63] applied CPA to medical records from a nationwide Hospital Discharge Register database. They found that a change point in the occurrence of GBS had probably already taken place before the oral poliovirus vaccine campaign started. They concluded that



there was a temporal association between poliovirus infection and increased occurrence of GBS, but no proof of the suspected causal relation between oral poliovirus vaccines and risk of GBS. This example shows how a precise investigation of the details of temporal associations can both refute some causal hypotheses and suggest others – in this case, that an increase in polio in the population was a common cause of both increased GBS risk and the provision of the vaccine. It also illustrates why a temporal association between an adverse effect and a suspected cause, such as the fact that administration of vaccines preceded increases in GBS risk, should not necessarily be interpreted as providing evidence to support the hypothesis of a causal relation between them.

---

## 4 Causal Analytics: Determining Whether a Specific Exposure Harms Human Health

Table 43.2 lists seven principles that have proved useful in various fields for determining whether available data provide valid evidence that some events or conditions cause others. They can be applied to epidemiological data to help determine whether and how much exposures to a hazard contribute causally to subsequent risks of adverse health outcomes in a population, in the sense that reducing exposure would reduce risk – for example, whether and by how much a given reduction in air pollution would reduce cardiovascular mortality rates among the elderly, whether and by how much reducing exposure to television violence in childhood would reduce propensity for violent behavior years later, or whether decreasing high-fat or high-sugar diets in youth would reduce risks

**Table 43.2** Principles of causal analytics

- 
1. **Conditional independence principle:** Causes and effects are informative about each other. Technically, there should be positive mutual information (measured in bits and quantified by statistical methods of information theory) between the random variables representing a cause and its effect. This positive mutual information cannot be removed by conditioning on the levels of other variables.
  2. **Granger principle:** Changes in causes should precede, and help to predict and explain, changes in the effects that they cause.
  3. **Transfer entropy principle:** Information flows from causes to their effects over time.
  4. **Counterfactual principle:** Changes in causes make future effects different from what they otherwise would have been.
  5. **Causal graph principle:** Changes in causes produce changes in effects by propagation via one or more paths (sequences of causal mechanisms) connecting them.
  6. **Mechanism principle:** Valid causal mechanisms are lawlike, yielding identically distributed outputs when the inputs are the same.
  7. **Quasi-experiment principle:** Valid causal relations produce differences (e.g., compared to relevant comparison groups) that cannot be explained away by noncausal explanations.
-

of heart attacks in old age. The following sections explain and illustrate these principles and introduce technical methods for applying them to data. They also address the fundamental questions of how to model causal responses to exposure and other factors, how to decide what to do to reduce risk, how to determine how well interventions have succeeded in reducing risks, and how to characterize uncertainties about the answers to these questions.

---

## 5 Causes and Effects Are Informative About Each Other: DAG Models, Conditional Independence Tests, and Classification and Regression Tree Algorithms

A key principle for causal analytics is that causes and their effects provide information about each other. If exposure is a cause of increased disease risk, then measures of exposure and of response (i.e., disease risk) should provide *mutual information* about each other, in the sense that the conditional probability distribution for each varies with the value of the other. Software for determining whether this is the case for two variables in a data set is discussed at the end of this section. In addition, if exposures are direct causes of responses, then the mutual information between them cannot be eliminated by conditioning on the values of other variables, such as confounders: a cause provides unique information about its effects. This provides the basis for using statistical *conditional independence tests* to test the observable statistical implications of causal hypotheses: *An effect should never be conditionally independent of its direct causes, given (i.e., conditioned on) the values of other variables.*

As a simple example, if both air pollution and elderly mortality rates are elevated on cold winter days, then if air pollution is a cause of increased elderly mortality rate, the mutual information between air pollution and elderly mortality rates should not be eliminated (“explained away”) by temperature, even though temperature may be associated with each of them. If both temperature and air pollution contribute to increased mortality rates (indicated in causal graph notation as  $temperature \rightarrow mortality\_rate \leftarrow pollution$ ), then conditioning on the level of temperature will not eliminate the mutual information between pollution and mortality rate. On the other hand, if the correct causal model were that temperature is a confounder that explains both mortality rate and pollution (e.g., because coal-fired power plants produce more pollution during days with extremely hot and cold weather, and, independently, these temperature extremes lead to greater elderly mortality), diagrammed as  $mortality\_rate \leftarrow temperature \rightarrow pollution$ , then conditioning on the level of temperature would eliminate the mutual information between pollution and mortality rate. Thus, tests that reveal conditional independence relations among variables can also help to discriminate among alternative causal hypotheses.

The notation in these graphs is as follows. Each node in the graph (such as *temperature*, *pollution*, or *mortality\_rate* in the preceding example) represents a random variable. Arrows between nodes reveal statistical dependencies (and,

implicitly, conditional independence relations) among the variables. The arrows are usually constrained to form a directed acyclic graph (DAG), meaning that no node can be its own predecessor in the partial ordering of nodes determined by the arrows. The probability distribution of each variable with inward-pointing arrows depends on the values of the variables that point into it, i.e., the conditional probability distribution for the variable at the head of an arrow is affected by the values of its direct “parents” (the variables that point into it) in the causal graph. Conversely, a random variable represented by a node is conditionally independent of all other variables, given the values of the variables that point into it (its parents in the DAG), the values of the variables into which it points (its children), and the values of any other parents of its children (its spouses) – a set of nodes collectively called its Markov blanket in the DAG model.

To illustrate these ideas, suppose that  $X$  causes  $Y$  and  $Y$  causes  $Z$ , as indicated by the DAG and  $X \rightarrow Y \rightarrow Z$ , where  $X$  is an exposure-related variable (e.g., job category for an occupational risk or location of a residence for a public health risk),  $Y$  is a measure of individual exposure, and  $Z$  is an indicator of adverse health response. Then even though each variable is statistically associated with the other two,  $Z$  is conditionally independent of  $X$  given the value of  $Y$ . But  $Z$  cannot be made conditionally independent of  $Y$  by conditioning on  $X$ . One way to test for such conditional independence relations in data is with classification and regression tree algorithms (see, e.g., <https://cran.r-project.org/web/packages/rpart/rpart.pdf> for a free R package and documentation). In this example, a tree for  $Z$  would not contain  $X$  after splitting on values of  $Y$ , reflecting the fact that  $Z$  is conditionally independent of  $X$  given  $Y$ . However, a tree for  $Z$  would always contain  $Y$ , provided that the data set is large and diverse enough so that the tree-growing algorithm can detect the mutual information between them.

For practitioners, algorithms are now freely available in R, Python, and Google software packages to estimate mutual information, the conditional entropy reduction in one variable when another is observed, and related measures for quantifying how many bits of information observations of one variable provide about another and whether one variable is conditionally independent of another gives the values of other variables ([74]; Ince et al. 2009). For example, free R software and documentation for performing these calculations can be found at the following sites:

<https://cran.r-project.org/web/packages/entropy/entropy.pdf>

<https://cran.r-project.org/web/packages/partykit/vignettes/partykit.pdf>.

---

## 6 Changes in Causes Should Precede, and Help to Predict and Explain, Changes in the Effects that They Cause

If changes in exposures always precede and help to predict and explain subsequent corresponding changes in health effects, this is consistent with the hypothesis that exposures cause health effects. The following methods and algorithms support formal testing of this hypothesis.

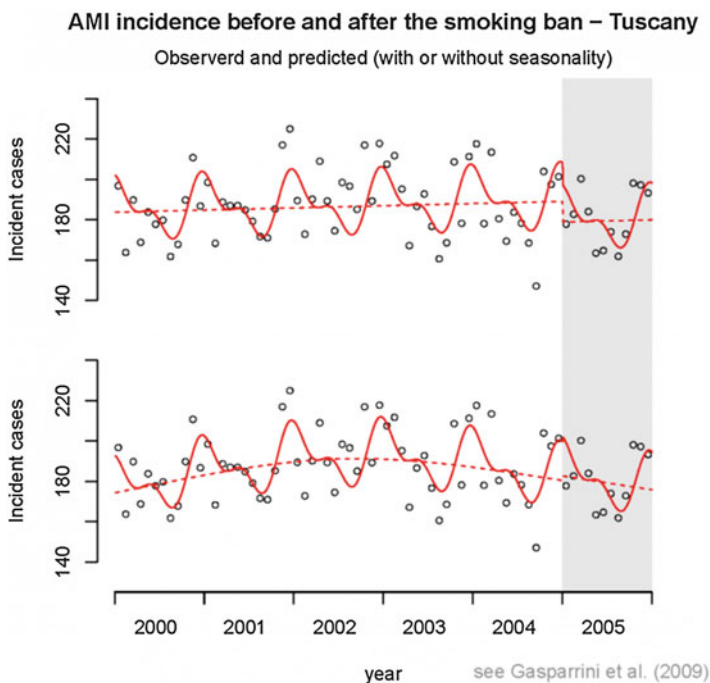
## 6.1 Change-Point Analysis Can Be Used to Determine Temporal Order

The change-point analysis (CPA) algorithms already discussed can be used to estimate when changes in effects time series occurred. These times can then be compared to the times at which exposures changed (e.g., due to passage of a regulation or to introduction or removal of a pollution source) to determine whether changes in exposures are followed by changes in effects. For example, many papers have noted that bans on public smoking have been followed by significant reductions in risks of heart attacks (acute myocardial infarctions). However, Christensen et al. [14], in a study of the effects of a Danish smoking ban on hospital admissions for acute myocardial infarctions, found that a significant reduction in admissions was already occurring a year *before* the bans started. Thus, the conclusion that bans caused the admissions reductions may be oversimplified. The authors suggest that perhaps some of the decline in heart attack risk could have been caused by earlier improvements in diets or by gradual enactment of smoking bans. Whatever the explanation, checking when reductions began, rather than only whether post-intervention risks are smaller than pre-intervention risks, adds valuable insight to inform potential causal interpretations of the data.

## 6.2 Intervention Analysis Estimates Effects of Changes Occurring at Known Times, Enabling Retrospective Evaluation of the Effectiveness of Interventions

How much difference exposure reductions or other actions have made in reducing adverse health outcomes or producing other desired outcomes is often addressed using *intervention analysis*, also called *interrupted time series* analysis. The basic idea is to test whether the best description of an effects time series changes significantly when a risk factor or exposure changes, e.g., due an intervention that increases or reduces it [47, 48, 68]. If the answer is yes, then the size of the changeover time provides quantitative estimates of the sizes and timing of changes in effects following an intervention. For example, an intervention analysis might test whether weekly counts of hospital admissions with a certain set of diagnostic codes or cardiovascular mortalities per person per year among people over 70 fell significantly when exposures fell due to closure of a plant that generated high levels of air pollution. If so, then comparing the best-fitting time series models (e.g., the maximum-likelihood models within a broad class of models, such as the autoregressive integrated moving average (ARIMA) models widely used in time series analysis) describing the data before and after the date of the intervention may help to quantify the size of the effect associated with the intervention. If not, then the interrupted time series does not provide evidence of a detectable effect of the intervention. Free software for intervention analysis is available in R (e.g., [80]; CausalImpact algorithm from Google 2015).

Two main methods of intervention analysis are *segmented regression*, which fits regression lines or curves to the effects time series before and after the intervention and then compares them to detect significant changes in slope or level, and *Box-Tiao analysis*, often called simply intervention analysis, which fits time series models (ARIMA or Box-Jenkins models with models of intervention effects, e.g., jumps in the level, changes in the slope, or ramp-ups or declines in the effects over time) to the effects data before and after the intervention and tests whether proposed effects of interventions are significantly different from zero. If so, the parameters of the intervention effect are estimated from the combined pre- and post-intervention data (e.g., [47, 48]). For effects time series that are stationary (meaning that the same statistical description of the time series holds over time) both before and after an intervention that changes exposure, but that have a jump in mean level due to the intervention, quantifying the difference in effects that the intervention has made can be as simple as estimating the difference in means for the effects time series before and after the intervention, similar to the CPA in Fig. 43.2, but for a known change point. The top panel of Fig. 43.5 shows a similar comparison for heart attack rates before and after a smoking ban. Based only on the lines shown, it appears that



**Fig. 43.5** Straight-line extrapolation of the historical trend for heart attack (AMI) rates over-predicts future AMI rates (upper panel) and creates the illusion that smoking bans were followed by reduced AMI rates, compared to more realistic nonlinear extrapolation (lower panel), which shows no detectable benefit from a smoking ban (Source: [33])

heart attack rates dropped following the ban. (If the effects of a change in exposure occur gradually, then distributed-lag models of the intervention's effects can be used to describe the post-intervention observations [47].) In nonstationary time series, however, the effect of the intervention may be obscured by other changes in the time series. Thus, the bottom panel of Fig. 43.5 considers nonlinear trends over time and shows that, in this analysis, any effect of the ban now appears to be negative (i.e., heart attack rates are increased after the ban compared to what is expected based on the nonlinear trend extrapolated from pre-ban data).

Intervention analyses, together with comparisons to time series for comparison populations not affected by the interventions, have been widely applied, with varying degrees of justification and success, to evaluate the impacts caused by changes in programs and policies in healthcare, social statistics, economics, and epidemiology. For example, Lu et al. [75] found that prior authorization policies introduced in Maine to help control the costs of antipsychotic drug treatments for Medicaid and Medicare Part D patients with bipolar disorder were associated with an unintended but dramatic (nearly one third) reduction in initiation of medication regimens among new bipolar patients, but produced no detectable switching of currently medicated patients toward less expensive treatments. Morriss et al. [84] found that the time series of suicides in a district population did not change significantly following a district-wide training program that measurably improved the skills, attitudes, and confidence of primary care, accident and emergency, and mental health workers who received the training. They concluded that "Brief educational interventions to improve the assessment and management of suicide for front-line health professionals in contact with suicidal patients may not be sufficient to reduce the population suicide rate." [55] used intervention analysis to estimate that the introduction of pedestrian countdown timers in Detroit cut pedestrian crashes by about two thirds. Jiang et al. [59] applied intervention analysis to conclude that, in four Australian states, the introduction of randomized breath testing led to a substantial reduction in car accident fatalities. Callaghan et al. [10] used a variant of intervention analysis, regression-discontinuity analysis, to test whether the best-fitting regression model describing mortality rates among young people changed significantly at the minimum legal drinking age, which was 18 in some provinces and 19 in others. They found that mortality rates for young men jumped upward significantly precisely at the minimum legal drinking, which enabled them to quantify the impact of drinking-age laws on mortality rates. In these and many other applications, intervention analysis and comparison groups have been used to produce empirical evidence for what has worked and what has not and to quantify the sizes over time of effects attributed to interventions when these effects are significantly different from zero.

Intervention analysis has important limitations, however. Even if an intervention analysis shows that an effects time series changed when an intervention occurred, this does not show whether the intervention *caused* the change. Thus, in applications from air pollution bans to gun control, initial reports that policy interventions had significant beneficial effects were later refuted by findings that equal or greater beneficial changes occurred at the same time in comparison populations not affected

by the interventions [44, 64]. Also, more sophisticated methods such as transfer entropy, discussed later, must be used to test and estimate effects in nonstationary time series, since both segmented regression models and intervention analyses that assume stationarity typically produce spurious results for nonstationary time series. For example, as illustrated in Fig. 43.5, Gasparrini et al. [33] in Europe and Barr et al. [8] in the United States found that straight-line projections of what future heart attack (acute myocardial infarction, AMI) rates would have been in the absence of an intervention that banned smoking in public places led to a conclusion that smoking bans were associated with a significant reduction in AMI hospital admission rates following the bans. However, allowing for nonlinearity in the trend, which was significantly more consistent with the data, led to the reverse conclusion that the bans had no detectable impact on reducing AMI admission rates. As illustrated in Fig. 43.5, the reason is that fitting a straight line to historical data and using it to project future AMI rates in the absence of intervention tend to overestimate what those future AMI rates would have been, because the real time series is downward-curving, not straight. Thus, estimates of the effect of an intervention based on comparing observed to model-predicted AMI admission rates will falsely attribute a positive effect even to an intervention that had no effect if straight-line extrapolation is used to project what would have happened in the absence of an intervention, ignoring the downward curvature in the time series. This example illustrates how model specification errors can lead to false inferences about effects of interventions. The transfer entropy techniques discussed later avoid the need for curve-fitting and thereby the risks of such model specification errors.

### 6.3 Granger Causality Tests Show Whether Changes in Hypothesized Causes Help to Predict Subsequent Changes in Hypothesized Effects

Often, the hypothesized cause (e.g., exposure) and effect (e.g., disease rate) time series both undergo continual changes over time, instead of changing only once or occasionally. For example, pollution levels and hospital admission rates for respiratory or cardiovascular ailments change daily. In such settings of ongoing changes in both hypothesized cause and effect time series, *Granger causality tests* (and the closely related Granger-Sims tests for pairs of time series) address the question of whether the former helps to predict the latter. If not, then the exposure-response histories provide no evidence that exposure is a (Granger) cause of the effects time series, no matter how strong, consistent, etc., the association between their levels over time may be. More generally, a time series variable  $X$  is not a Granger cause of a time series variable  $Y$  if the future of  $Y$  is conditionally independent of the history of  $X$  (its past and present values), given the history of  $Y$  itself, so that future  $Y$  values can be predicted as well from the history of  $Y$  values as from the histories of both  $X$  and  $Y$ . If exposure is a Granger cause of health effects but health effects are not Granger causes of exposures, then this provides evidence that the exposure time series might indeed be a cause of the effects time

series. If exposure and effects are Granger causes of each other, then a confounder that causes both of them is likely to be present. The key idea of Granger causality testing is to provide formal quantitative statistical tests of whether the available data suffice to reject (at a stated level of significance) the null hypothesis that the future of the hypothesized effect time series can be predicted no better from the history of the hypothesized cause time series together with the history of the effect time series than it can be predicted from the history of the effect time series alone. Data that do not enable this null hypothesis to be rejected do not support the alternative hypothesis that the hypothesized cause helps to predict (i.e., is a Granger cause of) the hypothesized effect.

Granger causality tests can be applied to time series on different time scales to study effects of time-varying risk factors. For example, [76] identified a Granger-causal association between fatty diet and risk of heart disease decades later in aggregate (national level) data. Cox and Popken [16] found a statistically significant historical association, but no evidence of Granger causation, between ozone exposures and elderly mortality rates on a time scale of years. Granger causality testing software is freely available in R (e.g., <http://cran.r-project.org/web/packages/MSBVAR/MSBVAR.pdf>).

Originally Granger causality tests were restricted to stationary linear (autoregressive) time series models and to only two time series, a hypothesized cause and a hypothesized effect. However, recent advances have generalized them to multiple time series (e.g., using vector autoregressive (VAR) time series models) and to nonlinear time series models (e.g., using nonparametric versions of the test or parametric models that allow for multiplicative as well as additive interactions among the different time series variables) ([6, 7, 111, 122]; Diks and Wolski 2014). These advances are now being made available in statistical toolboxes for practitioners ([7]). For nonstationary time series, special techniques have been developed, such as vector error-correction (VECM) models fit to first differences of nonstationary variables or algorithms that search for co-integrated series (i.e., series whose weighted averages show zero mean drift). However, these techniques are typically quite sensitive to model specification errors [91]. Transfer entropy (TE) and its generalizations, discussed next, provides a more robust analytic framework for identifying causality from multiple nonstationary time series based on the flow of information among them.

---

## 7 Information Flows From Causes to Their Effects over Time: Transfer Entropy

Both Granger causality tests and conditional independence tests apply the principle that causes should be informative about their effects; more specifically, changes in direct causes provide information that helps to predict subsequent changes in effects. This information is not redundant with the information from other variables and cannot be explained away by knowledge of (i.e., by conditioning on the values of) other variables. A substantial generalization and improvement of this information-



based insight is that information flows over time from causes to their effects, but not in the reverse direction. Thus, instead of just testing whether past and present exposures provide information about (and hence help to predict) future health effects, it is possible to quantify the rate at which information, measured in bits, flows from the past and present values of the exposure time series to the future values of the effects time series. This is the key concept of *transfer entropy* (TE) [81, 91, 99, 118]). It provides a nonparametric, or model-free, way to detect and quantify rates of information flow among multiple variables and hence to infer causal relations among them based on the flow of information from changes in causal variables (“drivers”) to subsequent changes in the effect variables that they cause (“responses”). If there is no such information flow, then there is no evidence of causality.

Transfer entropy (TE) is model-free in that it examines the empirically estimated conditional probabilities of values for one time series, given previous values of others, without requiring any parametric models describing the various time series. Like Granger causality, TE was originally developed for only two time series, a possible cause and a possible effect, but has subsequently been generalized to multiple time series with information flowing among them over time (e.g., [81, 91, 99]). In the special case where the exposure and response time series can be described by linear autoregressive (AR) processes with multivariate normal error terms, tests for TE flowing from exposure to response are equivalent to Granger causality tests (Barnett et al. 2009), and Granger tests, in turn, are equivalent to conditional independence tests for whether the future of the response series is conditionally independent of the history of the exposure series, given the history of the response series. Free software packages for computing the TE between or among multiple time series variables are now available for MATLAB [81] and other software (<http://code.google.com/p/transfer-entropy-toolbox/downloads/list>). Although transfer entropy and closely related information-theoretic quantities been developed and applied primarily within physics and neuroscience to quantify flows of information and appropriately defined causal influences [58] among time series variables, they are likely to become more widely applied in epidemiology as their many advantages become more widely recognized.

---

## **8 Changes in Causes Make Future Effects Different From What They Otherwise Would Have Been: Potential-Outcome and Counterfactual Analyses**

The insight that changes in causes produce changes their effects, making the probability distributions for effect variables different from what they otherwise would have been, has contributed to a well-developed field of *counterfactual (potential-outcome) causal modeling* [51]. A common analytic technique in this field is to treat the unobserved outcomes that would have occurred had causes (e.g., exposures or treatments) been different as missing data and then to apply missing-data methods for regression models to estimate the average difference

in outcomes for individuals receiving different treatments or other causes. The estimated difference in responses for treated compared to untreated individuals, for example, can be defined as a measure of the impact caused by treatment at the population level.

To accomplish such counterfactual estimation in situations where randomized assignments of treatments or exposures to individuals is not possible, counterfactual models and methods such as *propensity score matching* (PSM) and *marginal structural models* (MSMs) [96] construct weighted samples that attempt to make the estimated distribution of measured confounders the same as it would have been in a randomized control trial. If this attempt is successful, and if the individuals receiving different treatments or exposures are otherwise statistically identical (more precisely, *exchangeable*), then any significant differences between the responses of subpopulations receiving different treatments or exposures (or other combinations of causes) might be attributed to the differences in these causes, rather than to differences in the distributions of measured confounders [18, 96]. However, this attribution is valid only if the individuals receiving different treatments or exposures are exchangeable – a crucial assumption that is typically neither tested nor easily testable. If treated and untreated individuals differ on unmeasured confounders, for example, then counterfactual methods such as PSM or MSM may produce mistaken estimates of causal impacts of treatment or exposure. Differing propensities to seek or avoid treatment or exposure based in part on unmeasured differences in individual health status could create biased estimates of the impacts of treatment or exposure on subsequent health risks. In general, counterfactual methods for estimating causal impacts of exposures or treatments on health risks make assumptions that imply that estimated differences in health risks between different exposure or treatment groups are caused by differences in the exposures or treatments. The validity of these assumptions is usually unproved. In effect, counterfactual methods assume (rather than establishing) the key conclusion that differences in health risks are caused by differences in treatments or exposures, rather than by differences in unmeasured confounders or by other violations of the counterfactual modeling assumptions.

In marginal structural models (MSMs), the most commonly used sample-weighting techniques (called inverse probability weighting (IPW), as well as refined versions that seek to stabilize the variance of the weights) can be applied at multiple time points to populations for which exposures or treatments, confounders, and individuals entering or leaving the populations are all time-varying. This flexibility, together with emphasis on counterfactuals and missing observations, make MSMs particularly well suited to the analysis of time-varying confounders and effects of treatments or interventions that involve feedback loops, such as when the treatment that a patient receives depends on his or her responses so far and also to analysis of data in which imperfect compliance, attrition from the sample, or other practical difficulties drive a wedge between what was intended and what actually occurred in the treatment and follow-up of patients [96]. For example, MSMs are often applied to *intent-to-treat* data, in which the intent or plan to treat patients in a certain way is taken as the controllable causal driver of outcomes, and what happens next may depend in part on real-world uncertainties.

Despite their advantages in being able in principle to quantify causal impacts in complex time-varying data sets, MSMs have some strong practical limitations. Their results are typically very sensitive to errors in the specification of the regression models used to estimate unobserved counterfactual values, and the correct model specification is usually unknown. Therefore, MSMs are increasingly being used in conjunction with *model ensemble* techniques to address model uncertainty. Model ensemble methods (including Bayesian model averaging, various forms of statistical boosting, k-fold cross-validation techniques, and super-learning, as described next) calculate results using many different models and then combine the results. The use of diverse plausible models avoids the false certainty and potential biases created by selecting a single model. For example, in *super-learning algorithms*, no single regression model is selected. Instead multiple different standard machine-learning algorithms (e.g., logistic regression, random forest, support vector machine, naïve Bayesian classifier, artificial neural network, etc.) are used to predict unobserved values [86] and to estimate IPW weights [37]. These diverse predictions are then combined via weighted averaging, where the weights reflect how well each algorithm predicts known values that have been deliberately excluded (held out for test purposes) from the data supplied to the algorithms – the computational statistical technique known as model cross validation. Applied to the practical problem of estimating the mortality hazard ratio for initiation versus no initiation of combined antiretroviral therapy among HIV-positive subjects, such ensemble learning algorithms produced clearer effects estimates (hazard ratios further below 1, indicating a beneficial effect of the therapy) and narrower confidence intervals than traditional single-model (logistic regression modeling) analysis (*ibid*).

Yet even these advances do not overcome the fact that MSMs requires strong, and often unverifiable, assumptions to yield valid estimates of causal impacts. Typical examples of such assumptions are that there are no unmeasured confounders, that the observed response of each individual (e.g., of each patient to treatment or nontreatment) is in fact caused by the observed risk factors, and that every value for the causal variables occurs for every combination of levels of the confounding variables (e.g., there is no time period before exposures began but confounders were present) [18]. Assuming that these conditions hold may lead to an unwarranted inference that a certain exposure causes an adverse health effect, e.g., that ozone air pollution causes increased asthma-related hospitalizations, even if analyses based only on realistic, empirically verifiable assumptions would reveal no such causal relation [82].

Counterfactual models are often used to assess the effects on health outcomes of medical treatments, environmental exposures, or preventable risk factors by comparing what happened to people who receive the treatments to what models predict would have happened without the treatments. However, a limitation of such counterfactual comparisons is that they are seldom explicit about *why* treatments would not have occurred in the counterfactual world envisioned. Yet, the answer can crucially affect the comparison being drawn [35, 46]. For example, if it is assumed that a patient would not have received a treatment because the physician is confident that it would not have worked and the patient would have died anyway,

then the estimated effect of the treatment on mortality rates might be very different from what it would be if it is assumed that the patient would not have received the treatment because the physician is confident that there is no need for it and that the patient would recover anyway. In practice, counterfactual comparisons usually do not specify in detail the causal mechanisms behind the counterfactual assumptions about treatments or exposures, and this can obscure the precise interpretation of any comparison between what did happen and what is supposed, counterfactually, would have happened had treatments (or exposure) been different. Standard approaches estimate effects under the assumption that those who are treated or exposed are exchangeable with those who are not within strata of adjustment factors that may affect (but are not affected by) the treatment or the outcome, but the validity of this assumption is usually difficult or impossible to prove.

An alternative, increasingly popular, approach to using untested assumptions to justify causal conclusions is the *instrumental variable* (IV) method, originally developed in econometrics [112]. In this approach, an *instrument* is defined as a variable that is associated with the treatment or exposure of interest – a condition that is usually easily verifiable – and that also affects outcomes (e.g., adverse health effects) only through the treatment or exposure variable, without sharing any causes with the outcome. (In DAG notation, such an instrument would be a variable with arrows directed only into Y and Z in the DAG model  $X \rightarrow Y \rightarrow Z$ , where Z is the outcome variable, Y the treatment or exposure variable, and X a variable that affects exposure, such as job category, residential location, or intent-to-treat.) These latter conditions are typically assumed in IV analyses, but not tested or verified. If they hold, then the effects of unmeasured confounders on the estimated association between Y and Z can be eliminated using observed values of the instrument, and this is the potential great advantage of IV analysis. However, in practice, IV methods applied in epidemiology are usually dangerously misleading, as even minor violations of their untested assumptions can lead to substantial errors and biases in estimates of the effects of different exposures or treatments on outcomes; thus, many methodologists consider it inadvisable to use IV methods to support causal conclusions in epidemiology, despite their wide and increasing popularity for this purpose [112]. Unfortunately, within important application domains in epidemiology, including air pollution health effects research, leading investigators sometimes draw strong but unwarranted causal conclusions using IV or counterfactual (PSM or MSM) methods and then present these dubious causal conclusions and effects estimates to policy-makers and the public as if they were known to be almost certainly correct, rather than depending crucially on untested assumptions of unknown validity (e.g., [104]). Such practices lead to important-seeming journal articles and policy recommendations that are untrustworthy, potentially reflecting the ideological biases or personal convictions of the investigator rather than true discoveries about real-world causal impacts of exposures [103]. Other scientists and policy makers are well advised to remain on guard against such enthusiastic claims about causal impacts and effects estimates promoted by practitioners of IV and counterfactual methods who do not appropriately caveat their conclusions by emphasizing their dependence on untested modeling assumptions.

When the required assumptions for counterfactual modeling cannot be confidently determined to hold, other options are needed for counterfactual analyses to proceed. The simplest and most compelling approach is to use genuine randomized control trials (RCTs), if circumstances permit it. They rarely do, but the exceptions can be very valuable. For example, the state of Oregon in 2008 used a lottery system to expand limited Medicaid coverage for the uninsured by randomly selecting names from a waiting list. Comparing subsequent emergency department use among the randomly selected new Medicaid recipients to subsequent use by those still on the waiting list who had not yet received Medicaid revealed a 40% increase in emergency department usage over the next 18 months among the new Medicaid recipients, including visits for conditions that might better have been treated in primary care physician settings. Because the selection of recipients was random, this increase in usage could be confidently attributed to a causal effect of the Medicaid coverage on increasing emergency department use [114]. The main limitation of such RCTs is not in establishing the existence and magnitude of genuine causal impacts of an intervention in the studied population but rather in determining to what extent the result can be generalized to other populations. While conclusions based on valid causal laws and mechanisms can be transported across contexts, as discussed later, this is not necessarily true of aggregate population-level causal impacts, which may depend on specific circumstances of the studied population.

In the more usual case where random assignment is not an option, use of non-randomized control groups can still be very informative for testing, and potentially refuting, assumptions about causation. Indeed, analyses that estimate the impacts of changes in exposures by comparing population responses before and after an intervention that changes exposure levels can easily be misled unless appropriate comparison groups are used. For example, a study that found a significant drop in mortality rates from the six years prior to a coal burning ban in Dublin county, Ireland, to the six years following the ban concluded that the ban had caused a prompt, significant fall in all-cause and cardiovascular mortality rates [42]. This finding eventually led officials to extend the bans to protect human health. However, such a pre-post comparison study design cannot support a logically valid inference of causality, since it pays no attention to what would have happened to mortality rates in the absence of an intervention, i.e., the coal-burning ban. When changes in all-cause and cardiovascular mortality rates outside the ban area were later compared to those in areas affected by the ban, it turned out that there was no detectable difference between them: contrary to the initial causal inference, the bans appeared to have had no detectable impact on reducing these rates [44]. Instead, the bans took place during a decades-long period over which mortality rates were decreasing, with or without bans, throughout much of Europe and other parts of the developed world, largely due to improvements in early detection, prevention, and treatment of cardiovascular risks. In short, what would have happened in the absence of an intervention can sometimes be revealed by studying what actually did happen in appropriate comparison or control groups – a key idea developed and applied in the field of quasi-experimental (QE) studies, discussed next. Counterfactual causal inferences drawn without such comparisons can easily be misled.

## 9 Valid Causal Relations Cannot Be Explained Away by Noncausal Explanations

An older, but still useful, approach to causal inference from observational data, developed largely in the 1960s and 1970s, consists of showing that there is an association between exposure and response that cannot plausibly be explained by confounding, biases (including model and data selection biases and specification errors), or coincidence (e.g., from historical trends in exposure and response that move together but that do not reflect causation). *Quasi-experiment* (QE) design and analysis approaches originally developed in social statistics [12] systematically enumerate potential alternative explanations for observed associations (e.g., coincident historical trends, regression to the mean, population selection, and response biases) and provide statistical tests for refuting them with data, if they can be refuted. The interrupted time series analysis studies discussed earlier are examples of quasi-experiments: they do not allow random assignment of individuals to exposed and unexposed populations but do allow comparisons of what happened in different populations before and after an intervention that affects some of the populations but not others (the comparison groups).

A substantial tradition of *refutationist approaches* in epidemiology follows the same general idea of providing evidence for causation by using data to explicitly test, and if possible refute, other explanations for exposure-response associations [77]. As stated by Samet and Bodurow [100], “Because a statistical association between exposure and disease does not prove causation, plausible alternative hypotheses must be eliminated by careful statistical adjustment and/or consideration of all relevant scientific knowledge. Epidemiologic studies that show an association after such adjustment, for example through multiple regression or instrumental variable estimation, and that are reasonably free of bias and further confounding, provide evidence but not proof of causation.” This is overly optimistic, insofar as associations that are reasonably free of bias and confounding do not necessarily provide evidence of causation. For example, strong, statistically significant associations (according to usual tests, e.g., t-tests) typically occur in regression models in which the explanatory and dependent variables undergo statistically independent random walks. The resulting associations do not arise from confounding or bias but from spurious regression, i.e., coincident historical trends created by random processes that are not well described by the assumptions of the regression models. Nonetheless, the recommendation that “plausible alternative hypotheses must be eliminated by careful statistical adjustment and/or consideration of all relevant scientific knowledge” well expresses the refutationist point of view.

---

## 10 Changes in Causes Produce Changes in Effects via Networks of Causal Mechanisms

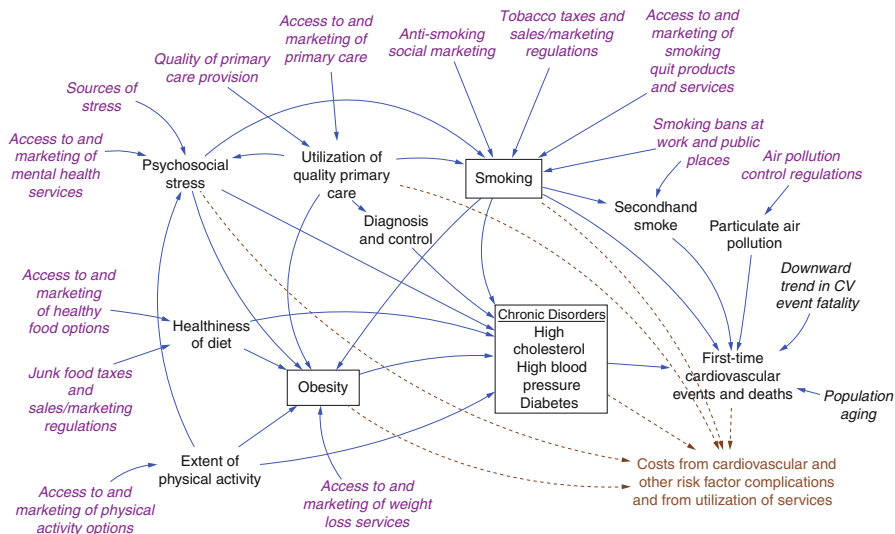
Perhaps the most useful and compelling valid evidence of causation, with the possible exception of differences in effects between treatment and control groups in well-conducted randomized control trials, consists of showing that changes in

exposures propagate through a network of validated lawlike structural equations or mechanisms to produce predictable changes in responses. For example, showing that measured changes in occupational exposures to a workplace chemical consistently produce a sequence of corresponding changes in lung inflammation markers, recruitment rates of activated alveolar macrophages and activated neutrophils to the chronically inflamed lung, levels of tissue-degrading enzymes released by these cell populations, and resulting rates of lung tissue destruction and scarring, leading to onset of lung pathologies and clinically detectable lung diseases (such as emphysema, silicosis, fibrosis, or inflammation-mediated lung cancer), would provide compelling evidence of a causal relation between changes in exposures and changes in those disease rates. Observing the network of mechanisms by which changes in exposures are transduced to changes in disease risks provides knowledge-based evidence of causation that cannot be obtained from any purely statistical analysis of observational data on exposures and responses alone.

Several causal modeling techniques are available to describe the propagation of changes through networks of causal mechanisms. *Structural equation models (SEMs)*, in which changes in right-hand side variables cause adjustments of left-hand side variables to restore all equalities in a system of structural equations, as in Fig. 43.1, provide one way to describe causal mechanisms for situations where the precise time course of the adjustment process is not of interest. Differential equation models, in which flows among compartments change the values of variables representing compartment contents over time (which in turn may affect the rates of flows), eventually leading to new equilibrium levels following an exogenous intervention that changes the compartment content or flow rates, provide a complementary way to describe mechanisms when the time course of adjustment is of interest. Simulation models provide still another way to describe and model the propagation of changes through causal networks. Figure 43.6 illustrates the structure of a simulation model for cardiovascular disease (CVD) outcomes. At each time step, the value of each variable is updated based on the values of the variables that point into it. The time courses of all variables in the model can be simulated for any history of initial conditions and exogenous changes in the input variables (those with only outward-pointing arrows), given the computational models that determine the change in the value of each variable at each time step from the values of its parents in the DAG.

## 10.1 Structural Equation and Path Analysis Models Model Linear Effects Among Variables

For most of the past century, DAG models such as those in Figs. 43.1 and 43.6, in which arrows point from some variables into others and there are no directed cycles, have been used to explicate causal networks of mechanisms and to provide formal tests for their hypothesized causal structures. For example, *path analysis* methods showing the dependency relations among variables in SEMs have been used for many decades to show how some variables influence others when all relations are assumed to be linear. Figure 43.7 presents an example involving several



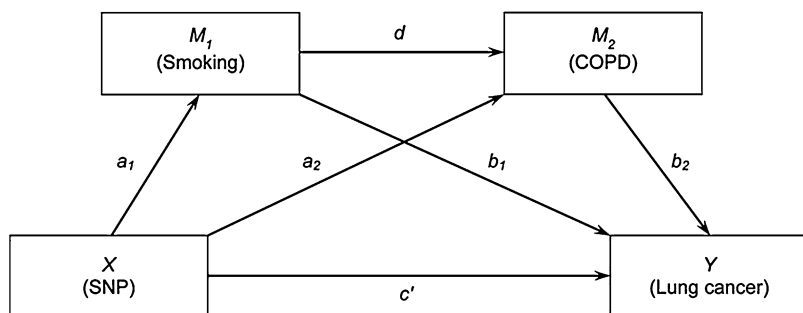
**Fig. 43.6** Simulation model for major health conditions related to cardiovascular disease (CVD) and their causes. Boxes represent risk factor prevalence rates modeled as dynamic stocks. Population flows among these stocks – including people entering the adult population, entering the next age category, immigration, risk factor incidence, recovery, cardiovascular event survival, and death – are not shown (Source: [52])

**Key:** *Blue solid arrows:* causal linkages affecting risk factors and cardiovascular events and deaths. *Brown dashed arrows:* influences on costs.

*Purple italics:* factors amenable to direct intervention.

*Black italics* (population aging, cardiovascular event fatality): other specified trends.

*Black nonitalics:* all other variables, affected by italicized variables and by each other



**Fig. 43.7** A path diagram with standardized coefficients showing linear effects of some variables on others (Source: [120])

variables that are estimated to significantly predict lung cancer risk: the presence of a particular single nucleotide polymorphism (SNP) (the CHRNA5-A3 gene cluster, a genetic variant which is associated with increased risk of lung cancer), smoking, and presence of chronic obstructive pulmonary disease (COPD) [120].



The path coefficient on an arrow indicates by how much (specifically, by how many standard deviations) the expected value of the variable into which it points would change if the variable at the arrow's tail were increased by one standard deviation, holding all other variables fixed and assuming that all relations are well approximated by linear structural equation regression models, i.e., that changing the variable at the arrow's tail will cause a proportional change in the variable at its head. In this example, the path coefficients are denoted by  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $c'$ , and  $d$ . These numbers must be estimated from data to complete the quantification of the path diagram model. Although such path analysis models are derived from correlations, the causal interpretation (i.e., that changing a variable at the tail of an arrow will change the variable at its head in proportion to the coefficient on the arrow between them) is an assumption. It is justified only if the regression equations used are indeed structural (causal) equations and if the assumptions required for multiple linear regression (e.g., additive effects, constant variance, normally distributed errors) hold. For the path diagram in Fig. 43.7, the authors found that the gene variant,  $X$ , affected lung cancer risk,  $Y$ , by increasing smoking behavior and, separately, by increasing COPD risk, as well as by increasing smoking-associated COPD risk: "The results showed that the genetic variant influences lung cancer risk indirectly through all three different pathways. The percent of genetic association mediated was 18.3% through smoking alone, 30.2% through COPD alone, and 20.6% through the path including both smoking and COPD, and the total genetic variant-lung cancer association explained by the two mediators was 69.1%."

Path diagrams reflect the fact that, if all effects of variables on each other are well approximated by linear regression SEMs, then correlations between variables should be stronger between variables that are closer to each other along a causal chain than between variables that are more remote, i.e., that have more intervening variables. Specifically, the effect of a change in the variable at the start of a path on a variable at the end of it that is transmitted along that path is given by the product of the path coefficients along the path. Thus, in Fig. 43.7, the presence of the SNP should be more strongly correlated with COPD than with COPD-associated lung cancer. Moreover, the effect of a change in an ancestor variable on the value of a remote descendent (several nodes away along one or more causal paths) can be decomposed into the effects of the change in the ancestor variable on any intermediate variables and the effects of those changes in intermediate variables, in turn, on the remote descendent variable. If one variable does *not* point into another, then the SEM/path analysis model implies that the first is not a direct cause of the second. For example, the DAG model  $X \rightarrow Y \rightarrow Z$  implies that  $X$  is an ancestor (indirect cause) but not a parent (direct cause) of  $Z$ . An implication of the causal ordering in this simple DAG model can be tested, as previously noted, by checking whether  $Z$  is conditionally independent of  $X$  given  $Y$ .

In linear SEM/path analysis models, conditional independence tests specialize to testing whether partial correlation coefficients between two variables become zero after conditioning on the values of one or more other variables (e.g., the partial correlation between  $X$  and  $Z$ , holding  $Y$  fixed, would be zero for the path  $X \rightarrow Y \rightarrow Z$ ). This makes information-theoretic methods unnecessary when the

assumptions of linear SEMs and jointly normally distributed error terms relating the value of each variable to the values of its parents hold; analyses based on correlations can then be used instead. Such consistency and coherence constraints can be expressed as systems of equations that can be solved, when identifiability conditions hold, to estimate the path coefficients (including confidence intervals) relating changes in parent variables to changes in their children. Summing these changes over all paths leading from exposure to response variables allows the total effect (via all paths) of a change in exposure on changes in expected responses to be estimated or predicted.

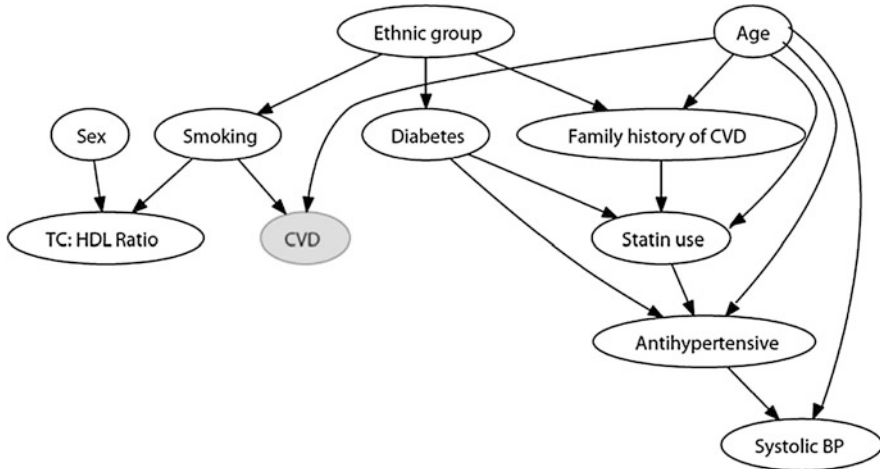
Path analysis and other SEM models are particularly valuable for detecting and quantifying the effects of unmeasured (“latent”) confounders based on the patterns of correlations that they induce among observed variables. SEM modeling methods have also been extended to include quadratic terms, categorical variables, and interaction terms [66]. Standard statistics packages and procedures, such as PROC CALIS in SAS, have made this technology available to modelers for the past four decades, and free modern implementations are readily available (e.g., in the R packages SEM or RAMpath).

## 10.2 Bayesian Networks Show How Multiple Interacting Factors Affect Outcome Probabilities

Path analysis, which is now nearly a century old, provided the main technology for causal modeling for much of the twentieth century. More recently, DAG models have been generalized so that causal mechanisms need not be described by linear equations for expected values, but may instead be described by arbitrary conditional probability relations. The nodes in such a graph typically represent random variables, stochastic processes, or time series variables in which a decision-maker may intervene from time to time by taking actions that cause changes in some of the time series [4, 23]).

*Bayesian networks* (BNs) are a prominent type of DAG model in which nodes represent constants, random variables, or deterministic functions [3]. Figure 43.8 shows an example of a BN for cardiovascular disease (CVD). As usual, the absence of an arrow between two nodes, such as between *ethnic group* and *CVD* in Fig. 43.8, implies that neither has a probability distribution that depends directly on the value of the other. Thus, *ethnic group* is associated with CVD, but the association is explained by *smoking* as a mediating variable, and the structure of the DAG shows no further dependence of *CVD* on *ethnic group*. Statistically, the random variable indicating cardiovascular disease, *CVD*, is conditionally independent of *ethnic group*, given the value of *smoking*.

Some useful causal implications may be revealed by the structure of a DAG, even before the model is quantified to create a fully specified BN (or other DAG) model. For example, if the DAG structure in Fig. 43.8 correctly describes an individual or population, then elevated systolic BP (blood pressure) is associated with CVD risk, since both have age and ethnicity as ancestors in the DAG. However, changes in statin use, which could affect systolic BP via the intermediate variable



**Fig. 43.8** Directed acyclic graph (DAG) structure of a Bayesian network (BN) model for cardiovascular disease (CVD) risk (Source: [115])

antihypertensive, would not be expected to have any effect on CVD risk. Learning the correct DAG structure of causal relations among variables from data – a key part of the difficult task of *causal discovery*, discussed later – can reveal important and unexpected findings about what works and what does not for changing the probabilities of different outcomes, such as CVD in Fig. 43.8. However, uncertainty about the correct causal structure can make sound inference about causal impacts (and hence recommendations about the best choice of actions to produce desired changes) difficult to determine. Such model uncertainty motivates the use of ensemble modeling methods, discussed later.

### 10.3 Quantifying Probabilistic Dependencies Among BN Variables

For quantitative modeling of probabilistic relations among variables, input nodes in a BN (i.e., nodes with only outward-pointing arrows, such as sex, age, and ethnic group in Fig. 43.8) are assigned marginal (unconditional) probability distributions for the values of the variables they represent. These marginal distributions can be thought of as being stored at the input nodes, e.g., in tables that list the probability or relative frequency of each possible input value (such as male or female for sex, age in years, etc.). They represent the prior probabilities that each input node will have each of its possible values for a randomly selected case or individual described by the BN model, before getting more information about a specific case or individual. For any specific individual to whom the BN model is applied, if the values of inputs such as sex, age, and ethnicity are known, then their values would be specified as inputs and conditioned on at subsequent nodes in applying the model to that individual. Figure 43.9 illustrates this concept.



**Fig. 43.9** BN model of risk of infectious diarrhea among children under 5 in Cameroon. The left panel shows the unconditional risk for a random child from the population (14.97%); the right panel shows the conditional risk for a malnourished child from a home in the lowest income quintile and with poor sanitation (20.00%) (Source: [87])

The left panel shows a BN model for risk of infectious diarrhea among young children in Cameroon. Each of three risk factors – income quintile, availability of toilets (“sanitation”), and stunted growth (“malnutrition”) – affects the probability that a child will ever have had diarrhea for at least two weeks (“diarrhea”). In addition, these risk factors affect each other, with an observation of low income making observations of poor sanitation and malnutrition more likely and observed poor sanitation also making observed malnutrition more likely at each level of income. The right panel shows an instance of the model for a particular case of a malnourished child from the poorest income quintile living with poor sanitation; these three risk factor values all have probabilities set to 100%, since their values are known. The result is that the risk of diarrhea, conditioned on this information, is increased from an average value of 14.97% in this population of children to 20% when all three risk factors are set to these values.

A BN model stores *conditional probability tables* (CPTs) at nodes with inward-pointing arrow. A CPT simply tabulates the conditional probabilities for the values of the node variable for each combination of values of the variables that point into it (its “parents” in the DAG). For example, the *malnutrition* node in Fig. 43.9 has a CPT with 10 rows, since its two parents, *Income* and *Sanitation*, have 5 and 2 possible levels, respectively, implying ten possible pairs of input values. For each of these ten combinations of input values, the CPT shows the conditional probability for each possible value of *Malnutrition* (here, just *Yes* and *No*, so that the CPT for *Malnutrition* has 2 columns); these conditional probabilities must sum to 1 for each row of the CPT. BN nodes can also represent deterministic functions by CPTs that assign 100% probability to a specific output value for each combination of input values. The conditional probability distribution for the value of a node (i.e., variable) thus depends on the values of the variables that point into it; it can be freely specified (or estimated from data, if adequate data are available) without making any

restrictive assumptions about linearity or normal errors. Such BN models greatly extend the flexibility of practical causal hypothesis testing and causal predictive modeling beyond traditional linear SEM and path analysis models.

In practice, CPTs can usually be condensed into relatively small tables by using classification trees or other algorithms (e.g., rough sets) to bin the potentially large number of combinations of values for a node's parents into just those that predict significantly different conditional probability distributions for the node's value. Instead of enumerating all the combinations of values for the parents, "don't care" conditions (represented by blanks in the CPT entries or by missing splits in a classification tree) can reduce the number of combinations that must be explicitly stored in the CPT. Alternatively, a logistic regression model or other statistical model can be used in place of a CPT at each node. For example, although the *Diarrhea* node in Fig. 43.9 could logically have a CPT with  $5 \times 2 \times 2 = 20$  rows, it may be that a simple regression model with only three coefficients for the main effects of the parents, and few or no additional terms for interactions, would adequately approximate the full CPT.

## 10.4 Causal vs. Noncausal BNs

Any joint probability distribution of multiple random variables can be factored into a product of marginal and conditional probability distributions and displayed in DAG form, usually in several different ways. For example, the joint probability mass function  $P(x, y)$  of two discrete random variables  $X$  and  $Y$ , specifying the probability of each pair of specific values  $(x, y)$  for random variables  $X$  and  $Y$ , can be factored as  $P(x)P(y|x)$  or as  $P(y)P(x|y)$  and can be displayed in a BN as  $X \rightarrow Y$  or as  $Y \rightarrow X$ , respectively. Here,  $x$  and  $y$  denote possible values of random variables  $X$  and  $Y$ , respectively;  $P(x, y)$  denotes the joint probability that  $X = x$  and  $Y = y$ ;  $P(x)$  denotes the marginal probability that  $X = x$ ;  $P(y)$  denotes the marginal probability that  $Y = y$ , and  $P(y|x)$  and  $P(x|y)$  denote conditional probabilities that  $Y = y$  given that  $X = x$  and that  $X = x$  given that  $Y = y$ , respectively. Thus, there is nothing inherently causal about a BN. Its nodes need not represent causal mechanisms that map values of inputs to probabilities for the values of outputs caused by those inputs. Even if they do represent such causal mechanisms, they may not explicate how or why the mechanisms work. For example, the direct link from *Income* to *Malnutrition* in Fig. 43.9 gives no insight into how or why changes in income affect changes in malnutrition – e.g., what specific decisions or behaviors are influenced by income that, in turn, results in better or worse nutrition. Thus, it is possible to build and use BNs for probabilistic inference without seeking any causal interpretation of the statistical dependencies among its variables.

However, BNs are often deliberately constructed and interpreted to mean that changes in the value of a variable at the tail of an arrow will cause a change in the probability distribution of the variable into which it points, as described by the CPT at that node. The effect of a change in a parent variable on the probability

distribution of a child variable into which it points may depend on the values of other parents of that node, thus allowing interactions among direct causes at that node to be modeled. For example, in Fig. 43.8, the effects of smoking on CVD risk may be different at different ages, and this would be indicated in the CPT for the CVD node by having different probabilities for the values of the CVD variable at different ages for the same value of the smoking variable. A *causal BN* is a BN in which the nodes represent stable causal mechanisms or laws that predict how changes in input values change the probability distribution of output values. The CPT at a node of a causal BN describes the conditional probability distribution for its value caused by each combination of values of its inputs, meaning that changes in one or more of its input values will be followed by corresponding changes in the probability distribution for the node's value, as specified by the CPT. This is similar to the concept of a causal mechanism in structural equation models, where a change in a right-hand side (explanatory or independent) variable in a structural equation is followed by a change in the left-hand side (dependent or response variable) to restore equality [119].

A causal BN allows changes at input nodes to be propagated throughout the rest of the network, yielding a posterior joint probability distribution for the values of all variables. (If the detailed time course of changes in probabilities is of interest, then differential equations or dynamic Bayesian networks (DBNs), discussed later, may be used to model how the node's probability distribution of values changes from period to period.) The order in which changes propagate through a network provides insight into the (total or partial) causal ordering of variables and can be used to help deduce network structures from time series data [119]. Similarly, in a set of simultaneous linear structural equations describing how equilibrium levels of variables in a system are related, the causal ordering of variables (called the *Simon causal ordering* in econometrics) is revealed by the order in which the equations must be solved to determine the values of all the variables, beginning with exogenous inputs (and assuming that the system of equations can be solved uniquely, i.e., that the values of all variables are uniquely identifiable from the data). Causality flows from exogenous to endogenous variables and among endogenous variables in such SEMs (*ibid*). Exactly how the changes in output probabilities (or in the expected values of left-side variables in an SEM) caused by changes in inputs are to be interpreted (e.g., as changes in the probability distribution of future observed values for a single individual or as changes in the frequency distribution of the variable values in a population of individuals described by the BN) depends on the situation being modeled.

## 10.5 Causal Mechanisms Are Lawlike, Yielding the Same Output Probabilities for the Same Inputs

A true causal mechanism that has been explicated in enough detail to make reliable predictions can be modeled as a conditional probability table (CPT) that gives the same conditional probabilities of output values whenever the input values are the same. Such a stable, repeatable relation, which might be described as lawlike,

can be applied across multiple contexts as long as the inputs to the node are sufficient to determine (approximately) unique probabilities for its output values. For example, a dose-response relation between radiation exposure and excess age-specific probability (or, more accurately, hazard rate) for first diagnosis with a specific type of leukemia might be estimated from data for one population and then applied to another with similar exposures, provided that the change in risk caused by exposure does not depend on omitted factors. If it depends on age and ethnic group, for example, then these would have to be included, along with exposure, as inputs to the node representing leukemia status. By contrast, unexplained heterogeneity, in which the estimated CPT differs significantly when study designs are repeated by different investigators, signals that a lawlike causal mechanism has not yet been discovered. In that case, the models and the knowledge that the BN represents need to be further refined to discover and express predictively useful causal relations that can be applied to new conditions. The key idea is that, to be transferable across contexts (e.g., populations), the probabilistic relations encoded in CPTs must include all of the input conditions that suffice to make their conditional probabilities accurate, given accurately measured or estimated input values.

A proposed causal relation that turns out to be very heterogeneous, sometimes showing significant positive effects and other times no effects or significant negative effects under the same conditions, does not correspond to a lawlike causal relation and cannot be relied on to make valid causal predictions (e.g., by using mean values averaged over many heterogeneous studies). Thus, the estimated CPTs at nodes in Fig. 43.9 may be viewed as averages of many individual-specific CPTs, and the predictions that they make for any individual case may not be accurate. CPTs that simply summarize historical data on conditional frequency distributions, but that do not represent causal mechanisms, may be no more than mixtures of multiple CPTs for the (perhaps unknown) populations and conditions that contributed to the historical data. They cannot necessarily be generalized to new populations or conditions (sometimes described as being transported to new contexts) or used to predict how outputs will change in response to changes in inputs, unless the relevant mixtures are known. For example, suppose that the *Sanitation* node has a value of 1 for children from homes with toilets and a value of 0 otherwise. If homes may have toilets either because the owners bought them or because a government program supplied them as part of a program along with a child food and medicine program, then the effect of a finding that *Sanitation* = 1 on the conditional probability distribution of *Malnutrition* may depend very much on which of these reasons resulted in *Sanitation* = 1. But this is not revealed by the model in Fig. 43.9. In such a case, the estimated CPTs for the nodes in Fig. 43.9 should not be interpreted as describing causal mechanisms, and the effects on other variables of setting *Sanitation* = 1 by alternative methods cannot be predicted from the model in Fig. 43.9.

## 10.6 Posterior Inference in BN Models

Once a BN has been quantified by specifying its DAG structure and the probability tables at its nodes, it can be used to draw a variety of useful inferences by applying

any of several well-developed algorithms created and refined over the past three decades [3]. The most essential inference capability of a BN model is that if observations (or “findings”) about the values at some nodes are entered, then the conditional probability distributions of all other nodes can be computed, conditioned on the evidence provided by these observed values. This is called “posterior inference.” In other words, the BN provides computationally practicable algorithms for accomplishing the Bayesian operation of updating prior knowledge or beliefs, represented in the node probability tables, with observations to obtain posterior probabilities. For example, if known values of a patient’s age, sex, and systolic blood pressure were to be entered for the BN in Fig. 43.8, then the conditional probability distributions based on that information could be computed for all other variables, including diabetes status and CVD risk, by BN posterior inference algorithms. In Fig. 43.9, learning that a child is from a home with inadequate sanitation would allow updated (posterior) probabilities for the possible income and nutrition levels, as well as the probability of diarrhea, to be computed using exact probabilistic inference algorithms. The best-known exact algorithm (the junction tree algorithm) is summarized succinctly by [3]. For large BNs, approximate posterior probabilities can be computed efficiently using Monte Carlo sampling methods, such as Gibbs sampling, in which input values drawn from the marginal distributions at input nodes are propagated through the network by sampling from the conditional distributions given by the CPTs, thus building up a sample distribution for any output variable(s) of interest.

The BN in Fig. 43.10 illustrates that different types of data, from demographics (age and sex) to choices and behaviors (smoking) to comorbidities (diabetes) to clinical measurements (such as systolic and diastolic blood pressures (SP and DBP)) and biomarkers (cholesterol levels) can be integrated in a BN model, here built using

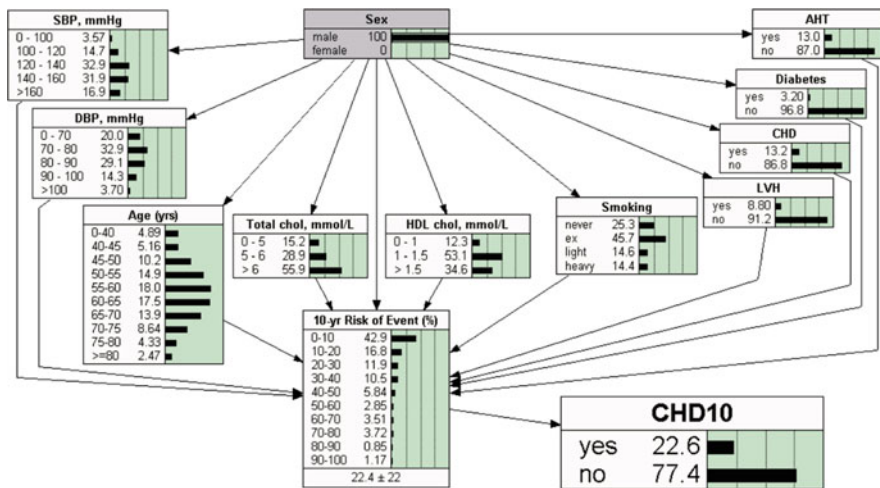


Fig. 43.10 BN model for predicting CHD risk from multiple types of data [117]



the popular *Netica* BN software product, to inform risk estimates for coronary heart disease (CHD) occurring in the next ten years. In addition to posterior inference of entire probability distributions for its variables, BNs can be used to compute the most likely explanations for observed or postulated outcomes (e.g., what are the most likely input values that lead to a specified set of output values) and to study the sensitivity of the probability of achieving (or avoiding) specified target sets of output values to changes in the probabilities of input values.

BN software products such as *Netica*, *Hugin*, or *BayesiaLab* not only provide algorithms to carry out these computations but also integrate them with graphic user interfaces for drawing BNs, populating their probability tables, and reporting results. For example, the DAG in Fig. 43.10, drawn in *Netica*, displays probability distributions for the values of each node. The probabilities are for a man, since the *Sex* node at the top of the diagram has its probability for the value “male” set to 100%. This node is shaded to show that observed or assumed values have been specified by the user, rather than being inferred by the BN model. If additional facts (“findings”) are entered, such as that the patient is a diabetic never-smoker, then the probability distributions at the *10-yr Risk of Event* node and the other nodes would all be automatically updated to reflect (condition upon) this information.

Free BN software packages for creating BNs and performing posterior inference are also available in both R and Python. In R, the *gRain* package allows BNs to be specified by entering the probability tables for their nodes. The resulting BN model can then be queried by entering the variables for which the posterior probability is desired, along with observed or assumed values for other variables. The package will return the posterior probabilities of the query variables, conditioned on the specified observations or assumptions. Both exact and approximate algorithms (such as the junction tree algorithm and Monte Carlo simulation-based algorithms, respectively) for such posterior inference in Bayesian networks are readily available if all variables are modeled as discrete with only a few possible values. For continuous variables, algorithms are available if each node can be modeled as having a normal distribution with a mean that is a weighted sum of the values of its parents, so that each node value depends on its parents’ values through a linear regression equation. Various algorithms based on Monte Carlo simulation are available for the case of mixed discrete and continuous BNs [13].

## 10.7 Causal Discovery of BNs from Data

A far more difficult problem than posterior inference is to infer or “learn” BNs or other causal graph models directly from data. This is often referred to as the problem of *causal discovery* (e.g., [54]). It includes the *structure discovery* problem of inferring the DAG graph of a BN from data, e.g., by making sure that it shows the conditional independence relations (treated as constraints), statistical dependencies, and order of propagation of changes [119] inferred from data. Structure learning algorithms are typically either constraint-based, seeking to find DAG structures that satisfy the conditional independence relations and other constraints inferred

from data or score-based, seeking to find the DAG structure that maximizes a criterion (e.g., likelihood or posterior probability penalized for complexity) [3,9,20], although hybrid algorithms have also been developed. Learning a BN from data also requires quantifying the probability tables (or other representations of the probabilistic input-output relation) at each node, but this is usually much easier than structure learning. Simply tabulating the frequencies of each output value for each combination of input values may suffice for large data sets if the nodes have been constructed to represent causal mechanisms. For smaller data sets, fitting classification trees or regression models to available data can generate an estimated CPT, giving the conditional probability of each output value for each set of values of the inputs. Alternatively, Bayesian methods can be used to condition priors (typically, Dirichlet priors for multinomial random variables) on available data to obtain posterior distributions for the CPTs [110].

Although many BN algorithms are now available to support learning BNs from data [105], a fundamental limitation and challenge remains that multiple different models often provide approximately equally good explanations of available data, as measured by any of the many scoring rules, information-theoretic measures, and other criteria that have been proposed, and yet they make different predictions for new cases or situations. In such cases, it is better to use an ensemble of BN models instead of any single one to make predictions and support decisions [3]. How best to use common-sense knowledge-based constraints (e.g., that death can be an effect but not a cause of exposure or that age can be a cause but not an effect of health effects) to extract unique causal models, or small sets of candidate models, from data is still an active area of research, but most BN packages allow users to specify both required and forbidden arrows between nodes when these knowledge-based constraints are available. Since it may be impossible to identify a unique BN model from available data, the BN-learning and causal discovery algorithms included in many BN software packages should be regarded as useful heuristics for suggesting possible causal models, rather than as necessarily reliable guides to the truth.

For practical applications, the *bnlearn* package in R [105] provides an assortment of algorithms for causal discovery, with the option of including knowledge-based constraints by specifying directed or undirected arcs that must always be included or that must never be included. For example, in Fig. 43.8, sex, age, and ethnic group cannot have arrows directed into them (they are not caused by other variables), and CVD deaths cannot be a cause of any other variable [115]. The DAG model for cardiovascular disease risk prediction in Fig. 43.8 was discovered using one of the *bnlearn* algorithms (the grow-shrink algorithm for structure learning), together with these knowledge-based constraints. On the other hand, the BN model in Fig. 43.10, which was developed manually based on an existing regression model, has a DAG structure that is highly questionable. Its logical structure is that of a regression model: for men, all other explanatory or independent variables point into the dependent variable *10-year Risk of Event*, and there are no arrows directed between explanatory variables, e.g., from smoking to risk of diabetes. Such additional structure would probably have been discovered had machine learning algorithms for causal discovery such as those in *bnlearn* been applied to the original data. If the

DAG structure of a BN model is incorrect, then the posterior inferences performed using it – e.g., inferences about risks (posterior probabilities) of disease outcomes, and how they would change if inputs such as smoking status were altered – will not be trustworthy. This raises a substantial practical challenge when the correct DAG structure of a BN is uncertain.

## 10.8 Handling Uncertainty in Bayesian Network Models

BNs and other causal graph model are increasingly used in epidemiology to model uncertain and multivariate exposure-response relations. They are particularly useful for characterizing uncertain causal relations, since they can represent both uncertainty about the appropriate causal structure (DAG model), via the use of multiple DAGs (“ensembles” of DAG models), and uncertainties about the marginal and conditional probabilities at the input and other nodes. As noted by Samet and Bodurow [100], “The uncertainty about the correct causal model involves uncertainty about whether exposure in fact causes disease at all, about the set of confounders that are associated with exposure and cause disease, about whether there is reverse causation, about what are the correct parametric forms of the relations of the exposure and confounders with outcome, and about whether there are other forms of bias affecting the evidence. One currently used method for making this uncertainty clear is to draw a set of causal graphs, each of which represents a particular causal hypothesis, and then consider evidence insofar as it favors one or more of these hypotheses and related graphs over the others.”

An important principle for characterizing and coping with uncertainty about causal models is not to select and use any single model when there is substantial uncertainty about which one is correct [3]. It is more effective, as measured by many performance criteria for evaluating predictive models, such as mean squared prediction error, to combine the predictions from multiple models that all fit the data adequately (e.g., that all have likelihoods at least 10% as large as that of the most likely model). Indeed, the use of multiple models is often essential for accurately depicting model uncertainty when quantifying uncertainty intervals or uncertainty sets for model-based predictions. For example, Table 43.3 presents a

**Table 43.3** A machine learning challenge: What outcome should be predicted for case 7 based on the data in cases 1–6?

Case	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Outcome
1	1	1	1	1	1
2	0	0	0	0	0
3	0	1	1	0	1
4	1	1	0	0	0
5	0	0	0	0	0
6	1	0	1	1	1
7	1	1	0	1	?

small hypothetical data set to illustrate that multiple models may provide equally good (in this example, perfect) descriptions of all available data and yet make very different predictions for new cases. For simplicity, all variables in this example are binary (0–1) variables.

Suppose that cases 1–6 constitute a “training set,” with 4 predictors and one outcome column (the right most) to be predicted from them. The challenge for predictive analytics or modeling in this example is to predict the outcome for case 7 (the value, either 0 or 1, in the “?” cell in the lower right of the table). For example, predictors 1–4 might represent various features (1 = present, 0 = absent) of a chemical, or perhaps results of various quick and inexpensive assays for the chemical (1 = positive, 0 = negative) and the outcome might indicate whether the chemical would be classified as a rodent carcinogen in relatively expensive two-year live-animal experiments. A variety of machine-learning algorithms are available for inducing predictive rules or models from training data, from logistic regression to classification trees (or random forest, an ensemble-modeling generalization of classification trees) to BN learning algorithms. Yet, no algorithm can provide trustworthy predictions for the outcome in case 7 based on the training data in cases 1–6, since many different models fit the training data equally well and yet make opposite predictions. For example, the following two models each describe the training data in rows 1–6 perfectly, yet they make opposite prediction for case 7:

Model 1: Outcome = 1 if the sum of predictors 2, 3, and 4 exceeds 1, else 0

Model 2: Outcome = value of Predictor 3.

Likewise, these two models would make opposite predictions for a chemical with predictor values of (0, 0, 1, 0). Model 1 can be represented by a BN DAG structure in which predictors 2, 3, and 4 are the parents of the outcome node (and the CPT is a deterministic function with probabilities of 1 or 0 that the outcome = 1, depending on the values of these predictors). Model 2 would be represented by a BN in which only node 3 is a parent of the outcome node. The following are additional models or prediction rules, e.g.:

Model 3: Outcome is the greater of the values of predictors 1 and 2 except when both equal 1, in which case the outcome is the greater of the values of predictors 3 and 4.

Model 4: Outcome is the greater of the values of predictors 1 and 2 except when both equal 1, in which case the outcome is the lesser of the values of predictors 3 and 4.

These models also provide equally good fits to, or descriptions of, the training data, but make opposite predictions for case 7 and imply yet another BN structure. Thus, it is impossible to confidently identify a single correct model structure from the training data (the data-generating process is *non-identifiable* from the training data), and no predictive analytics or machine learning algorithm can determine from these data a unique model (or set of prediction rules) for correctly predicting the outcome for new cases or situations.

This example illustrates that successful classification or description of reference cases in a training set is a different task from successful prediction of outcomes for new cases outside the training set. It is possible for a computational procedure to have up to 100% accuracy on the former task, while making predictions with no better than a random (50–50) probability of being correct for the latter task. Yet, it is the latter that should be the goal of chief interest to practitioners who want to make predictions or decisions for cases other than those used in building the model. Using ensembles of models can help to characterize the range or set of predicted outcomes for new cases that are consistent with the training data, in the sense of being predicted by models that describe the training data well. They can also provide a basis for procedures that adaptively improve predictions (or decisions) as new cases are observed.

One way to implement this model ensemble approach is via weighted averaging of model-specific predictions, with weights chosen to reflect each model's performance, e.g., how well it explains the data, as assessed by its relative likelihood [3, 78, 79]. Such *Bayesian model averaging* (BMA) of multiple causal graphs avoids the risk of betting predictions on a single model. It demonstrably leads to superior predictions and to reduced model-selection and over-fitting biases in many situations [79]. Similar ideas are included in *super-learning* algorithms, already discussed, which assess model performance and weights via cross validation rather than via likelihood and adaptive learning approaches that learn to optimize not just predictions, but decision rules for making sequences of interventions as outcomes are gradually observed over time (e.g., the *iqLearn* algorithm of Linn et al. [72]). An important application of such decision rule learning algorithms is in sequential multiple assignment randomized trial (SMART) designs for clinical trials. These designs allow treatments or interventions for individual patients to be modified over time as their individual response and covariate histories are observed, in order to increase the probabilities of favorable outcomes for each patient while learning what intervention sequences work best for each type of patient [71].

When the probabilities to be entered into BN node probability tables are unknown, algorithms that propagate *imprecise probabilities* through BN models can be used (e.g., [29, 30]). Both the marginal probabilities at input nodes and the resulting probabilities of different outcomes (or the values at particular output nodes) will then be intervals, representing imprecise probabilities. More generally, instead of specifying marginal and conditional probability tables at the nodes of a BN, uncertainty about the probabilities can be modeled by providing a (usually convex) *set* of probability distributions at each node. BNs generalized in this way are called *credal networks*. Algorithms for propagating sets of probabilities through credal networks have been developed [15] and extended to support optimization of risk management decisions [20].

Alternatively, *second-order probability distributions* (“probabilities of probabilities”) for the uncertain probabilities at BN nodes can be specified. If these uncertainties about probabilities are well approximated by Dirichlet or beta probability distributions (as happens naturally when probabilities or proportions are estimated from small samples using Bayesian methods with uniform or Dirichlet

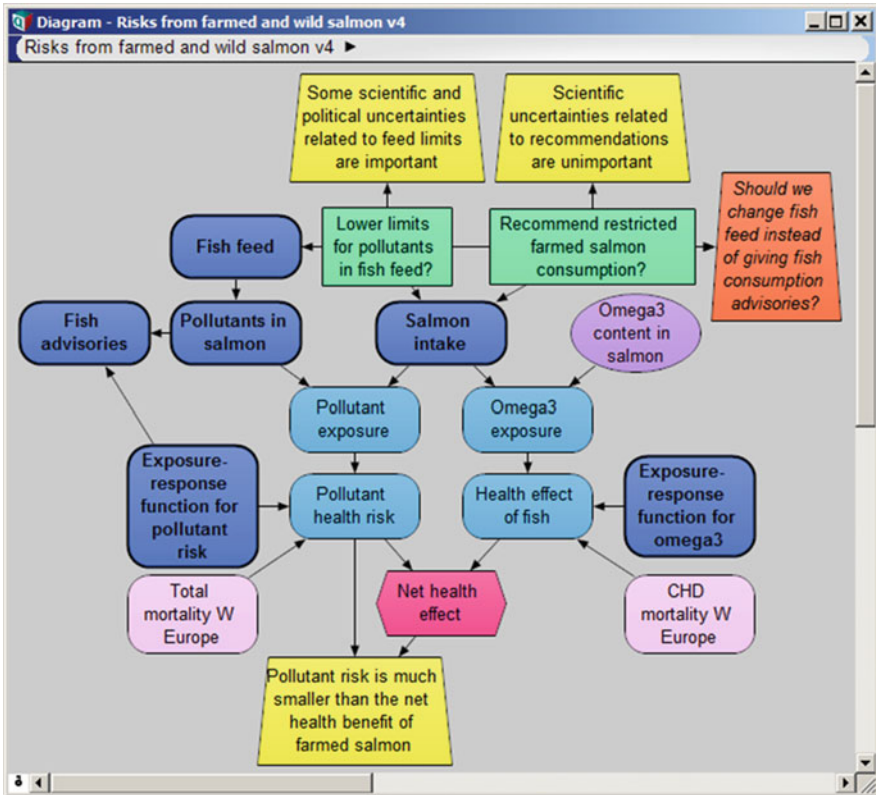
priors), then Monte Carlo uncertainty analysis can be used to propagate the uncertain probabilities efficiently through the BN model, leading to uncertainty distributions for the posterior probabilities of the values of the variables in the BN (Kleiter 1996). Imprecise Dirichlet models have also been used to learn credal sets from data, resulting in upper and lower bounds for the probability that each variable takes on a given value [15].

Rather than using sets or intervals for uncertain probabilities, it is sometimes possible to simply use best guesses (point estimates) and yet to have confidence that the results will be approximately correct. Henrion et al. (1996) note that, in many situations, the key inferences and insights from BN models are quite insensitive (or “robust”) to variations in the estimated values in the probability tables for the nodes. When this is the case, best guesses (e.g., MLE point estimates) of probability values may be adequate for inference and prediction, even if the data and expertise used to form those estimates are scarce and the resulting point estimates are quite uncertain.

## 10.9 Influence Diagrams Extend BNs to Support Optimal Risk Management Decision-Making

The BN techniques discussed so far are useful for predicting how output probabilities will change if input values are varied, provided that the DAG structure can be correctly interpreted as showing how changes in the inputs propagate through networks of causal mechanisms to cause changes in outputs. (As previously discussed, this requires that the network is constructed so that the CPTs at nodes represent not merely statistical descriptions of conditional probabilities in historical data but causal relations determining probabilities of output values for each combination of input values.) Once the probabilities of different outputs can be predicted for different inputs, it is natural to ask how the controllable inputs should be set to make the resulting probability distribution of outputs as desirable as possible. This is the central question of decision analysis, and mainstream decision analysis provides a standard answer: choose actions to maximize the expected utility of the resulting probability distribution of consequences.

To modify BN models to support optimal (i.e., expected utility-maximizing) risk management decision-making, the BNs must be augmented with two types of nodes that do not represent random variables or deterministic functions. There is a *utility node*, also sometimes called a *value node*, which is often depicted in DAG diagrams as a hexagon and given a name such as “Decision-maker’s utility.” There must also be one or more *choice nodes*, also called *decision nodes*, commonly represented by rectangles. The risk management decision problem is to make choices at the decision nodes to maximize the expected value of the utility node, taking into account the uncertainties and conditional probabilities described by the rest of the DAG model. Input decision nodes (i.e., decision nodes with only outward-directed arrows) represent inputs whose values are controlled by the decision-maker. Decision nodes with inputs represent *decision rules*, i.e., tables or functions specifying how the decision node’s value is to be chosen, for each



**Fig. 43.11** An influence diagram (ID) model with two decision nodes (green rectangles) and with *Net health effect* as the value node. (Questions and comments in trapezoids on the periphery are not parts of the formal ID model, but help to interpret it for policy makers) (Source: [www.lumina.com/case-studies/farmed-salmon/](http://www.lumina.com/case-studies/farmed-salmon/))

combination of values of its inputs. BNs with choice and value nodes are called *influence diagram* (ID) models. BN posterior inference algorithms can be adapted to solve for the best decisions in an ID, i.e., the choices to make at the choice nodes in order to maximize the expected value of the utility node [3, 123].

Figure 43.11 shows an example of an ID model developed and displayed using the commercial ID software package *Analytica*. Its two decision nodes represent choices about whether to lower the allowed limits for pollutants in fish feed and whether to recommend to consumers that they restrict consumption of farmed salmon, respectively [49]. The two decision nodes are shown as green rectangles, located toward the top of the ID. The value or utility node in Fig. 43.11, shown as a pink hexagon located toward the bottom of the diagram, is a measure of net health effect in a population. It can be quantified in units such as change in life expectancy (added person-year of life) or change in cancer mortality rates caused by different decisions and by the other factors shown in the model. Many of these

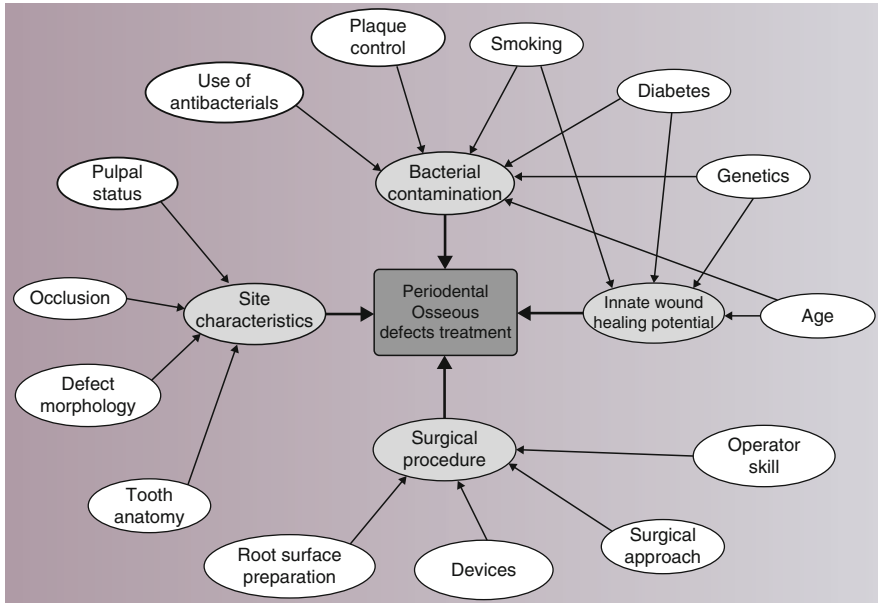
factors, such as (a) the estimated exposure-response relations for health harm caused by consuming pollutants in fish and (b) the health benefits caused by consuming omega three fatty acids in fish, are uncertain. The uncertainties are represented by random variables (the dark blue oval-shaped nodes throughout the diagram) and by modeling assumptions that allow other quantities (the light blue oval-shaped nodes) to be calculated from them.

An example of a modeling assumption is that pollutants increase mortality rates in proportion to exposure, with the size of this slope factor being uncertain. Different models (or expert opinions) for relevant toxicology, dietary habits, consumer responses to advisories and recommendations, nutritional benefits of fish consumption, and so forth can contribute to developing the CPTs for different parts of the ID model and characterizing uncertainties about them. IDs thus provide a constructive framework for coordinating and integrating multiple submodels and contributions from multiple domains of expertise and for applying them to help answer practical questions such as how different policies, regulations, warnings, or other actions will affect probable health effects, consumption patterns, and other outcomes of interest.

If multiple decision makers with different jurisdictions or spans of control attempt to control the same outcome, however, then coordinating their decisions effectively may require resolving game-theoretic issues in which each decision maker's best decision depends on what the others do. For example, in Fig. 43.11, if the regulators in charge of setting allowed limits for pollutant contamination levels in fish feed are different from the regulators or public health agencies issuing advisories about what to eat and what not to eat, then each might decide not to take additional action to protect public health if it mistakenly assumes that the other will do so. Problems of risk regulation or management with multiple decision-makers can be solved by generalizing IDs to *multi-agent influence diagrams* (MAIDs) [67, 89, 107]. MAID algorithms recommend what each decision-maker, each with its own utility function and decision variables, should do, taking into account any information it has about the actions of others, when their decisions propagate through a DAG model to jointly determine probabilities of consequences.

Although the idea of extending BNs to include decision and value nodes seems straightforward in principle, understanding which variables are controllable by whom over what time interval may require careful thought in practice. For example, consider a causal graph model (Fig. 43.12) showing factors affecting treatment of tooth defects (central node), such as patient's *age*, *genetics*, *smoking* status, *diabetes*, *use of antibacterials*, *pulpal status*, available surgical *devices*, and *operator skill* [1]. These variables have not been pre-labeled as chance or choice nodes. Even without expertise in dentistry, it is clear that some of the variables, such as *genetics* or *age*, should not ordinarily be modeled as decision variables. Others, such as *Use of antibacterials* and *Pulpal status* (reflecting oral hygiene) may result from a history of previous decisions by patients and perhaps other physicians or periodontists. Still others, such as available surgical *devices* and *operator skill*, are fixed in the short run, but might be considered decision variables over intervals long enough to include the operator's education, training, and experience or if the decisions to be





**Fig. 43.12** Different variables can be treated as decision variables on different time scales (Source: [1])

made include hiring practices and device acquisition decisions of the clinic where the surgery is performed. *Smoking* and *diabetes* indicators might also be facts about a patient that cannot be varied in the short run, but that might be considered as at least in part determined by past health and lifestyle decisions. In short, even if a perfectly accurate causal graph model were available, the question of who acts upon the world how and over what time frame via the causal mechanisms in the model must still be resolved in formulating an ID or MAID model from a causal BN. In organizations or nations seeking to reduce various risks through policies or regulations, who should manage what, which variables should be taken as exogenously determined, and which should be subjected to control must likewise be resolved before ID or MAID models can be formulated and solved to obtain recommended risk management decisions.

### 10.10 Value of Information (VOI), Dynamic Bayesian Networks (DBNs), and Sequential Experiments for Reducing Uncertainties Over Time

Once a causal ID model has been fully quantified, it can be used to predict how the probability distributions for different outcomes of interest (such as net health effect in Fig. 43.11) and expected utility will change if different decisions are

made. This what-if capability, in turn, allows decision optimization algorithms to identify which specific decisions and decision rules maximize expected utility and to calculate how sensitive the recommended decisions are to other uncertainties and assumptions in the model. ID software products such as *Analytic* and *Netica* support construction of IDs and automatically solve them for the optimal decisions. For the example in Fig. 43.11, a robust optimal decision is to not recommend restrictions in fish consumption to consumers, as the estimated health benefits of greater fish consumption far outweigh the estimated health risks. This conclusion is unlikely to be reversed by further reductions in uncertainty, i.e., there is little doubt that it is true. By contrast, whether it is worth lowering allowed levels of pollutants in fish feed is much less clear, with the answer depending on modeling assumptions that are relatively uncertain. This implies a positive *value of information* (VOI) for reducing these uncertainties, meaning that doing so might change the best decision and increase expected utility. ID models can represent the option of collecting additional information before making a final decision about what actions to take, such as lowering or not lowering allowed pollutant levels, by including one or more additional decision nodes to represent information acquisition, followed by chance nodes showing what the additional information might reveal.

In an ID with options for collecting more information before taking a final action, the optimal next step based on presently available information might turn out to be to collect additional information before committing to final regulations or other costly actions. This will be the case if and only if the costs of collecting further information next, including any costs of delay that this entails, are less than the benefits from better-informed subsequent decisions, in the sense that collecting more information before acting (e.g., implementing a regulation or issuing a warning in Fig. 43.11) has greater expected utility than taking the best action now with the information at hand. Optimal delay and information acquisition strategies based on explicit VOI calculations often conflict with more intuitive or political criteria. Both individuals and groups are prone to conclude prematurely that there is already sufficient information on which to act and that further delay and information collection are therefore not warranted, due to narrow framing, overconfidence and confirmation biases, groupthink, and other psychological aspects of decision-making [61]. Politicians and leaders may respond to pressure to exhibit the appearance of strong leadership by taking prompt action without first learning enough about their probable consequences. VOI calculations can help to overcome such well-documented limitations of informal decision-making by putting appropriate weight on the value of reducing uncertainty before acting.

To explicitly model the sequencing of information collection, action selection, and resulting changes in outcomes over time, consecutive period-specific BNs or IDs can be linked by information flows, meaning that the nodes in each period's network (or "slice" of the full multi-period model) are allowed to depend on information received in previous periods. The resulting *dynamic Bayesian networks* (DBNs) or dynamic IDs provide a very convenient framework for predicting and optimizing decisions and consequences over time as initial uncertainties are gradually reduced or resolved. They have proved valuable in medical decision-

making for forecasting in detail the probabilities of different time courses of diseases and related quantities, such as probability of first diagnosis with a disease or adverse condition within a specified number of months or years [121], survival times and probabilities for patients with different conditions and treatments, and remaining days of hospitalization or remaining years of life for individual patients being monitored and treated for complex diseases, from cancers to multiple surgeries to sequential organ failures [5, 101].

DBN estimation software is freely available in R packages [69, 94]. It has been developed and used largely by the systems biology community for interpreting time series of gene expressions in systems biology. Biological and medical researchers, electrical engineers, computer scientists, artificial intelligence researchers, and statisticians have recognized that DBNs generalize important earlier methods of dynamic estimation and inference, such as Hidden Markov Models and Kalman filtering for estimation and signal processing [34]. DBNs are also potentially extremely valuable in a wide range of other engineering, regulatory, policy, and decision analysis settings where decisions and their consequences are distributed over time, where feedback loops or other cycles make any static BN inapplicable, or where detailed monitoring of changing probabilities of events is desired so that midcourse changes in actions can be made in order to improve final outcomes.

Development and application of DBN algorithms and various generalizations are fruitful areas of ongoing applied research. Key concepts of DBNs and multi-agent IDs have been successfully combined to model *multi-agent control* of dynamic random processes (modeled as multi-agent partially observable Markov decision processes, POMDPs) [93]. More recently, DBN methods have been combined with ideas from change-point analysis for situations where arcs in the DAG model are gained or lost at certain times as new influences or mechanisms start to operate or former ones cease [97]. These advances further extend the flexibility and realism of DBN models (and dynamic IDs based on them) to apply to description and control of nonstationary time series.

As already discussed, value of information (VOI) calculations, familiar from decision analysis, can be carried out straightforwardly in ID models. Less familiar, but still highly useful, are methods for optimizing the sequential collection of information to better ascertain correct causal models. The best available methods involve design of experiments [116] and of time series of observations [90]. When the correct ID model describing the relation between decisions and consequence probabilities is initially uncertain, then collecting additional information may have value not only for improving specific decisions (i.e., changing decisions or decision rules to increase expected utility) within the context of a specified ID model but also for discriminating among alternative ID models to better ascertain which ones best describe reality. New information can help in learning IDs from data by revealing how the effects of manipulations develop in affected variables over time [119]. For example, Tong and Koller [116] present a Bayesian approach to sequential experimentation in which a distribution of BN DAG structures and CPTs is updated by experiments that set certain variables to new values and monitor the changes in

values of other variables. At each step, the next experiment to perform is selected to most reduce expected loss from incorrect inferences about the presence and directions of arcs in the DAG model. Even in BNs without decision or utility nodes, designing experiments and time series of observations to facilitate accurate learning of BN descriptions can be very valuable in creating and validating models with high predictive accuracy [90].

---

## 11 Causal Analytics

The preceding sections have discussed how causal Bayesian networks and other DAG and time series algorithms provide constructive methods for carrying out many risk assessment and risk management tasks, even when there is substantial initial uncertainty about relevant cause-and-effect relations and about the best (expected utility-maximizing) courses of action. Other graphical formalisms for risk analysis and decision-making, such as decision trees, game trees, fault trees, and event trees, which have long been used to model the propagation of probabilistic events in complex systems, can all be converted to equivalent IDs or BNs, often with substantial reductions in computational complexity and with savings in the number of nodes and combinations of variable values that must be explicitly represented [107]. Thus, BNs and IDs provide an attractive unifying framework for characterizing, quantifying, and reducing uncertainties and for deciding what to do under the uncertainties that remain. They, together with time series techniques and machine learning techniques, provide a toolkit for using data to inform inference, prediction, and decision-making with realistic uncertainties. These methods empower the following important and widely used types of analytics for using data to inform decisions:

- *Descriptive analytics*: BNs and IDs describe how the part of the world being modeled probably works, showing which factors influence or determine the probability distributions for which other variables and quantifying the probabilistic relations among variables. If a BN or ID has CPTs that represent the operation of lawlike causal mechanisms – i.e., if it is a causal BN or ID – then it can be used to describe how changes in some variables affect the probability distributions of others and hence how probabilistic causal influences propagate to change the probabilities of outcomes.
- *Predictive analytics*: A BN can be used to predict how the probabilities of future observations change when new evidence is acquired (or assumed). A causal BN or ID predicts how changes made at input nodes will affect the future probabilities of outputs. Dynamic Bayesian networks (DBNs) are used to forecast the probable sequences of future changes that will occur after observed changes in inputs, culminating in a new posterior joint probability distribution for all other variables over time (calculated via posterior inference algorithms). BNs and DBNs are also used to predict and compare the probable consequences (changes in probability distributions of outputs and other variables) caused by alternative hypothetical (counterfactual) scenarios for changes in inputs,

including alternative decisions. Conversely, BNs can predict the most likely explanation for observed data, such as the most likely diagnosis explaining observed symptoms or the most likely sequence of component failures leading to a real or hypothesized failure of a complex system. By predicting the probable consequences of alternative policies or decisions and the most likely causes for undesired outcomes, BNs can inform risk management decision-making and help to identify where to allocate resources to repair or forestall likely failure paths.

- *Uncertainty analytics.* Both BNs and IDs are designed to quantify uncertainties about their predictions by using probability distributions for all uncertain quantities. When model uncertainty is important, model ensemble methods allow the predictions or recommendations from multiple plausible models to be combined to obtain more accurate forecasts and better-performing decision recommendations [3]. DBNs provide the ability to track event probabilities in detail as they change over time, and dynamic versions of MAIDs allow uncertainties about the actions of other decision-makers to be modeled.
- *Prescriptive analytics.* If a net benefit, loss, or utility function for different outcomes is defined, and if the causal DAG relating choices to probabilities of consequences is known, then ID algorithms can be used to solve for the best combination of decision variables to minimize expected loss or maximize expected utility. If more than one decision-maker or policy maker makes choices that affect the outcome, then MAIDs or dynamic versions of MAIDs can be used to recommend what each should do.
- *Evaluation and learning analytics.* Ensembles of BNs, IDs, and dynamic versions and extensions of these can be learned from data and experimentation. Value of information (VOI) calculations determine when a single decision-maker in a situation modeled by a known ID should stop collecting information and take action. Dynamic causal BNs and IDs can be learned from time series data in many settings (including observed responses to manipulations or designed experiments) and current decision rules or policies can be evaluated and improved during the learning process, via methods such as low-regret learning with model ensembles, until no further improvements can be found [107]. Learning about causal mechanisms from the observed time series of responses to past interventions, manipulations, decisions, or policies provides a promising technical approach to using past experience to deliberately improve future decisions and outcomes.

Table 43.4 shows how these various components, which might collectively be called *causal analytics*, provide constructive methods for answering the fundamental questions raised in the introduction. For event detection and consequence prediction, DBNs (especially, nonstationary DBNs) and change-point analysis (CPA) algorithms are well suited for detecting changes in time series of observations and occurrences of unobserved events based on their observable effects. DBNs and causal simulation models, as well as time series models that accurately describe how impacts of changes are distributed over time, are also useful for predicting the probable future consequences of recent changes or “shocks” in the inputs to a system.

**Table 43.4** Causal analytics algorithms address fundamental risk management questions under realistic uncertainties

Fundamental questions	Causal analytics algorithms and methods for answering the questions
<i>Event detection: What has changed recently in disease patterns or other adverse outcomes, by how much, when?</i>	<ul style="list-style-type: none"> <li>• Change-point analysis (CPA) algorithms</li> <li>• Dynamic Bayesian networks (DBNs)</li> </ul>
<i>Consequence prediction: What are the implications for what will probably happen next if different actions (or no new actions) are taken?</i>	<ul style="list-style-type: none"> <li>• Dynamic Bayesian networks (DBNs)</li> <li>• Simulation modeling</li> <li>• Time series forecasting</li> </ul>
<i>Risk attribution: What is causing current undesirable outcomes? Does a specific exposure harm human health, and, if so, who is at greatest risk and under what conditions?</i>	<ul style="list-style-type: none"> <li>• Causal DAG models (e.g., BNs, IDs)</li> <li>• Ensembles of DAG models</li> <li>• Granger causality and transfer entropy (TE) for time series</li> </ul>
<i>Response modeling: What combinations of factors affect health outcomes, and how strongly? How would risks change if one or more of these factors were changed?</i>	<ul style="list-style-type: none"> <li>• Causal DAG models, e.g., BN models</li> </ul>
<i>Decision making: What actions or interventions will most effectively reduce uncertain health risks?</i>	<ul style="list-style-type: none"> <li>• Influence diagram (ID) algorithms</li> <li>• MAIDs for multiple decision makers</li> <li>• Adaptive learning methods, e.g., iqLearn, if the ID model is uncertain</li> </ul>
<i>Retrospective evaluation and accountability: How much difference have exposure reductions or other policies actually made in reducing adverse health outcomes?</i>	<ul style="list-style-type: none"> <li>• Quasi-experimental (QE) studies</li> <li>• Intervention analysis for time series</li> <li>• Ensemble learning algorithms such as iqLearn for continuous improvement</li> </ul>

For risk attribution, causal graph models (such as BNs, IDs, and dynamic versions of these) or ensembles of such models can be learned from data and used to quantify the evidence that suspected hazards indeed cause the adverse effects attributed to them (i.e., that there is, with high confidence, a directed arc pointing from a node representing exposure to a hazard into a node representing the effect). If so, the CPT for the effect node quantifies how changes in the exposure node change probabilities of effects, given the levels of other causes with which exposure may interact. Multivariate response modeling, in which the joint distributions of one or more responses vary with the levels of one or more factors that probabilistically cause them, can readily be modeled by DAGs that include the different causal factors and effects. For risk management or regulation under uncertainty, if utility nodes and decision nodes are incorporated into the causal graph models to create known causal ID or MAID models, then the best decisions for risk management (i.e., for inducing the greatest achievable expected utilities) can be identified by well-developed ID solution algorithms, and VOI calculations can be used to optimize costly information collection and the timing of final decisions.

Finally, for retrospective evaluation and accountability, quasi-experiments and intervention analysis of interrupted time series provide traditional methods of analysis, although they require using data (or assumptions) to refute noncausal explanations for changes in time series. More recently developed ensemble-learning methods [3, 107] and adaptive learning algorithms (such as iqLearn for learning to optimize treatment sequences) can be used to continually evaluate and improve the success of current decision rules, policies, or regulations for managing uncertain risks, based on their performance to date and on relative expected costs of switching among them and of failing to do so. Such adaptive evaluation and improvement is possible provided that the consequences of past actions (probably) are monitored and the data are made available and used to update causal IDs, MAIDs, or dynamic versions of such models to allow ongoing learning and optimization. Thus, causal graph methods (including ensemble methods, when appropriate models are uncertain, and time series methods that uncover DAG structures relating time series variables) provide a rich set of tools for addressing fundamental challenges of uncertainty quantification and decision-making under uncertainty.

---

## 12 Summary and Conclusions: Applying Causal Graph Models to Better Manage Risks and Uncertainties

The power and maturity of the technical methods in Table 43.4 have spurred their rapid uptake and application in fields such as neurobiology, systems biology, econometrics, artificial intelligence, control engineering, game theory, signal processing, and physics. However, they have so far had relatively limited impact on the practice of uncertainty quantification and risk management in epidemiology, public health, and regulatory science, perhaps because these fields give great deference to the use of subjective judgments informed by weight-of-evidence considerations – an approach widely used and taught since the 1960s, but of unproved and doubtful probative value [83]. Previous sections have illustrated some of the potential of more modern methods of causal analytics, but the vast majority of applied work in epidemiology, public health, and regulatory risk assessment unfortunately still uses older association-based methods and subjective opinions about the extent to which statistically significant differences between risk model coefficients for differently exposed populations might have causal interpretations.

To help close the gap between these poor current practices and the potentially much more objective, reliable, accurate, and sensitive methods of causal analytics in Table 43.4, the following checklist may prove useful in judging the adequacy of policy analyses or quantitative risk assessments (QRAs) that claims to have identified useful predictive causal relations between exposures to risk factors or hazards and resulting risks of adverse effects (responses), i.e., causal exposure-response (E-R) relations.

1. *Does the QRA show that changes in exposures precede the changes in health effects that they are said to cause? Are results of appropriate technical analyses (e.g., change-point analyses, intervention analyses and other quasi-experimental comparisons, and Granger causality tests or transfer entropy results) presented, along with supporting data? If effects turn out to precede their presumed causes, then unmeasured confounders or residual confounding by confounders that the investigators claim were statistically “controlled for” may be at work.*
2. *Does the QRA demonstrate that health effects cannot be made conditionally independent of exposure by conditioning on other variables, especially potential confounders? Does it present the details, data, and results of appropriate statistical tests (e.g., conditional independence tests and DAGs) showing that health effects and exposures share mutual information that cannot be explained away by any combination of confounders?*
3. *Does the QRA present and test explicit causal graph models, showing the results of formal statistical tests of the causal hypotheses implied by the structure of the model (i.e., which variables point into which others)? Does it identify which alternative causal graph models are most consistent with available data (e.g., using the Occam’s Window method of [78])? Most importantly, does it present clear evidence that changes in exposure propagate through the causal graph, causing successive measurable changes in the intermediate variables along hypothesized causal paths? Such coherence, consistency, and biological plausibility demonstrated in explicit causal graph models showing how hypothesized causal mechanisms dovetail with each other to transduce changes in exposures to changes in health risks can provide compelling objective evidence of a causal relation between them, thus accomplishing what older and more problematic WoE frameworks have long sought to provide [95].*
4. *Have noncausal explanations for statistical relations among observed variables (including exposures, health effects, and any intermediate variables, modifying factors, and confounders) been explicitly identified and convincingly refuted using well-conducted and reported statistical tests? Especially, have model diagnostics (e.g., plots of residuals and discussions of any patterns) and formal tests of modeling assumptions been presented that show that the models used appropriately describe the data to which the QRA applies them and that claimed associations are not caused by model selection biases or specification errors, failures to model errors in exposure estimates and other explanatory variables, omitted confounders or other latent variables, uncorrected multiple testing bias, or coincident historical trends (e.g., spurious regression, if the exposure and health effects time series in longitudinal studies are not stationary)?*
5. *Have all causal mechanisms postulated in the QRA modeling been demonstrated to exhibit stable, uniform, lawlike behavior, so that there is no substantial unexplained heterogeneity in estimated input-output (e.g., E-R or C-R) relations? If the answer is no, then missing factors may need to be identified and their effects modeled before valid predictions can be made based on the assumption that changes in causes will yield future changes in effects that can be well described and predicted based on estimates of cause-effect relations from past data.*



If the answers to these five diagnostic questions are all yes, then the QRA has met the burden of proof of showing that the available data are consistent with a causal relation and that other (noncausal) explanations are not plausible. It can then proceed to quantify the estimated changes in probability distributions of outputs, such as future health effects, that would be caused by changes in controllable inputs (e.g., future exposure levels) using the causal models developed to show that exposure causes adverse effects. The effort needed to establish valid evidence of a causal relation between historical levels of inputs and outputs by being able to answer yes to questions 1–5 pays off at this stage. Causal graph models (e.g., Bayesian networks with validated causal interpretations for their CPTs), simulation models based on composition of validated causal mechanisms, and valid path diagrams and SEM causal models can all be used to predict quantitative changes in outputs that would be caused by changes in inputs, e.g., changes in future health risks caused by changes in future exposure levels, given any scenario for the future values of other inputs.

Conversely, if the answer to any of the preceding five diagnostic questions is no, then it is premature to make causal predictions based on the work done so far. Either the additional work needed to make the answers yes should be done or results should be stated as contingent on the as-yet unproved assumption that this can eventually be done.

---

## Cross-References

► [Rare-Event Simulation](#)

---

## References

1. Alpiste Illueca, F.M., Buitrago Vera, P., de Grado Cabanilles, P., Fuenmayor Fernandez, V., Gil Loscos, F.J.: Periodontal regeneration in clinical practice. *Med. Oral Patol. Oral Cir. Bucal.* **11**(4), e3:82–e3:92 (2006)
2. Angrist, J.D., Pischke, J.-S.: *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton (2009)
3. Ashcroft, M.: Performing decision-theoretic inference in Bayesian network ensemble models In: Jaeger, M., Nielsen, T.D., Viappiani, P. (eds.) *Twelfth Scandinavian Conference on Artificial Intelligence*, Aalborg, vol. 257, pp. 25–34 (2013)
4. Arnold, A., Liu, Y., Abe, N.: Temporal causal modeling with graphical Granger methods. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD-07)*, San Jose, 12–15 Aug 2007. ACM, New York. <http://dl.acm.org/citation.cfm?id=1281192&picked=prox>
5. Azhar, N., Ziraldo, C., Barclay, D., Rudnick, D.A., Squires, R.H., Vodovotz, Y., Pediatric Acute Liver Failure Study Group: Analysis of serum inflammatory mediators identifies unique dynamic networks associated with death and spontaneous survival in pediatric acute liver failure. *PLoS One.* **8**(11), e78202 (2013). doi:10.1371/journal.pone.0078202
6. Bai, Z., Wong, W.K., Zhang, B.: Multivariate linear and nonlinear causality tests. *Math. Comput. Simul.* **81**(1), 5–17 (2010)
7. Barnett, L., Seth, A.K.: The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J. Neurosci. Methods* **223** (2014)

8. Barr, C.D., Diez, D.M., Wang, Y., Dominici, F., Samet, J.M.: Comprehensive smoking bans and acute myocardial infarction among Medicare enrollees in 387 US counties: 1999–2008. *Am. J. Epidemiol.* **176**(7), 642–648 (2012). Epub 17 Sep 2012
9. Brenner E, Sontag D. (2013) SparsityBoost: a new scoring function for learning Bayesian network structure. In: 29th Conference on Uncertainty in Artificial Intelligence (UAI2013). Westin Bellevue Hotel, Washington, DC, 11–15 July 2013. <http://auai.org/uai2013/prints/papers/30.pdf>
10. Callaghan, R.C., Sanches, M., Gatley, J.M., Stockwell, T.: Impacts of drinking-age laws on mortality in Canada, 1980–2009. *Drug Alcohol Depend.* **138**, 137–145 (2014). doi:10.1016/j.drugalcdep.2014.02.019
11. Cami, A., Wallstrom, G.L., Hogan, W.R.: Measuring the effect of commuting on the performance of the Bayesian Aerosol Release Detector. *BMC Med. Inform. DecisMak.* **9**(Suppl 1), S7 (2009)
12. Campbell, D.T., Stanley, J.C.: *Experimental and Quasi-experimental Designs for Research*. Rand McNally, Chicago (1966)
13. Chang, K.C., Tian, Z.: Efficient inference for mixed Bayesian networks. In: Proceedings of the Fifth International Conference on Information Fusion, Annapolis, vol. 1, pp 527–534, 8–11 July 2002. IEEE. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1021199>
14. Christensen, T.M., Møller, L., Jørgensen, T., Pisinger, C.: The impact of the Danish smoking ban on hospital admissions for acute myocardial infarction. *Eur. J. PrevCardiol.* **21**(1), 65–73 (2014). doi:10.1177/2047487312460213
15. Corani, G., Antonucci, A., Zaffalon, M.: Bayesian networks with imprecise probabilities: theory and application to classification. In: Holmes, D.E., Jaim, C. (eds.) *Data Mining: Foundations and Intelligent Paradigms*. Intelligent Systems Reference Library, vol. 23, pp. 49–93 (2012)
16. Cox, L.A. Jr., Popken, D.A.: Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States? *Ann. Epidemiol.* **25**(3), 162–173 (2015)
17. Cox, L.A. Jr., Popken, D.A., Berman, D.W.: Causal versus spurious spatial exposure-response associations in health risk analysis. *Crit. Rev. Toxicol.* **43**(Suppl 1), 26–38 (2013)
18. Crowson, C.S., Schenck, L.A., Green, A.B., Atkinson, E.J., Therneau, T.M.: The basics of propensity scoring and marginal structural models. Technical report #84, 1 Aug 2013. Department of Health Sciences Research, Mayo Clinic, Rochester. <http://www.mayo.edu/research/documents/biostat-84-pdf/doc-20024406>
19. Dash, D., Druzdzel, M.J.: A note on the correctness of the causal ordering algorithm. *Artif. Intell.* **172**, 1800–1808 (2008). <http://www.pitt.edu/~druzdzel/psfiles/aij08.pdf>
20. De Campos C.P., Ji, Q.: Efficient structure learning of Bayesian networks using constraints. *J. Mach. Learn. Res.* **12**, 663–689 (2011)
21. Dominici, F., Greenstone, M., Sunstein, C.R.: Science and regulation. Particulate matter matters. *Science.* **344**(6181), 257–259 (2014). doi:10.1126/science.1247348
22. The Economist: Trouble at the Lab: scientists like to think of science as self-correcting. To an alarming degree, it is not. [www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble](http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble), 19 Oct 2013
23. Eichler, M., Didelez, V.: On Granger causality and the effect of interventions in time series. *Lifetime Data Anal.* **16**(1), 3–32 (2010). Epub 26 Nov 2009. <http://www.ncbi.nlm.nih.gov/pubmed/19941069>
24. EPA (U.S. Environmental Protection Agency): The Benefits and Costs of the Clean Air Act from 1990 to 2020. Final Report – Rev. A. Office of Air and Radiation, Washington, DC (2011)
25. EPA: Expanded expert judgment assessment of the concentration-response relationship between PM2.5 exposure and mortality. [www.epa.gov/ttn/ecas/regdata/Uncertainty/pm\\_ee\\_report.pdf](http://www.epa.gov/ttn/ecas/regdata/Uncertainty/pm_ee_report.pdf) (2006)
26. Exarchos, K.P., Goletsis, Y., Fotiadis, D.I.: A multiscale and multiparametric approach for modeling the progression of oral cancer. *BMC Med. Inform. DecisMak.* **12**, 136 (2012). doi:10.1186/1472-6947-12-136.

27. Ezzati, M., Hoom, S.V., Lopez, A.D., Danaei, G., Rodgers, A., Mathers, C.D., Murray, C.J.L.: Comparative quantification of mortality and burden of disease attributable to selected risk factors. In: Lopez, A.D., Mathers, C.D., Ezzati, M., Jamison, D.T., Murray, C.J.L. (eds.) *Global Burden of Disease and Risk Factors*, chapter 4. World Bank, Washington, DC (2006)
28. Fann, N., Lamson, A.D., Anenberg, S.C., Wesson, K., Risley, D., Hubbell, B.J.: Estimating the national public health burden associated with exposure to ambient PM<sub>2.5</sub> and Ozone. *Risk Anal.* **32**(1), 81–95 (2012)
29. Ferson, S., Donald, S.: Probability bounds analysis. In: Mosleh, A., Bari, R.A. (eds.) *Probabilistic Safety Assessment and Management*, pp. 1203–1208. Springer, New York (1998)
30. Ferson, S., Hajagos, J.G.: Arithmetic with uncertain numbers: rigorous and (often) best possible answers. In: Helton, J.C., Oberkampf, W.L. (eds.) *Alternative Representations of Epistemic Uncertainty*. *Reliability Engineering & System Safety*, vol. 85, pp. 135–152; 1–369 (2004)
31. Freedman, D.A.: Graphical models for causation, and the identification problem. *Eval. Rev.* **28**(4), 267–293 (2004)
32. Friedman, N., Goldszmidt, M.: Learning Bayesian networks with local structure. In: Jordan, M.I. (ed.) *Learning in Graphical Models*, pp. 421–459. MIT, Cambridge (1998)
33. Gasparini, A., Gorini, G., Barchielli, A.: On the relationship between smoking bans and incidence of acute myocardial infarction. *Eur. J. Epidemiol.* **24**(10), 597–602 (2009)
34. Ghahramani, Z.: Learning dynamic Bayesian networks. In: Giles, C.L., Gori, M. (eds.) *Adaptive Processing of Sequences and Data Structures*. International Summer School on Neural Networks "Caianiello, E.R." Vietri sul Mare, Salerno, 6–13 Sept 1997. Tutorial Lectures. *Lecture Notes in Computer Science*, vol. 1387 (1998). <http://link.springer.com/sbook/10.1007/BFb0053992>, <http://link.springer.com/bookseries/558>, <http://mlg.eng.cam.ac.uk/zoubin/SAL/D/learnDBNs.pdf> (1997)
35. Greenland, S.: Epidemiologic measures and policy formulation: lessons from potential outcomes. *Emerg. Themes Epidemiol.* **2**, 5 (2005)
36. Greenland, S., Brumback, B.: An overview of relations among causal modelling methods. *Int. J. Epidemiol.* **31**(5), 1030–1037 (2002). <http://www.ncbi.nlm.nih.gov/pubmed/12435780>
37. Gruber, S., Logan, R.W., Jarrín, I., Monge, S., Hernán, M.A.: Ensemble learning of inverse probability weights for marginal structural modeling in large observational datasets. *Stat. Med.* **34**(1), 106–117 (2015)
38. Grundmann, O.: The current state of bioterrorist attack surveillance and preparedness in the US. *Risk Manag. Health Policy.* **7**, 177–187 (2014)
39. Hack, C.E., Haber, L.T., Maier, A., Shulte, P., Fowler, B., Lotz, W.G., Savage, R.E., Jr.: A Bayesian network model for biomarker-based dose response. *Risk Anal.* **30**(7), 1037–1051 (2010)
40. Harris, A.D., Bradham, D.D., Baumgarten, M., Zuckerman, I.H., Fink, J.C., Perencevich, E.N.: The use and interpretation of quasi-experimental studies in infectious diseases. *Clin. Infect Dis.* **38**(11), 1586–1591 (2004)
41. Harris, A.D., McGregor, J.C., Perencevich, E.N., Furuno, J.P., Zhu, J., Peterson, D.E., Finkelstein, J.: The use and interpretation of quasi-experimental studies in medical informatics. *J. Am. Med. Inform. Assoc.* **13**(1), 16–23 (2006)
42. Harvard School of Public Health: 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](http://www.hsph.harvard.edu/news/press-releases/archives/2002-releases/press10172002.html) (2002)
43. Health Effects Institute (HEI): Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes. HEI Research Report #148 (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>

44. Health Effects Institute (HEI): Did the Irish Coal Bans Improve Air Quality and Health? HEI Update. <http://pubs.healtheffects.org/getfile.php?u=929> (Summer, 2013). Last Retrieved 1 Feb 2014
45. Helfenstein, U.: The use of transfer function models, intervention analysis and related time series methods in epidemiology. *Int. J. Epidemiol.* **20**(3), 808–815 (1991)
46. Hernán, M.A., Taubman, S.L.: Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int. J. Obes. (Lond.)* **32**(Suppl 3), S8–S14 (2008)
47. Hibbs, D.A., Jr.: On analyzing the effects of policy interventions: Box-Jenkins and Box-Tiao vs. structural equation models. *Sociol. Methodol.* **8**, 137–179 (1977). <http://links.jstor.org/sici?sici=0081-1750%281977%298%3C137%3AOATEOP%3E2.0.CO%3B2-K>
48. Hipel, K.W., Lettenmaier, D.P., McLeod, I.: Assessment of environmental impacts part one: *Interv. Anal. Environ. Manag.* **2**(6), 529–535 (1978)
49. Hites, R.A., Foran, J.A., Carpenter, D.O., Hamilton, M.C., Knuth, B.A., Schwager, S.J.: Global assessment of organic contaminants in farmed salmon. *Science.* **303**(5655), 226–229 (2004)
50. Hoeting, J., Madigan, D., Raftery, A., Volinsky, C.: Bayesian model averaging. *Stat. Sci.* **14**, 382–401 (1999)
51. Höfler, M.: The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg. Themes Epidemiol.* **2**, 11 (2005)
52. Homer, J., Milstein, B., Wile, K., Trogdon, J., Huang, P., Labarthe, D., et al.: Simulating and evaluating local interventions to improve cardiovascular health. *Prev. Chronic Dis.* **7**(1), A18 (2010). [www.cdc.gov/pcd/issues/2010/jan/08\\_0231.htm](http://www.cdc.gov/pcd/issues/2010/jan/08_0231.htm). Accessed 3 Nov 2015
53. Hora, S.: Eliciting probabilities from experts. In: Edwards, W., Miles, R.F., von Winterfeldt, D. (eds.) *Advances in Decision Analysis: From Foundations to Applications*, pp. 129–153. Cambridge University Press, New York (2007)
54. Hoyer, P.O., Hyvärinen, A., Scheines, R., Spirtes, P., Ramsey, J., Lacerda, G., Shimizu, S.: Causal discovery of linear acyclic models with arbitrary distributions. In: *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence - UAI, Helsinki, Conference held 9–12 July 2008*, pp. 282–289. <http://arxiv.org/ftp/arxiv/papers/1206/1206.3260.pdf>
55. Huitema, B.E., Van Houten, R., Manal, H.: Time-series intervention analysis of pedestrian countdown timer effects. *Accid Anal Prev.* **72**, 23–31 (2014). doi:10.1016/j.aap.2014.05.025
56. Ioannidis, J.P.A.: Why most published research findings are false. *PLoS Med.* **2**(8), e124 (2005). doi:10.1371/journal.pmed.0020124
57. James, N.A., Matteson, D.S.: ecp: an R package for nonparametric multiple change point analysis of multivariate data. *J. Stat. Softw.* **62**(7) (2014). <http://www.jstatsoft.org/v62/i07/paper>
58. Janzing, D., Balduzzi, D., Grosse-Wentrup, M., Scholkopf, B.: Quantifying causal influences. *Ann. Stat.* **41**(5), 2324–2358 (2013). doi:10.1214/13-AOS1145
59. Jiang, H., Livingston, M., Manton, E.: The effects of random breath testing and lowering the minimum legal drinking age on traffic fatalities in Australian states. *Inj. Prev.* **21**(2), 77–83 (2015). doi:10.1136/injuryprev-2014-041303
60. Joffe, M., Gambhir, M., Chadeau-Hyam, M., Vineis, P.: Causal diagrams in systems epidemiology. *Emerg. Themes Epidemiol.* **9**(1), 1 (2012). doi:10.1186/1742-7622-9-1
61. Kahneman, D.: *Thinking, Fast and Slow*. Farrar, Straus, and Giroux, New York (2011)
62. Kass-Hout, T.A., Xu, Z., McMurray, P., Park, S., Buckeridge, D.L., Brownstein, J.S., Finelli, L., Groseclose, S.L.: Application of change point analysis to daily influenza-like illness emergency department visits. *J. Am. Med. Assoc.* **19**(6), 1075–1081 (2012). doi:10.1136/amiajnl-2011-000793
63. Kinnunen, E., Junttila, O., Haukka, J., Hovi, T.: Nationwide oral poliovirus vaccination campaign and the incidence of Guillain-Barré Syndrome. *Am. J. Epidemiol.* **147**(1), 69–73 (1998)
64. Kleck, G., Britt, C.L., Bordua, D.: The emperor has no clothes: an evaluation of interrupted time series designs for policy impact assessment. *J. Firearms Public Policy* **12**, 197–247 (2000)

65. Klein, L.R.: Regression systems of linear simultaneous equations. In: *A Textbook of Econometrics*, 2nd edn, pp. 131–196. Prentice-Hall, Englewood Cliffs (1974). ISBN:0-13-912832-8
66. Kline, R.B.: *Principles and Practice of Structural Equation Modeling*. Guilford Press, New York (1998)
67. Koller, D., Milch, B.: Multi-agent influence diagrams for representing and solving games. In: *Proceedings of the 17th International Joint Conference on Artificial Intelligence* (2001)
68. Lagarde, M.: How to do (or not to do) . . . Assessing the impact of a policy change with routine longitudinal data. *Health Policy Plan.* **27**(1), 76–83 (2012). doi: 10.1093/heapol/czr004.
69. Lebre, S.: Package 'G1DBN': a package performing dynamic Bayesian network inference. CRAN repository, 19 Feb 2015. <https://cran.r-project.org/web/packages/G1DBN/G1DBN.pdf>
70. Lehrer, J.: Trials and errors: why science is failing us. *Wired*. <http://www.wired.co.uk/magazine/archive/2012/02/features/trials-and-errors?page=all>, 28 Jan 2012
71. Lei, H., Nahum-Shan, I., Lynch, K., Oslin, D., Murphy, S.A.: A “SMART” design for building individualized treatment sequences. *Ann. Rev. Clin. Psychol.* **8**, 14.1–14.28 (2012)
72. Linn, K.A., Laber, E.B., Stefanski LA.: iqLearn: interactive Q-learning in R. <https://cran.r-project.org/web/packages/iqLearn/vignettes/iqLearn.pdf> (2015)
73. Lipsitch, M., Tchetgen Tchetgen, E., Cohen, T.: Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* **21**(3), 383–388 (2010)
74. Lizier, J.T.: JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. *Front. Robot. AI* **1**, 11 (2014); doi:10.3389/frobt.2014.00011 (preprint: arXiv:1408.3270), <http://arxiv.org/pdf/1408.3270.pdf>
75. Lu, C.Y., Soumerai, S.B., Ross-Degnan, D., Zhang, F., Adams, A.S.: Unintended impacts of a Medicaid prior authorization policy on access to medications for bipolar illness. *Med Care.* **48**(1), 4–9 (2010). doi:10.1097/MLR.0b013e3181bd4c10.
76. Lynch, W.D., Glass, G.V., Tran, Z.V.: Diet, tobacco, alcohol, and stress as causes of coronary artery heart disease: an ecological trend analysis of national data. *Yale J. Biol. Med.* **61**(5), 413–426 (1988)
77. Maclure, M.: Taxonomic axes of epidemiologic study designs: a refutationist perspective. *J. Clin. Epidemiol.* **44**(10), 1045–1053 (1991)
78. Madigan, D., Raftery, A.: Model selection and accounting for model uncertainty in graphical models using Occam’s window. *J. Am. Stat. Assoc.* **89**, 1535–1546 (1994)
79. Madigan, D., Andersson, S.A., Perlman, M.D., Volinsky, C.M.: Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Commun. Stat. Theory Methods* **25**, 2493–2519 (1996)
80. McLeod et al. (2011) Time series analysis with R. <http://www.stats.uwo.ca/faculty/aim/tsar/tsar.pdf>
81. Montalto, A., Faes, L., Marinazzo, D.: MuTE: a MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy. *PLoS One* **9**(10), e109462 (2014). doi:10.1371/journal.pone.0109462
82. Moore, K.L., Neugebauer, R., van der Laan, M.J., Tager, I.B.: Causal inference in epidemiological studies with strong confounding. *Stat Med.* (2012). doi:10.1002/sim.4469
83. Morabia, A.: Hume, Mill, Hill, and the sui generis epidemiologic approach to causal inference. *Am. J. Epidemiol.* **178**(10), 1526–1532 (2013)
84. Morriss, R., Gask, L., Webb, R., Dixon, C., Appleby, L.: The effects on suicide rates of an educational intervention for front-line health professionals with suicidal patients (the STORM project). *Psychol. Med.* **35**(7), 957–960 (2005)
85. Nakahara, S., Katanoda, K., Ichikawa, M.: Onset of a declining trend in fatal motor vehicle crashes involving drunk-driving in Japan. *J. Epidemiol.* **23**(3), 195–204 (2013)
86. Neugebauer, R., Fireman, B., Roy, J.A., Raebel, M.A., Nichols, G.A., O’Connor, P.J.: Super learning to hedge against incorrect inference from arbitrary parametric assumptions in marginal structural modeling. *J. Clin. Epidemiol.* **66**(8 Suppl):S99–S109 (2013). doi:10.1016/j.jclinepi.2013.01.016

87. Nguefack-Tsague, G.: Using Bayesian networks to model hierarchical relationships in epidemiological studies. *Epidemiol. Health* **33**, e2011006 (2011). doi:10.4178/epih/e2011006. Epub 17 Jun 2011. <http://e-epih.org/journal/view.php?doi=10.4178/epih/e2011006>
88. Nuzzo, R.: Scientific method: statistical errors. P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume. *Nature* **506**, 150–152 (2014). doi:10.1038/506150a
89. Owczarek, T.: On modeling asymmetric multi-agent scenarios. In: *IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Rende (Cosenza)*, 21–23 Sept 2009
90. Page, D., Ong, I.M.: Experimental design of time series data for learning from dynamic Bayesian networks. *Pac. Symp. Biocomput.* **2006**, 267–278 (2006)
91. Papan, A., Kyrtsov, C., Kugiumtzis, D., Cees, D.: Detecting causality in non-stationary time series using partial symbolic transfer entropy: evidence in financial data. *Comput. Econ.* **47**(3), 341–365 (2016). <http://link.springer.com/article/10.1007%2Fs10614-015-9491-x>
92. Pearl, J.: An introduction to causal inference. *Int. J. Biostat.* **6**(2), Article 7 (2010). doi:10.2202/1557-4679.1203
93. Polich, K., Gmytrasiewicz, P.: Interactive dynamic influence diagrams. In: *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*. ACM, New York. Article No. 34. <http://dl.acm.org/citation.cfm?id=1329166>
94. Rau, A.: Package 'ebdbNet': empirical Bayes estimation of dynamic Bayesian networks. CRAN repository, 19 Feb 2015. <https://cran.r-project.org/web/packages/ebdbNet/ebdbNet.pdf>
95. Rhomberg, L.: Hypothesis-based weight of evidence: an approach to assessing causation and its application to regulatory toxicology. *Risk Anal.* **35**(6), 1114–1124 (2015)
96. Robins, J.M., Hernán, M.A., Brumback, B.: Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**(5), 550–560 (2000)
97. Robinson, J.W., Hartemink, A.J.: Learning non-stationary dynamic Bayesian networks. *J. Mach. Learn. Res.* **11**, 3647–3680 (2010)
98. Rothman, K.J., Lash, L.L., Greenland, S.: *Modern Epidemiology*, 3rd edn. Lippincott, Williams, & Wilkins, New York (2012)
99. Runge, J., Heitzig, J., Petoukhov, V., Kurths, J.: Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.* **108**, 258701. Published 21 June 2012
100. Samet, J.M., Bodurow, C.C. (eds.): *Improving the Presumptive Disability Decision-Making Process for Veterans*. Committee on Evaluation of the Presumptive Disability Decision-Making Process for Veterans, Board on Military and Veterans Health, Institute of Medicine. National Academies Press, Washington, DC (2008)
101. Sandri, M., Berchiolla, P., Baldi, I., Gregori, D., De Blasi, R.A.: Dynamic Bayesian networks to predict sequences of organ failures in patients admitted to ICU. *J. Biomed. Inform.* **48**, 106–113 (2014). doi:10.1016/j.jbi.2013.12.008
102. Sarewitz, D.: Beware the creeping cracks of bias. *Nature* **485**, 149 (2012)
103. Sarewitz, D.: Reproducibility will not cure what ails science. *Nature* **525**(7568), 159 (2015)
104. Schwartz, J., Austin, E., Bind, M.A., Zanolotti, A., Koutrakis, P.: Estimating causal associations of fine particles with daily deaths in Boston. *Am. J. Epidemiol.* **182**(7), 644–650 (2015)
105. Scutari, M.: Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* **35**(3) (2010). [www.jstatsoft.org/v35/i03/paper](http://www.jstatsoft.org/v35/i03/paper). Last accessed 5 May 2015
106. Shen, Y., Cooper, G.F.: A new prior for Bayesian anomaly detection: application to biosurveillance. *Methods Inf. Med.* **49**(1), 44–53 (2010)
107. Shoham, Y., Leyton-Brown, K.: *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge (2010)
108. Skrvøseth, S.O., Bellika, J.G., Godtliebsen, F.: Causality in scale space as an approach to change detection. *PLoS One.* **7**(12), e52253 (2012). doi:10.1371/journal.pone.0052253
109. Stebbings, J.H., Jr.: Panel studies of acute health effects of air pollution. II. A methodologic study of linear regression analysis of asthma panel data. *Environ. Res.* **17**(1), 10–32 (1978)

110. Steck, H.: Learning the Bayesian network structure: Dirichlet prior versus data. In: Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI2008), University of Helsinki City Centre Campus, Helsinki, 9–12 July 2008
111. Sun, X.: Assessing nonlinear granger causality from multivariate time series. *Mach. Learn. Knowl. Discov. Databases. Lect. Notes Comput. Sci.* **5212**, 440–455 (2008)
112. Swanson, S.A., Hernán, M.A.: How to report instrumental variable analyses (suggestions welcome). *Epidemiology* **24**(3), 370–374 (2013)
113. Tashiro, T., Shimizu, S., Hyvärinen, A., Washio T.: ParceLiNGAM: a causal ordering method robust against latent confounders. *Neural Comput.* **26**(1), 57–83 (2014)
114. Taubman, S.L., Allen, H.L., Wright, B.J., Baicker, K., Finkelstein, A.N.: Medicaid increases emergency-department use: evidence from Oregon’s health insurance experiment. *Science*. **343**(6168), 263–268 (2014). doi:10.1126/science.1246183
115. Thornley, S., Marshall, R.J., Wells, S., Jackson, R.: Using directed acyclic graphs for investigating causal paths for cardiovascular disease. *J. Biomet. Biostat.* **4**, 182 (2013). doi:10.4172/2155-6180.1000182
116. Tong, S., Koller, D.: Active learning for structure in Bayesian networks. In: International Joint Conference on Artificial Intelligence (IJCAI), Seattle (2001)
117. Twardy, C.R., Nicholson, A.E., Korb, K.B., McNeil, J.: Epidemiological data mining of cardiovascular Bayesian networks. *J. Health Inform.* **1**(1), e3:1–e3:13 (2006)
118. Vicente, R., Wibral, M., Lindner, M., Pipa, G.: Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **30**(1), 45–67 (2011)
119. Voortman, M., Dash, D., Druzdzel, M.J.: Learning causal models that make correct manipulation predictions with time series data. In: Guyon, I., Janzing, D., Schölkopf, B. (eds.) *JMLR Workshop and Conference Proceedings*, vol. 6, pp. 257–266. NIPS 2008 Workshop on Causality. <http://jmlr.csail.mit.edu/proceedings/papers/v6/voortman10a/voortman10a.pdf> (2008)
120. Wang, J., Spitz, M.R., Amos, C.I., et al.: Method for evaluating multiple mediators: Mediating effects of smoking and COPD on the association between the CHRNA5-A3 Variant and Lung Cancer Risk. de Torres JP, ed. *PLoS One.* **7**(10), e47705 (2012). doi:10.1371/journal.pone.0047705
121. Watt, E.W., Bui, A.A.: Evaluation of a dynamic Bayesian belief network to predict osteoarthritic knee pain using data from the osteoarthritis initiative. *AMIA Annul. Symp. Proc.* **2008**, 788–792 (2008)
122. Wen, X., Rangarajan, G., Ding, M.: Multivariate Granger causality: an estimation framework based on factorization of the spectral density matrix. *Philos. Trans. R. Soc. A* **371**, 20110610 (2013). <http://dx.doi.org/10.1098/rsta.2011.0610>
123. Zhang, N.L.: Probabilistic inference in influence diagrams. *Comput. Intell.* **14**, 475–497 (1998)