Robins J M 1997 Causal inference from complex longitudinal data. In: Berkane M (ed.) *Latent Variable Modeling with Applications to Causality*. Springer-Verlag, New York, pp. 69–117

Rubin D B 1974 Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**: 688–701

Rubin D B 1978 Bayesian inference for casual effects: The role of randomization. *Annals of Statistics* **6**: 34–58

Shafer G 1996 *The Art of Causal Conjecture*. MIT Press, Cambridge, MA

P. W. Holland

Causal Inference and Statistical Fallacies

1. Generalities

The pairing of causality and fallacies may seem idiosyncratic. In fact it nicely captures the point that many statistical fallacies, i.e., plausible-seeming arguments that give the wrong conclusion, hinge on the overinterpretation or misinterpretation of statistical associations as implying more than they properly do. The article begins by discussing three main views of causality, briefly indicating the scope for fallacious arguments and then at the end returns to discuss some fallacies in slightly more detail. See *Graphical Models: Overview*.

The very long history in the philosophical literature of discussions of causality is largely irrelevant for these purposes. It typically regards a cause as necessary and sufficient for an effect: all smokers get lung cancer, all lung cancer patients smoke. Here the concern is with situations with multiple causes, even if one is predominant, and where explicit or implicit statistical or probabilistic considerations are needed.

2. Notions of Causality

2.1 Causality as Stable Association

Suppose that a study, or preferably several different but related studies, shows that two features, C and R, of the individuals (people, firms, communities, households, etc.) under investigation are associated. That is, if we take, to be explicit, positive monotone association, individuals with high values of C tend to have high values of R and vice versa. For example C and R might be test scores at a given age in arithmetic and language, or level of crime and unemployment rate in a community.

Under what circumstances might one reasonably conclude that C is a cause of a response R, or at least make some step in the direction of that conclusion? And what would such a statement of causality mean?

2.1.1 Symmetric and directed relations.
Association is a symmetric relation between two, or possibly more, features. Causality is not symmetric. That is, if C is associated with R then R is associated with C, but if C is a cause of R then R is not a cause of C. Thus the first task, given any two features C and R, is to distinguish the cases where:

(a) C and R are to be regarded as in some sense on an equal footing and treated in a conceptually symmetric way in any interpretation.

(b) One of the variables, say C, is to be regarded as explanatory to the other variable, R, regarded as a response. That is, if there is a relation, it is regarded asymmetrically.

Often significance tests for the existence of association and of dependency are identical. The distinction being studied here is a substantive one of interpretation. Failure to observe this distinction leads to the fallacy of the overinterpreted association.

2.1.2 Graphical representation.
A useful graphical representation shows two variables X_1 and X_2, regarded on an equal footing, if associated, as connected by an undirected edge, whereas two variables such that C is explanatory to R, if connected, are done so by a directed edge. See Fig. 1a and Fig. 1b.

There are two possible bases for the distinction between explanatory and response variables. One is that features referring to an earlier time point are explanatory to features referring to a later time point. The second is a subject-matter working hypothesis based for example on theory or on empirical data from other kinds of investigation. Thus the weight of a child at one year is a response to maternal smoking behavior during pregnancy. In such situations the relevant time is not the time when the observation is made but the time to which the features refer, although of course observations recorded retrospectively are especially subject to recall biases.

As an example of the second type of explanatory-response relation, suppose that data are collected on diabetic patients assessing their knowledge of the disease and of their success in managing their disease, as measured by glucose control. These data may well refer to the same time point and it is not inconceivable that, for example, patients with poor glucose control are thereby encouraged to learn more about their disease. Nevertheless, as a working hypothesis, one might interpret the data assuming that knowledge, C, is explanatory to glucose control, R, considered as a response. This is represented in simple graphical form in Fig. 1b by the directed edge from C to R.
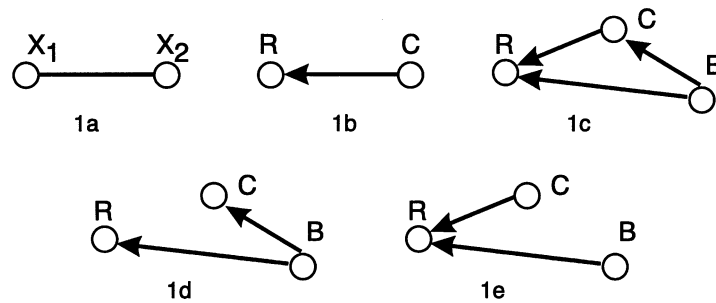
Figure 1

(a) Undirected edge between two variables X_1, X_2 on an equal footing. (b) Directed edge between explanatory variable C and response variable R. (c) General dependence of response R on B, C. (d) Special situation with $R \perp\!\!\!\perp C \mid B$. (e) Special situation with $B \perp\!\!\!\perp C$ corresponding in particular to randomization of C

In summary the first step towards causality is to require good reasons for regarding C as explanatory to R as a response and that any notion of causal connection between C and R, and there may be none, is that C is a cause of R, not the other way round.

We may talk about the 'fallacy of the incorrect direction' when the explanatory-response relation is identified in the wrong direction.

2.1.3 Common explanatory variables. Next, consider the possibility of one or more common explanatory variables. For this, suppose that B is potentially explanatory to C and hence also to R. There are a number of possibilities of which the most general is shown in Fig. 1c with directed edges from B to C, from C to R, and also directly from B to R. On the other hand, if the relation were that represented schematically in Fig. 1d, the only dependence between C and R is that induced by their both depending on B. Then C and R are said to be conditionally independent given B, sometimes conveniently written $R \perp\!\!\!\perp C \mid B$. There is no direct path from C to R that does not pass via B. Such relations are typically assessed empirically by some form of regression analysis. In such a situation, one would not regard C as a cause of R, even though in an analysis without the background variable B there is a statistical dependence between the two.

This discussion leads to one definition used in the literature of C being a cause of R, namely that there is a dependence between C and R and that the sign of that dependence is unaltered whatever variables B_1, B_2 etc., themselves explanatory to C, are considered simultaneously with C as possible sources of dependence. This definition has a long history but is best articulated by I. J. Good and P. Suppes. A corresponding notion for time series is due to N. Wiener and C. W. J. Granger. This definition underlies much empirical statistical analysis in so far as it aims towards causal explanation.

The definition entertains all possible alternative explanatory variables. In implementation via an observational study one can at best check that the measured background variables B do not account for the dependence between C and R. The possibility that the dependence could be explained by variables explanatory to C that have not been measured, i.e., by so-called unobserved confounders, is less likely the larger the apparent effect, and can be discounted only by general plausibility arguments about the field in question. Sensitivity analysis may be helpful in this: that is, it may be worth calculating what the properties of an unobserved confounder would have to be to explain away the dependence in question. For further details see Rosenbaum (1995) and the entry *Observational Studies: Overview*.

Mistaken conclusions reached via neglect of confounders, observed or unobserved, may be called 'fallacies of the neglected confounder.'

2.1.4 Role of randomization. The situation is substantially clarified if the potential explanatory variable C is a randomized treatment allocated by the investigator. Then in the scheme sketched in Fig. 1e there can be no edge between the B's and C since such dependence would be contrary to randomization, i.e., to each individual under study being equally likely to receive each treatment possibility. In this situation an apparent dependence between C and R cannot be explained by a background variable as in Fig. 1d. It is in this sense that it is sometimes stated, especially in the statistical literature, that causality can be inferred from randomized experiments and not from observational studies. It is argued here that while other things being equal, randomized experiments are greatly to be preferred to
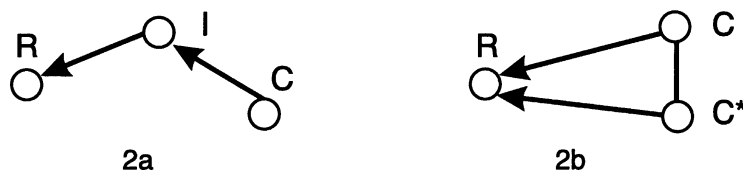
Figure 2
(a) Intermediate variable I accounting for overall effect of C after ignoring I; $R \perp\!\!\!\perp C \mid I$. (b) Correlated variables C, C^* on an equal footing and both explanatory to response R

observational studies, difficulties of interpretation, sometimes serious, remain.

2.1.5 Intermediate variables. In Sect. 2.1.4 the variables B have been supposed explanatory to C and hence to R. For judging a possible causal effect of C it would be wrong to consider in the same way variables intermediate between C and R, i.e., variables I that are responses to C and explanatory to R. They are valuable in clarifying the nature of any indirect path between C and R, but the use of I in a regression analysis of R and C would not be correct in assessing whether such a path exists. If R is independent of C given an intermediate variable I, but dependent on I, then C may still have caused I and I may be a cause of R.

For instance suppose that C represents assignment to a new medical regimen as compared with a control regimen and that the former, but not the latter, eventually induces lower blood pressure, I, which in turn induces a reduced cardiac event rate, R; see Fig. 2a. Does the new regime cause a reduced cardiac event rate? If R is conditionally independent of C given I, it would be reasonable to say that the regimen does cause a reduction in R and that this reduction appears to be explained via improved blood-pressure control.

2.1.6 Explanatory variables on an equal footing. In some ways an even more delicate situation arises if we consider the role of variables C^* on an equal footing with the variable C whose causal status is under consideration; see Fig. 2b. If the role of C is essentially the same whether or not C^* is conditioned, i.e., whether or not C^* is included in the regression equation, there is no problem, at least at a qualitative level. On the other hand, consider the relatively common situation where there is clear dependence on (C, C^*) as a pair but that either variable on its own is sufficient to explain the dependence. There are then broadly three routes to interpretation:

(a) To regard (C, C^*) collectively as the possibly causal variables.

(b) To present at least two possibilities for interpretation, one based on C and one on C^*.

(c) To obtain further information clarifying the relation between C and C^*, establishing for instance

that C^* is explanatory to C and that the appropriate interpretation is to fix C^* when analysing variations of C.

For example, suppose that C and C^* are respectively measures of educational performance in arithmetic and language of a child, both measured at the same age, and that the response is some adult feature. Then the third possibility is inapplicable; the first possibility is to regard the variables as a two-dimensional measure of educational performance and to abandon, at least temporarily, any notion of separating the role of arithmetic and language.

In summary, this first broad notion of causality is that of a statistical dependency that cannot be explained away via an eligible alternative explanation.

2.2 Causality as the Effect of Intervention

2.2.1 Counterfactuals. While the notion of causality discussed in Sect. 2.1.6 is certainly important and is strongly connected with the approach adopted in many empirical statistical studies, it does not, however, directly capture a stronger interpretation of the word causal. This is connected with the idea of hypothetical intervention or modification. Suppose for simplicity of exposition that C takes just two possible forms, to be called presence and absence. Thus presence might be the implementation of some program of intervention, and absence a suitable control state. For monotone relations, one may say that the presence of C causes an increase in the response R if an individual with C present tends to have a higher R than that same individual would have had if C had been absent, other things being equal.

Slightly more explicitly, let B denote all variables possibly explanatory to C and suppose that there are no variables C^* to be considered on an equal footing to C. Consider for each individual two possible values of R, R_{pres}, R_{abs} that would arise as C takes on its two possible values, present and absent, and B is fixed. Then presence of C causes, say, an increase in R if R_{pres} is in some sense systematically greater than R_{abs}.

We now discuss this notion, which has its origins at least in part in J. Neyman's and R. A. Fisher's work on design of experiments and in the studies of H. A. Simon and has been systematically studied and fruitfully applied by D. B. Rubin.

For a given individual, only one of R_{pres} and R_{abs} can be observed, corresponding to the value of C actually holding for that individual. The other value of R is a so-called counterfactual whose introduction seems, however, essential to capture the notion hinted at above of a deeper meaning of causality.

2.2.2 Differences in counterfactuals. The simplest and least demanding relation between the two values of R is that over some populations of individuals under study the average of R_{pres} exceeds that of R_{abs}. This is a notion of an average effect and is testable empirically in favorable circumstances. A much stronger requirement is that the required inequality holds for every individual in the population of concern. Stronger still is the requirement that the difference between the two values of R is the same for all individuals, i.e., that for all individuals

$$R_{\mathrm{pres}} - R_{\mathrm{abs}} = \Delta$$

This is called, in the language of the theory of the design of experiments, the assumption of unit-treatment additivity.

Now these last two assumptions are clearly not directly testable and can be objected to on that account. The assumptions are indirectly testable, to a limited extent at least. If the individuals are divided into groups, for example on the basis of one or more of the background variables B, the assumptions imply for each individual observed that the difference between the two levels of R has the same sign in the first case and that it is the same, except for sampling errors, in the second. In Sect. 2.2.3 the consequences of the possibly causal variable C having a very different effect on different individuals are discussed.

2.2.3 Ignorable treatment allocation. In Sect. 2.2.2 it has been tacitly assumed throughout that the two possible values of R for each individual depend on the value of C for that individual and would be unaffected by reallocation of C to other individuals. That is, the effects of C act independently as between different individuals. These considerations have a strong bearing on the appropriate definition of a unit of study. For example, in a comparison of different methods of class teaching of children, the unit would primarily be a whole class of students, i.e., the whole group who are taught and work together.

2.2.4 Intrinsic variables. There is an important restriction implicit in this discussion. It has to be meaningful in the context in question to suppose that C

for an individual might have been different from how it in fact is. This is relevant only to variables that appear solely as explanatory variables. For example they may be variables measured at base-line, i.e., at entry into a study. Any intermediate variable by its nature of being at some point a response is potentially manipulable. Purely explanatory variables can be divided into intrinsic variables, essentially defining characteristics of the individual, and potential explanatory variables, which might therefore play the role of C in the present discussion. Intrinsic variables should not be regarded as even potentially causal in the present sense. For example the gender of an individual would in most contexts be regarded as an intrinsic characteristic. The question 'what would R have been for this woman had she been a man, other things being held fixed?' is in many, although not quite all, contexts meaningless.

2.2.5 Variables to be held fixed. Finally, care is essential in defining what is to be held fixed under hypothetical changes of C. Certainly responses I to C are not fixed. Variables, B, explanatory to C are held fixed. There is an essential ambiguity for variables C^* on an equal footing with C. This is strongly connected with the issue of which explanatory variables to include in empirical regression analyses.

2.3 Causality as Explanation of a Process

There is a third notion of causality that is in some ways more in line with normal scientific usage. This is that there is some understanding, albeit provisional, of the process that leads from C to R. This understanding typically comes from theory, or often from knowledge at a hierarchical level lower than the data under immediate analysis. Sometimes, it may be possible to represent such a process by a graph without directed cycles and to visualize the causal effect by the tracing of paths from C to R via variables I intermediate between C and R. Thus the effect of diet in an epidemiological study may be related to the physiological processes underlying the disease under study, the effect of pharmaceutical products related to the pharmaco-dynamics of their action, the effect of interventions at a community level related to ideas of individual psychology, and so on.

This last notion of causality as concerned with generating processes is to be contrasted with the second view of causality as concerned with the effects of intervention and with the first view of causality as stable statistical dependence. These views are complementary, not conflicting. Goldthorpe (1998) has argued for this third view of causality as the appropriate one for sociology, with explanation via

1557

Rational Choice Theory as an important route for interpretation.

To be satisfactory there needs to be evidence, typically arising from studies of different kinds, that such generating processes are not merely hypothesized.

3. Special Issues

3.1 Interaction Involving a Potentially Causal Variable

We now turn to the issue of interactions with a potentially causal variable. The graphical representations used in Sect. 2.3 to show the structure of various kinds of dependency and independency holding between a set of variables have the limitation, at least in the form used here, that they do not represent interaction, in particular that an effect of C may be systematically different for different groups of individuals.

This affects the role in analysis and interpretation especially of variables B that are themselves possibly explanatory to the variable C whose causal status is under consideration. So far we have been largely concerned with whether such variables could explain the effect on response of C. In more detailed discussion we should consider the possibility that the effect of C is substantially different at different levels of B. For example if B is an intrinsic feature such as gender, we consider whether the effect of C is different for men and for women. In particular, if the effects of C are in opposite directions for different levels of B we say there is a qualitative interaction, a possibility of special importance for interpretation.

Note especially, that even when C represents a randomized treatment which is automatically decoupled from preceding variables B, the possibility of serious interactions with B cannot in general be ignored; see Sect. 5.1.

Viewed slightly differently, absence of interaction is important not only in simplifying interpretation but also in enhancing generalizability and specificity. That is, an effect that has been shown to have no serious interaction with a range of potential variables is more likely to be reproduced in some new situation and more likely to have a stable subject-matter interpretation.

3.2 Unwanted Unobserved Intermediate Variable

Consider further the role of variables I referring to time points after the implementation of C. A subject-matter distinction can be drawn between on the one hand intermediate variables that are responses to C

and that are explanatory to R and are part of some natural process, and on the other hand interventions into the system that may depend on C and which may be explanatory to R, but which in some sense are unwanted or inappropriate for interpretation. In the context of clinical trials, an example is the failure of patients to comply with a treatment assigned to them. Ignoring such noncompliance can lead to inappropriate intention-to-treat analysis.

Another example is in evaluations of study programs, whenever students in only one of the programs receive intensive encouragement during the evaluation period.

3.3 Aggregation

So far, little has been said about the choice of observational units for study. At a fundamental research level it may be wise to choose individuals showing the effects of interest in their simplest and most striking form. More generally, however, the choice has to be considered at two levels. There is the level at which ultimate interpretation and action is required and the level at which careful observation of largely decoupled individuals is available. For example, a criminologist comparing different sentencing or policing policies is interested in individual offenders but may be able to observe only different communities or policing areas. A nutritional epidemiologist comparing different diets is interested in individual people but may have to rely, in part at least, on consumption and mortality data from whole countries. The assumption that a dependence established on an aggregate scale, for example at a country level, has a similar interpretation at a small-scale level, for example for individual persons, involves the assumption that there are no confounders B at a person level that would account for the apparent dependency. This will typically be very hard or even impossible to check at all carefully from country-level data.

Errors rising as a result of over-aggregated units of analysis are called 'ecological fallacies' or in econometrics 'aggregation biases.' See *Ecological Inference*.

4. Bradford Hill's Conditions

The above discussion implicitly emphasizes that, while causal understanding is the aim of perhaps nearly all research work, a cautious approach is essential, especially, but by no means only, in observational studies. The most widely quoted conditions tending to make a causal interpretation more likely are those of Bradford Hill (1965), put forward in connection with the interpretation of epidemiological studies. Bradford Hill emphasized their tentative character.

For a critical discussion of these conditions, see Rothman and Greenland (1999) and for a slightly revised version of them Cox and Wermuth (1996, Sect. 8.7).

Koch gave conditions for inferring causality when the potential cause can be applied, withdrawn, and reapplied in a relatively controlled way and the pattern of response observed.

5. Some Fallacies in More Detail

The previous discussion has mentioned various points at which fallacious arguments are not only possible, but relatively common. This in no way covers the wealth of fallacious arguments possible in a statistical context, perhaps the most pervasive being the comparison of rates using inappropriate or ill-defined denominators. There is not space to discuss all these possibilities. This article is therefore concluded with three specific examples related to the main discussion of causality.

5.1 Inappropriate Reference Set for Probabilities

The first fallacy to be discussed, unlike the others, does not center around misuse of notions connected with causality, but rather with mathematically correct but inappropriate calculations of probability. Cornfield (1969) discussed a criminal case in California, People versus Collins, in which a couple had been found guilty partly on the basis of the following argument. Eye-witnesses described the crime as committed by a man with a moustache, by an Afro-American man with a beard, by an inter-racial couple in a car, by a couple in a partly yellow car, by a girl with blond hair, and by a girl with a ponytail.

Probabilities were assigned to these six features and, assuming independence, multiplied to give a probability $P = 0.8 \times 10^{-7}$, the smallness of this being

argued as evidence of guilt. The California Court of Appeal rejected this argument as inappropriate, primarily because they regarded the issue as being whether there could be one or more further matching couples in the Los Angeles area. For this they argued in effect that if N is the population of the greater Los Angeles area then the number of couples with the assigned properties has a Poisson distribution of mean NP and, since one such couple is known to exist, the number has a zero-truncated Poisson distribution with parameter NP. From this the Court of Appeal calculated a probability of about 0.4 of there being one or more further couples with matching features, too high to justify a safe conviction of the original couple.

A more formal argument would use Bayes's theorem to calculate the posterior probability of guilt, assuming that the only evidence is that stated above and that the numerical assignments are reasonably accurate.

5.2 Interactive Effect Involving an Unobserved Explanatory Variable

It has been reported (Zivin and Choi 1991) that early controlled clinical trials with stroke patients were discouraging because they appeared to give differing answers to the question: do thrombolytic agents effectively improve the status of stroke patients? These are substances to dissolve blood clots, i.e., thrombi. Nowadays it is known that a stroke can be caused by a thrombus or by a burst vein. Thus, a thrombolytic agent may improve the patient's status or it may worsen it considerably, depending on the reason for the stroke. With patients of both types in any study 'treatment-unit' additivity will not hold, nor can there be 'strongly ignorable treatment allocation,' even if the study is a controlled clinical trial with random allocation of patients to one group treated with a thrombolytic agent and the other with a placebo.

Instead, observations like those in Table 1 are to be anticipated. The main response is success of treatment

Table 1
Counts, percentages and odds-ratios; the two explanatory variables are independent, $C \perp\!\!\!\perp B$, due to randomized allocation of treatments, C, to patients; depending on the patient's unobserved status, B, the thrombolitic agent has a very different chance of treatment success

| R, Success of treatment | B = 1, Burst vein | | B = 2, Thrombus | |
| | C, Thrombolytic agent | | C, Thrombolytic agent | |
	yes	no	yes	no
yes	6 (2 percent)	60 (20 percent)	1425 (95 percent)	300 (20 percent)
no	294	240	75	1200
sum	300	300	1500	1500
odds-ratio	0.08		16	

Table 2
Counts, percentages and the same association between R and C as measured in terms of relative chances given each level combination of A and B; strong three-factor interaction between A, B, C; response depends on each of the explanatory variables A, B, C

	$A = 1$				$A = 1$			
	$B = 1$		$B = 2$		$B = 1$		$B = 2$	
	$C = 1$	$C = 2$	$C = 1$	$C = 2$	$C = 1$	$C = 2$	$C = 1$	$C = 2$
$R = 1$	604	40	300	2000	301	2015	600	40
	(30 percent)	(20 percent)	(75 percent)	(50 percent)	(75 percent)	(50 percent)	(30 percent)	(20 percent)
$R = 2$	1396	160	100	2000	99	1985	1400	160
Sum	2000	200	400	4000	400	4000	2000	200

Relative chances for $R = 1$ comparing $C = 1$ to $C = 2$ given A, B

1.5	1.5	1.5	1.5

Table 3
Counts, percentages and relative chances obtained from Table 2 by summing over the levels of A show seemingly replicated associations for R and C given B; the actual associations between R and C, as shown in the previous table, appear reversed and $(R, C) \amalg B$

	$B = 1$		$B = 2$	
	$C = 1$	$C = 2$	$C = 1$	$C = 2$
$R = 1$	905	2055	900	2040
	(38 percent)	(49 percent)	(38 percent)	(49 percent)
$R = 2$	1495	2145	1500	2160
Sum	2400	4200	2400	4200

Relative chances for $R = 1$ comparing $C = 1$ to $C = 2$ given B

0.78	0.78

and the main explanatory variable is the type of treatment. In this example there is a ratio of 1:5 for the number of stroke patients with a burst vein to those with a thrombus. With successful randomization the patient's status will be independent of treatment, but the strong interactive effect of status and treatment on outcome cannot be avoided. Whenever it is not feasible to observe the patient's status, then the reported results of any clinical trial will strongly depend on the actual percentage of patients in the study having a thrombus as cause of the stroke. In any case, nowadays it would be unethical to include stroke patients known to have a burst vein in a clinical trial designed to study a thrombolytic agent.

5.3 Dependence Reversal

The artificial data in Table 2 illustrate that a strong three-way interaction among explanatory variables A, B, C can lead to replicated dependence reversals whenever the response R depends on each of the explanatory variables.

Here, treatment $C = 1$ is consistently better under four different conditions, since the chances for successful treatment $R = 1$ are higher for $C = 1$ if compared with $C = 2$; relative chances are even identical in the four conditions. But, this treatment appears to be consistently worse under two replications $B = 1$ and $B = 2$, i.e., when A is unobserved. In addition, in an analysis of only the three variables R, B, C, it appears as if randomization had been used successfully since $C \amalg B$.

These effects are an extreme example of what is often called the Yule-Simpson paradox. Although in a sense the word paradox is inappropriate, the possibility of this kind of dependence reversal reinforces the need for either carefully studying relations among explanatory variables or for using effective randomization procedures.

If a statistical association is to be judged as evidence for a causal hypothesis, then one should be certain that the observed associations do not mislead us about the actual associations. This is impossible without assumptions about unobserved variables even in trials

with randomization (see also Stone 1993). Therefore it appears that substantial progress in establishing causes can be expected only via understanding and description of the processes which generate observed effects. See *Linear Hypothesis: Fallacies and Interpretive Problems (Simpson's Paradox)*.

6. Suggested Further Reading

The statistical aspects of causality are best approached via the discussion paper of Holland (1986), where in particular, key references to the earlier literature will be found; see also Cox and Wermuth (1996, Sect. 8.7). For general issues about observational studies, see Cochran (1965) and Rosenbaum (1995). For a philosophical perspective, see Simon (1972) and Cartwright (1989). For an interventionist view, see Rubin (1974) and for a more formal analysis still from a social science viewpoint Sobel (1995). For a development based on directed acyclic graphs, see Pearl (2000) and for the general connections with graph theory Lauritzen (2000). For an approach based on a complete specification of all independencies between a set of variables, followed by a computer-generated listing of all directed acyclic graphs consistent with those independencies, see Spirtes et al. (1993). The use of counterfactuals is criticized by Dawid (2000). The reader should be aware that many rather different interpretations of causality are involved in these discussions.

An elementary account of fallacies is given, with many interesting examples, by Huff (1954). Good (1978) gives a detailed classification of types of fallacy, again with excellent examples. See also Agresti (1983).

See also: Causation (Theories and Models): Conceptions in the Social Sciences; Causation: Physical, Mental, and Social; Explanation: Conceptions in the Social Sciences; Scientific Reasoning and Discovery, Cognitive Psychology of

Bibliography

Agresti A 1983 Fallacies, statistical. In: Kotz S, Johnson N L (eds.) *Encyclopedia of Statistical Sciences*. Wiley, New York, Vol. 3, pp. 24–8

Bradford Hill A 1965 The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine* **58**: 295–300

Cartwright N 1989 *Nature's Capacities and their Measurement*. Clarendon Press, Oxford, UK

Cochran W G 1965 The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society* A **128**: 234–265

Cornfield J 1969 The Bayesian outlook and its application (with discussion). *Biometrics* **25**: 617–57

Cox D R, Wermuth N 1996 *Multivariate Dependencies—Models, Analysis and Interpretation*. Chapman and Hall, London

Dawid A P 2000 Causality without counterfactuals (with discussion). *Journal of the American Statistical Association* **95**: 407–48

Goldthorpe J G 1998 *Causation, Statistics and Sociology*. Economic and Social Research Institute, Dublin, Republic of Ireland

Good I J 1978 Fallacies, statistical. In: Kruskal W H, Tanur J M (eds.) *Encyclopedia of Statistics*. Free Press, New York, Vol. 1, pp. 337–49

Holland P 1986 Statistics and causal inference (with discussion). *Journal of the American Statistical Society* **81**: 945–70

Huff D 1954 *How to Lie with Statistics*. Norton, New York

Lauritzen S L 2000 Causal inference from graphical models. In: Klüppelberg C, Barndorff-Nielsen O E, Cox D R (eds.) *Complex Stochastic Systems*. Chapman and Hall/CRC, London

Pearl J 2000 *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, UK

Rosenbaum P R 1995 *Observational Studies*. Springer, New York

Rothman K J, Greenland S (eds.) 1998 *Modern Epidemiology*, 2nd edn. Raven-Lippincott, Philadelphia, PA

Rubin D B 1974 Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Studies* **66**: 688–701

Simon H A 1972 Causation. In: Kruskal W H, Tanur J M (eds.) *Encyclopedia of Statistics*. Free Press, New York, Vol. 1, pp. 35–41

Sobel M E 1995 Causal inference in the social and behavioral sciences. In: Arminger G, Clogg C C, Sobel M E (eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Plenum, New York

Stone R 1993 The assumptions on which causal inferences rest. *Journal of the Royal Statistical Society* B **55**: 455–66

Spirtes P, Glymour C, Scheines R 1993 *Causation, Prediction and Search*. Springer, New York

Zivin J A, Choi D W 1991 Neue Ansätze zur Schlaganfall-Therapie. *Spektrum der Wissenschaft* **Sept**: 58–66

D. R. Cox and N. Wermuth

Causation (Theories and Models): Conceptions in the Social Sciences

Many, perhaps most problems and hypotheses in social and behavioral research concern *causal relations*. What caused the fall of communism? What caused Peter's depression? What causes aggression in general? Atkinson and Birch (1978, p. 30) consider as one of the fundamental questions in the study of motivation: 'What causes the strength of tendencies to change?' However, the idea or concept of causation is also involved in many cases in which people do not use the word 'cause.' Many concepts include the idea of causation as part of their meaning, for example, a

1561