International Statistical Review (2004), **72,** 3, 285–305, Printed in Wales by Cambrian Printers

Causality: a Statistical View

D.R. Cox[1] and Nanny Wermuth[2]

[1]*Nuffield College, Oxford, UK* [2]*University of Gothenburg, Sweden*

Summary

Statistical aspects of causality are reviewed in simple form and the impact of recent work discussed. Three distinct notions of causality are set out and implications for densities and for linear dependencies explained. The importance of appreciating the possibility of effect modifiers is stressed, be they intermediate variables, background variables or unobserved confounders. In many contexts the issue of unobserved confounders is salient. The difficulties of interpretation when there are joint effects are discussed and possible modifications of analysis explained. The dangers of uncritical conditioning and marginalization over intermediate response variables are set out and some of the problems of generalizing conclusions to populations and individuals explained. In general terms the importance of search for possibly causal variables is stressed but the need for caution is emphasized.

Key words: Chain block graph; Counterfactual; Explanation; Instrumental variable; Interaction; Markov graph; Observational study; Overview; Regression analysis; Surrogate variable; Unit-treatment additivity; Unobserved confounder.

1 Introduction

Statisticians concerned with the interpretation of their analyses have implicitly always been interested in causality even if they have been sparing in the use of the word. Thus Yule (1900) emphasized, especially in a time series context, the distinction between correlation and causation. Fisher (1926, 1935) showed that randomization could yield causal inference about treatment effects in which uncertainty could be assessed probabilistically on the basis of the randomization without special assumptions about the structure of the uncontrolled variation.

Cochran (1965) gave a penetrating discussion of many aspects of the analysis of observational studies and in particular pointed to the need to extend Sewall Wright's path analysis to address issues of possible causality, thus anticipating the thrust of much recent work. In addition Cochran quoted Fisher's reply to a question that Cochran had asked him about how to make observational studies more likely to yield causal answers: the answer was "Make your theories elaborate". This might be achieved in various ways, for example by assembling evidence of different types or by obtaining somewhat similar evidence under a wide range of conditions.

Hill (1965) gave guidelines. Satisfaction of some or all of them would strengthen the case for causality inferred from observational studies; he did not state explicitly what he meant by the term causal, although it seems very likely that it was what is termed below first-level causality. Although formulated in an epidemiological context his guidelines are widely relevant. He emphasized that they were indeed guidelines not criteria.

Box (1966) stressed the care needed in giving in effect a causal interpretation to regression equations fitted to observational data. While his illustration was set in a chemical engineering context the argument was again of broad applicability.

Rubin (1974), in an influential paper, adapted notions of causality from the design of experiments to observational studies via a representation similar to Fisher's which, without the essential element

of physical randomization, had been given by Neyman (1923). Subsequently Rubin developed and applied these ideas notably in social science contexts. His and much other previous work is best approached through the review paper of Holland (1986).

Cox & Snell (1981, pp. 84, 85), in an elementary account of regression, outlined five different interpretations of regression equations and coefficients. One was to examine the effect of imposed changes in one or more variables and the care needed, especially in observational studies, in specifying what is held fixed under the imposed changes was emphasized.

Robins, in a long series of papers, in effect explores notions of causality in a clinical trial and epidemiological setting. For problems where treatments or interventions are applied in sequence, see, for example, Robins (1997) and in more detail van der Laan & Robins (2002).

Rosenbaum (2002) has given a searching discussion of the conceptual and methodological issues involved in the analysis of observational studies.

The above work can be regarded as in a main-stream statistical tradition. In this the central idea is that of regression analysis, taken in a very general sense as meaning the study of the dependence of one or more response variables on explanatory variables. The key issues are broadly as follows:

- to choose an appropriate general form of regression relation
- to determine which explanatory variables can legitimately be included in the relation additional to those that have a potentially causal interpretation
- to examine possible nonlinear and interactive effects that may be central to correct interpretation
- to combine evidence from several studies.

There are some situations where causality is clear. The effect may be large and the consequence of a major perturbation of the system or may be firmly related to long and broad experience or to well-established theory. Our discussion, however, is largely focused on situations where establishing causality is more delicate, either because the effect under study is small or because of the possibility of competing explanations of the data. Freedman (2003) has warned against overinterpretation of statistical analyses, giving examples especially from epidemiology and sociology; see also Dempster (1988). Doll (2002) has emphasized that causality can be inferred from empirical epidemiological studies but that considerable care is needed if the effect is only a modest one.

There are many examples where successful search for a causal effect has involved a chain of studies of different types. It might start with the observation of several unusual events, followed by retrospective and prospective studies and evidence from other sources, for example animal studies in a human health context. One prominent example concerns a particular malformation of the eye. It was first noted by an Australian physician (Gregg, 1941) as a common feature of several newborns with this malformation that the mothers had been in early pregnancy during the height of a rubella epidemic. It took a large number of additional studies to establish that the malformation can only occur if the mother had not been exposed to rubella before the pregnancy and then only if she had been in contact with rubella during the first three months of pregnancy. Major reports, in particular on health issues, such as that of the U.S. Surgeon-General (U.S. Department of Health, Education and Welfare, 1964) concluding that smoking is a cause of lung cancer, are typically based on a wide range of evidence.

Deterministic notions of causality have a long history. More probabilistic notions of causality have received much recent attention in the philosophical and computer science literature on knowledge and belief systems and in particular there is both the important early work of Spirtes *et al.* (1993), for a review of which see, for example, Scheines (1997), and a book by Pearl (2000) summarizing and extending his earlier work. This work is in a sense more formalized than most of the statistical ideas summarised above and one of the aims of the present paper is to examine the relation between the two strands of work; see especially Section 4.2.

2 Some Definitions

We now sketch three different notions of causality. It is important to distinguish causality as a property of the physical or biological or social world from its representation in statistical models. We aim for statistical models that permit interpretations in the former sense; to call such models causal models is, however, potentially misleading.

We start with a view of causality, to be called here zero-level causality, and used often in the statistical literature. This is a statistical association, i.e. non-independence, with clearly established ordering from cause to response and which cannot be removed by conditioning on *allowable* alternative features. A crucial aspect concerns the term allowable. For example, in assessing the possible causal effect of an intervention on the occurrence of a cardiac event, blood pressure three months after starting treatment would not be an allowable conditioning feature because it itself may be affected by the intervention under study.

What is termed here the zero-level view of causality was studied by Good (1961, 1962) and comprehensively developed by Suppes (1970) and in a time-series context by Granger (1969) and in a more general stochastic process formulation by Schweder (1970) and by Aalen (1987).

We next introduce a different formulation, to be called first-level causality. This broad approach seems most immediately relevant in many applications of concern to statisticians.

For this, faced with two or more possible interventions in a system, we may aim to compare the outcomes that would arise under the different interventions. For example, consider two possible medical interventions, C_1 and C_0, a new treatment and a control, only one of which can be used on a particular patient. We aim to compare the outcome observed, say with C_1. with the outcome that would have been observed on that patient had C_0 been used, *other things being equal*. Evidence of a systematic difference would be evidence that use of C_1 rather than C_0 causes a change in outcome. This viewpoint may have a decision-making objective although this is by no means necessary. For example, when considering whether an anomalous gene causes some disease, the intervention as between the abnormal and normal version of the gene is hypothetical and moreover no immediate decision-making process is typically involved. This definition of causality is explicitly comparative.

One of the delicate aspects of this formulation is that it is most immediately formulated as concerning individuals but its verification and often its real meaning involve aggregate or statistical issues, i.e. involve average effects over some set of individuals. In that case explicit specification of a reference population of individuals may be important.

Finally we introduce what we name second-level causality. In a scientific context suppose that careful design and analysis have established a pattern of dependencies or associations or have provided reasonable evidence of first- or zero-level causality. The question then arises of explaining how these dependencies or associations arose. What underlying generating process was involved, i.e. what is underlying the structure observed? Often this will involve incorporating information from many different sources, for example in a physical science context establishing connections with basic principles of classical or quantum physics and perhaps between observational and laboratory-scale observations. Goldthorpe (1998) has argued for such a broad notion of causality also in sociology and Hoover (2002) in macroeconomics. A methodological distinction between epidemiological and sociological research is that in the former the possible causal effect of specific risk factors is often of concern as a potential base for public health recommendations. In sociological work interest may often lie in the whole process linking say parental socio-economic class and individual life-features.

In all fields, explanations via a generating process are inevitably to some extent provisional and the process hardly lends itself to very formal characterization. In this view it is important to distinguish between different types of explanation. Some are merely hypothesized, and these can be a valuable preliminary and a source of stimulating research questions. Others are reasonably solidly evidence-based. Moreover some such evidence-based explanations are formulated before the examination of some data to be analysed, whereas others may be retrospectively constructed in the light of that

analysis. The former are typically more immediately convincing and the latter will often call for independent confirmation. This view of causality does not imply a notion of ultimate causation; any proposed generating process may itself have a further explanation at some deeper level.

Use of terms varies substantially between individuals and fields. Nevertheless the notion of evidence-based process seems to correspond broadly but not exclusively to usage in the natural sciences. The first-level notion seems, however, to be most frequently involved in statistical work, especially in such fields as epidemiology with a relatively applied focus. For further discussion of these distinctions, see, for example, Cox (1992), Holland (1986) and Cox & Wermuth (1996, pp. 219–227).

Because of the need for care in interpretation, it is often convenient to use the following terminology. We call C a candidate cause if it makes sense in the context in question to consider C as a possible cause of R, for example in the sense of level-one causality. We call C a potential cause if there is evidence of a possibly causal effect, for example that the notional responses to alternative levels, for example C_1 and C_0, are systematically different. We omit the qualification potential when the evidence is convincing that there is no alternative explanation, and especially when the developmental process is well understood. We use this cautious approach not to discourage the search for causality, but rather to rule out the possibility that real associations can be deemed causal merely by naming them so.

A referee has pointed out a possible connection with the notions of Suppes (1970) of prima facie, genuine and spurious causes. The first of these corresponds broadly to what we have called possible and potential causes. The third of Suppes's types deals with variables whose possible causal effect is explained via other allowable variables.

Many of the essential points in the paper are concerned with putting into perspective the three different views of causality outlined above and with showing them in a framework of probability models. For this it is for most purposes enough to consider a system with four variables measured on each individual, a response R, an intermediate variable I, the potential causal variable C and a background variable B; see Figure 1 for a graphical representation in which we suppose each of the four variables to have two components.

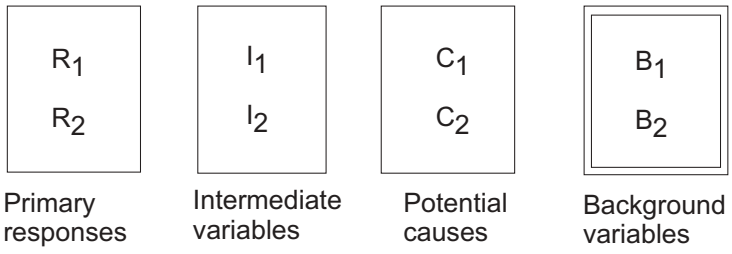

Figure 1. *Graphical representation with four types of variable. In statistical analysis the background variables B, shown in a double-lined box, would usually be considered conditionally on their observed values.*

The role of I will be discussed later in the paper and will not contribute to the first part of the discussion. We have, however, introduced it in the initial formulation because of its conceptual importance. A primary role of B is to specify what is held fixed under notional changes of the variable C.

It is assumed that the variables can at the start be arranged so that a joint probability distribution

is defined recursively. In a simplified notation for densities we write

$$f_{RICB} = f_{R|ICB}\, f_{I|CB}\, f_{C|B}\, f_B. \tag{1}$$

To ignore I, i.e. marginalize over it, we integrate (1) over all values i of I.

3 Level-one Causality

Level-one causality (Rubin, 1974) involves for the simple situation of Section 2 the idea that for each individual there are two notional responses R_1 and R_0 depending on whether C_1 or C_0 is used. Only one of these notional responses can be observed and the other thus is in principle not observable and therefore called a counterfactual. This formulation is combined with an assumption that any difference between R_1 and R_0 is systematic, in an extreme form that

$$R_1 - R_0 = \Delta, \tag{2}$$

a constant, i.e. the same for all individuals in the study.

An important aspect hidden in this definition is that as C notionally changes other relevant aspects are fixed; we shall see the more formal expression of this later via the role of background variable B.

A form equivalent to (2), that of so-called unit-treatment additivity, specifies that if unit of study s receives C_i for $i = 0, 1$ the resulting response is

$$\alpha_s + i\Delta, \tag{3}$$

with a direct extension if the potential cause takes more than two possible forms.

There is the further assumption that the response on unit s does not depend on the assignments of C to other units. We shall not address this issue here but clearly there are contexts where this consideration either dictates the size and nature of the appropriate unit of study or requires elaboration of (2) and (3) and the resulting statistical analysis. Thus in an agricultural fertiliser trial if the plot size were too small, quite apart from technical difficulties in implementation and harvesting, fertiliser might diffuse from one plot to another and make the yield on one plot depend in part on the treatment applied to an adjacent plot.

Note that the formulation (2) and (3), which is directly adapted from one used in the theory of experimental design, is put deterministically at an individual level. We discuss later a different formulation in which a population of individuals is involved and a stochastic element enters.

The assumption (2) is misleading even in an average sense if, for example, there are two different types of individual responding very differently to the causal variable C. For instance a blood-thinning agent used in the treatment of stroke could be beneficial to some patients and fatal to others, depending on the nature of the stroke.

Use of counterfactuals has been criticized by Dawid (2000) and defended in the resulting discussion. It is clear that (2) and (3) can be tested only indirectly via the stability of estimated differences, i.e. by the absence of interaction with meaningful features of the individuals. Further the parameter Δ can be directly estimated only as an average rather than as an individual effect. For some purposes, however, the individual interpretation of (2) is helpful.

This is not the place for an extended discussion of the role of counterfactuals. While it is clearly important that crucial assumptions in a statistical argument are not merely capable of being tested in principle but are subject to adequate test, there seems ample evidence that assumptions and formulations open at best to indirect test can be helpful aids to concept formulation and interpretation.

In some contexts (2) and (3) would be better formulated by regarding any causal effect as operating proportionally, or equivalently by taking (2) on a log scale.

The null hypothesis that there is no causal effect takes in this formulation the very strong form that the response observed on any individual is totally unaffected by the choices about C_0 and C_1. In

randomized experiments this leads to a test based solely on the randomization. For binary data, this is the exact hypergeometric test for a 2×2 table. The form (2) cannot in the nonnull case apply to binary responses and then estimation via randomization theory of the magnitude of an effect is more complicated; see Copas (1973).

We return to the issue of the individual versus the aggregate definition in Section 8.2.

4 Some Recent Work on Causality

4.1 Preliminary Results

We now review some recent work on statistical aspects of causality, especially stemming from that of Pearl (2000).

That work comes from a different background from that of most statisticians. There are, however, three accounts of it from a more statistical position. Lauritzen (2000) has placed Pearl's work in the context of the theory of graphical models in the form given in his book (Lauritzen, 1996). Lindley (2002) has reviewed Pearl's book and given a lucid account of some essential ideas and made important comments. Finally Dawid (2002) has reformulated the discussion using influence diagrams.

We first repeat two of Lindley's comments. The value of Pearl's formulation does not depend on the particular view of probability taken. Thus while much is formulated in terms of probability as assessing judgement or knowledge, the discussion is equally relevant to those concerned with probability as representing say physical or biological processes. Secondly, while Pearl's results do establish conditions under which first level causal conclusions are possible, checking of these conditions may be difficult; there is no suggestion that Pearl would disagree.

4.2 Conditioning and Intervening

A central theme in Pearl's discussion is the distinction between *conditioning* on C and *setting* or *intervening* on C. We start with the joint distribution of R, C, B, having integrated out I, taking it in the recursive form

$$f_{RCB} = f_{R|CB}\, f_{C|B}\, f_B. \tag{4}$$

In graphical representations of these systems, conditional independencies are represented by missing edges. In particular, absence of an effect of a potential cause C on a response R given B would be represented by a missing edge between C and R; see Figure 2b. A key issue in the formulation (4) is the assumption that the variables can on *a priori* grounds be placed in order so that each variable is a response to the subsequent variables in the sequence.

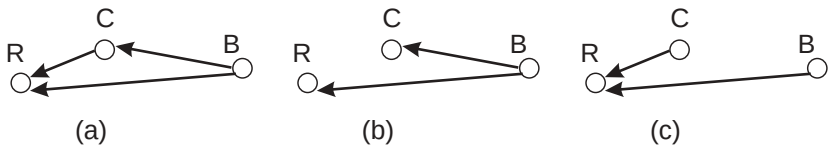

Figure 2. *(a) Graphical representation of general dependence of R on C and B and of C on B in initial system. (b) Absence of effect of C on R given B shown by missing edge, implying $R \perp\!\!\!\perp C \mid B$. (c) Modified system with explanatory variables acting independently shown as the missing edge between B and C.*

Conditioning in Pearl's sense is the standard conditioning calculation in (4), given only that $C = c$. We consider the resulting conditional distribution of R having marginalized over B. That is,

$$f_{R|C} = \int f_{R|CB} f_{B|C} db, \tag{5}$$

where $f_{B|C} = f_{CB}/f_C$. It would be appropriate to use $f_{R|C}$ for constructing an empirical prediction of R given only $C = c$. It corresponds to the total regression of R on C omitting B, i.e. allowing B to change with c in accordance with its conditional distribution given $C = c$.

To represent the effect of a notional or actual intervention to set $C = c$ in a system in which the directions of dependency in (4) are meaningful and thus can only act in one direction, we must express the notion that intervening on C has no backward effect on B, i.e. the value of B is unchanged and hence the distribution of B after the intervention remains f_B. That is, in (5) $f_{B|C}$ is replaced by f_B; see Figure 2c. This in general defines a different distribution for R having intervened to make $C = c$ and various notations are in use to describe this. Lauritzen (2000) used $||$ to replace the usual conditioning sign, leading to

$$f_{R||C} = \int f_{R|CB} f_B db. \tag{6}$$

This is Pearl's definition of a causal effect, interest focusing on how this distribution changes with c, having marginalized over B. The relation of this to the counterfactual notion involved in level-one causality is as follows. An individual has a given value $C = c$ and level-one causality concerns how R would change if c were to change by intervention.

The distinction between the two probability distributions $f_{R|C}$ and $f_{R||C}$ is crucial to the discussion. The former in (5) may sometimes have a useful interpretation but is inappropriate for examining the effect of intervention on C in that unrealistic changes in B are involved, i.e. changes in the past before the intervention.

In both (5) and (6) it is assumed that the conditional distribution of R and I given C and B remains unaffected by the intervention. This is not a trivial assumption. For example, the idea that serious interventions may distort all the relations in an economic system is the essence of the Lucas critique in econometric theory (Lucas, 1976).

Dawid (2002) introduces a unifying synthesis in which there is an augmented variable C^*, a decision node, with a directed edge only to C and which indicates whether conditioning (Figure 2a) or intervention (Figure 2c) is involved for computing an effect of C on R marginalizing over B. An advantage of this new formulation is that the usual properties of directed acyclic graphs apply in both cases. Dawid also shows the possibility of representing counterfactuals via functional relations involving error random variables represented by additional nodes and stresses, in effect, the impossibility of distinguishing an individual level version of (2) and (3) from an aggregate or population level form.

For a wide-ranging series of papers on causality, see McKim & Turner (1997).

4.3 The Linear Case

The representation in Section 4.2 has been framed for general distributions and centres on notions of statistical dependence and independence. It is, however, useful to set out the corresponding discussion for linear systems. These are formed from linear least squares regression equations, that is equations in which a response variable is expressed as a linear combination of explanatory variables plus a residual term of zero mean uncorrelated with the relevant explanatory variables. Such a relation is always possible subject to the existence of variances but its statistical relevance depends on nonlinearities being relatively unimportant.

Thus with just three variables, R, C, B, measured as deviations from their means, we may write,

corresponding to Figure 2a,

$$\begin{aligned} R &= \beta_{RC.B}C + \quad \beta_{RB.C}B + \quad \epsilon_R, \\ C &= \qquad\qquad \beta_{CB}B + \quad \epsilon_C, \\ B &= \qquad\qquad\qquad\quad \epsilon_B. \end{aligned}$$

Here, for example, $\beta_{RC.B}$ denotes the least squares linear regression coefficient of R on C adjusting for B, whereas β_{RB} would denote the regression coefficient of R on B marginalizing over, i.e. ignoring, C. This is easily calculated by substitution of the second equation into the first, noting that the resulting equation is indeed a linear least squares relation and hence giving (Cochran, 1938)

$$\beta_{RB} = \beta_{RC.B}\beta_{CB} + \beta_{RB.C}. \tag{7}$$

Similarly

$$\beta_{RC} = \beta_{RC.B} + \beta_{RB.C}\beta_{BC}. \tag{8}$$

A conditional independence statement such as $R \perp\!\!\!\perp C \mid B$ in the general formulation of Section 4.2 corresponds in the linear theory to $\beta_{RC.B} = 0$ and $C \perp\!\!\!\perp B$ corresponds to $\beta_{CB} = 0 = \beta_{BC}$. For multivariate Gaussian distributions this implies conditional independence. In general it implies the weaker property of no relation detectable by analysis linear in the relevant variables.

Thus in the linear case (5) corresponds to computing the overall regression coefficient of R on C marginalizing over B, referring to the graph in Figure 2a. On the other hand (6) corresponds to the overall regression coefficient of R on C in the modified system of Figure 2c in which B has been decoupled from C, i.e. B and C are nonadjacent in the graph. Therefore B does not change when there is an intervention on C. From equation (8) it follows for $\beta_{BC} = 0$ that $\beta_{RC} = \beta_{RC.B}$, i.e. the partial effect coincides with the overall effect by the assumptions of a notional intervention and treatment-unit additivity.

If by design or otherwise $\beta_{CB} = 0$ there is no difference between the two formulations. That is, $\beta_{RC.B} = \beta_{RC}$ or, in general, if $C \perp\!\!\!\perp B$, then $f_{R|C} = f_{R||C}$.

4.4 Relation with Statistical Practice

There are strong connections and an important difference between the discussion summarized above and mainstream statistical thinking. A concern common to the two fields is about what should be regarded as held fixed under hypothetical changes in the cause C. In regression terminology, which explanatory variables should be included in any regression equation for R additional to C itself? There is no disagreement that for assessment of a potential causal effect of C on R, background variables B are to be included, i.e. conditioned on, whereas any variables intermediate between the cause C and the response R should be excluded, i.e. marginalized over.

A major difficulty in many specific applications concerns whether all appropriate background variables have been included in B to ensure that the relevant regression coefficient captures the effect of C itself, so that the term cause is appropriately applied to C. This issue is distinct from the purely statistical uncertainty in estimating the effect from limited data.

The general discussion in terms of arbitrary densities leaves quite open the special assumptions of functional and distributional form that are often so important in serious statistical work. Of more general concern, however, is the notion of averaging an effect over the distribution of B. While this is sometimes convenient, in general the marginalization is a bad idea, notably because it discourages the study of interactions between C and additional features included in B. Such interactions may be crucial for interpretation. Also, as will be discussed in Section 8.1, verifying the absence of important interactions may give important security in interpretation.

In summary, marginalizing in (5) deals with the following question: given a probability distribution

over a set of variables (estimated from appropriate data) and given only $C = c$, what can be inferred about R? This question is remote from discussion of causality and is relevant in contexts where the objective is exclusively empirical prediction and in particular excludes the study of pathways of dependence; see Section 5.2.

Setting or intervention in (6) deals with the issue of estimating the effect of modifying the system by imposing a change on C that has no impact on a background variable B in the past and which leaves other statistical relations unchanged. The objective is to assess the effect on R of such a change in C and thereby to compare the effect of different interventions, i.e. different values of c. This is expressed in (6) by B retaining its distribution f_B independently of the intervention on C.

As already noted and as will be discussed further in Section 8.2, marginalizing over B is in general unwise and the appropriate distribution for causal interpretation is $f_{R|CB}$, as a function of both c and b, and not $f_{R||C}$.

The distinctions set out here essentially formalize via the variable or variables B the ideas mentioned in Sections 2 and 3 of respectively *allowable* alternative explanations and of *other things being equal* in connection with zero-level and with first-level causality. When the intermediate variables I are marginalized, it is implicitly assumed that the conditional distribution of R given C, B is not changed by intervention except via the implied change in C. Similarly, when background variables are marginalized interactions between C and B are ignored.

5 Intermediate and Surrogate Variables

5.1 General Discussion

Up to now variables intermediate between C and R have been ignored; there are, however, a number of important roles that they may play, including the following:

- to suggest pathways of development between the potential cause and the response and thereby to link with the second-level definition of causality in Section 3
- in further studies or in the presence of missing responses to serve as a surrogate response variable
- to monitor the correct application of the intervention
- to record any important unanticipated further effect that occurs between the potential cause and the response.

The first two of these reasons are in a sense the most interesting. Figure 3a shows a general dependence and Figures 3b and 3c are special cases of interest.

5.2 Study of Pathways

We turn now to second-level causality. As already stated, to find convincing evidence about the generating process in general, in line with Fisher's dictum as quoted by Cochran, requires assembly of evidence of various kinds. Nevertheless an important first step towards level-two causality may often be analysis involving the intermediate variable or variables I which in the previous discussion have been marginalized. These may indicate possible pathways between potential causal variables C and the response R, following the original motivation of Sewall Wright's path analysis and, for example, its introduction into sociology by Duncan (1975). Detailed interpretation will have the limitations of observational studies discussed above but nevertheless may be the primary objective of investigation. Even in the simpler discussion of potential causes it may sometimes be dangerous to disregard I totally, for this may indicate some unexpected and in a sense unwanted consequence of the intervention for which some account needs to be taken.

We give a simple outline example.

Example. Suppose in an agricultural fertiliser trial different levels of C represent different fertilisers, R is the yield of crop and that I is the number of plants per square metre all measured for each plot, the last half-way through the growing season. An increased yield might arise from the support of an increased number of plants per plot or from an increased yield per plant or from some combination of effects. In estimating the effect of C on yield, I would be ignored. The role of I is then to point to possible explanation of any fertiliser differences established. The case where the effect of C on R can be totally explained via I is shown in Figure 3b.

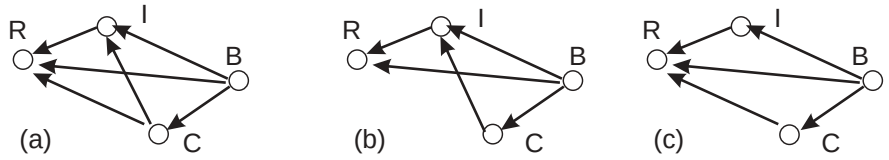

Figure 3. *(a) General dependence of R on I, C, B. (b) Given B and I, response R depends on C only via I. (c) Variable I conditionally independent of C given B and hence may be treated as an explanatory variable in addition to B when studying possible causal dependence of R on C.*

5.3 Surrogate Response

The possibility of an intermediate variable acting as a surrogate response can arise in two ways and raises important fresh issues. In one context, some individuals have missing response variables but measured surrogate. In another only the potential surrogate is recorded and its suitability has to be judged from background knowledge and previous data. If in the former case the missing responses are missing at random and the intermediate variable is measured in a comparable way on all individuals, fairly straightforward analysis should usually be possible. Essentially a regression equation in which the response is regressed on a surrogate response (and possibly other explanatory variables) can be used to predict the missing responses.

Strong conditions for a surrogate variable R_S, say, to be suitable as a total substitute for R were formulated by Prentice (1989). They are equivalent to $R \perp\!\!\!\perp C \mid R_B$. The additional requirement that R and R_S are not conditionally independent given C, B, i.e. that there is some dependence, hopefully a strong one, between real and surrogate responses is implied in every graphical formulation in which an edge present corresponds to an association of substantive interest (Wermuth & Lauritzen, 1990). For a further discussion of surrogates and related issues, see Frangakis & Rubin (2002) and Lauritzen (2003).

A condition weaker than that of Prentice is that in tracing paths from B, C to R the dependence in the relation of R_S to C given B is in the same direction as that when R itself is used instead of R_S (Cox, 1999). In terms of linear representations we require that R and R_S are measured in such a way that a positive effect of C on R_S implies a positive effect on R and that zero effect on R_S implies zero effect on R. In terms of linear representations, we have that

$$\beta_{RC.B} = \beta_{RC.R_SB} + \beta_{RR_S.CB}\beta_{R_SC.B}.$$

To preserve a qualitative interpretation we want $\beta_{RC.B}$ and $\beta_{R_SC.B}$ to have the same sign. Simple conditions for this when $\beta_{RR_S.CB} > 0$ are that $\beta_{RC.B}$ and $\beta_{RC.B} - \beta_{RC.R_SB}$ have the same sign. This condition is appreciably weaker and more realistic than requiring $R \perp\!\!\!\perp C \mid R_S B$.

A major difficulty with conditions for the appropriateness of surrogates is that the conditions need to hold for a broad range of circumstances or to be justified by some evidence-based knowledge of process; verification in one set of pilot data would on its own give little security for their future use.

This means that suggestions of causality for R based in fact on the surrogate R_S are likely to be especially tentative unless the pathway from R_S to R is well understood.

There is a difference of emphasis depending on whether the surrogate variable is of some intrinsic interest as compared with situations in which it is of no concern except in its surrogate role.

Example. In industrial life-testing accelerated testing in extreme environments is commonly used as a surrogate assessing reliability in a working context and justified explicitly or implicitly by some such proportionality assumptions as that if R and R_S are failure times in natural and accelerated modes then $R_S = R/\alpha$, where α is an acceleration factor assumed relatively stable across the various situations to be considered, i.e. in particular independent of C.

In this instance the surrogate variable is likely to be of no intrinsic interest. On the other hand in some medical applications, symptomatic improvement may be an intrinsically interesting surrogate for longer term response.

5.4 Other Roles

In some very limited circumstances it is reasonable to condition on an intermediate variable as if it were explanatory, namely if $I \perp\!\!\!\perp C \mid B$; see Figure 3c. That is, I is independent of the potential cause given the background information. For example, I might represent some important aspect of environment known *a priori* to be independent of C. Thus in an industrial experiment in which each day corresponded to a different experimental unit, the temperature and relative humidity occurring on a particular day might very well be treated as independent of C (Cox, 1958, p. 49).

In a linear representation

$$\beta_{RC.B} = \beta_{RC.BI} + \beta_{RI.CB}\beta_{IC.B}$$

and the second term vanishes if $\beta_{IC.B} = 0$.

More generally, the possibility of additional intervention or deviation from the protocol of the investigation bears, in particular, on the issue of non-compliance, sometimes called non-adherence, in clinical and other trials, i.e. of failure of patients to follow the treatment regime to which they have been assigned. In this case I serves a warning that the individual in question may not be informative about the effect of C in the way that was originally envisaged. Thus Cox & Wermuth (1996, p. 224) describe an only partly apocryphal agricultural trial in which the intermediate variable I was the severity of attack by birds. This acted selectively by treatment allocation and to ignore this would lead to quite misleading conclusions, judged either scientifically or technologically.

In general, however, the variables intermediate between C and R should not be included as explanatory variables in the primary analysis of the potential causal effect of C on R.

Example. Violanti (1998) has used police records of traffic accidents in Oklahoma to study the possible impact of mobile phones in vehicles on accidents. In one of the studies the occurrence or non-occurrence of a fatality was taken as the outcome variable. That is, in effect the paper studied the possible impact of a mobile phone on the seriousness of an accident, given that an accident occurred. It used logistic regression of the outcome on a considerable number of explanatory variables of which presence of a mobile phone was one. Another was a record that a vehicle ended on the wrong side of the road. It can, however, plausibly be argued that this is an intermediate response between possible mobile phone use and a fatality and as such should not be included in the regression equation for assessing the potential causal impact of a mobile phone on the occurrence of a fatality.

6 Unobserved Background Variables

6.1 Confounders in the Presence of Independencies

The main limitation to the interpretation of observational studies is often the possible presence of unobserved confounders, i.e. variables, U, whose omission seriously distorts the dependence of interest, but which were not observed, perhaps because their existence and nature were not appreciated.

That is, we would like to have studied $f_{R|CBU}$ but in fact can only estimate $f_{R|CB}$. In this discussion we again ignore possible intermediate variables I; see Figure 4a.

To study the relation between these distributions we return first to the linear case, writing now

$$\beta_{RC.B} = \beta_{RC.BU} + \beta_{RU.CB}\beta_{UC.B}, \tag{9}$$

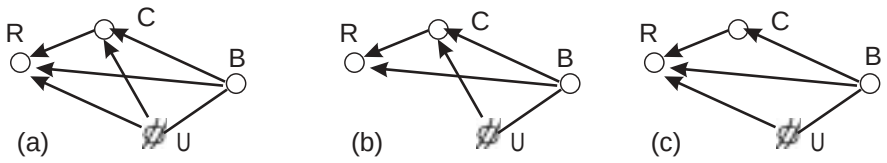

Figure 4. *(a) Relations between R, C and B in presence of unobserved confounder U; (b) Missing edge between U and R, i.e. $R \perp\!\!\!\perp U \mid CB$; (c) Missing edge between U and C, i.e. $U \perp\!\!\!\perp C \mid B$.*

The two terms on the right-hand side of (9) correspond to the two paths between C and R not passing through B in Figure 4a. It follows that inclusion of U has no effect on the regression coefficient if and only if the second term on the right-hand side vanishes, i.e. either $\beta_{RU.CB} = 0$ or $\beta_{CU.B} = 0 = \beta_{UC.B}$. The first condition is shown in Figure 4b; there is no direct edge from U to R. The second condition is shown in Figure 4c; there is no edge between U and C given B. If C is a randomized treatment the second condition is satisfied in virtue of the design even were the randomization probabilities to depend on B; see Figure 4c. In observational studies, the assumption, if made, amounts to supposing that the value of C is determined in a way that is essentially equivalent to such randomization, an assumption not directly checkable in the absence of observation of U. It may sometimes be rather less problematic if the variable U is a feature expected to be important but which is not observed in the study under analysis, although it has been observed in other studies.

It is immaterial whether U is a response to or explanatory to B and in general both variables may be multidimensional and the ordering relation between them a partial ordering, in that some pairs of variables may be on an equal footing in a sense to be explained in Section 7.1. Therefore no direction need be attached to the edge between U and B.

The above discussion is for linear systems. For general distributions, the condition that $R \perp\!\!\!\perp U \mid CB$ implies directly that $f_{R|CBU} = f_{R|BC}$, corresponding to $\beta_{RU.CB} = 0$. That is, inclusion of U in a study of the dependence of R on explanatory variables would, in large samples, induce no change.

First if $R \perp\!\!\!\perp C \mid BU$ and $C \perp\!\!\!\perp U \mid B$, then $(R, U) \perp\!\!\!\perp C \mid B$, so that in the null case of no effect of C on R given BU no spurious effect is induced by omitting U.

Secondly when there is dependence of R on C given BU, but $C \perp\!\!\!\perp U \mid B$, as in Figure 4c, the form of the relation is changed by marginalizing over U, but it can be shown (Cox & Wermuth, 2003) that there is qualitative invariance in the following sense. If R is stochastically increasing with C in the conditional distribution given B, U then it remains stochastically increasing after marginalization over U. Thus, so long as $U \perp\!\!\!\perp C \mid B$, marginalizing over U cannot induce an effect reversal, showing

the strong consequences of randomization in inducing qualitatively similar dependencies of R on C given B and of R on C given B and U.

However, even if $C \perp\!\!\!\perp BU$, as in the case of randomization, there may be an unobserved interactive effect of U and C on the response R. This is, for instance, the case in the example of stroke patients mentioned in Section 3. There C is a blood-thinning treatment and U, the unobserved status of the patient, could have two levels, corresponding to a burst vessel or a thrombosis. The omission of this distinction had led to inconclusive and contradictory results in early controlled clinical trials with thrombolytic agents (Zivin & Choi, 1991).

One important and traditional approach to the possible effect of unobserved confounders is by sensitivity analysis. That is, one considers how strong an effect an unobserved confounder would have to exert to explain an apparent dependence and then, if that effect is strong, one examines what possible unobserved features might exert such an effect. Detailed discussion of this is given by Rosenbaum (2002).

6.2 Confounders and Instrumental Variables

We have seen in the previous subsection some very special circumstances in which no confounding is induced by unobserved background variables. There is another possibility of correcting for bias induced by an unobserved confounder. We develop this in outline for the simple system of four variables R, C, Z, U, that is omitting B purely to simplify the notation. Here U is again unobserved. In the system shown in Figure 5a, the variable Z is called an instrumental variable. It is marginally independent of U and it exerts an influence on R via C.

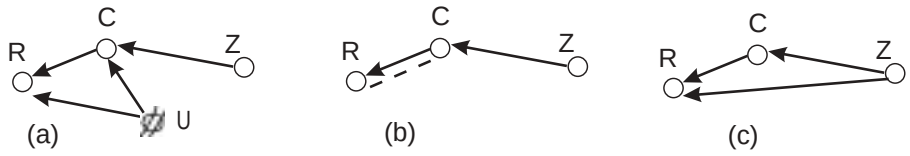

Figure 5. *(a) Graphical representation of dependence of R on C and unobserved U, involving instrumental variable Z. (b) Equivalent structural equation model with dashed edge denoting correlated errors. (c) Equivalent saturated system.*

In the linear case this gives for variables measured from their mean

$$\begin{array}{llll} R = \beta_{RC.U} C + & + & \beta_{RU.C} U + & \epsilon_R, \\ C = & \beta_{CU} U + & \beta_{CZ} Z + & \epsilon_C, \\ Z = & & & \epsilon_Z, \\ U = & & & \epsilon_U, \end{array} \tag{10}$$

where the ϵ's are error terms uncorrelated with the explanatory variables on the right-hand side of the relevant equation. The variables are measured from their means. The special assumptions about Z have been used to simplify the notation. Elimination of U from the above equations shows that the system R, C, Z is saturated, i.e. has an arbitrary covariance matrix. This implies that the special independence assumptions made in formulating these equations cannot be empirically tested from R, C, Z alone; they can be justified only on subject-matter grounds. It follows that on investigating the system in which U is unobserved

$$\mathrm{cov}(R, Z) = \beta_{RC.U} \beta_{CZ} \mathrm{var}(Z), \quad \mathrm{cov}(C, Z) = \beta_{CZ} \mathrm{var}(Z),$$

from which it follows that the coefficient of interest, namely $\beta_{RC.U}$, can be estimated via $\text{cov}(R, Z)/\text{cov}(C, Z) = \beta_{RZ}/\beta_{CZ}$.

This argument has a long history in more general form in econometrics (Goldberger, 1991) but until recently appears to have been little used in other fields and possibly is less frequently employed also in its original context. This is partly because the assumptions are strong and not directly checkable and partly because the resulting estimate has low precision unless the denominator β_{CZ} is well determined, i.e. the relation between C and Z is quite strong.

The instrumental variable formulation in (10) with U unobserved is equivalent to the structural equation model

$$R = \alpha C + \eta_R, \quad C = \beta Z + \eta_C,$$

summarized in Figure 5b. In this Z is uncorrelated with η_C but C is correlated with η_R, so that the first equation is not a least squares regression equation. There are six parameters in this system equivalent to the saturated system for R, C, Z shown in Figure 5c.

7 Joint Responses and Joint Causes

7.1 General Formulation

The discussion in Sections 3–6 has hinged on the assumption that all variables may be ordered so that for any pair of variables one is explanatory to the other considered as a response. While whenever B and I are sets of variables with several components ordering of the variables within the sets may be largely irrelevant, the set-up is too restrictive for many purposes and we therefore sketch a more general formulation, thereby returning to Figure 1.

For each individual we suppose that a number of features or variables are recorded. These can be classified in various ways that are context-specific. Typically one group will be one or more response variables, representing in some sense outcomes. Another group will be explanatory to those response variables and also can be regarded as candidate causal variables, in particular as conceivably taking values for that individual different from those actually obtaining. A further set of variables is regarded as intrinsic in that their values are essential to the definition of the individual in question. Intrinsic variables are not regarded as potentially causal. Finally there may be intermediate responses, sometimes used as surrogate markers, between the explanatory variables and the responses of main interest.

In our graphical representation we place the intrinsic variables and other background variables in a box to the right enclosed with double lines to indicate that they are not represented probabilistically and are not potential causal variables in the context considered. Indeed the only reason to represent them probabilistically would be to see whether their distribution matches that in some target population, an issue we do not address here.

For all other variables we assume the following. For any pair of variables, say X_i, X_j either

- X_i is explanatory to X_j or vice versa
- X_i and X_j are to be considered on an equal footing.

More detailed distinctions can be drawn. The explanatory-response relation may be based on temporal ordering, a strong sense, or on a subject-matter working hypothesis, the latter being the only possibility in those cross-sectional studies in which the variables measured all refer to the same time point. Two or more variables which are somewhat arbitrary coordinates specifying a single multivariate feature are naturally regarded on an equal footing. In other cases it may just be a noncommittal view of the direction of dependency.

It then follows under mild additional assumptions that the variables can be grouped in blocks in such a way that

- all variables in the same block are on an equal footing
- the blocks are ordered with all variables in one block representing potential responses to variables in subsequent blocks.

In the graphical representation of dependencies, directed edges are used between nodes in different blocks and undirected edges between nodes in the same block, missing edges denoting conditional independencies. To cover the possibilities encountered in applications it is necessary to distinguish two types of conditioning (Cox & Wermuth, 1993, 1996) but here we consider only the possibility that in considering the relation between two nodes in the same block g we always condition on nodes in subsequent blocks and marginalize over any additional nodes in block g.

Example. Cox & Wermuth (1996, Chapter 6) discussed a cross-sectional study of the factors influencing diabetic patients in controlling their disease. Because of the cross-sectional nature of the study the progression of variables from explanatory to response, shown in graphical form in Figure 6, is based to some extent on working hypothesis; for example, it is possible that success at control is explanatory to knowledge of the disease rather than vice versa. This raises the interesting issue of the implications of the independencies implied by one ordering of the variables were the blocking of the variables to be rearranged (Wermuth & Cox, 2004).

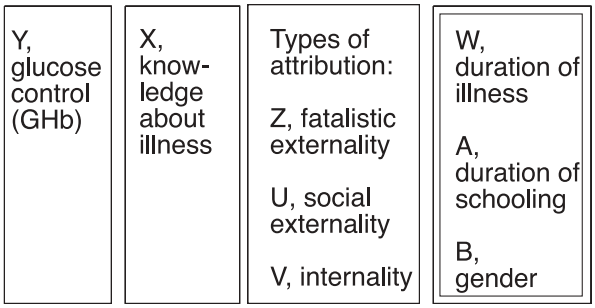

Figure 6. *Schematic representation of dependencies in study of diabetes.*

Details of the analysis are given in the reference cited. The essence is that the primary outcome variable is regressed on all other variables by linear regression with some checks for interactions and nonlinearities. Then the next variable is regressed on all other variables, excluding the primary response and so on. In this instance no special complications arose from variables on an equal footing. An outline summary of the resulting analysis is given in Figure 7a with Figure 7b showing the structure after conditioning on A, duration of schooling, used as a binary variable. An important conclusion of the analysis was that there was an interaction between A and duration of illness, W, studied by examining the dependencies of Y, X and Z separately at the two levels of A. Such an interaction is not easily shown in the graphical representations used here. In fact, while the same type of generating process is suggested at the two levels of A the direction and strengths of the effects differ.

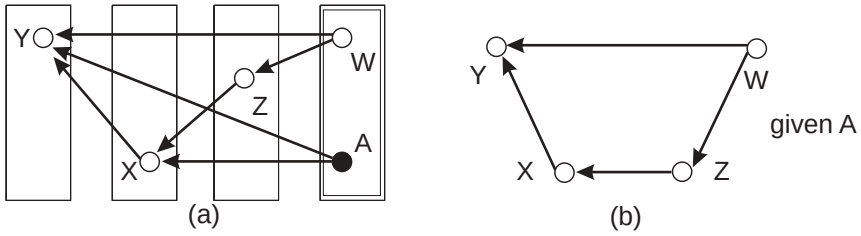

Figure 7. *(a) Detailed representation of dependencies between variables listed and grouped in Figure 6. (b) Representation conditionally on A, i.e. for two given levels of formal schooling.*

7.2 Causal Variables on an Equal Footing

This more general formulation allows us to address further issues. Very particularly, suppose that there are two potential causal variables C_1 and C_2 on an equal footing. When we notionally intervene on C_1 what happens to C_2? There are several possibilities

- C_2 may be unaffected, i.e. for this particular purpose be treated as a background variable.
- C_2 may change as specified by the generating distribution, i.e. for this particular purpose be treated as an intermediate response.
- C_2 may change in a way that is governed by a different process from that involved in the original generating process, possibly but not necessarily a situation intermediate between the first two.
- It may ultimately be more informative to regard C_1, C_2 as two factors defining a factorial "treatment" structure to be assessed simultaneously rather than separately.

Example. Suppose that C_1 and C_2 are respectively sodium and potassium levels in the blood and R is some response, perhaps blood pressure or perhaps occurrence of a cardiac event. In the following discussion it is important to distinguish the blood level of, say, sodium from the intake of sodium. The latter is in principle controllable whereas the former is the outcome of a complex process.

If for a particular individual we consider imposing a change in sodium level to a new value, or perhaps consider imposing a change of a certain magnitude, it is unclear what will happen to the level of potassium. It would be conceptually possible to manipulate potassium intake rather than blood level directly so that potassium blood level remained constant and this would be the first possibility listed above.

The second possibility would be that potassium changes, consequent on the change in sodium, in the same way as in the data under analysis; of course the reasonableness of this depends strongly on how the data are collected and if the analysis involves inter-personal comparisons the assumption is unreasonable.

The third possibility would involve collecting special data to study the effect of imposed changes of sodium level on potassium level. This might include the study of the dynamics of the processes involved.

The fourth possibility of treating sodium and potassium levels as factors defining an explicit or implicit factorial experiment would imply interventions in which both variables were manipulated to preset levels and, while in principle more informative about the effect on ultimate response, would be even more remote from direct observation.

The third possibility listed above requires for its implementation a separate set of data or theoretical calculation estimating the effect on C_2 of changing the prescribed level of C_1 and use of a generalization of (7) in the form

$$\beta^*_{RC_1.B} = \beta_{RC_1.BC_2} + \beta_{RC_2.C_1B}\gamma_{C_2C_1.B}.$$

Here $\gamma_{C_2C_1.B}$ is a regression coefficient for an investigation in which C_1 is varied and the consequent changes of C_2 are measured.

8 Some More Detailed Issues

8.1 Choice of Candidate Causal Variables

We now deal more briefly with some specific issues. For a variable C to be a potential causal variable C it needs to be reasonable to consider at least notionally the idea that an individual with $C = c$ might have had a different value of c without changing the essential nature of that individual. This consideration is context-specific. Thus in most situations gender would not be considered as a candidate cause. For to do so would involve the notion of considering the value of R resulting for, say, a male if that individual were female, all other aspects remaining unchanged, and this usually makes no sense. In contexts of possible discriminatory employment practices, however, the comparison of, say, pay for a man with given work experience, skills, etc. as compared with a woman with the same work experience, etc. is the central issue (Dempster, 1988).

Another example is that passage of time is not to be considered as causal in itself, only processes that develop in time. This is because the notional intervention in which passage of time does not occur, other things being equal, makes no sense. Processes that develop in time may be considered as potentially causal.

In principle in the more general formulation of Section 7 any variable that is not considered as intrinsic might be considered as potentially causal for the response R. Which are actually viewed as causal and which as background variables depends crucially on the objectives of the investigation, the most ambitious objective being to analyse the whole set of pathways from initial explanatory variables to response. Since implicitly causality is regarded, in the contexts of most statistical interest, as a multiple process there is no conflict in regarding for particular purposes variables that could be causal as part of the background variables B in assessing the effect of a variable C of primary concern.

In approaching a system from first principles it would be sensible to regard variables far back in time, or in the representation in question, as in some sense initial causes and then to estimate the additional information provided by each new stage as it arises. An instance is the so-called foetal origins hypothesis, where foetal events are claimed to have a life-long health impact. Of course causal variables well separated from the response will often show relatively weak dependency.

In such studies the role of interaction effects may be very important and this is especially important in genetic epidemiology. For example, suppose that in studying a clinical outcome both clinical and genetic variables are considered explanatory. It might well happen that genetics is explanatory for disease occurrence and indeed for some current clinical aspects, even if its overall explanatory power for outcome is relatively small compared with current clinical status. Another important possibility is of interaction between genetic and clinical features, in extreme cases that genetics separates the disease into distinct types for which the interpretation of given clinical features is different. The study of Wilm's tumour (Beckwith *et al.*, 1990) is an important example of this.

8.2 Basis of Generalization

Suppose next that a potentially causal difference is established between, say, two treatments on the basis of a well-conducted randomized trial. Under what circumstances is it reasonable to conclude that similar conclusions will apply in the future in inevitably somewhat different circumstances? Also what basis is there for concluding that the conclusion will apply to a single individual?

Even if the conclusions are replicated in independent studies, any notion of generalization based on regarding the studies as a random sample from a population of studies seems very artificial (Yates & Cochran, 1938), even though any such replication is clearly reassuring at a qualitative level at least. Basis for generalizing may better rest partly on second-level causality, i.e. on some understanding of underlying process, and partly on absence of interaction with important intrinsic variables describing the study individuals. Subject to essential stability of effect, the basis for generalization can be achieved either by synthesis of conclusions from different studies, or by initial design to ensure a broad range of validity; see, for example, Cox (1958, p. 17).

The same considerations apply also to specificity. A randomized experiment establishes an average treatment effect over the study individuals. To conclude something for a new specific individual, for example for a new patient, requires both generalization, often to a new environment, and the assumption that there is relatively little treatment by individual interaction. Part of the advantage of independent replication of studies with a broadly similar objective as contrasted with increasing the size of single studies is that the range of explanatory features involved is likely to be increased.

The formulation (2) and (3), which is directly adapted from one used in the theory of experimental design, is initially formulated deterministically at an individual level. The addition to the notional responses of independent and identically distributed random variables representing measurement error has no immediate impact on the resulting analysis and conclusions. A different interpretation of such an extended model is to regard the potential causal effect as defined only at an aggregate level over some population of individuals (Cox, 1958, sections 2.1–2.3). In the original formulation, however, the conclusions refer to the individuals actually studied.

The population-based formulation appears to give a broader base to the conclusions but unless the individuals studied are a random sample, or at least a representative sample, of a target population of interest the extension has little direct force. If, indeed, the population is purely hypothetical then it is unclear that any real basis for meaningful generalization has been achieved.

Example. In a clinical trial setting the conclusions might be regarded as applying fairly directly to the population of individuals from the regions in question and giving informed consent to participation. This may well differ appreciably from the target population of, for example, all patients with a particular condition. If there are special features in which these populations differ, it becomes especially important to check that any treatment effect does not depend, i.e. interact with, those features. Thus in randomized clinical trials it is desirable to check not only that the features agree reasonably well as between the treatment arms, i.e. check on the effectiveness of the randomization, but more importantly that any major discrepancies with the presumed target population are uncovered.

For specificity the individual level formulation of Section 8.1 is more appropriate but as is clear this can be checked only partially.

We do not, even in the discussion of Section 7, allow the possibility that two variables C_1 and C_2 on an equal footing are each a cause of the other and hence in effect responses. Such representations are studied in linear form in the econometric literature as simultaneous equation models in which cyclic dependencies are permitted such as that R_2 depends on R_1 and R_1 depends on R_2. Such dependencies are best studied by the explicit introduction of time.

8.3 Design Issues

We do not in this paper discuss details of study design and statistical analysis important though these ideas are. Implicitly we have taken the form of most studies to be randomized experiments or their approximate observational equivalent, a cohort study. If applied to cross-sectional data particularly strong subject-matter knowledge is essential to give any plausibility to the ordering of variables that is essential to the present analysis. In some fields, especially those studying relatively rare outcomes, retrospective studies, broadly of the case-control form, are common. They are best analysed and interpreted by considering the questions: what is the corresponding cohort study and to what extent does the retrospective data allow conclusions about such a cohort study to be drawn? As such, no special issues of principle concerning the nature of causality appear, although there are more detailed and often major concerns about data quality, especially concerning the possibility of recall bias, and about the appropriate choice of control group.

9 Discussion

The object of the present paper is to review the concepts and assumptions involved in attaching a causal interpretation to statistical dependencies. Especially in the context of observational studies the role of unobserved confounders is probably the most critical aspect. We have ignored the more technical statistical issues. These include key concerns about data quality, the formulation of representations that capture empirical dependencies in interpretable form, the assessment of the magnitude of random errors of estimation and the dealing with biases and random errors of measurement, missing values and any consequences of unusual design structures.

The main broad implications for statistical work are simple but important and are as follows:

- Studies of dependence with a causal objective are not to be confused with the construction of empirical prediction systems.
- Only some variables may be treated as potentially causal and their choice is critical.
- Choice of explanatory variables for inclusion or exclusion in principle from regression-like calculations is crucial.
- This choice may be clarified by a chain block representation of the variables involved corresponding to a process in single or joint variables.
- Checks for possible interaction between the effect of a potential cause and intrinsic features of the study individuals are essential, in particular in connection with generalizability and specificity.
- Especially in observational studies, some description, even if only qualitative, of the possible role of unobserved explanatory variables is desirable in general and essential if they represent confounders.

Our attitude is that the search for causality is of key importance in many contexts but that the goal is hard to achieve except when large effects are involved. Then sensitivity analysis may reasonably establish that some of the complications discussed here are unlikely to affect the conclusions materially and that delicate statistical analysis is likely to be unnecessary. The approach sketched above is designed to encourage the uncovering of causal structure while at the same time being realistic about the assumptions involved. In more applied contexts, especially biomedical ones, there is some empirical evidence that false claims of causality undermine the credibility of other careful studies where causality is indeed reasonably firmly established. The case for reasoned and optimistic caution is then particularly clear.

Acknowledgement

We are grateful to the referees for their meticulous reading of the paper and for very constructive comments.

References

Aalen, O. (1987). Dynamic modelling and causality. *Scand. Actuarial J.*, **13**, 177–190.
Beckwith, J.B., Kiviat, N.B. & Bondodio, J.F. (1990). Nephrogenic rests, nephroblastomatosis and pathogenesis of Wilm's tumor. *Pediatric pathology*, **10**, 1–25.
Box, G.E.P. (1966). Use and abuse of regression. *Technometrics*, **8**, 625–629.
Cochran, W.G. (1938). The omission or addition of an independent variable in multiple linear regression. *Suppl. J.R.Statist. Soc.*, **5**, 171–176.
Cochran, W.G. (1965). The planning of observational studies in human populations (with discussion). *J.R. Statist. Soc. A*, **128**, 234–265.
Copas, J.B. (1973). Randomization models for the matched and unmatched 2×2 tables. *Biometrika*, **60**, 467–476.
Cox, D.R. (1958). *Planning of experiments.* New York: Wiley.
Cox, D.R. (1992). Causality: some statistical aspects. *J.R. Statist. Soc. A*, **155**, 291–301.
Cox, D.R. (1999). Some remarks on failure-times, surrogate markers, degradation, wear, and the quality of life. *Lifetime Data Analysis*, **5**, 307–314.
Cox, D.R. & Snell, E.J. (1981). *Applied statistics.* London: Chapman and Hall.
Cox, D.R. & Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion, p. 283). *Statistical Science*, **8**, 204–218.
Cox, D.R. & Wermuth, N. (1996). *Multivariate dependencies.* London: Chapman and Hall.
Cox, D.R. & Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalization. *J.R. Statist. Soc. B*, **65**, 937–941.
Dawid, A.P. (2000). Causal inference without counterfactuals (with discussion). *J. Amer. Statist. Assoc.*, **95**, 407–448.
Dawid, A.P. (2002). Influence diagrams for causal modelling and inference. *Int. Statist. Rev.*, **70**, 161–189.
Dempster, A.P. (1988). Causality and statistics. *J.Statistical Planning and Inference*, **25**, 261–278.
Doll, R. (2002). Proof of causality. *Perspectives in biology and medicine*, **45**, 499–515.
Duncan, O.D. (1975). *Introduction to structural equation models.* New York: Academic Press.
Fisher, R.A. (1926). The arrangement of field experiments. *J. Ministry of Agric.*, **33**, 503–513.
Fisher, R.A. (1935). *Design of experiments.* Edinburgh: Oliver and Boyd. And subsequent editions.
Frangakis, C.B. & Rubin, D.B. (2002). Principal strata in causal inference. *Biometrics*, **58**, 21–29.
Freedman, D. (2003). From association to causation: some remarks on the history of statistics. In *Stochastic musings*, Eds. J. Panaretos, pp. 45–71. Mahwah, NJ: Lawrence Erlbaum.
Goldberger, A.S. (1991). *A course in econometrics.* Harvard University Press.
Goldthorpe, J. (1998). *Causation, statistics and sociology.* 29th Geary lecture. Dublin: ESRI.
Good, I.J. (1961). A causal calculus, I. *British J. Philosophy of Science*, **11**, 305–318.
Good, I.J. (1962). A causal calculus, II. *British J. Philosophy of Science*, **12**, 43–51.
Granger, C.W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.
Gregg, N.N. (1941). Congenital cataract following German measles in the mother. *Transactions of the Ophthalmological Society of Australia*, **3**, 35–46.
Hill, A. B. (1965). The environment and disease: association or causation. *Proc. R. Soc. Medicine*, **58**, 295–300.
Holland, P.W. (1986). Statistics and causal inference (with discussion). *J. Amer. Statist. Assoc.*, **81**, 945–970.
Hoover, K.D. (2002). *Causality in macroeconomics.* Cambridge University Press.
Lauritzen, S.L. (1996). *Graphical models.* Oxford University Press.
Lauritzen, S.L. (2000). Causal inference from graphical models. In *Complex stochastic systems*, Eds. O.E. Barndorff-Nielsen *et al.*, pp. 63–107. London: Chapman and Hall.
Lauritzen, S.L. (2003). Graphical models for surrogates. Invited paper for ISI session, Berlin. *Bulletin Internat. Statist. Inst. 54th Session*, Vol. **60**, book 1, 144–147.
Lindley, D.V. (2002). Seeing and doing: the concept of causation. *Int. Statist. Rev.*, **70**, 191–214.
Lucas, R.E. (1976). Econometric policy evaluation: a critique. In *Studies in business-cycle theory*, Ed. R.E. Lucas, pp. 104–130. Cambridge, Mass: MIT Press.
McKim, V.R. & Turner, S.P., Eds. (1997). *Causality in crisis?* University of Notre Dame Press.
Neyman, J. (1923). On the application of probability theory to agricultural experiments. English translation from the Polish original plus commentary: *Statistical Science* (1990) **5**, 465–480.
Pearl, J. (2000). *Causality.* Cambridge University Press.
Prentice, R.L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine*, **8**, 431–440.
Robins, J. (1997). Causal inference in complex longitudinal data. In *Latent variable modeling with applications to causality*, Ed. M. Berkane, pp. 69–117. New York: Springer.
Rosenbaum, P.R. (2002). *Observational studies.* Second ed. New York: Springer.

Rubin, D.B. (1974). Estimating causal effect of treatments in randomized and nonrandomized studies. *J. Educational Psychol.*, **66**, 688–701.
Scheines, R. (1997). An introduction to causal inference. In *Causality in crisis?*, Eds. V.R. McKim and S.P. Turner, pp. 185–199. University of Notre Dame Press.
Schweder, T. (1970). Composable Markov processes. *J. Appl. Prob.*, **7**, 400–410.
Spirtes, P., Glymour, C. & Scheines, R. (1993). *Causation, prediction and search.* New York: Springer.
Suppes, P. (1970). *A probabilistic theory of causation.* Amsterdam: North Holland.
U.S. Department of Health, Education and Welfare (1964). Smoking and health. Report of the advisory committee to the Surgeon-General of the public health service. Washington DC: U.S. Government Printing Office.
van der Laan, M.J. & Robins, J.M. (2002). *Unified methods for censored longitudinal data and causality.* New York: Springer.
Violanti, J.M. (1998). Cellular phones and fatal traffic collisions. *Accid. Anal. Prev.*, **30**, 519–528.
Wermuth, N. & Cox, D.R. (2004). Joint response graphs and separation induced by triangular systems. *J.R. Statist.Soc. B*, **66**, 687–717.
Wermuth, N. & Lauritzen, S.L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J.R. Statist. Soc. B*, **52**, 21–72.
Yates, F. & Cochran, W.G. (1938). The analysis of groups of experiments. *J. Agric. Science*, **28**, 556–580.
Yule, G.U. (1900). On the association of attributes in statistics. *Phil. Trans. Roy. Soc. (London)* A, **194**, 257–319.
Zivin, J.A. & Choi, D.W. (1991). Neue Ansaetze zur Schlaganfalltherapie. *Spektrum der Wissenschaft.* September, pp. 58–66.

Résumé

On fait une revue critique de la causalité statistique. On presente trois definitions de la causalité et on discute les consequences pour l'analyse statistique et l'interpretation.

[*Received October 2003, accepted June 2004*]