Causality: Some Statistical Aspects

D. R. Cox

Journal of the Royal Statistical Society. Series A (Statistics in Society), Volume 155, Issue 2 (1992), 291-301.

Stable URL:
http://links.jstor.org/sici?sici=0964-1998%281992%29155%3A2%3C291%3ACSSA%3E2.0.CO%3B2-O

Journal of the Royal Statistical Society. Series A (Statistics in Society)

J. R. Statist. Soc. A (1992)
155, *Part* 2, *pp.* 291–301

Causality: Some Statistical Aspects

By D. R. COX†

Nuffield College, Oxford, UK

[Received May 1991]

SUMMARY

After some brief historical comments on statistical aspects of causality two current views are outlined and their limitations sketched. One definition is that causality is a statistical association that cannot be explained away by confounding variables and the other is based on a link with notions in the design of experiments. The importance of underlying processes or mechanisms is stressed. Implications for empirical statistical analysis are discussed.

Keywords: EPIDEMIOLOGY; ERROR OF MEASUREMENT; EXPERIMENT; INTERACTION; INTERVENTION; LATENT VARIABLE; OBSERVATIONAL STUDY; REGRESSION

1. INTRODUCTION

There is a very extensive philosophical literature on causality, some of it rather negative in tone. The object of the present paper is to review recent more statistical thinking on the topic, taking the viewpoint that there is certainly some sense in which causality is central to the scientist's efforts to understand the real world. The implications will be pointed out for statistical analysis, especially for the empirical study of dependences via regression analysis, using that term in a broad sense to include logistic regression, regression analysis of survival data, etc.

In Section 2 a brief historical review is given of some statistical work on causality. Section 3 discusses attempts to define causality as statistical association which cannot be explained away via confounding variables, stressing both the value and the limitations of that approach. Section 4 outlines a way in which notions from experimental design are applied in an extended context. The remainder of the paper deals with some miscellaneous issues connected with applied statistical work. For a general introduction to the topic, see the encyclopaedia article of Barnard (1982).

2. SOME HISTORICAL COMMENTS

Yule (1903) (see also Stuart (1971)) made a careful analysis of the relations between several variables, both discrete and continuous, developed a notation for partial and total regression coefficients which is again coming into favour, introduced the term nonsense or spurious correlation and discussed what is now often called Simpson's paradox, i.e. the notion that there can marginally be a positive association between two binary variables C and B, even though conditional on a third variable A the association between C and B may be negative (at both levels of A). These issues underlie much later discussion especially that summarized in Section 3.

†*Address for correspondence*: Nuffield College, Oxford, OX1 1NF, UK.

 0035–9238/92/155291 $2.00

Bradford Hill (1937, 1965) discussed the circumstances under which an effect obtained in an observational study is relatively likely to have a causal interpretation. Such conditions include that the effect

(a) is large,
(b) is reproduced in independent studies,
(c) shows a monotone relation with 'dose',
(d) corresponds to a 'natural experiment',
(e) behaves appropriately when the potential cause is applied, removed and then reinstated,
(f) is consistent with subject-matter knowledge or
(g) is, for example, predicted by reasonably well-established theory.

Cochran in a series of papers (in particular Cochran (1965, 1972)) described in largely qualitative terms the design and analysis of observational studies and in particular discussed what is often called Fisher's dictum. Cochran reports that R. A. Fisher, asked at a conference for his comments on the step from association to causation, replied: make your theories elaborate. We return to this below.

The work reviewed above provides important qualitative background to discussions of causality. We turn now to more specific developments beginning with two rather contrasting definitions. In some ways it is remarkable how relatively little causality is mentioned in the general statistical literature, except in a social science context, perhaps because causality is regarded as an essentially subject-matter issue. If there is a general statistical view on the matter, it is presumably that causal inference is possible from randomized experiments and that if attempted from observational studies these had better be longitudinal rather than cross-sectional.

3. CAUSALITY VIA ASSOCIATION

One definition of causality used in the philosophical literature requires that, if C is to be the cause of an effect E, then C must happen if E is to be observed. This is clearly inappropriate in, for example, most epidemiological contexts, settings where some probabilistic notion seems essential, involving usually also some idea of multiple causes. Thus smoking is neither a necessary nor a sufficient condition for lung cancer.

The first such notion that we shall discuss is essentially that causation is a statistical association that cannot be explained as in fact a dependence on other features. This can be formalized in various ways and arises in particular in the work of Good, Suppes, Wiener and Granger, the last two in a time series context, in the last specifically econometric; for references on these and other sources, see the important review by Holland (1986).

For the simplest form of the above notion, let C and E be binary events and B be a third variable or collection of variables. We may say that C is a candidate cause of E if C and E are positively associated, i.e. if

$$P(E|C) > P(E|\text{not } C).$$

We can regard the cause as spurious if B explains the association in that

$$P(E|C \text{ and } B) = P(E|\text{not } C, \text{ and } B),$$

i.e. if E and C are conditionally independent given B. The cause is confirmed if C is a candidate cause that is not spurious, i.e. which cannot be so explained via any B.

More generally a variable x_C is a cause of the response y_E if it occurs in all regression equations for y_E whatever other variables x_B are included. This is an important notion and corresponds closely to common statistical practice. Nevertheless the use of the word causal seems unwise, because of a considerable gap between this and common scientific interpretation. Apart from questions of terminology, there are at least two major qualifications to be made.

First in any particular application B is restricted to features observed. To some extent this point is alleviated by the notion of propensity score (Rosenbaum and Rubin, 1983; Rosenbaum, 1984a, b, 1987) in which, in effect, the properties are judged that an unobserved variable would need to provide an explanation of the effect under study.

Secondly and crucially it is necessary to restrict the variables C, E and B, not least to express the essential asymmetry between cause and effect. There are various ways in which this can be done. The simplest is to insist that B occurs before C in time which in turn occurs before E. We leave aside the difficult issue of simultaneous causality, noting only its central role in discussions connected with the bases of quantum theory. The strict temporal ordering condition serves both to establish a clear asymmetry between cause and effect and also to ensure that a causal effect of say C is not removed by some subsequent B that was itself a consequence of C. Some such restriction on the allowable B is clearly essential.

Sometimes spatial proximity can be used as a basis for ordering effects instead of temporal ordering.

A third possibility is that specific subject-matter knowledge is used to establish a presumed causal ordering of variables. For example, in a psychological study measurements of trait and state may be available for properties such as anxiety and it would be natural to treat trait as prior to state. Again in a study of hypertensive patients data might be available for each patient on biochemical variables, on blood pressure, on performance in some 'objective' physical exercise test and on subjective self-assessed well-being (quality of life). We might then treat those four kinds of variable as in the stated order causally, although clearly non-trivial assumptions are involved in so doing.

These approaches all involve the injection of external information to assess the direction of causality. Another approach is to use simplicity of structure to achieve that end. This is the approach of Pearl and Verma (1991); they have developed a powerful computer algorithm to examine the conditional independence structures among a large number of binary variables to find the simplest graph-theoretic representation of the conditional independences involved, thereby in particular making causal inference in their sense possible from cross-sectional data without *a priori* assumptions about the nature of the variables.

As a simpler although different example of the same kind of argument, suppose that in several independent studies each with a bivariate normal distribution of two variables i the regression lines of y on x are identical throughout whereas the regression lines of x on y, although parallel, require a separate intercept for each set of data. The dependence of y on x is thus simpler than that of x on y and might be claimed to show that y causes x rather than vice versa.

An even more extreme example is provided by some simple non-linear systems

which are deterministic in one direction of time and stochastic in time reversed (Rosenblatt, 1980) and which might, if the simplicity argument were to be applied, be claimed to show the direction of the 'arrow of time'. The view taken here is that, although it may be of considerable interest to establish these simplicities of structure, use of the word causal is unwise for the reason stated above.

There are several morals for statistical analysis. First we should aim for models that are at least potentially causal. For this, account should be taken of the nature of variables as explanatory, intermediate response variables (possibly at several stages) and response variables, i.e. to recognize relations between variables of the type discussed above whenever it is reasonable on subject-matter grounds to take these relations as at least a provisional basis for interpretation.

Further it seems reasonable that models should be specified in a way that would allow direct computer simulation of data, the argument being that if we cannot simulate data directly from the model how can nature have used this form to generate the data under analysis? This, for example, precludes the use of y_2 as an explanatory variable for y_1 if at the same time y_1 is an explanatory variable for y_2. Or again, specification of a multivariate normal set of variables by their mean and covariance matrix is, from this particular viewpoint, unacceptable. Of course, in both cases mathematically equivalent acceptable specifications are possible. The point, however, is that, if the model equations themselves and the individual coefficients in them are to be given substantive interpretations, then different mathematically equivalent specifications are not scientifically equivalent. There is an extensive discussion of alternative formulations of models in the econometrics literature.

Care in the simultaneous use of variables of the same status arises even when these are purely explanatory, i.e. at the same conceptual level, especially when the variables are of very similar nature. For example, suppose that log-systolic and log-diastolic blood pressures are used as explanatory variables in a regression equation, say for survival time. Then the regression coefficient on, say, log-diastolic blood pressure gives the effect of a unit increase in log-diastolic pressure with systolic pressure held fixed. This interpretation is, however, very artificial because of the intimate relation (not the same as high correlation) between the two components.

We may perhaps re-express the equation in derived explanatory variables, such as the average and difference of the log-systolic and log-diastolic pressures, hoping that way to capture all or most of the dependence on the first and easily interpreted component. If this does not work we may have to abandon a meaningful interpretation of the individual coefficients and concentrate on the combined dependence.

In other contexts, it may be that the ratio of regression coefficients is more easily interpreted than the regression coefficients themselves. For example, the ratio of the regression coefficients on x_2 and on x_1 specifies the increase necessary in x_1 to achieve the same effect as a unit increase in x_2, in each case all other variables being fixed. This may sometimes be more directly interpretable than the regression coefficients themselves and is in any case relatively insensitive to the detailed specification of the model. An example (Solomon, 1984; Struthers and Kalbfleisch, 1986) is that fitting a proportional hazards regression model to survival data when an accelerated life regression model is appropriate gives virtually proportional regression coefficients when the same explanatory variables are used, as Gore *et al.* (1984) had found empirically in a study of survival of breast cancer patients, taking as explanatory variables tumour size, menopausal state, etc. This suggests that ratios of coefficients may sometimes be

a valuable guide in trying to bridge the gap between empirical studies of the kind just described and causal processes.

Aims of the study of associations include the obtaining of bases for empirical prediction, an aspect not discussed here, and the isolation of conditional independence structures which may prove the basis for developing causal explanations.

4. CAUSALITY VIA QUASI-EXPERIMENTS

We now turn to a different approach to causality based on ideas from the theory of randomized experimental design. See Rubin (1974) and for an introductory account and discussion see Holland (1986).

Suppose initially that we have a single potential causal variable C, sometimes called a treatment or quasi-treatment, which can take one of two forms 0 and 1, say, and that there are individuals or units each of whom experiences one of the forms of C followed by the measurement of a response y. To begin with, we suppose y to be a continuous variable. Now suppose that conceptually any unit might have received either level of C and that it is reasonable to assume unit–treatment additivity, i.e. the response that would be observed on any unit under $C=1$ differs by a constant Δ from the response that would be observed on that same unit were it to receive $C=0$. Especially if $C=0$ represents a control, it is convenient to call the hypothetical response at $C=0$ the base-line response. We can then call Δ, the difference between the response at $C=1$ and the base-line response, the causal effect of changing C from 0 to 1. For some of the considerations in a comparable discussion for binary responses, see Copas (1973). There are many generalizations, such as to C with more than two levels or having factorial structure, the incorporation of explanatory variables and multivariate responses, and so on, but the essential points are best illustrated in the simplest situation as will be done in the following.

The assumption of unit–treatment additivity is not directly checkable, i.e. it involves non-observables, because it is of the essence that the response can be observed on any unit only for one level of C. The estimation of Δ involves the comparison of different units; note, for example, that in a crossover design a unit is a combination of a physical individual (patient or experimental animal etc.) and a time slot. There are thus two rather different initial issues. Is the assumption of unit–treatment additivity reasonable and, even if it is, can Δ be estimated from the data available?

Over the first issue, although the unit–treatment additivity is not directly checkable, it is at least partly so. If suitable explanatory variables are available on each individual, a check for treatment by explanatory variable interaction can be carried out. Even without such additional variables, under unit–treatment additivity, the variance of response should be the same in the two treatment arms and indeed the distribution functions should be translations of one another. If that condition is not satisfied, a non-linear transformation of the response will be called for if unit–treatment additivity is to be achieved, although if the distribution functions intersect the failure of the assumption is not correctable by transformation.

The more difficult aspect concerns the estimation of Δ. By a familiar argument, in a randomized trial in which the allocation of units to treatments is under the investigator's control and involves appropriate randomization, the two groups of individuals receiving different treatments differ only by the accidents of random

sampling and the associated uncertainty in the standard estimate of Δ can be calculated. There is a very powerful qualitative argument for randomization as a device for achieving initial comparability of groups of individuals to be treated differently.

If the allocation of individuals to treatments is under the investigator's control and is done in a clearly defined way not involving randomization, then again estimation of Δ is possible, a general subject-matter judgment being necessary that the allocation procedure is equivalent to randomization or at least involves no serious bias. The standard probability calculations do not cover this other than by assumption. In an observational study, however, the process determining treatment allocation will typically be largely or wholly unknown. Some strong and largely untestable assumption is therefore necessary if a parameter representing a causal effect is to be estimated, so that the uncertainty in any final estimate exceeds, often substantially, that assessed by the usual statistical calculations. More precisely, the usual procedure is to adjust the comparison between treatment arms for imbalance on explanatory variables x and then the assumption involved is that treatment allocation is conditionally independent of base-line response given x. Essentially this amounts to saying that there is no further explanatory variable, possibly unobserved, which differs appreciably between treatment arms and which is residually correlated with the base-line response after adjustment for x. When this condition is satisfied to a reasonable approximation, estimation of the causal parameter Δ is possible.

In this discussion, only those variables which in the context in question can conceptually be manipulated are eligible to represent causes, i.e. it must make sense, always in the context in question, that for any individual the causal variable might have been different from the value actually taken. Thus in most situations gender is not a causal variable but rather an intrinsic property of the individual. The study of sex-change operations and of possible discriminatory employment practices would be exceptions. Again, the passage of time as such is not a causal variable. For example, in the study of the stress–strain–time relation in textile fibres, to say what would have been the response (strain) at the end of a period of time if time had not passed is meaningless; to ask what the response would have been after that time if certain processes of molecular rearrangement had been inhibited would, however, be sensible, i.e. processes going on in time can in this sense be regarded as causal.

Often it clarifies the interpretation in studies of dependence to classify the explanatory variables as

(a) treatments (or quasi-treatments), i.e. as potentially causal,
(b) intrinsic properties of the individuals under study or
(c) non-specific, representing blocks, replicates, centres, countries, etc. likely to correspond to differences but typically having many different identifying features.

When intermediate response variables are used as explanatory variables the same classification may be useful. The terminology used in making such distinctions varies quite widely.

The classification depends strongly on the context. For example, in a randomized trial comparing alternative treatments for hypertension, response being the occurrence of a critical event in say 2 years, blood pressure at entry would be an intrinsic variable characterizing individual patients and possibly as a basis for study of

treatment × patient interaction. In-study blood pressure would be an intermediate response variable and might also be regarded as a quasi-treatment pointing towards but not providing a definitive answer to the role of blood pressure control in avoiding critical events, i.e. it would aim to answer questions about the possible effect on the occurrence of a critical event of an individual's blood pressure being changed.

The definition of causality in this section, with its interventionist emphasis, seems to capture a deeper notion than that outlined in Section 3. Nevertheless there remains a major limitation and a major lack of clarity.

The limitation concerns the absence of an explicit notion of an underlying process or understanding at an observational level that is deeper than that involved in the data under immediate analysis. This seems to be an important part of the general scientific notion of causality. It is not that such an explanation will be 'ultimate' but rather that it should relate the phenomenon under study with some knowledge at a different level, for example an epidemiological finding to some biochemical or immunological process. It is at least in part this thought that presumably lies behind Fisher's dictum mentioned in Section 2.

The lack of clarity arises from the need to specify what is being held constant when the hypothetical change in causal variable is made. This is, of course, directly connected with the question of what variables are included in a regression equation in addition to the variable representing the potential causal variable. Several choices may be needed. For example, if the causal variable represents some aspect of alcohol intake, the question of food changes when the alcohol intake changes is clearly relevant. In a randomized clinical trial the usual 'intention to treat' analysis assesses the effect of one treatment compared with another, where the imposition of a treatment carries with it any other changes that the experimental set-up allows. For example, if supplementary medication is not controlled and is very different in the different treatment arms, then the causal effect of a treatment includes the consequences of the supplementary medication associated with that treatment. The issue of what is allowed to vary as treatment varies can be crucial to the interpretation.

In summary, we have discussed in a little detail two definitions of causality. The first, that developed in Section 3, is very useful as a guide to the kind of empirical model that is suitable for summarizing especially of observational studies; it has, however, been argued that it is too far from the underlying explanation for the use of the word causal to be wise. The second based on a notion of hypothetical intervention is closer to the physical notion of causality. My preference, however, is to restrict the term to situations where some explanation in terms of a not totally hypothetical underlying process or mechanism is available. As noted above, the need to specify precisely the nature of the intervention contemplated is, in any case, often of key importance.

5. SOME MISCELLANEOUS ISSUES

The previous sections have been devoted to rather general issues and to their immediate statistical implications. We now turn to some more detailed matters.

5.1. *Latent Variables*

It is a consequence of the wish to establish a link with underlying mechanisms that it may often be necessary to invoke latent explanatory variables, i.e. variables that are

not directly observed. These are of broadly two types: variables measured with an error that is so large that a non-trivial distortion of the dependence on the underlying notional true value is introduced and variables that are constructs from several sources and in the data under analysis not observable even with error.

An allowance for measurement error is a technical statistical issue that is relatively straightforward for standard linear models provided that information is available about the covariance matrix of the errors. In other situations similar corrections are more complicated; see, for example, Armstrong (1985), Stefanski and Carroll (1987) and the brief discussion by Cox and Snell (1989).

In all cases two rather different effects are involved. Individual regression coefficients are attenuated and, more seriously for qualitative interpretation, the relative importance of different explanatory variables may be distorted, the importance of variables measured with relatively high error being downgraded in favour of correlated variables with relatively smaller error. Dr Valerie Beral has pointed out a possible instance of this. In some of the early work on Aids, numbers of sexual partners and the use of 'poppers' (amyl nitrite pills) were used as explanatory variables for the progression to Aids and the second variable appeared, presumably wrongly, as the predominant one in a logistic regression analysis (Vandenbroucke and Pardoel, 1989). It is conceivable that this effect was obtained because of very different measurement errors in the two variables.

The second type of latent variable is a construct usually from a linear combination of observed variables with unknown coefficients plus an error term. Such variables are then considered the basis for relationships that are usually linear; see, for example, Joreskög (1977). Such models are sometimes called causal, but it should be clear that they are causal only by explicit prior assumption and do not establish causality empirically from data. Care is also needed in the interpretation of the coefficients in the derived latent variables in that they are not in general regression coefficients.

Another important role for latent variables is in the provision of hypothetical processes for the explanation of particular patterns of conditional independence that may be observed empirically but which have no simple direct interpretation in terms of recursive systems of regression relations (Wermuth and Cox, 1991).

5.2. *Hierarchical Variation*

In some situations, observational and experimental, variation is encountered at several levels. For example, in an epidemiological, sociological or educational study, data might be obtained on individuals, grouped by areas within countries and by countries. The relation between response and explanatory variables might then be different between individuals within areas, between areas within countries and between countries. This raises two different issues. The first is that in the presence of hierarchical error structure, typically unbalanced, an efficient estimation of regression parameters and variance components requires special techniques and software and raises technical problems if the analysis is not essentially based on normal theory. Note, however, that simpler methods in which in effect each stratum of error is given a separate unweighted analysis will often give insight and be a valuable starting point for a notionally more efficient analysis.

A more important matter, however, is that it will quite often be wise to see whether the regression coefficient on a particular explanatory variable takes the same value in

the various strata. There can be many reasons why differences in regression coefficient between the strata might arise, e.g. because of unmeasured explanatory variables operating at the higher levels or because measurement errors in the explanatory variables could have different effects on the different variables at different levels. Other things being equal, covariation at a low hierarchical level is most likely to be linked to a causal process. The overinterpretation of a regression at a high hierarchical level is sometimes called the ecological fallacy.

5.3. *Causality in Randomized Experiments*

If the ideal conditions of randomization theory are closely realized in practice then the causal effect of treatments in the sense of Section 4 can be estimated giving what in clinical trial terminology is called the analysis by intention to treat. This is a powerful argument in favour of randomization. However, the causal consequences of a treatment arm stem from all the effects of being allocated to that arm. Thus if, in a clinical trial comparing two treatments, compliance is very poor in a new treatment it may well be required to separate the effect of non-compliance from the genuine biological effect of the new treatment but, because non-compliance is a non-controlled intermediate response that is unlikely to be independent of the main end response of the clinical trial, assumptions will be involved that are very similar to those involved in analysing observational studies. See, for example, Efron and Feldman (1991).

5.4. *Generalizability*

Somewhat connected with the issue of causality is that of the generalization of conclusions especially in a somewhat applied context. Yates and Cochran (1938), in a thorough examination of the principles of combining evidence from several studies, stressed the importance of independent replication, in their case of replication across sites and years. This emphasizes the significance at an empirical level of establishing stability of effect across studies and more broadly of, if possible, showing the absence of an important interaction with intrinsic explanatory variables as a basis for generalization; if important non-removable interactions are found these may establish explicit limits on some otherwise superficially broadly applicable conclusion or recommendation. Even if homogeneity is established across replicate studies, it is possible that the effect of interest is very heterogeneous with respect to an important but unobserved explanatory variable. Thus it is possible that a treatment effect shows no interaction with observed intrinsic variables and is stable across replicate studies, yet has a strong interaction with an unobserved explanatory variable. If further that variable has a distribution in a target population that is very different from that in the data, then seriously misleading conclusions will result.

The relevance of causal processes is that, if one such is reasonably well understood for the situation under study, it is likely to give a clearer understanding of when conclusions from a study or set of studies can be applied more widely.

5.5. *Role of Regression Analysis*

In the statistical discussion above, primary emphasis has been placed on regression analysis in the study of the dependence of one or more response variables on

explanatory variables. This includes the use of chains of relations in which certain variables enter first as (intermediate) response variables and then as explanatory variables for the final responses. The object of such analyses is broadly to examine and in some cases to suggest substantive research hypotheses (Wermuth and Lauritzen, 1990) concerning dependences and independences. From these, suggestions may arise about potential causal relations.

In regression analysis the explanatory variables are regarded as held fixed at their observed values, even if they are randomly distributed, as indeed they normally would be in an observational study, i.e. the distribution of explanatory variables is not used in the analysis, except for its role in determining the precision of the estimated regression coefficients. In fact, however, some consideration of the distribution of the explanatory variables is desirable, in particular to obtain an idea of the conditions under which the validity of primary conclusions has been established. Thus any discrepancy between the distribution of explanatory variables as between replicate sets of data or between the data and known features of a target population should be noted. In particular an appreciable discrepancy may point to selection biases in the data having major implications for interpretation.

ACKNOWLEDGEMENTS

I am very grateful to Nanny Wermuth for many conversations on the topic of this paper and for comments on a preliminary version and to Michael Sobel for a copy of an unpublished paper of his on causality.

REFERENCES

Armstrong, B. (1985) Measurement error in the generalized linear model. *Communs Statist. Simuln Computn*, **14**, 529–544.

Barnard, G. A. (1982) Causation. In *Encyclopedia of Statistical Sciences* (eds S. Kotz and N. L. Johnson), vol. 1, pp. 387–389. New York: Wiley.

Bradford Hill, A. (1937) *Principles of Medical Statistics*. London: Arnold.

——— (1965) The environment and disease: association or causation. *Proc. R. Soc. Med.*, **58**, 295–300.

Cochran, W. G. (1965) The planning of observational studies of human populations (with discussion). *J. R. Statist. Soc.* A, **128**, 234–265.

——— (1972) Observational studies. In *Statistical Papers in Honor of George W. Snedecor* (eds H. A. David and H. T. David), pp. 77–90. Ames: Iowa State College Press.

Copas, J. B. (1973) Randomization models for the matched and unmatched 2×2 tables. *Biometrika*, **60**, 467–476.

Cox, D. R. and Snell, E. J. (1989) *Analysis of Binary Data*, 2nd edn, sect. 3.4. London: Chapman and Hall.

Efron, B. and Feldman, D. (1991) Compliance as an explanatory variable in clinical trials (with discussion). *J. Am. Statist. Ass.*, **86**, 9–26.

Gore, S. M., Pocock, S. J. and Kerr, G. R. (1984) Regression models and non-proportional hazards in the analysis of breast cancer survival. *Appl. Statist.*, **33**, 176–195.

Holland, P. W. (1986) Statistics and causal inference (with discussion). *J. Am. Statist. Ass.*, **81**, 945–970.

Joreskög, K. G. (1977) Structural equation models in the social sciences: specification, estimation and testing. In *Applications of Statistics* (ed. P. R. Krishnaiah), pp. 267–287. Amsterdam: North-Holland.

Pearl, J. and Verma, T. S. (1991) A theory of inferred causation. In *Principles of Knowledge Representation and Reasoning* (eds J. A. Allen, R. Fikes and E. Sandewall). San Mateo: Morgan Kaufmann.

Rosenbaum, P. R. (1984a) From association to causation in observational studies. *J. Am. Statist. Ass.*, **79**, 41–48.

——(1984b) Conditional permutation tests and the propensity score in observational studies. *J. Am. Statist. Ass.*, **79**, 565–574.

——(1987) Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, **74**, 13–26.

Rosenbaum, P. R. and Rubin, D. B. (1983) The central role of the propensity score in observational studies for causal effect. *Biometrika*, **70**, 41–55.

Rosenblatt, M. (1980) Linear processes and bispectra. *J. Appl. Probab.*, **17**, 265–270.

Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, **66**, 688–701.

Solomon, P. J. (1984) Effect of misspecification of regression models in the analysis of survival data. *Biometrika*, **71**, 291–298.

Stefanski, L. A. and Carroll, R. J. (1987) Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74**, 703–716.

Struthers, C. A. and Kalbfleisch, J. D. (1986) Misspecified proportional hazard models. *Biometrika*, **73**, 363–369.

Stuart, A. (ed.) (1971) *Statistical Papers of George Udny Yule*. High Wycombe: Griffin.

Vandenbroucke, J. P. and Pardoel, V. P. A. M. (1989) An autopsy of epidemiological methods: the case of "poppers" in the early epidemic of the acquired immunodeficiency syndrome (AIDS). *Am. J. Epidem.*, **129**, 455–457.

Wermuth, N. and Cox, D. R. (1991) Explanations for nondecomposable hypotheses in terms of univariate regressions including latent variables. To be published.

Wermuth, N. and 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. and Cochran, W. G. (1938) The analysis of groups of experiments. *J. Agric. Sci. Camb.*, **28**, 556–580.

Yule, G. U. (1903) Notes on the theory of association of attributes in statistics. *Biometrika*, **2**, 121–134.