8 The limits of statistics: causal explanation

While statistically informed methods of data collection and analysis are foundational in establishing the probabilistic population regularities that constitute sociological explananda, statistical analysis alone cannot lead to causal explanations of these regularities.

In preceding chapters, the emphasis has been on the crucial part played by statistically informed methods of data collection and analysis in forming the objects of study – the proper explananda – of sociology as a population science: that is, probabilistic population regularities. Such methods would appear essential – no viable alternative to them has been demonstrated – as the means of accommodating the variability of human social life, while at the same time enabling regularities within it to be discovered and adequately described. In the present chapter, however, the emphasis changes. It now falls on what statistical methodology by itself *cannot* achieve within sociology, and should not be expected to achieve: that is, the provision of causal explanations of the regularities that this methodology serves to establish. Dudley Duncan (1992: 668) once spoke of sociology's 'Faustian bargain' with statistics; it is some aspects of the darker side of this bargain that have now to be considered.

As noted in Chapter 7, in early applications of multivariate data analysis in sociology it was often supposed that such analyses could not only reveal various regularities of association in social data but also lead to the demonstration of causal relationships. In particular, forms of regression analysis, culminating in causal path modelling, were seen as the prime means of moving from association to causation. However, over recent decades this view has become subject to increasing doubts and criticism from within statistics and sociology alike, and it would by now appear to receive rather little overt support. At most, it might be claimed that regression analyses allow for a 'causal interpretation'.[1]

From the side of statistics, perhaps the most telling interventions have been those of David Freedman (1991, 1992, 1997, 2010). Freedman explicitly rejects the idea that causal relationships can, in any field of research, be cranked out of data on the basis of statistical technique alone. Thus, in the case of regression, he stresses that if a regression model is to be appropriately specified for purposes of causal explanation, a prior subject-matter input is always required regarding the processes by which the data under analysis have actually been generated. This is necessary in order to determine the variables that are to be included in the analysis, their supposed causal ordering, the functional form of the relations between them, the properties of error terms and so on. If the regression model adopted is not consistent with the processes in question – if, to take the most obvious example, a relevant variable is omitted – then all causal inferences drawn from the model will be vitiated. Freedman accepts that the problems that here arise may be less consequential if regression is taken simply as a means of description: that is, as a means of summarising the associations that happen to hold in the data analysed among the variables that are in fact recognised. But if causal inference is the aim of the analysis, the crucial issue that arises is that of whether the regression coefficients returned can be taken to have what Freedman (1997: 117) aptly calls 'a life of their own': that is, outside of the data from which in any particular case they are estimated. This will only be so if the regression model is correctly specified, and only then will counterfactuals implying causation be licensed. In other words, only then

[1] A good deal of confusion, or equivocation, is in this regard made possible by the fact that the language in which regression is discussed is permeated with terms with apparent causal implications: 'effects', 'determinants', 'dependence' and so on. It would be difficult at this stage to introduce any alternative terminology, but it would be helpful if more sociologists were to follow the practice of making it clear when they are using such apparently causal language simply as a *façon de parler.* The term 'statistical effects' is sometimes used to indicate that 'causal effects' are not being implied.

will it be possible to claim that if a particular independent variable were to be changed, a corresponding change – of a size indicated by the coefficient for that variable – would be caused in the dependent variable of the analysis.

Freedman (2010: 11–15) illustrates this line of argument with the example of regression analysis as applied in the case of Hooke's Law. This law states the qualitative nature of a relationship: that is, that, up to a limit, the extension of a spring is in direct linear proportion to the load added to it. Using experimental or observational data in order to regress extension on load then serves to quantify the law in the case of a particular spring or type of spring. The law itself reflects the physical process through which the data will have been produced, and the form of the law is in turn embodied in the regression model. In this case, then, a high R^2 can be expected, with the error term reflecting simply measurement error – from, say, deficiencies in the experimental set-up or faulty observation; and the regression coefficient estimated will be interpretable as capturing *a specific property* of the spring or type of spring: that is, as having in this sense a 'life of its own' rather than one conditional on the particular data analysed. In turn, therefore, this coefficient could be used, within a limit, in order to predict what change in the extension of the spring would be caused by a change in load.

Here, one has in fact a clear example of what, following Xie (see pp. 92–3), one could characterise as the Gaussian conception of regression. What has, however, to be recognised is that, in sociology, regression can rarely, if ever, be applied under this conception: the kind of theory exemplified by Hooke's Law is simply not available as a basis. Consider, for example, analyses that, using data for a particular place and time, regress the earnings or occupational status of individuals in different ethnic groups on their levels of educational attainment (with, say, various control variables being also included). What has here to be accepted is, first of all, that there are no strong grounds for supposing that the coefficients estimated will be replicated with data for other places and times – they may or may not be.

But further, and more seriously, it has also to be accepted that if, in the original case, the distribution of education were to change – if, say, educational differences among ethnic groups were to be narrowed – the changes in ethnic differences in earnings or status that would be predicted under the regression model would not necessarily follow. It could just as well be that change would occur *in the regression coefficients themselves*: for example, that the earnings or status returns to education would fall (see Lieberson, 1987: 166–7, 186–8). The underlying problem is that no general and compelling theory exists of the processes whereby such returns to education are actually produced that could guide the specification of the regression model.[2]

A growing realisation of the difficulties involved in seeking to derive causal explanations from regression of the kind highlighted by Freedman is, then, one source of the re-evaluation of regression as essentially a descriptive method that, as noted in the previous chapter, has become evident in sociology – and that in turn lends importance to Xie's argument that a Galtonian rather than a Gaussian conception of regression is that which sociologists do now in fact mostly apply and should explicitly adopt.[3] However, criticism of the idea of moving from association to causation via regression has been made not only from a statistical standpoint but further, and also influentially, from within sociology itself.

What is in this regard notable is the degree to which such criticism has run in parallel to that made by Freedman: that is, in

[2] Some economists would appear to believe that, so far as the earnings returns to education are concerned, human capital theory fills the bill. However, how exactly educational attainment is to be related to the concept of human capital and what control variables – such as, say, cognitive ability or various non-cognitive attributes – it is appropriate to introduce appear far from clear. In the evaluation of human capital theory, this then allows, in Blaug's (1992: 218) words, 'the persistent resort to ad hoc auxiliary assumptions to account for every perverse result'.

[3] An important turning-point in this process was perhaps marked by Freedman's (1992) critique of causal path analysis as deployed in Hope (1984), although, as was generally recognised (see Duncan, 1992), this was simply a restatement of an earlier critique of this methodology in Blau and Duncan (1967) that had been widely circulated as a working paper.

emphasising the need for an understanding of the processes that generate the data that are under analysis. While, for Freedman, such an understanding is necessary if a regression model is to be properly specified, for many sociologists, revealing such processes is in fact what establishing causation in sociology essentially entails, independently of any statistical procedures. It is this view that motivates the objection to 'variable sociology', referred to in Chapter 7, that it reduces sociological explanation simply to showing how far dependent variables can be statistically 'accounted for' by those treated as independent, without any attention being given to the social processes that underlie the results obtained. And, at least for sociologists working within the individualistic paradigm, these processes are ones that need to be ultimately understood in terms of the actions and interactions of the individuals involved. Thus, Coleman (1986: 1314–15) has pointed to the paradox that sociologists engaged in 'empirical, statistical survey research' largely analyse individual-level data yet very frequently do so without any explicit reference to the individual action from which these data derive. Likewise, Boudon (1987: 61–2) has objected that in such research it is too often variables rather than individuals that are in effect taken as the units of analysis, and that demonstrations of the statistical effect of one variable on another are considered as 'final results' without any attempt being made to show how these statistical relations derive from their 'real causes', which can lie only in the actions of individuals.

Thus, to revert, for purposes of illustration, to regression analyses of the effects of individuals' education on their earnings or occupational status, what would be maintained from the position taken up by Coleman and Boudon is that if education could, on this basis, be said to 'cause' earnings or occupational status, this could only be in a very elliptical and sociologically uninformative sense. What would be further needed for an adequate demonstration of causation would be some account of why the statistical results come out as they do. For example, some account would be needed, on the one hand, of the processes – involving choice and constraint – through which individuals attain certain educational levels, with perhaps their economic futures in mind (see further pp. 117–19); and, on the other hand, of the processes through which this attainment then conditions their chances of entry, via the actions of employers or their agents, into occupations affording differing levels of earnings or status (Goldthorpe, 2014).

In sum, the argument is, again as with Freedman, that establishing causation cannot result from statistical procedures alone but must be dependent upon some subject-matter theoretical input relating to how the data under analysis are produced. The main difference is that while, for Freedman, causal inference from statistical analysis requires such input from the start, for Coleman and Boudon such input has to follow on from the results of statistical analysis, in themselves essentially descriptive, if a causal explanation is to be provided of the regularities that are demonstrated.

The question of how sociologists can in fact best seek to move from the description to the explanation of population regularities will be the specific concern of Chapter 9. In the remainder of the present chapter, attention needs to be given to a further reaction that has developed within sociology to the ending of 'the age of regression' – but one which would, again, appear to lead to undue expectations about the part that can be played in establishing causation by statistical methodology alone. This reaction derives from a conception of causation that differs significantly from that underlying regression and related forms of analysis, and one which, its proponents would maintain, is of a stronger and 'deeper' kind.

With regression, as indeed with Lazarsfeld's attempts to derive causation from contingency-table analysis, causation is in effect equated with the existence of an association between the dependent variable of the analysis and the independent variable or variables *that can be shown to be robust*: that is, that can be shown to persist when other conceivably explanatory variables are introduced. It is this robust dependence that serves as the evidence that a causal relationship prevails. The main weakness of this understanding of causation is, of course, the difficulty involved in ruling out the possibility that the supposed causal relationship might be shown to be spurious if some further, hitherto unconsidered, perhaps entirely unsuspected factor – a 'lurking' variable – were to be taken into account. The alternative conception of causality that has of late been advocated for sociology seeks to avoid this difficulty. It derives from primarily applied research in such fields as agriculture, medicine and education, where it is possible for experimental or quasi-experimental methods to be pursued and where interest centres on the effectiveness, or otherwise, of some kind of *intervention*.

Under this alternative conception, causation, or a 'causal effect', is understood in terms of *the change that is produced* in a dependent or outcome variable of interest as the result of an intervention or of what is often referred to as a 'treatment'. More specifically, a causal effect is the difference found in the outcome variable as between randomly selected experimental units that receive a treatment and those that do not – that is, that serve as controls: for example, as between crop yields on plots of land given or not given a fertiliser, between the recovery rates of patients given or not given a drug or between the examination results of students subject or not subject to a particular pedagogical method.[4] This difference in outcomes can be quantified by calculating the 'average treatment effect', or some variant thereof, with the counterfactual and causally significant implication then being that, absent the treatment, no such difference would be observed. It is therefore crucial for the applicability of this 'potential outcomes' understanding of causation that an intervention in some form or other should occur – or, at least, could be envisaged as occurring: in Holland's (1986: 958) phrase, 'no causation without manipulation'.[5]

[4] The random allocation of units to the treatment or control 'arms' of the experiment is regarded as the crucial means of controlling for all possibly confounding variables, known or unknown.

[5] I have discussed this and other understandings of causation in sociology at greater length in Goldthorpe (2007: vol. 1, ch. 9).

An impressive body of statistical technique has evolved in connection with the potential outcomes approach: not only concerning the design of experiments to test the effectiveness of interventions – in particular 'randomised controlled trials' (RCTs) – but further, and of main relevance for present purposes, concerning the extension of the approach to non-experimental, observational studies, including those based on sample surveys of populations (see e.g. Rosenbaum, 1995). In this case, what is in effect entailed is considering these studies *as if* they were experiments, although ones not carried out under the control of the researcher, and then seeking means of counteracting the effects of departures from an appropriate experimental design that could follow from this lack of control. The major work reviewing these developments from the standpoint of sociology, and advocating their use by sociologists, is Morgan and Winship (2007). These authors specifically contrast the potential outcomes approach to causation with the regression or 'equation-based' approach and see the latter as being inimical to 'careful thinking about how the data in hand differ from what would have been generated by the ideal experiments one might wish to have conducted' (Morgan and Winship, 2007: 13).[6]

Taken on its own terms, the line of argument that Morgan and Winship pursue is a forceful one, and in attempts to develop causal arguments on the basis of data from observational studies, to adopt

[6] Morgan and Winship actually prefer to speak of the 'counterfactual' rather than the 'potential outcomes' approach to causation, and are also clearly influenced by the work of Pearl (2000), which can be seen as an attempt to re-express and develop the potential outcomes approach by introducing ideas from computer science, implemented through directed acyclic graphs (DAGs). However, much controversy in this regard persists. See, for example, the fierce exchanges in *Statistics in Medicine*, 2007–09. For a different approach to the use of DAGs, more consistent with the understanding of causation to be developed in Chapter 9, see Cox and Wermuth (1996: 219–27 esp.). These authors regard graphical methods as being able to provide representations of data that are 'potentially causal' – that is, that are 'consistent with or suggestive of causal interpretation' – while recognising that causality has itself to be established in terms of some 'underlying process or mechanism' derived from theory in substantive research areas.

the standpoint of experimental design may often provide a valuable discipline. However, the potential outcomes approach has not, to date, been at all widely taken up in sociology, and insofar as sociology is to be understood as a population science, this situation would seem unlike to change in view of the fact that, in this context, the approach gives rise to at least three significant difficulties.

First of all, what is necessarily involved in pursuing the potential outcomes approach is – to take up an important distinction already found in the work of John Stuart Mill (1843/1973–74) – a focus on *the effects of causes* rather than on *the causes of effects.* Some putative cause of an outcome of interest is selected – whatever the motivation for this selection might be – and the aim is then to estimate its effect. This implies a very different orientation from starting out from effects, as, say, established population regularities, and then seeking a causal explanation of them. As already noted, the potential outcomes approach derives from, and has an apparent appropriateness in, applied research, including applied social research, where the effect of a given intervention is typically of prime interest: in other words, where the aim is to evaluate whether, or how far, some form of intervention – the cause – has achieved its objective in the sense of producing the effect that was sought.

However, the applicability of this approach to the central concerns of sociology as a population science would seem limited. Morgan and Winship (2007: 280) do in fact themselves acknowledge that if a researcher is primarily concerned with the causes of an observed effect, such as a demonstrated empirical regularity within a population, then the potential outcomes understanding of causation will be 'less helpful' (see also Gelman, 2011). And it is of further interest in this regard that, in a paper devoted to the treatment of causation in demography, as an established population science, Ní Bhrolcháin and Dyson (2007: 1–4) take the view that the 'interventionist' approach 'is not often applicable', because in demography 'the big questions are those about the causes of effects – what causes fertility change? what induces mortality decline?'.[7] And it might indeed be more generally maintained, following Popper (1972: 115), that 'In all sciences the ordinary approach is from the effects to the causes. The effect raises the problem – the problem to be explained . . . – and the scientist tries to solve it by constructing an explanatory hypothesis.'

Second, it has to be recognised that the potential outcomes approach does not escape from the criticism raised against attempts at demonstrating causation via regression that no account is provided of how the causal effect claimed is actually produced. Thus, Cox (1992: 297) would see it as a 'major limitation' of the approach that 'no explicit notion of an underlying process' is introduced 'at an observational level that is deeper than that involved in the data under immediate analysis'. And similar reservations would appear to underlie an increasing scepticism over recent years about whether *even in applied research* the potential outcomes approach – in particular as exemplified in RCTs – has necessarily to be taken as the 'gold standard' so far as causal inference is concerned.

An important social science example of such scepticism is provided by Deaton's (2010) discussion of methodologies for the evaluation of economic development assistance projects, as, say, in regard to poverty, health or education. Deaton is critical of the idea – as apparently now favoured by the World Bank – that RCTs should be given a

[7] Mahoney and Goertz (2006: 230–1) claim the 'causes of effects' orientation as a distinctive feature of qualitative sociology, in contrast to the 'effects of causes' orientation prevailing in quantitative sociology. This is, however, a quite groundless claim, and again illustrates the extremely limited – and self-serving – view of quantitative sociology (and political science) that proponents of logical methods of analysis characteristically take (see Achen's comment on Ragin in Chapter 4, n. 8). It may be added here that the approach to establishing the causes of effects in terms of INUS conditions (Mackie, 1974) that is favoured by Mahoney and Goertz comes up against exactly the same problems of moving from association to causation as occur in the case of regression. As Cartwright (2007: 34–5) puts it, '. . . INUS conditions are not causes. The INUS formula represents an association of features, a correlation, and we know that correlations may well be spurious'. For those unfamiliar with Mackie's work, an INUS condition is an *i*nsufficient but *n*ecessary part of an *u*nnecessary but *s*ufficient condition for some outcome to be realised. Of course, insofar as the social world, or at least our knowledge of it, has to be regarded as probabilistic, there are no necessary or sufficient causes anyway.

privileged position in such evaluation; and the main basis of his criticism is that, while RCTs may provide reliable information on *whether* particular projects succeeded, they can in themselves say little about *why*. In other words, they can say little about the 'underlying process', or the mechanisms, involved in success. It is, however, knowledge of these mechanisms, Deaton argues, that is crucial for determining the external validity of an RCT, or, that is, for determining how far a project that has been shown to 'work' in one population will work in another, perhaps quite different, population. For it is not the actual results of an RCT that can 'travel' – that can be generalised – from one such context to another, but only an understanding of the way in which these results were produced, and with then the requirement that due consideration be given to the conditions under which the mechanism in question is, or is not, likely to be maintained. (For further, more general and developed statements of essentially this position in regard to social policy formation and evaluation, see Pawson and Tilley, 1997; Cartwright and Hardie, 2012.)[8]

Third, where in sociology as a population science the attempt is made to give an account of the processes or mechanisms that create a causal relationship, this account is required, under the individualistic paradigm, to be one expressed ultimately in terms of individual action and interaction. However, this requirement then comes into direct conflict with Holland's maxim, basic to the potential outcomes approach, of 'no causation without manipulation'. This point is best

[8] Arguments running on much the same lines as Deaton's are also being advanced in the medical field in questioning whether clinical trials – often taken as the prime exemplars of the potential outcomes approach – should be viewed as setting the gold standard for evidence-based medicine (see e.g. Worrall, 2007; Steel, 2008; Thompson, 2011). Clinical trials do, however, tend to be more theoretically informed than RCTs carried out in the social field, and – as David Cox has pointed out to me – could often be regarded as attempts to further test ideas about mechanisms (see Chapter 9) that already have some empirical support, as, say, from laboratory work. A more appropriate extension of Deaton's argument would be to the predictions made from big data through entirely inductive, correlational analyses, where the concern is, quite explicitly, only with *what* and not with *why* (Mayer-Schönberger and Cukier, 2013: 4). Such predictions are, of course, heavily dependent on the future being like the past so far as underlying causal processes are concerned.

brought out by reference to the illuminating discussion that Holland himself provides of possible causal statements that are, and are not, compatible with this maxim. Holland (1986: 954–5) considers the following three statements (for present purposes, I have changed the order in which they appear in the original):

> She did well on the exam because she was coached by her teacher.
> She did well on the exam because she is a woman.
> She did well on the exam because she studied for it.

The first statement presents no problems from the standpoint of the potential outcomes approach: manipulation or an intervention – the coaching – occurred, and this can be taken as the cause of the woman doing well. With the second statement, some difficulty arises insofar as being a woman could be regarded as an 'intrinsic attribute' and thus one that is not open to manipulation. But, in sociology at least, it may often be possible to finesse such a difficulty by reinterpretation: for example, in the case in point, by taking 'because she is a woman' to refer not to unalterable biological sex but rather to socioculturally variable, and thus conceivably manipulable, gender. It is, however, the third statement that leads to a fundamental problem. This is so because, instead of there being any manipulation, the woman made what, to revert to the discussion of Chapter 3, could be regarded as an informed choice in pursuing a particular end: she wished to do well on the exam; she believed that studying for it was the best means to this end; she acted on this belief; and, her belief being correct, she did well.

As Holland observes, it is 'the voluntary aspect of the supposed cause' that here leads to incompatibility with the potential outcomes approach; and, he goes on, 'The voluntary nature of much of human activity makes causal statements about these activities difficult in many cases' (Holland, 1986: 955). But what, in effect, Holland has to be taken as saying here is 'difficult *given* the potential outcomes approach'. To which the response may be made that this in a further way indicates the limited relevance of this approach for sociology, or at all events for sociology understood as a population science. For, in this case, as has been maintained throughout, the ultimate aim is to give causal explanations of established population regularities in terms of social processes that are grounded in individual action – action in which a significant autonomous element, as expressed in informed choice and its implied rationality, has to be recognised.

An underlying issue here is a long-standing philosophical one of whether reasons can be causes. It was once fashionable to argue (see e.g., with specific reference to social science, Winch, 1958) that they could not be, since a cause has to be logically distinct from its effect and a reason is not logically distinct from the action to which it leads. But this view has subsequently lost favour to one in which individuals' reasons are seen as providing the basis for, perhaps a special, but still a quite legitimate form of causal explanation for their actions (see e.g. Davidson, 1980: ch. 1). It is, in effect, this latter view that I accept, and that I elaborate on in Chapter 9.