The Linacre Quarterly

Volume 36 | Number 4 Article 7

November 1969

Medical Data and Applied Ethics: Part II: The Sources of Data

Edmond A. Murphy

Medical Data and Applied Ethics

Part II*

The Sources of Data

Edmond A. Murphy, M.D., Sc.D.

In exploring the second of our problems – the analysis of the sources of information – it will be necessary to brush aside all the semantic difficulties considered previously. To illustrate certain principles about the sources of data and how they operate, it will be necessary to pretend that there are no epistemological difficulties over the use of such words as "cancer", or "disease".

It will also, I hope, be evident that the kinds of difficulties connected with determining the facts about disease, apply equally to what has been discovered about psychology, education or moral problems by the same method of inference. It is of the first importance to realize that these rules apply to all knowledge defended on empirical grounds alone.

It must not be supposed that any exclusive claims are being made for the value or the cogency of scientific evidence. There are doubtless many other channels through which knowledge can be attained. But the epistemological defense for the soundness of a conclusion must rest squarely in some field. A painter might claim that Rembrandt was a great artist because of his development of chiaroscuro; and a political theorist might claim he was a great artist because of his recognition of the dignity of the common man. To a scientist – in the abstract sense of a narrow specialist – both contentions would be unintelligible, now, and perhaps permanently. That this kind of rigid compartmentalizing of knowledge does not occur in practice reflects the broadmindedness of the specialist. But the rigorous examination of such a claim must be undertaken within some scheme of criticism. It may be that a system of defense could be elaborated which lay not in science or esthetics but at the interface between them; but for this to be possible such a system would have to be developed in its own right.

Suppose that some pastoral theologian expresses some opinion about, let us say, the effects of Catholic education or the dangers of mixed marriages. Insofar as his opinion depends on, for example, infused knowledge or on some deductions from a scriptural source, then he will meet no criticism from me and I must leave it to his peers to decide whether his conclusions are sound. But if he claims that his conclusion is based on scientific inference from experience, then in its empirical mode his conclusion must be subject to the kinds of discipline discussed here. I do not deny (neither do I affirm) that certain men of great holiness or percipience can arrive at pastoral wisdom which transcends the scientifically imperfect sources of their information. But such arguments leave me a little nervous if for no other reason then that formal demonstration of the truth of these conclusions is wanting. It has been a common belief within the medical profession that an analgous kind of transcendental wisdom can be attained in clinical matters. But where formal scientific studies have subsequently been performed, the beliefs arrived at by this method have distressingly often proved false. I hope it will not be thought an unduly hostile comment to say that most, perhaps all, of the experience and the pastoral wisdom which the Church has built up over the ages, *viewed strictly as scientific evidence* is almost all unsound.

The cardinal problem which besets the scientist who wishes to establish demonstrable truth, arises from the inductive nature of his discipline. The scientist capitalizes on the fact that there is some, not necessarily perfect, uniformity in the behavior of sensible phenomena. In principle all the facts that he needs to know can be found by direct observation except, of course, those facts which are beyond the resolving power of the available means of observation. An engineer wishing to decide whether a particular design of bridge will stand up or fall down, can always build it and see. But, apart from the expense and the danger of this method, any self-respecting engineer would view it with professional dissatisfaction; he would feel that such an approach does not do justice to the extent of the organization in the theoretical aspects of his field. Now clearly any alternative to this "try-it-and-see" approach implies a belief that natural phenomenon can be described in terms which are less numerous than the facts collected i.e. data can be *reduced.* This belief cannot be demonstrated, though it seems evident that if it were false, policies predicated on it would lead, sooner or later, to failure. But if this belief be granted, the recognition of truth in any particular field by empirical means is based on two conditions:–

1) That there be sufficient experience in the field; and

2) That what experience there is, be representative.

The violation of these principles leads on the one hand to problems of sampling error, and on the other hand to problems of bias. The two are not mutually exclusive in that they may coexist, but they differ fundamentally. The difference can perhaps be best illustrated by simple if rather transparent examples.

**Part I was printed in the August, 1969 Linacre Quarterly.*

Dr. Murphy received his M.D. degree from Queen's University, Belfast, Ireland. John Hopkins University awarded him the Sc.D in Biostatistics and he is Associate Professor of Medicine and Biostatistics at John Hopkins. His principal areas of interests have been in vascular disease, genetics, and the theoretical aspects of scientific inference.

Example I. A surgeon after performing some new and elaborate operation successfully three times might be led to claim that the mortality rate is zero and that the operation is without risk to life. Now one needs no sophisticated scientific training to recognize that such a claim is excessive. Even assuming that the surgeon has not selected his patients, it may well be that just by change he has each time picked patients with a good operative risk. His sample by common consent is too small for such a confident generalization. The only remedy is a more extensive experience. In his further cases, of course, the surgeon may go on being lucky. But the larger his experience the less likely it is that his success is attributable to good luck alone. There is always some possibility of this explanation; but with large numbers and uniform success, the reasonable interpretation would be that the risk is in fact small. Formal exploration of this point by statistical theory sustains the conclusions of common sense.

However, it is important not to get carried away in one's criticism. A commonplace comment is that, "You can tell nothing from three cases". This is manifest nonsense. However uncertain conclusions based on such a limited experience, *any* data are a vast improvement on none at all. It is nonetheless true that the uncertainties associated with sampling can be progressively attenuated by increasing the size of the sample.

Example II. Now by contrast, consider the therapeutic nihilist – an internist perhaps – who is convinced that some common operation is extremely dangerous. Since he does not have direct access to the surgeon's patients, he arranges with a pathologist to be notified every time one of these patients, who underwent the operation, comes to autopsy. On the basis of the information obtained from this source he proposes to asses the risks of the operation. It is surely obvious that all those coming to autopsy are dead; and since the nihilist gains no experience of any other kind of case, he concludes that the mortality rate is 100%.

(Of course, if his informant had been the administrator who arranges for taxis to transport the patient home, then he would be aware only of the patients who survive; and he would conclude that the mortality rate is zero percent.)

Such methods of ascertainment, like the observations from small samples, also lead to erroneous conclusions, but for quite different reasons. Here there is a *systematic* error in the selection of cases, one which large numbers will do nothing to correct. Mortality rates based on 10,000 patients coming to autopsy will be just as erroneous as those based on three patients only, and for precisely the same reason. Such a systematic error is called a *bias.* The basic fault is that the sample on which the conclusions are being based is not representative of the population about which the generalization is being made. It may be accidentally true that the autopsy patients are representative because it may be that the particular operation is invariably fatal. But this will rarely be true and the whole point at issue is whether such is the case or not.

This problem of obtaining representative (as distinct from adequately large) samples must be met by exact ideas and methods. To decide the fate of patients undergoing operation A, at a particular hospital* it would be necessary to consider all those admitted or some kind of a systematically collected sample of them. If the mortality rate is 10% then for every case going to the morgue, nine will be discharged eventually: and the sampling procedure should reflect these proportions. The investigator cannot simply collect his data in a hap-hazard way and hope that the proportions will come out right.

How a sampling method is best used in an individual case is a complicated subject (1,2) which could not possibly be dealt with here even in outline. The points that matter are to be aware of the importance of representativeness and to recognize that non-representativeness (or bias) is not cured by taking a large sample; the statistician may be able to adjust for the effects of bias if the nature of the bias is known. But there are many cases which cannot be cured by any analytical finesse.

To bring all this a little closer to home, let us apply it to a hypothetical pastoral example. Suppose that the pastoral theologian wanted to express an opinion as to whether the miniskirt is a scandalous garment. He could, of course, base his opinion on his own individual reactions. But generally, I think his opinion would be formed in the light of what the average person's reaction was. Now how would he get information about the reactions of the average person? The confessional would be a poor source. The matter will be raised in the confessional for the most part by people to whom it has been either an occasion or a source of sin. Such people may represent a small proportion of the population only. If such a garment were a source of *virtue* to some, perhaps large, segment of the population – and as someone trying to make a balanced judgment, the theologian could hardly dismiss this possibility out of hand – then the means of collecting information will be silent on the point. Graham Greene has made the point well.

> "... A priest only knows the unimportant things".
> "Unimportant"?
> "Oh, I mean the sins", he said impatiently. "A man doesn't come to us and confess his virtues". (3)

The collection of information in the confessional is closely allied to that of the nihilist who is opposed to the surgical operation. The only information obtained at all is to the detriment of the garment and the defect is not remedied by collecting a vast amount of information. The experiences of fifty pastors hearing confessions will merely be the bias of one magnified fifty times.

The second line of information might be those who come to consult the priest outside the confessional. This source would be perhaps a little better since it is not quite so closely concerned with sin; but the improvement would be only slight. A priest in such a situation might build up an extensive experience; but it would be an extensive experience of certain kinds of person. There is no reason to suppose them representative and indeed a good deal of reason to suppose the contrary.

Much of this, of course, does not matter. In the difficult area of purity the usual practice has been to treat penitents individually, recognizing that what may be quite harmless to one may not be to another. Yet the problem of modesty in dress (which does not involve one person only) cannot

**There is, of course, no reason for supposing that the experience in this hospital is representative of that in hospitals generally.*

be dealt with in these terms. And, at least in the past, there has been an almost total disregard for this principle of individuality in the censorship of films and books. In neither case, to my knowledge, has there ever been any pretence of consulting representative members of the laity before passing judgment.

The next phase of the problem arises out of this very notion of individualism. However necessary it may be to "tailor-make" spiritual direction in practice, it would be unmanageable in theory. If books on morals are to be written at all, then I suppose that they must contain broad principles.

But a generalization always involves some distortion of the truth and in general, the more extensive it is the greater the distortion. The statement, "University professors are highly educated people", is only roughly true: there are exceptions. But in the generalization "Most white-collar workers are moderatley well educated", the exceptions are more numerous and differ more widely from the average state. The second generalization of course deals with a larger and, necessarily, more heterogeneous group of people.

Thus the objectives of generality (simplicity) and accuracy are in conflict; conditions which favor the one are inimical to the other. In the nature of things a compromise is necessary. Grouping or "stratification" may be used in an attempt to produce groups which are more or less homogeneous and yet large enough to avoid those problems arising from small samples considered earlier. They should be numerous enough for the generalizations to be reasonably accurate, and yet not so numerous as to make it difficult to detect any broa[illegible]er-all trends.

For example, the preval[illegible]s of coronary disease are not the[illegible]ne at all ages. A common practi[illegible]s to compute prevalences in [illegible]ople grouped by age at intervals [illegible]ve or ten years. This arrangement m[illegible]s the requirements pretty well; and [illegible] it is of sufficient simplicity to all[illegible] such general statements as that th[illegible]sease is rare before the age of 30 [illegible] that the prevalence increases stead[illegible] with age. Both these facts would [illegible]ainly have been missed if prevale[illegible] had been computed without regard[illegible] age; and if no grouping had been em[illegible]oyed at all, but individual patients [illegible] had been studied, the relationship[illegible] age would be much more diffi[illegible]t to appreciate by inspection espe[illegible]ly in view of the fact that the num[illegible]rs of persons at risk also changes w[illegible] age.

Again, I am sure pastoral a[illegible]ogies to this heterogeneity of behav[illegible] will leap to mind. Men's moral p[illegible]lems are different from womens', ch[illegible]rens' from adults' and longshoremen[illegible] from those of professors of d[illegible]natic theology. In particular, if one[illegible]ished in the light of scientific critieria[illegible]o lay down broad principles of c[illegible]duct deduced from empirical data, [illegible]en it would be necessary to collect in[illegible]rmation in such a way that by anal[illegible]s the risk for each group, and ev[illegible] for each patient*, could be dete[illegible]ined separately.

A common statement is to the [illegible]ffect that "you can prove anything[illegible] with statistics:" and the claim is [illegible]ound provided that what has been pr[illegible]ved is capable of statistical exploratio[illegible] and also of course, provided that it i[illegible] true.

There are elaborate statistical methods which allow risk to be assessed without gr[illegible]uping. (4,5)

But the implication that one can prove things which are untrue can only be based on prejudice or on having experienced the results of the perversion of statistical method. Probably the commonest perversion is the misuse of generalizations about heterogeneous populations. This is not to say that no generalizations are possible in such cases. But they are apt to be brittle.

For example, official figures may show that the prevalence of lung cancer in country A is 2%. This may be a sound figure, and from certain standpoints, a useful one. A health administrator in deciding how extensive the provision which must be made for the problems of lung cancer is, would find such information useful. In country B, on the other hand, the prevalence may be 4% and this figure will also be of value to the administration.

But to push these figures futher is to court danger. The propagandist in country A, either through stupidity or dishonesty, may attempt to prove, by comparing these figures, that A is a healthier country to live in than B. But it may be that there are more elderly people in country B; and since cancer affects older rather than younger patients, it is hardly to be wondered at that lung cancer is also more common. Such would be the case even if the two countries did not differ from each other in any other respect whatsoever than age composition.

But there is more to it than that. The question arises as to why people in country B should be older; and there are all sorts of possible reasons. It may be that the younger people in country B are moving to country A, thus changing the composition of the two populations. It may be that A has a bigger birthrate than B. But it may even be that B is a less healthy country than A and that more people are dying young from other diseases; in consequence they do not live long enough to acquire cancer.

It must be obvious that no useful conclusions can be arrived at about the healthiness of the two countries unless one of two conditions is met: –

1) The two populations are identical in all other respects

or

2) Extraneous differences have been allowed for.

Even where this requirement has been met the scientist can rarely know it or can even be fairly sure. The exception is where in a deliberate study he has randomly assigned members from the same population to the two "subpopulations" A and B and thereafter has introduced the difference of interest between the two experimental groups.

Now, to revert to the problem of cancer rates in the two populations, if there is a free system of emigration, A and B are more-or-less self-selected populations: there is no reason to believe that they are equivalent to random assignations and good grounds to believe the contrary.

Those who emigrate from B to A may be the more ambitious and intelligent, and therefore likely to be successful. In consequence of both intelligence and prosperity, they may maintain better standards of personal hygiene and thus have a lower risk of cancer. But this is not a reflection of the environment of country A, merely of the fruits of intelligence and wealth; and the prevalence of cancer might be no higher in a comparable group in A.

For this reason the analysis of spontaneously formed groups is always more treacherous, and the conclusions arrived at surrounded by much more uncertainty, than for experimental studies of subjects who have been deliberately grouped in accordance with some systematic scheme of allocation.

CONCLUSIONS

What are the important lessons from all this?

First, in any exercise in empirical inference, it is necessary to define the population of interest. This must be done explicitly and before the event. Deciding on the population of interest after the data have been inspected, creates problems of the logic of inference which cannot be discussed here.

Secondly, where only part of the population is studied, the random sampling procedure must be carefully and explicitly thought out. The word "random" has a precise technical meaning – for instance it implies that before the sample has been selected the probability of any one person being included is specified. Anything less must be labelled "haphazard" and cannot be used as the basis for any sound non-trivial scientific inference.

Certain trivial inferences can be made. A physician in practice, deriving his patients from an unspecified source, who sees a patient with hypernephroma in country C, can thereafter claim that this cancer is "not unknown in C". Or a priest may know from the confessional, that drinking vinegar may excite some people to simony because he had encountered such a case. But if the representativeness of the samples is unknown no useful conclusions about frequencies are possible.

Thirdly, in comparing two populations, account must be taken of any extraneous differences. Thus, if we wish to explore the effect of smoking on lung cancer by comparing populations A and B, then adjustment must be made for all pertinent differences between A and B other than their smoking habits. Of course in any real situation it is not known what differences are petinent. In consequence logically certain or compelling conclusions are not possible and all results of this kind must be accepted with reservations. This logical difficulty can be circumvented if the two populations can be assigned randomly from the same population. But in many, perhaps most, situations either for moral, legal, political or logistic reasons it is not possible to do so, and the scientist must needs have recourse to second class evidence. But the exigencies arising from the system should never lead one to represent the scientific soundness of that evidence available as better than it is.

REFERENCES

1. Deming, W. E. "Some Theory of Sampling." Wiley, New York, 1950
2. Cochran, W. G. "Sampling Techniques." Wiley, New York, 1953
3. Green, G. "The Heart of the Matter." The Viking Press, New York, 1957, Part 3, Chapter 3.
4. Walker, S. H. and Duncan D. B. "Estimation of the probability of an event as a function of several independent variables." Biometrika 54, 167, 1967
5. Truett, J., Cornfield J., and Kannel, W. "A multivariate analysis of the risk of coronary heart disease in Framingham." J. Chron. Dis. 20, 511, 1967

Part III
The Interpretation of Evidence

In this third and last discussion attention will be centered on the interpretation of data. To some slight extent, this matter has been discussed in the second paper of the series. But the motivation is different. Comparison of cancer rates in countries A and B was used as an illustration of the importance of defining the population and specifying the sampling procedure.

But in this paper we will assume not only that there are no semantic and ontological difficulties but also that the sources of the data are beyond reproach. We still have the further problem of how we are to interpret what we have found. The subject is a vast one; but by way of illustration we shall consider two major sources of difficulty and error – on the one hand confounding, on the other correlation and the implications it has for causation.

CONFOUNDING

Suppose some new drug for the treatment of rheumatoid arthritis has been introduced. We give it to some patient with the disease and twenty-four hours later he feels better and has less pain; his fever has abated; objective measurements show that he has appreciably greater mobility in his joints and certain abnormal blood tests are more nearly normal. The unwary might be led to conclude that this is a triumphant demonstration of the therapeutic value of the drug; and they would be more than ever convinced if this experience was repeated in say 14 out of the first 20 patients on whom the drug was tried.

But this conclusion is logically unsound. Suppose we consider another more transparent example. To a group of 50 normal adult people we administer 50 mg of Vitamin C at 11 p.m. and 45 of them have a sound night's sleep. This proves that Vitamin C is a successful hypnotic in most cases. The fallacy in this conclusion is, needless to say, that most of the 50 people would have had a sound night's sleep anyway, and for all we know the preparation might be completely inert. The example demonstrates what should be our watchword in all such cases.

Do we know what would have happened if the treatment had *not* been given?

If so, does what has happened on treatment differ from what happens in the spontaneous course of the disease? It comes as something of a surprise to most lay people and to many doctors to find that in few cases can we answer these questions with any accuracy. At least two major difficulties exist. First, readers of textbooks commonly have concealed from them just how little hard fact there is about the natural history of disease and the factors which modify it: this is true even for common diseases. A second and more insidious matter is that the very existence of an effective treatment modifies the kind of patient who comes to the doctor's attention. It is a general (and not unexpected) truth that doctors are more likely to be on the alert for diseases which can be cured or at least adequately treated.

Tuberculosis for instance offers the best chance of cure with the minimum of scarring if diagnosis and treatment are early. But it is also known that most tuberculous infections, if left to themselves, will heal spontaneously. Thus by treating early, many patients the physician is treating would have recovered in any case. By diagnosing late, the physician is recognizing those cases who have not healed spontaneously and are unlikely to do so. Thus the two kinds of patients are not comparable. If the duration of survival is to be the criterion, of course, a more subtle bias is at work. Other things being equal if diagnosis is made six months earlier, then survival *from diagnosis* will be six months longer.

In the face of all this, could we say what would have happened to the patients with rheumatoid arthritis if they had not received the new treatment? Have we any evidence that what we have observed is not a spontaneous remission? The *logical structure* of the experiment is such that the effects of the drug cannot be distinguished from spontaneous changes in the disease. This relationship is a special case of confounding of effects. The word is of course cognate with "confusion" and the statement is equivalent to saying that two sets of effects are confused one with another: they are "poured together" like whiskey and soda. Without information from outside sources it is logically impossible to tell whether the whiskey or the soda is the intoxicating ingredient.

It is generally the case that in a study, one variable is of interest while the other variables are "contaminating" or "nuisance" variables. For the most part, confounding is difficult to avoid and considerable ingenuit ind perceptiveness may be requir to circumvent it. A few instance ay make it clear.

1) Injections of compound X r eve asthmatic attacks and this may e a property of X. But it may not he physical process of sticking a nee in a patient may have this effect; the physician, by his manner, ma influence patients by suggestion; it may be the solution in which is dissolved which is the active part. ere there are at least three variables onfounded with the variable of int est, (compound X).

2) A patient with rheum oid arthritis is instructed to take wo drugs: X before breakfast and Y ter. The treatment is only mode ely effective. After one month the physician reverses the order in ich the drugs are to be taken an the benefit is greatly increased. he change may be spontaneous nd nothing to do with either regim of treatment; or remission may ave occurred because Y is being ven early; or X late; or both; or it m be that X is what really is producin the effect but it must in some wa be protected by the prior action Y. Here at least four effects are ing confounded.

3) Surgeon M treats his pati nts' peptic ulcers by gastrectomy, wh eas surgeon N treats them by comb ned gastrojejunostomy and vagotom M gets the better results. But this oes not necessarily mean that gastrect my is the better operation. Perhaps I is the better surgeon. Perhaps N has referred to him the more serious y ill patients; it is maybe that N prefe s to do the less radical operation and does

gastrectomies only in those patients in whom the lesser operation fails; or the hospital in which N works may be less well equipped and staffed; or N may take mainly patients of some special racial group, or social group, or a religious group who do not approve of blood transfusion.

4) As an illustration of the pertinence of this phenomenon in other fields, we might consider again the effects of Catholic education. Suppose it could be shown that the lapse rate is lower among the alumni than in those attending undenominational schools. Superficially it might appear that this demonstrates that Catholic education is safer. But what kind of people attend Catholic schools? Inasmuch as the general teaching of the Church has been that Catholic education if not compulsary is at least warmly recommended, the parents who are more loyal and more respectful of authority will tend to send their children to Catholic schools. So will more prosperous parents. Moreover, it may be that the future careers of the alumni may be different because, in some countries at least, there has been more emphasis on the humanities and less on the sciences in Catholic schools. There are thus several sources of confounding: sufficient, in my opinion, to make the whole inference, from the empirical standpoint, doubtful unless at the very least, suitable adjustments can be made for these distortions.

What can be done about confounding?

Perhaps the most important step is to be aware of the phenomenon and to bear it continually in mind. The watchword must alway be, "Is there any factor other than the one in which we are interested which could account for the effect which has been observed?" Alternative explanations, however implausible, must be discarded with circumspection and, in strict practice, only in the light of clear positive evidence.

Secondly, at least in experimental work, random allocation of cases to the various treatments should be used wherever possible. In default of this, the evidence must *always* be regarded as imperfect.

Thirdly, except in those rare cases where the natural behavior of the disease (or whatever the phenomenon being studied) *is well known for the circumstances in which the present study is to be conducted* then a "control" sample should be studied. Thus, if we wish to find the effect of cortisone injections on asthmatic patients, we should have, simultaneously and under the same conditions, a comparable group of asthmatics who receive injections of some inert preparation such as the fluid in which the cortisone has been dissolved. Where possible, the asthmatics should be randomly assigned to the two groups; and for preference neither the patient nor the doctor assessing his progress should know which preparation the patient is receiving ("a double-blind experiment.") By this means confounding of the effects of the drug with the effects of suggestion can be largely avoided.

Failing these niceties, adjustment should be made in the analysis for variables confounded with the main one, or at least such of these variables as are believed important. Classical factors which are considered are differences in age, sex, race, education, social class, occupation and stage of the disease. Other factors will suggest themselves in the individual case. The technic of adjustment may be very complicated and need not concern us here. The main point is that such methods do exist; and the services of a competent statistician should be enlisted where necessary.

ASSOCIATION AND CORRELATION

Much of inference in certain fields such as epidemiology depends on recognizing that an association exists between two phenomena. By an "association" we mean that they occur together more often than can be accounted for by chance. Thus, there is a positive association between tonsilitis and rheumatic fever: those with sore throats are more likely to get rheumatic fever. Conversely those who have had vaccinia (i.e. have been "vaccinated") are *less* likely to get smallpox: thus there is a negative association between these diseases.

Where the entities being considered are continuous variables, non-independence between them is usually expressed as a correlation. Thus there is a positive correlation between height and weight. On the average, tall men are heavier than short men. Conversely the size of the heart is negatively correlated with expectation of life: on the average, the larger the heart the worse the prognosis. Note that these statements refer to averages only and are highly fallible in the individual cases.

The actual technics by which significant associations and correlations are demonstrated are beside the point. What is important is their interpretation. All the above examples quoted have been fairly well worked out. But it will be illuminating to consider examples in which the nature of the relationship is in doubt.

It has been repeatedly shown that if a group of men age, say, 40 to 60, who have had a coronary thombosis, be compared with a group of similar men of the same age, they have on the average a higher blood cholesterol level. This association can be interpreted according to three classes of theories.

1) A high serum cholesterol (C) promotes the development of thrombosis (T) or perhaps of the arterial disease which gives rise to it. To test this kind of theory we could explore whether C was present before T.

2) Coronary thrombosis causes certain body reaction among which high blood cholesterol is one, much as pneumonia causes a fever.

3) Finally we could argue that both C and T are separate consequences of some other unspecified factor. For instance it might be that the kind of person who gets coronary disease also has high cholesterol. There is reason to believe that emotional stress affects both; so, probably, does social class; so does exercise.

Now logically there is nothing to choose between these three and I personally am not convinced in the case of coronary disease and blood cholesterol, that the matter has been resolved though I know many scientists would not agree with me. But if we represent these diagrammatically denoting "A causes B" by "A→B" and indicating the unknown primordial mechanism by X then it

will be clear that we have exhausted the possibilities:

1) $X \rightarrow C \rightarrow T$

2) $X \rightarrow T \rightarrow C$

3) $X \begin{smallmatrix} \nearrow T \\ \searrow C \end{smallmatrix}$

The dimensions of the problem rapidly mounts in complexity. Three phenomena which have been shown to be associated (or correlated) can be arranged in sixteen causal pathways; four phenomena in 125 ways; six in nearly seventeen thousand. The lesson is clearly that the interpretation of such relationships is complex. It might of course be that many, perhaps most, of such schemes could not be squared with the subsidiary evidence. But the only means of ensuring that no plausible hypothesis is overlooked is to write down all possible combinations and consider them individually.

It would be a mistake, however, to think that this is the method which any sensible scientist would use to resolve the uncertainty. Something very like this problem is dealt with by the biochemist who is trying to determine a metabolic pathway i.e. the series of steps by which one compound is converted to another. The two questions have in common that they deal with the order of phenomena related to one another; in the one case it is a matter of which is the anterior to which in time, in the other case which is anterior etiologically. How exactly the latter notion is to be interpreted philosophically I do not know. But certain it is that the scientist always behaves as if "etiology" had a perfectly definite meaning. If one can make the statement "A→B", then manipulating A should change B, but the converse is not true. And the classical notion of causality implies at least this.

Now the biochemist in his metabolic problem reduces its complexity by fragmenting the tree and studying the small fragments. The man exploring causal pathways must needs do the same. The tree-drawing approach has the defect that it is purely static; the efficient approach and indeed the only one which can provide cogent evidence on causality is of its very nature dynamic i.e. causality is inferred from watching wheter changes in A (for preference deliberate manipulations) are followed by changes in B or *vice versa.*

Spontaneous dynamic changes provide more information than static, but they are still suspect. We would not argue that because, to the observer, lightening preceeds thunder it is the cause and thunder the effect.

So much of the contentions of the historians are, I suspect, to be traced to their failure to grasp the inadequacy of their evidence on dynamic relationships which are merly spontaneous and not deliberately manipulated. The resources of the historian, like those of the astronomer and the archeologist, are of course limited by the nature of the subject.

CONCLUSIONS

I think perhaps I will leave it at that. The conclusions are clear and easily stated. First that there is usually an embarrassment of hypotheses. Secondly, that no sound conclusions about causality can be made in a static system. Tentative conclusions may be reached from studying spontaneous dynamic changes. But cogent conclusions can only be reached from deliberate experimental manipulations.

The heavy emphasis on training in the so-called exact sciences as a preliminary to biological and medical training has the disadvantage that it provides no education in dealing with the intricacies of biological systems. The hiatus arises in two ways. In the first place, elementary courses in physics and chemistry deal much more in fact than in method. The centimeter-gram-second system of measurement which is taken for granted by the student, in fact represents a triumph of analysis arrived at by centuries of grappling with the constructs of physics and challenged again by the developments of relativity. The student too often gains no insight into such reductions and is perhaps left with the illusion that they are easy to make. In the second place, the basic sciences are dealing with structurally very simple ideas. The physiologist deals, or attempts to deal, with description and analysis of the flow of blood in the arteries, a problem which the physicist views with horror. The histochemist has as his objective the description of the chemistry of the cell, a matter which the organic chemist would dismiss as intractably complex.

A 'high level' exploration of problems is not necessarily a wasted effort, but it must be conducted in its own terms and in accordance with its own disciplines. These disciplines many biologists and physicians never learn. In consequence they commonly misconstrue the evidence presented to them and take or recommend incorrect courses of action. No physician would argue that fever causes pneumonia simply because the two are associated; and though he may take steps to cool his patient it is not intended as a curative measure but as a means of alleviating a distressing and sometimes dangerous manifestation of the disease. Likewise he would not recommend a low calcium diet in tuberculosis simply because calcium is commonly present in tuberculous lesions. Yet a large number of physicians (it seems to me on no more cogent basis) treat atherosclerosis by a low cholesterol diet.

As illustrations of two common sources of erroneous inference we have discussed confounding and association. There is some simularity between them; they both give rise to multiple interpretations of results which can be distinguished only (if at all) by appeal to outside information. In both cases, however, ethical or legal obstacles may preclude the critical experiment. Both have their analogies in the pastoral as well as in the scientific sphere.