Meeting January 14 1965

President's Address

The Environment and Disease: Association or Causation?

by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS
(Professor Emeritus of Medical Statistics, University of London)

Amongst the objects of this newly-founded Section of Occupational Medicine are firstly 'to provide a means, not readily afforded elsewhere, whereby physicians and surgeons with a special knowledge of the relationship between sickness and injury and conditions of work may discuss their problems, not only with each other, but also with colleagues in other fields, by holding joint meetings with other Sections of the Society'; and, secondly, 'to make available information about the physical, chemical and psychological hazards of occupation, and in particular about those that are rare or not easily recognized'.

At this first meeting of the Section and before, with however laudable intentions, we set about instructing our colleagues in other fields, it will be proper to consider a problem fundamental to our own. How in the first place do we detect these relationships between sickness, injury and conditions of work? How do we determine what are physical, chemical and psychological hazards of occupation, and in particular those that are rare and not easily recognized?

There are, of course, instances in which we can reasonably answer these questions from the general body of medical knowledge. A particular, and perhaps extreme, physical environment cannot fail to be harmful; a particular chemical is known to be toxic to man and therefore suspect on the factory floor. Sometimes, alternatively, we may be able to consider what *might* a particular environment do to man, and then see whether such consequences are indeed to be found. But more often than not we have no such guidance, no such means of proceeding; more often than not we are dependent upon our observation and enumeration of defined events for which we then seek antecedents. In other words we see that the event B is associated with the environmental feature A, that, to take a specific example, some form of respiratory illness is associated with a dust in the environment. In what circumstances can we pass from this observed *association* to a verdict of *causation*? Upon what basis should we proceed to do so?

I have no wish, nor the skill, to embark upon a philosophical discussion of the meaning of 'causation'. The 'cause' of illness may be immediate and direct, it may be remote and indirect underlying the observed association. But with the aims of occupational, and almost synonymously preventive, medicine in mind the decisive question is whether the frequency of the undesirable event B will be influenced by a change in the environmental feature A. *How* such a change exerts that influence may call for a great deal of research. However, before deducing 'causation' and taking action we shall not invariably have to sit around awaiting the results of that research. The whole chain may have to be unravelled or a few links may suffice. It will depend upon circumstances.

Disregarding then any such problem in semantics we have this situation. Our observations reveal an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?

(1) *Strength.* First upon my list I would put the strength of the association. To take a very old example, by comparing the occupations of patients with scrotal cancer with the occupations of patients presenting with other diseases, Percival Pott could reach a correct conclusion because of the *enormous* increase of scrotal cancer in the chimney sweeps. 'Even as late as the second decade of the twentieth century', writes Richard Doll (1964), 'the mortality of chimney sweeps from scrotal cancer was some 200 times that of workers who were not specially exposed to tar or mineral oils and in the eighteenth century the relative difference is likely to have been much greater.'

To take a more modern and more general example upon which I have now reflected for over fifteen years, prospective inquiries into smoking have shown that the death rate from cancer of the lung in cigarette smokers is nine to ten times the rate in non-smokers and the rate in heavy cigarette smokers is twenty to thirty times

CX 91, Page 2 of 6
Docket No. TSCA-HQ-2010-5022

as great. On the other hand the death rate from coronary thrombosis in smokers is no more than twice, possibly less, the death rate in non-smokers. Though there is good evidence to support causation it is surely much easier in this case to think of some features of life that may go hand-in-hand with smoking – features that might conceivably be the real underlying cause or, at the least, an important contributor, whether it be lack of exercise, nature of diet or other factors. But to explain the pronounced excess in cancer of the lung in any other environmental terms requires some feature of life so intimately linked with cigarette smoking and with the amount of smoking that such a feature should be easily detectable. If we cannot detect it or reasonably infer a specific one, then in such circumstances I think we are reasonably entitled to reject the vague contention of the armchair critic 'you can't prove it, there *may* be such a feature'.

Certainly in this situation I would reject the argument sometimes advanced that what matters is the absolute difference between the death rates of our various groups and not the ratio of one to other. That depends upon what we want to know. If we want to know how many extra deaths from cancer of the lung will take place through smoking (i.e. presuming causation), then obviously we must use the absolute differences between the death rates – 0·07 per 1,000 per year in non-smoking doctors, 0·57 in those smoking 1–14 cigarettes daily, 1·39 for 15–24 cigarettes daily and 2·27 for 25 or more daily. But it does not follow here, or in more specifically occupational problems, that this best measure of the effect upon mortality is also the best measure in relation to ætiology. In this respect the ratios of 8, 20 and 32 to 1 are far more informative. It does not, of course, follow that the differences revealed by ratios are of any practical importance. Maybe they are, maybe they are not; but that is another point altogether.

We may recall John Snow's classic analysis of the opening weeks of the cholera epidemic of 1854 (Snow 1855). The death rate that he recorded in the customers supplied with the grossly polluted water of the Southwark and Vauxhall Company was in truth quite low – 71 deaths in each 10,000 houses. What stands out vividly is the fact that the small rate is 14 times the figure of 5 deaths per 10,000 houses supplied with the sewage-free water of the rival Lambeth Company.

In thus putting emphasis upon the strength of an association we must, nevertheless, look at the obverse of the coin. We must not be too ready to dismiss a cause-and-effect hypothesis merely on the grounds that the observed association appears to be slight. There are many occasions in medicine when this is in truth so. Relatively few persons harbouring the meningococcus fall sick of meningococcal meningitis. Relatively few persons occupationally exposed to rat's urine contract Weil's disease.

(2) *Consistency:* Next on my list of features to be specially considered I would place the *consistency* of the observed association. Has it been repeatedly observed by different persons, in different places, circumstances and times?

This requirement may be of special importance for those rare hazards singled out in the Section's terms of reference. With many alert minds at work in industry today many an environmental association may be thrown up. Some of them on the customary tests of statistical significance will appear to be unlikely to be due to chance. Nevertheless whether chance is the explanation or whether a true hazard has been revealed may sometimes be answered only by a repetition of the circumstances and the observations.

Returning to my more general example, the Advisory Committee to the Surgeon-General of the United States Public Health Service found the association of smoking with cancer of the lung in 29 retrospective and 7 prospective inquiries (US Department of Health, Education & Welfare 1964). The lesson here is that broadly the same answer has been reached in quite a wide variety of situations and techniques. In other words we can justifiably infer that the association is not due to some constant error or fallacy that permeates every inquiry. And we have indeed to be on our guard against that.

Take, for instance, an example given by Heady (1958). Patients admitted to hospital for operation for peptic ulcer are questioned about recent domestic anxieties or crises that may have precipitated the acute illness. As controls, patients admitted for operation for a simple hernia are similarly quizzed. But, as Heady points out, the two groups may not be *in pari materia*. If your wife ran off with the lodger last week you still have to take your perforated ulcer to hospital without delay. But with a hernia you might prefer to stay at home for a while – to mourn (or celebrate) the event. No number of exact repetitions would remove or necessarily reveal that fallacy.

We have, therefore, the somewhat paradoxical position that the different results of a different inquiry certainly cannot be held to refute the

CX 91, Page 3 of 6
Docket No. TSCA-HQ-2010-5022

original evidence; yet the same results from precisely the same form of inquiry will not invariably greatly strengthen the original evidence. I would myself put a good deal of weight upon similar results reached in quite different ways, e.g. prospectively and retrospectively.

Once again looking at the obverse of the coin there will be occasions when repetition is absent or impossible and yet we should not hesitate to draw conclusions. The experience of the nickel refiners of South Wales is an outstanding example. I quote from the Alfred Watson Memorial Lecture that I gave in 1962 to the Institute of Actuaries:

'The population at risk, workers and pensioners, numbered about one thousand. During the ten years 1929 to 1938, sixteen of them had died from cancer of the lung, eleven of them had died from cancer of the nasal sinuses. At the age specific death rates of England and Wales at that time, one might have anticipated one death from cancer of the lung (to compare with the 16), and a fraction of a death from cancer of the nose (to compare with the 11). In all other bodily sites cancer had appeared on the death certificate 11 times and one would have expected it to do so 10–11 times. There had been 67 deaths from all other causes of mortality and over the ten years' period 72 would have been expected at the national death rates. Finally division of the population at risk in relation to their jobs showed that the excess of cancer of the lung and nose had fallen wholly upon the workers employed in the chemical processes.

'More recently my colleague, Dr Richard Doll, has brought this story a stage further. In the nine years 1948 to 1956 there had been, he found, 48 deaths from cancer of the lung and 13 deaths from cancer of the nose. He assessed the numbers expected at normal rates of mortality as, respectively 10 and 0·1.

'In 1923, long before any special hazard had been recognized, certain changes in the refinery took place. No case of cancer of the nose has been observed in any man who first entered the works after that year, and in these men there has been no excess of cancer of the lung. In other words, the excess in both sites is uniquely a feature in men who entered the refinery in, roughly, the first 23 years of the present century.

'No causal agent of these neoplasms has been identified. Until recently no animal experimentation had given any clue or any support to this wholly statistical evidence. Yet I wonder if any of us would hesitate to accept it as proof of a grave industrial hazard?' (Hill 1962).

In relation to my present discussion I know of no parallel investigation. We have (or certainly had) to make up our minds on a unique event; and there is no difficulty in doing so.

(3) *Specificity:* One reason, needless to say, is the specificity of the association, the third characteristic which invariably we must consider. If, as here, the association is limited to specific workers and to particular sites and types of disease and there is no association between the work and other modes of dying, then clearly that is a strong argument in favour of causation.

We must not, however, over-emphasize the importance of the characteristic. Even in my present example there is a cause and effect relationship with two different sites of cancer – the lung and the nose. Milk as a carrier of infection and, in that sense, the cause of disease can produce such a disparate galaxy as scarlet fever, diphtheria, tuberculosis, undulant fever, sore throat, dysentery and typhoid fever. Before the discovery of the underlying factor, the bacterial origin of disease, harm would have been done by pushing too firmly the need for specificity as a necessary feature before convicting the dairy.

Coming to modern times the prospective investigations of smoking and cancer of the lung have been criticized for not showing specificity – in other words the death rate of smokers is higher than the death rate of non-smokers from many causes of death (though in fact the results of Doll & Hill, 1964, do not show that). But here surely one must return to my first characteristic, the strength of the association. If other causes of death are raised 10, 20 or even 50% in smokers whereas cancer of the lung is raised 900–1,000% we have specificity – a specificity in the magnitude of the association.

We must also keep in mind that diseases may have more than one cause. It has always been possible to acquire a cancer of the scrotum without sweeping chimneys or taking to mule-spinning in Lancashire. One-to-one relationships are not frequent. Indeed I believe that multi-causation is generally more likely than single causation though possibly if we knew all the answers we might get back to a single factor.

In short, if specificity exists we may be able to draw conclusions without hesitation; if it is not apparent, we are not thereby necessarily left sitting irresolutely on the fence.

(4) *Temporality:* My fourth characteristic is the temporal relationship of the association – which is the cart and which the horse? This is a question which might be particularly relevant with diseases of slow development. Does a particular diet lead to disease or do the early stages of the disease lead to those peculiar dietetic habits? Does a

CX 91, Page 4 of 6
Docket No. TSCA-HQ-2010-5022

particular occupation or occupational environment promote infection by the tubercle bacillus or are the men and women who select that kind of work more liable to contract tuberculosis whatever the environment – or, indeed, have they already contracted it? This temporal problem may not arise often but it certainly needs to be remembered, particularly with selective factors at work in industry.

(5) *Biological gradient:* Fifthly, if the association is one which can reveal a biological gradient, or dose-response curve, then we should look most carefully for such evidence. For instance, the fact that the death rate from cancer of the lung rises linearly with the number of cigarettes smoked daily, adds a very great deal to the simpler evidence that cigarette smokers have a higher death rate than non-smokers. That comparison would be weakened, though not necessarily destroyed, if it depended upon, say, a much heavier death rate in light smokers and a lower rate in heavier smokers. We should then need to envisage some much more complex relationship to satisfy the cause-and-effect hypothesis. The clear dose-response curve admits of a simple explanation and obviously puts the case in a clearer light.

The same would clearly be true of an alleged dust hazard in industry. The dustier the environment the greater the incidence of disease we would expect to see. Often the difficulty is to secure some satisfactory quantitative measure of the environment which will permit us to explore this dose-response. But we should invariably seek it.

(6) *Plausibility:* It will be helpful if the causation we suspect is biologically plausible. But this is a feature I am convinced we cannot demand. What is biologically plausible depends upon the biological knowledge of the day.

To quote again from my Alfred Watson Memorial Lecture (Hill 1962), there was

> '. . . no biological knowledge to support (or to refute) Pott's observation in the 18th century of the excess of cancer in chimney sweeps. It was lack of biological knowledge in the 19th that led a prize essayist writing on the value and the fallacy of statistics to conclude, amongst other "absurd" associations, that "it could be no more ridiculous for the stranger who passed the night in the steerage of an emigrant ship to ascribe the typhus, which he there contracted, to the vermin with which bodies of the sick might be infected". And coming to nearer times, in the 20th century there was no biological knowledge to support the evidence against rubella.'

In short, the association we observe may be one new to science or medicine and we must not dismiss it too light-heartedly as just too odd. As Sherlock Holmes advised Dr Watson, 'when you have eliminated the impossible, whatever remains, *however improbable*, must be the truth.'

(7) *Coherence:* On the other hand the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease – in the expression of the Advisory Committee to the Surgeon-General it should have coherence.

Thus in the discussion of lung cancer the Committee finds its association with cigarette smoking coherent with the temporal rise that has taken place in the two variables over the last generation and with the sex difference in mortality – features that might well apply in an occupational problem. The known urban/rural ratio of lung cancer mortality does not detract from coherence, nor the restriction of the effect to the lung.

Personally, I regard as greatly contributing to coherence the histopathological evidence from the bronchial epithelium of smokers and the isolation from cigarette smoke of factors carcinogenic for the skin of laboratory animals. Nevertheless, while such laboratory evidence can enormously strengthen the hypothesis and, indeed, may determine the actual causative agent, the lack of such evidence cannot nullify the epidemiological observations in man. Arsenic can undoubtedly cause cancer of the skin in man but it has never been possible to demonstrate such an effect on any other animal. In a wider field John Snow's epidemiological observations on the conveyance of cholera by the water from the Broad Street pump would have been put almost beyond dispute if Robert Koch had been then around to isolate the vibrio from the baby's nappies, the well itself and the gentleman in delicate health from Brighton. Yet the fact that Koch's work was to be awaited another thirty years did not really weaken the epidemiological case though it made it more difficult to establish against the criticisms of the day – both just and unjust.

(8) *Experiment:* Occasionally it is possible to appeal to experimental, or semi-experimental, evidence. For example, because of an observed association some preventive action is taken. Does it in fact prevent? The dust in the workshop is reduced, lubricating oils are changed, persons stop smoking cigarettes. Is the frequency of the associated events affected? Here the strongest

CX 91, Page 5 of 6
Docket No. TSCA-HQ-2010-5022

support for the causation hypothesis may be revealed.

(9) *Analogy:* In some circumstances it would be fair to judge by analogy. With the effects of thalidomide and rubella before us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy.

Here then are nine different viewpoints from all of which we should study association before we cry causation. What I do not believe – and this has been suggested – is that we can usefully lay down some hard-and-fast rules of evidence that *must* be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?

Tests of Significance

No formal tests of significance can answer those questions. Such tests can, and should, remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the 'proof' of our hypothesis.

Nearly forty years ago, amongst the studies of occupational health that I made for the Industrial Health Research Board of the Medical Research Council was one that concerned the workers in the cotton-spinning mills of Lancashire (Hill 1930). The question that I had to answer, by the use of the National Health Insurance records of that time, was this: Do the workers in the card-room of the spinning mill, who tend the machines that clean the raw cotton, have a sickness experience in any way different from that of other operatives in the same mills who are relatively unexposed to the dust and fibre that were features of the cardroom? The answer was an unqualified 'Yes'. From age 30 to age 60 the cardroom workers suffered over three times as much from respiratory causes of illness whereas from non-respiratory causes their experience was not different from that of the other workers. This pronounced difference with the respiratory causes was derived not from abnormally long periods of sickness but rather from an excessive number of repeated absences from work of the cardroom workers.

All this has rightly passed into the limbo of forgotten things. What interests me today is this: My results were set out for men and women separately and for half a dozen age groups in 36 tables. So there were plenty of sums. Yet I cannot find that anywhere I thought it necessary to use a test of significance. The evidence was so clear-cut, the differences between the groups were mainly so large, the contrast between respiratory and non-respiratory causes of illness so specific, that no formal tests could really contribute anything of value to the argument. So why use them?

Would we think or act that way today? I rather doubt it. Between the two world wars there was a strong case for emphasizing to the clinician and other research workers the importance of not overlooking the effects of the play of chance upon their data. Perhaps too often generalities were based upon two men and a laboratory dog while the treatment of choice was deduced from a difference between two bedfuls of patients and might easily have no true meaning. It was therefore a useful corrective for statisticians to stress, and to teach the need for, tests of significance merely to serve as guides to caution before drawing a conclusion, before inflating the particular to the general.

I wonder whether the pendulum has not swung too far – not only with the attentive pupils but even with the statisticians themselves. To decline to draw conclusions without standard errors can surely be just as silly? Fortunately I believe we have not yet gone so far as our friends in the USA where, I am told, some editors of journals will return an article because tests of significance have not been applied. Yet there are innumerable situations in which they are totally unnecessary – because the difference is grotesquely obvious, because it is negligible, or because, whether it be formally significant or not, it is too small to be of any practical importance. What is worse the glitter of the *t* table diverts attention from the inadequacies of the fare. Only a tithe, and an unknown tithe, of the factory personnel volunteer for some procedure or interview, 20% of patients treated in some particular way are lost to sight, 30% of a randomly-drawn sample are never contacted. The sample may, indeed, be akin to that of the man who, according to Swift, 'had a mind to sell his house and carried a piece of brick in his pocket, which he showed as a pattern to encourage purchasers'. The writer, the editor and the reader are unmoved. The magic formulæ are there.

Of course I exaggerate. Yet too often I suspect we waste a deal of time, we grasp the shadow and

CX 91, Page 6 of 6
Docket No. TSCA-HQ-2010-5022

lose the substance, we weaken our capacity to interpret data and to take reasonable decisions whatever the value of P. And far too often we deduce 'no difference' from 'no significant difference'. Like fire, the χ^2 test is an excellent servant and a bad master.

The Case for Action

Finally, in passing from association to causation I believe in 'real life' we shall have to consider what flows from that decision. On scientific grounds we should do no such thing. The evidence is there to be judged on its merits and the judgment (in that sense) should be utterly independent of what hangs upon it – or who hangs because of it. But in another and more practical sense we may surely ask what is involved in our decision. In occupational medicine our object is usually to take action. If this be operative cause and that be deleterious effect, then we shall wish to intervene to abolish or reduce death or disease.

While that is a commendable ambition it almost inevitably leads us to introduce differential standards before we convict. Thus on relatively slight evidence we might decide to restrict the use of a drug for early-morning sickness in pregnant women. If we are wrong in deducing causation from association no great harm will be done. The good lady and the pharmaceutical industry will doubtless survive.

On fair evidence we might take action on what appears to be an occupational hazard, e.g. we might change from a probably carcinogenic oil to a non-carcinogenic oil in a limited environment and without too much injustice if we are wrong. But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like. In asking for very strong evidence I would, however, repeat emphatically that this does not imply crossing every 't', and swords with every critic, before we act.

All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time.

Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8.30 next day.

REFERENCES

Doll R (1964) In: Medical Surveys and Clinical Trials. Ed. L J Witts. 2nd ed. London; p 333

Doll R & Hill A B (1964) *Brit. med. J.* i, 1399, 1460

Heady J A (1958) *Med. World, Lond.* 89, 305

Hill A B
(1930) Sickness amongst Operatives in Lancashire Spinning Mills. Industrial Health Research Board Report No. 59. HMSO, London
(1962) *J. Inst. Actu.* 88, 178

Snow J (1855) On the Mode of Communication of Cholera. 2nd ed, London (Reprinted 1936, New York)

US Department of Health, Education & Welfare (1964) Smoking and Health. Public Health Service Publication No. 1103. Washington