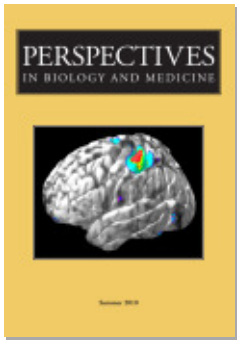
Proof of Causality: Deduction from Epidemiological Observation

Doll, Richard.

Perspectives in Biology and Medicine, Volume 45, Number 4, Autumn 2002, pp. 499-515 (Article)

Published by The Johns Hopkins University Press
DOI: 10.1353/pbm.2002.0067



Proof of Causality

deduction from epidemiological observation

Sir Richard Doll

It is a particular pleasure to have the honor of giving a Fisher memorial lecture, for it gives me the opportunity to acknowledge the formative influence that Sir Ronald's book, *Statistical Methods for Research Workers* (1934) had on my career.

I had, as a boy, developed a love for mathematics and when, in October 1931, I started to study medicine, I sought ways in which I could apply numerical methods in my work. Nothing productive emerged until, a few years later, I discovered his book and realized how relevant it was to medical research. About that time, one of the teaching staff drew our attention to a report of the use of pituitary gonadotropic hormone in the treatment of undescended testes in adolescent boys, which the authors said had a smaller percentage of failure than had been achieved by reliance on spontaneous descent (Spence and Scowen 1935). There were, of course, many difficulties in making such a comparison; but suffice it to say that even if the series were properly comparable, no consideration had been paid to the possibility that the difference observed might have been due to chance. Fisher's book led me to Pearson's χ^2 test, which showed that with the numbers observed as big or bigger differences would have been expected to

Clinical Trial Service Unit and Epidemiological Studies Unit, Harkness Building, Radcliffe Infirmary, Oxford OX2 6HE, United Kingdom.
Email: secretary@ctsu.ox.ac.uk (Cathy Harwood).
The text is that of the 23rd Fisher Memorial Lecture, given at Oxford on 29 October 2001.

Perspectives in Biology and Medicine, volume 45, number 4 (autumn 2002):499–515

occur by chance about as many as six times out of 10—a finding that I published in the *St. Thomas's Hospital Gazette* and that is recorded as my first scientific publication (Doll 1936).

Another 20 years had to pass, however, before I met Sir Ronald personally, when we both took part in a debate organized by Cambridge University Medical students about the validity of Bradford Hill's and my claim that our case-control study of patients with and without lung cancer justified the statement that cigarette smoking was a cause, and an important cause, of the disease (Doll and Hill 1950, 1952).

MEANING OF CAUSALITY

Whether our statement was justified depends on what is meant by "cause" and the distinction between *a* cause and *the* cause. Reference to *the* cause is appropriate when considering the mechanism by which a disease is produced, when a specific biological change may meet the philosophical requirement that it is both necessary and sufficient. The term is also used appropriately in reference to agents that are not sufficient but that satisfy Koch's postulates. These postulates, one of which requires that the agent should be found in all cases of the disease, were very helpful at the time that new discoveries were being made about bacteria and the origin of infectious disease, but were subsequently taught as essential requirements for a conclusion about causality, even though they were often quite inappropriate. I can remember, for example, being told, when a medical student, that smoking could not cause Buerger's disease because my teacher had seen a case in a non-smoker, and the same objection was occasionally raised some 15 years later to the idea that smoking might cause lung cancer. In both examples, the critics had confused *the* cause with *a* cause and had not appreciated what is now obvious: that many diseases have several causes, in the sense that we now give the word—that is, agents that are associated with a disease in such a way that increased exposure to them is followed by an increased risk of the disease and decreased exposure by a decreased risk.[1] That asbestos is a cause of lung cancer in this practical sense is incontrovertible, but we can never say that asbestos was responsible for the production of the disease in a particular patient, as there are many other etiologically significant agents to which the individual may have been exposed, and we can speak only of the extent to which the risk of the disease was increased by the extent of his or her exposure.

Whether epidemiology alone can, in strict logic, ever prove causality, even in this modern sense, may be questioned, but the same must also be said of laboratory experiments on animals. What can be achieved, even sometimes on epi-

[1]Such agents are distinguished from "risk factors," which may include them, but which also include agents whose association with a disease is due to confounding with some other agent that is an actual cause (Last 1994).

TABLE 1 GUIDES TO CAUSALITY OF OBSERVED ASSOCIATIONS

Strength of the association	
Consistency	Specificity
Dose-response relationship	Coherence of evidence
Temporal relationship	Experiment
Biological plausibility	Analogy

NOTE: AFTER HILL (1965).

demiological evidence alone, is proof beyond reasonable doubt, which in a court of law is sufficient to condemn the accused to the utmost penalty and which in society is sufficient to justify action to reduce risk. This is easier to achieve if one has both epidemiological and laboratory evidence, but it can be achieved on epidemiological evidence alone, as it was by Snow's observations on the distribution of cases of cholera, has been with many occupational hazards, and was by the International Agency for Research on Cancer, when it first classified arsenic, alcohol, and wood dust as causes of cancer in humans (IARC, 1980, 1981, 1988).

REQUIREMENTS FOR PROOF

The conditions required to reach such conclusions have often been discussed, and I shall only summarize them now, devoting most of my time to three problems with which I have been personally concerned, in which decisions about causality have had to be taken. In summary, we have to show, first, that the association cannot reasonably be explained by chance (bearing in mind that extreme chances do turn up from time to time or no one would buy a ticket in a national lottery), by methodological bias (which can have many sources), or by confounding (which needs to be explored but should not be postulated without some idea of what it might be). Second, we have to see whether the available evidence gives positive support to the concept of causality: that is to say, how it matches up to Hill's (1965) guidelines (Table 1). These are *not* criteria, as Hill insisted, but aids to thought. Only one is sine qua non: namely, the occurrence of the event at an appropriate interval after the exposure. All the others can, in some circumstances, be dispensed with, even plausibility and consistency, if the strength of the association is as great as it has been with some occupational hazards and the lack of consistency results from the lack of other evidence, rather than the presence of conflicting evidence.

CIGARETTE SMOKING AND LUNG CANCER

The first time I had to contend with the problem of causality was the subject of my debate with Sir Ronald. At the request of the Medical Research Council,

TABLE 2 DISTRIBUTION OF SMOKING HABITS: MALE PATIENTS WITH LUNG CARCINOMA AND CONTROLS

Category	*Lifelong Non-Smokers*	*Most Recent Amount Smoked by Smokers Before Onset of Disease (cigarettes per day)**			
		1-	*5-*	*15-*	*≥25*
Lung Carcinoma	2	33	250	196	168
Controls	27	55	293	190	84
Relative Risk	1	8	12	14	27

*1g tobacco in pipes or cigars classed as one cigarette.

NOTE: AFTER DOLL AND HILL (1950).

Bradford Hill and I had carried out a case-control study of patients with and without lung cancer in 20 London hospitals during 1948 to 1949 (Doll and Hill 1950). We had sought to interview patients admitted to hospital with a suspected diagnosis of lung, stomach, or large bowel cancer. We had obtained data on 709 patients, who on discharge were regarded as having lung cancer, and had compared this information with that obtained for 709 sex- and age-matched controls attending the same hospitals. The results showed, in each sex, a progressive increase in the ratio of cancer to control patients with increase in the amount said to have been smoked most recently before the onset of the disease that brought them into hospital, mounting in men who had smoked 25 or more cigarettes a day—or the equivalent in cigars or pipes—to 27 times that in lifelong non-smokers (Table 2).

Chance as an explanation seemed ruled out (with p-values $<1\times10^{-6}$ for smokers compared to lifelong non-smokers and $<1\times10^{-3}$ for the trend with the amount smoked daily by smokers). And so it seemed was bias. For we were able to exclude selection bias in the case of the controls selected by the interviewers, by showing that their smoking habits were similar to those of patients with gastric or large bowel cancer, who had been notified by the admitting hospitals, and we could exclude both interviewer and recall bias, by showing that the habits of patients in whom the diagnosis of lung cancer was not sustained by the discharge diagnosis were similar to those of all the patients we had seen without lung cancer.

Confounding had not, in 1950, been specifically named as an epidemiological problem, but the concept was recognized. We asked whether there could have been any common cause that had led both to the development of smoking and to the development of the disease and answered somewhat summarily by saying that we could not ourselves envisage any—a response we enlarged on two years later, when we tested for the possible effects of urban residence and socioeconomic status in a larger series (Doll and Hill 1952).

We noted the existence of a dose-response relationship for men (though we did not give it that name); that a similar, but much weaker, relationship was found for women; that lung cancer patients tended to have started smoking earlier and to have been less likely to have given up than their controls; that smoking had started many years before the disease presented; and that the relationship seemed to be specific for lung cancer, not being present to a significant extent, in our data, for other respiratory diseases or other cancers (from which cancers of the upper respiratory and digestive tracts had been excluded). One possibly anomalous finding was the lack of a positive relationship with inhaling; but as nothing was known of the size of the particles that carried the carcinogen, we had been assured that nothing could be said about the effect that any alteration in the rate and depth of respiration would have on the extent and site of deposition of the smoke droplets (Davies 1949).

The only carcinogen then known to have been found in tobacco smoke was arsenic (Daff and Kennaway 1950) for Roffo's (1940) detection of benzoapyrene was discounted in the United Kingdom because of the atypically high temperature at which he had burnt the tobacco (Cooper et al. 1932; Kennaway 1924), and the evidence that arsenic could cause cancer of the lung was suggestive rather than conclusive (Hill and Faning 1948). This we mentioned in our discussion, but we made no reference to the lack of evidence that tobacco tars were carcinogenic in laboratory animals, for the human evidence was so compelling.

Our conclusion that smoking was an important factor in the production of carcinoma of the lung was accepted by Sir Harold Himsworth, secretary of the Medical Research Council, but by few others. Some objections were puerile, like that put forward by a professor of physics that similar proportions of patients with and without lung cancer smoked the equivalent of five to 24 cigarettes a day (90.5 and 89.0 percent), ignoring the difference between the extremes. Those put forward by the tobacco industry were no more impressive—that histories of smoking habits were too unreliable to use for scientific purposes, that the correlation between national cigarette consumption and the mortality from lung cancer in different countries was only 0.5, and that the disease was obviously due to atmospheric pollution, for which they offered no evidence other than the well-known fact that it was somewhat more common in towns than in the country (Doll 1999). A later argument, that the increase in lung cancer mortality was much greater than the increase in tobacco consumption, was more sophisticated, but it failed to distinguish between the effects of cigarettes, pipes, and cigars and, most importantly, ignored the need for a prolonged latent period.

More weighty criticisms were voiced by Fisher, at first privately, but later in the *British Medical Journal*, *Nature*, the *Centennial Review*, and a booklet, which reproduced his earlier comments and added some more (Fisher 1957a, 1957b, 1958a, 1958b, 1958c, 1959). What stimulated Fisher to enter the debate was less our conclusion than an editorial in the *British Medical Journal* (Dangers of smok-

ing 1957), which followed the Medical Research Council's advice to government that smoking, and particularly cigarette smoking, should be accepted as the cause of a major part of the increase in the incidence of lung cancer (Medical Research Council 1957), and urged that the dangers of cigarette smoking "must be brought home to the public by all the modern devices of publicity." This was anathema to Fisher, reminding him of the unscrupulous use of such devices in the recent war. In his publications he raised four objections. First, he noted that it was logically possible that the association between smoking and lung cancer could have arisen because lung cancer caused smoking, rather than the other way around, if, as he suggested, the development of the disease was preceded by malignant changes in the mucosa that caused irritation that was relieved by smoking. Second, he noted that there might have been some common factor that was responsible both for the individual's smoking habits and his (or her) risk of developing the disease. This, he postulated, could be genetic, and he supported his hypothesis by showing that the smoking habits of pairs of monozygous twins were more similar than those of dizygous pairs and (on small numbers) appeared to be similar, irrespective of whether they had been raised together or apart (Fisher 1958a, 1958b). Third, he thought that our finding that reports of inhaling by smokers were less common in the patients with lung cancer (62 percent) than in the controls without (67 percent) weighed heavily against causation, unless it were also concluded that "inhaling cigarette smoke was a practice of considerable prophylactic value in preventing the disease." And lastly, he argued that secular changes in smoking habits could not be related to the increase in lung cancer, since "lung cancer has been increasing more rapidly in men relatively to women" and because "it is notorious, and conspicuous in the memory of most of us, that over the last 50 years the increase of smoking among women has been great, and that among men (even if positive) certainly small" (Fisher 1957b).

The first criticism was not very persuasive, as most patients had begun smoking in their teens some 40 years before the development of the disease, and the last was invalid, for he had ignored the generation effects whereby the risks among successive generations are directly determined not only by their recent smoking history but also by the predisposition to lung cancer imprinted on them by their smoking habits in the distant past. When this is allowed for and mortality rates and smoking habits are compared at appropriate ages, the trends in the sex ratio are not discrepant with the trends in cigarette consumption by sex over the relevant periods (Doll and Peto 1976).

Whether our findings with regard to inhaling, referred to in Fisher's third point, conflicted with our conclusion was moot, but he never took account of the difficulty of interpreting these findings without knowing where the smoke droplets were deposited and this could not be predicted without direct observation, as Davies (1949) had previously warned us. When, moreover, our study was completed and we had more data, we found that while inhaling was associated

with a diminished risk of cancer of the large bronchi, it was associated with an increased risk of developing cancer in the periphery of the lung (Doll and Hill 1952), which could make biological sense.

Fisher's second point was by far the most important, for confounding was certainly a theoretical possibility. We had not ourselves been able to envisage a factor that would meet the situation. He, however, did—though it was difficult to see how the genetic factor that he postulated would fit quantitatively and with the geographical and social distribution of the disease, quite apart from the fact that it failed to explain the increase in its incidence. It was, however, several years before it was possible to rule out a genetic explanation by the changes that took place in whole populations when sections of them gave up smoking (Doll and Peto 1976), and by the findings in monozygotic twins with different smoking habits (Carmelli and Page 1996; Floderus, Cederlöf, and Friberg 1988; Kaprio and Koskenvuo 1985).

Why Fisher took the view he did and adhered to it so strongly, basing his arguments solely on the data in our 1950 paper and ignoring our total data with the more understandable findings on inhaling (Doll and Hill 1952), the results of our prospective study of British doctors (Doll and Hill 1956), and the accumulation of evidence all over the world, is difficult to understand. It has been discussed in detail in a perceptive article by Stolley (1991), who cites the following passage from the memoir that Yates and Mather (1962) wrote for the Royal Society after Fisher's death: "In his own work, Fisher was at his best when confronted with small self-contained sets of data, and many of his solutions of such problems showed great elegance and originality. He was never much interested in the assembly and analysis of large amounts of data from varied sources bearing on a given issue"—and it was, of course, precisely the analysis of such data that allowed us to reach the conclusion that cigarette smoking was an important cause of the disease. According to Sir Walter Bodmer, a graduate student of Fisher's who visited him shortly before his death, he had, however, come to accept that smoking was a "co-factor" in the production of lung cancer and had intended to make a public statement of his revised position had he survived (Bodmer, personal communication).

TOXIC OIL SYNDROME

My second example is the determination of the cause of a disease that affected some 20,000 people in northwestern Spain in the spring and summer of 1981 and has become known as the "toxic oil syndrome." The disease was characterized by an acute episode of fever, cough, and dyspnea, and was often accompanied by myalgia, skin rashes, radiographic changes suggesting non-cardiogenic pulmonary edema and a marked eosinophilia in over 90 percent of patients by the third week. Most patients recovered spontaneously within a few weeks, but

the eosinophilia sometimes persisted for months, and some 15 to 20 percent progressed to a chronic phase, with peripheral neuropathy, sclerodermatous changes, and severe salivary and lachrymal hyposecretion. Some 2 percent of patients died. No evidence had been found of any of the infectious causes of eosinophilia, and there seemed no alternative to a toxic origin. No similar case was known to have occurred before or after the epidemic (with a few exceptions to which I will refer later), and none occurred outside Spain.

The idea that the disease might be due to the consumption of a particular type of oil occurred to Dr. J. M. Tabuenca Oliver (1981) as a result of enquiries about children admitted to the Niño Jesus Hospital. By 1 June 1981, he had come to believe that the disease was due to food poisoning, and he administered a questionnaire to the parents of 62 affected children and 62 children with other diseases. The questionnaire results revealed a striking difference (100 percent against 6 percent) in the proportions who had regularly consumed oil that was sold as olive oil in five-liter plastic containers bearing no trademark or seal (Casado-Flores et al. 1982). This led to a public announcement on 10 June that oil of this sort was responsible, and to an offer, on 26 June, to exchange all such oils for pure olive oil at government expense. Subsequently, however, it became clear that the oil with which the disease had been linked did not always have these characteristics, but was also sold in other containers or in markets or shops that had bought supplies from street vendors, and it came to be described simply as "street oil."

The hypothesis that the disease was due to some toxic substance in what was sold as "olive oil" became understandable when it was discovered that some rapeseed oil had been imported into Spain as industrial oil, to which, consequently, aniline had been added, thus avoiding a heavy import duty. and that some of the rapeseed oil had been treated at a refinery in Seville to remove the aniline, mixed with other oils, and passed on to street traders to sell as olive oil. Twelve other case-control studies were carried out after Tabuenca's announcement, and all supported his original claim. The differences were less marked, but still substantial: the pooled difference being 94 percent street oil purchased by case families, against 40 percent by unaffected families, giving an odds ratio of approximately 25:1, with a lower 95 percent confidence limit of 17:1. Other differences were also described in some of the studies—in, for example, other items of food and a particular shampoo sold by itinerant vendors—but none was anything like as great or as consistent as the differences in street oil.

Particularly striking was the occurrence of cases in four families in Seville, some 300 km away from the affected region. Members of two of the affected families, it turned out, had visited the epidemic area at the time of the epidemic, when they had consumed oil of the suspect type. Two families had not. The heads of both these families had, however, worked in the refinery that refined denatured rapeseed oil for the distributor with whose products the epidemic had been most strongly linked, and they were the only two out of a total of 24 work-

ers to whom supplies of the refinery's oil had been allotted—reminding one of the lady on Hampstead Hill who developed cholera in the London epidemic of 1854, having sent for drinking water from the Broad Street pump, the center of the epidemic area.

There were, however, aspects of the epidemic that gave rise to doubt, most notably the failure to identify any agent in samples of the oil that caused similar reactions in animals, and suggestions were made that the disease was due to a pesticide that had been used on tomatoes in southern Spain. The issue became highly politicized, and I came into the picture when I was asked by the European regional office of the World Health Organization (WHO) and the Spanish Ministry of Health to give an independent opinion on the cause of the disease. I agreed to do so on the condition that I could visit Madrid and take evidence from scientists who contested the belief that street oil was the cause, as well as receiving the mass of original reports that were provided by the Ministry. This proved well worthwhile, for the two clinicians I met were, as far as I could tell, sincere disciples of a professor who had blamed the epidemic on the tomato pesticides but had unfortunately died without leaving any written evidence for his belief. I besought them to send me evidence in support of his alternative hypothesis, but they never did.

The available evidence provided no reason to suppose that the association between the disease and the consumption of street oil bought between April and June 1981 was due to any form of bias or to confounding with any other factor with which the purchase of street oil was associated, but the failure to identify a toxic substance in samples of the oil was disturbing.

The geographical and temporal distributions of the disease generally supported the idea but were insufficiently clearly defined for definite proof. Cheap brands of so-called olive oil had been sold by street vendors throughout Spain for many years, so that any hazardous oil must have been produced in a new way or supplied from a new source. This was compatible with the vast majority of information obtained from affected families, practically all of whom had bought and consumed new supplies shortly before the symptoms of illness began, but it was not compatible with the history of the epidemic in one of the four convents in which epidemics occurred. In this convent, which was situated in a village that was closely connected with the owners of the suspect refinery in Seville, oil was purchased in the first "decena" of February, a term that is best translated as "the first ten or so days" of the month, whereas records show that the first shipment of suspect oil was received for refining on 11 February. No symptom suggestive of the disease, moreover, occurred in the convent in March or April, when the oil was used for salads and vegetables, but an outbreak did occur in May, after it also began to be used in cooking.

What was originally thought to be strong evidence implicating the oil was the decline of the epidemic following the announcement that the disease was due to adulterated oil and the subsequent exchange of suspect samples for pure oil at

government expense. In fact, however, the decline in incidence had begun a week or more before the announcement was made, when the idea that street oil might be responsible was known only to a small group of research workers.

The geographical limitation of the epidemic to one part of Spain was striking, and the histories given by affected families about specific street vendors implicated three suppliers. Many other sources, however, came to be suspected, and neither of the special studies that were undertaken succeeded in delineating a clear network of supplies corresponding to the affected areas. The initiation of legal proceedings had, however, clouded the issue, as it was against the interests of both vendors and suppliers to be associated with the sale of any suspect batches. All that could be said was that, if the oil was the source, toxicity was associated with a relatively small number of batches that were imported, processed, and widely distributed in the first half of 1981, some of which were imported by a named firm at San Sebastian, handled by another named firm, and refined by a third in Seville and a fourth in Madrid.

Less than 200 cases were registered in people who lived outside the 14 affected provinces, the majority of whom were found to have had meals in the affected region before the onset of their illness or, as described previously, had consumed oil direct from the refinery in Seville that had been most clearly implicated.

In sum, the evidence in favor of causality was the strength of the recorded association, the generally close temporal relationship between purchase and consumption of oil and the occurrence of disease, and the fact that so many of the sporadic cases occurring outside the region were found to have been exposed to the suspect oil by their personal behavior. Against was the lack of any clear quantitative relationship between the dose of oil and the risk of the disease, the failure to demonstrate geographic limits to the sale of the oil corresponding to the region in which the epidemic occurred and to demonstrate exposure for all affected subjects, and the time relationship between the purchase of oil and the development of typical symptoms in one convent.

The evidence against causality was, I thought, inconclusive, but there were too many gaps in the evidence to allow the conclusion that street oil was definitely the cause, and I outlined some further research that might settle the issue, including in particular a more detailed account of the few cases outside the region in people who had not apparently been exposed (Doll 2000).

Later Evidence

Twenty months later, more evidence had accumulated that changed the situation. First, intensive efforts had been made to review all cases that met strict criteria for the diagnosis of the disease that had occurred outside the 14 affected provinces. Altogether, 268 cases were identified from the lists provided by the Ministry of Justice, the Ministry of Health, and the official census; on review, these were classified as shown in Table 3. The clinical records were sought for the 41 cases that became ill outside the epidemic area and were exposed only to oil

TABLE 3 CASES RECORDED AS OCCURRING OUTSIDE EPIDEMIC AREA

Category	*Number of Cases*
Became ill outside area and exposed only to oil obtained outside area	41
Became ill outside area, but exposed to oil from within area	39
Became ill inside epidemic area and exposed to oil within area, but moved out before diagnosis made	158
Incorrectly included in list because of a coding error	3
Unclassifiable	3
Waiting review	24
All categories	**268**

NOTE: AFTER POSADA DE LA PAZ, ABAITUA BORDA, AND KILBOURNE (1987).

obtained outside it; 29 sets were obtained. Nineteen of these failed to satisfy the clinical definition of a case; of the remaining 10, five were of cases that occurred on the north border of the area, and three constituted the cluster of cases in Seville in people who had used oil provided by the suspect refinery. No link with the suspect oil had been discovered for only two confirmed cases.

Second, Kilbourne and his colleagues at the U.S. Centers for Disease Control in Atlanta had made a further attempt to seek a dose-response relationship (Kilbourne et al. 1988). Samples of oil were obtained from two warehouses, in which oils were stored that had been obtained from households in two towns in Madrid province during the government's oil exchange program in 1981. Ninety-three of the samples came from typical containers and were connected to families in which the occurrence or nonoccurrence of a case was firmly established. The oils were analyzed for their content of aniline and three fatty acid anilides. Each chemical showed major differences in the amounts found in the oils associated with the affected and unaffected families (p, in each case, <0.0001). Those for oleic acid anilide show a progressive increase in risk with the amount of anilide and a 19-fold increase when the samples contained more than 600 μg (Table 4).[2]

With the addition of this new evidence, I felt justified in concluding that adulterated oil *was* the cause of the syndrome (Doll 2000), a conclusion with which the court agreed when the case came to trial.

[2]Subsequent investigation has shown that the distribution of the disease fits better with other anilides—3-(phenylamino)propane-1,2-diol and its mono-, di-, and triacyl derivates—than with fatty acid anilides, and that the amounts of these anilides correlate even more closely with the risk of the disease (Posada de la Paz et al. 1999).

TABLE 4 CONCENTRATION OF OLEIC ACID ANILIDE IN OILS FROM AFFECTED AND UNAFFECTED FAMILIES

Type of Household	*Number of Samples Containing Different Amounts of Oleic Acid Anilide (μg per g)*					*Number of Households*
	0	*1-*	*100-*	*600-*	*≥1201*	
Affected	11	2	3	6	7	29
Unaffected	48	6	7	3	0	64

NOTE: AFTER KILBOURNE, ET AL. (1988).

ALCOHOL AND ISCHEMIC HEART DISEASE

My third example—the nature of the relationship between the consumption of ethanol and the risk of ischemic heart disease—shares the characteristic that a decision had to be taken on the epidemiological evidence without experimental evidence in humans or laboratory animals. In this case, however, there was laboratory evidence of a plausible mechanism.

That alcohol could have harmful effects on behavior if consumed in large amounts over a short time was obvious, but it was not until early in the 19th century that it became known that it could also have harmful effects on the brain and the liver, if consumed regularly in large amounts over a long period. That it could have beneficial effects other than by the provision of calories, easing social contacts, and giving pleasure was, however, less easy to detect, although a belief in their existence had been widespread in the 19th century.

The first good evidence of benefit to health was provided by Pearl (1926). His interest arose from observations of fowls, which lived longer if maintained for hours on end, day after day, in an atmosphere containing ethanol than control birds who were not exposed. Pearl consequently calculated actuarial survival curves for people with different drinking habits from data he had obtained in a study of the family health of two groups of individuals. His findings for men showed that heavy drinking was, not surprisingly, harmful throughout life, but moderate drinking was associated with a slightly better survival than abstention over about 65 years of age, moderate drinking being defined as small amounts at any one time, never leading to intoxication, but including the consumption of a "daily pint or two of beer, a bottle of claret, or a few glasses of whisky and soda." For women, there was no category of heavy drinking, but moderate drinking was associated with longer expectation of survival throughout life.

Vascular Disease and Total Mortality

Pearl's work made little impact, and it was not until after the Second World War that further evidence of a beneficial effect of alcohol began to be published. Then, in 1953, reports began to appear of an unusually low prevalence of coro-

nary artery disease in patients found to have cirrhotic livers at autopsy (Hall, Olsen, and Davis 1953; see Moore and Pearson 1986 for review). Autopsy series are, however, subject to many biases, and these reports excited little interest. Even in the 1970s, when case-control studies of people with and without myocardial infarcts and then cohort studies of people with different drinking habits reported a reduced risk of myocardial infarction in those who drank small or moderate amounts of alcohol in comparison with non-drinkers (Klatsky, Friedman, and Siegelaub 1974; Stason et al. 1976), scant attention was paid to them.

Now, however, the evidence for a beneficial effect is massive (Marmot and Brunner 1991; Moore and Pearson 1986; Poikolainen 1995; Rimm et al. 1999). It includes not only a reduction of about a third in the risk of vascular disease in cohort studies in several different countries, including a reduction in cerebral thrombosis as well as in myocardial infarction, but also, because vascular disease is such an important cause of death in middle and old age, a reduction in total mortality. Chance can be excluded by the consistency of the findings in so many different studies and the low p-value for the difference between the mortality in non-drinkers and light drinkers (which in our study of middle-aged and elderly British doctors was <0.001 for all causes combined [Doll et al. 1994]), and bias can be excluded by the results of cohort studies in which the causes of death had been recorded after the individuals' drinking habits.

Confounding, however, is another matter. One possibility, that the non-drinkers included ex-drinkers who might have given up because of ill health, was excluded by cohort studies in which ex-drinkers and lifelong non-drinkers had been classed separately and by data like our own, which included information on past medical histories and showed the same proportional reduction in risk in drinkers irrespective of any previous history of vascular disease. Another possibility was that drinkers might have differed from non-drinkers in other ways that would affect the risk of the disease, by, for example, including lower proportions of cigarette smokers, having a lower mean blood pressure and body mass index, a higher level of physical activity, or higher socioeconomic status (Shaper 1995). In fact, an association with smoking has the opposite effect to that postulated, as smoking is more prevalent in drinkers than in non-drinkers (Jarvis 1994), and allowance for it actually *increases* the evidence of benefit associated with drinking. In many studies it has been taken into account, as have all the other suspect factors in some of the more detailed studies, such as those by Stampfer, et al. (1988), and Thun, et al. (1997). When these factors were all allowed for, the observed differences were hardly altered. Confounding, often a serious concern when risks vary by less than two-fold, has in this case been tested and found wanting (Doll et al. 1997).

There remained the need for positive evidence for causality. Judged by the amount said to be drunk, there is no strong evidence of a dose-response effect, as closely similar reductions in risk are reported over a fairly wide range of one to four drinks a day. A quantitative relationship is, however, seen for regularity of

consumption, the benefit increasing progressively with the greater the number of days per week that drinks are consumed, given the same total weekly amount, as is shown in Peto's large case-control study of myocardial infarction (personal communication). A plausible explanation of how a benefit might be produced is, moreover, provided by the experimental finding that ethanol by mouth increases the blood level of high-density lipoprotein and lipoproteins A1 and A2, reduces slightly the blood levels of low-density lipoprotein and fibrinogen, and reduces the aggregability of platelets. You could hardly ask any antithrombotic drug to do more. It is not surprising, therefore, that despite the initial resistance of many public health authorities, expert advisers to the European Office of the WHO agreed that "drinking moderate amounts of alcoholic beverages is likely to reduce the risk of CHD [coronary heart disease] for some populations" (Edwards et al. 1994). That the inverse relationship between ischemic heart disease and the consumption of small or moderate amounts of alcohol is, for the most part, causal should, I believe, now be regarded as proved (Doll et al. 1997).

CONCLUSION

That the three associations I have described reflect causality may be thought banal, but it was not when it was originally claimed.

With the experience that we now have of thousands of epidemiological studies, we can conclude that large relative risks—of the order of >20:1—with evidence of a dose-response relationship, that cannot be explained by methodological bias or reasonably be attributed to chance (with p-levels of $<1\times10^{-6}$) are in themselves adequate proof of a causal relationship. Such was the situation with cigarette smoking and lung cancer in 1950, but similar situations are rare. Small relative risks of the order of 2:1 or even less are what are likely to be observed, like the risk now recorded for childhood leukemia and exposure to magnetic fields of 0.4 μT or more (Ahlbom et al. 2000) that are seldom encountered in the United Kingdom. And here the problems of eliminating bias and confounding are immense. It can be done, as we have seen with breast cancer and the use of steroid contraceptives, where an increased risk of 24 percent has been demonstrated during use and for a short period afterwards (Collaborative Group on Hormonal Factors in Breast Cancer 1996). Massive data were, however, required, and even with the amount now available, it may be doubted if the risk would be regarded as proved, had it not been for the knowledge that the chemicals concerned could cause breast cancer in animals.

These last two examples are, I fear, the sort of problems with which we are principally going to be faced in the future. But they can be solved in time, even in the absence of laboratory evidence, if we can use objective data, undertake cohort as well as case-control studies, and collaborate with colleagues worldwide to obtain evidence of sufficient weight.

REFERENCES

Ahlbom, A., et al. 2000. A pooled analysis of magnetic fields and childhood leukaemia. *Br. J. Cancer* 60:793–98.

Carmelli, D., and W. F. Page. 1996. Twenty-four-year mortality in World War II male veteran twins discordant for cigarette smoking. *Int. J. Epidemiol.* 25:554–59.

Casado-Flores, J., et al. 1982. Sindrome toxico for consumero de aceite adulterado. *CICLO Ensayos Medicos* 31:9–12.

Collaborative Group on Hormonal Factors in Breast Cancer. 1996. Breast cancer and hormonal contraceptives: Collaborative reanalysis of individual data on 53,297 women with beast cancer and 100,239 women without breast cancer from 54 epidemiological studies. *Lancet* 347:1713–27.

Cooper, E. A., et al. 1932. The role of tobacco-smoking in the production of cancer. *J. Hyg. Lond.* 32:293–300.

Daff, M. E., and E. L. Kennaway. 1950. Arsenic content of tobacco and of tobacco smoke. *Br. J. Cancer* 4:173–82.

Dangers of cigarette smoking. 1957. *BMJ* 1:1518–20.

Davies, C. N. 1949. Inhalation risk and particle size in dust and mist. *Brit. J. Industr. Med.* 6:245–53.

Doll, R. 1936. Medical statistics. *St. Thomas's Hosp. Gaz.*, 294–97.

Doll, R. 1999. Tobacco: A medical history. *J. Urban Health* 76:289–313.

Doll, R. 2000. The aetiology of the Spanish Toxic Oil Syndrome: Interpretation of the epidemiological evidence. *Gac. Sanit.* 14 (suppl.):77–98.

Doll, R., and A. B. Hill. 1950. Smoking and carcinoma of the lung: Preliminary report. *BMJ* 2:739–48.

Doll, R., and A. B. Hill. 1952. A study of the aetiology of carcinoma of the lung. *BMJ* 2: 1271–86.

Doll, R., and A. B. Hill. 1956. Lung cancer and other causes of death in relation to smoking: A second report on the mortality of British doctors. *BMJ* 2:1071–76.

Doll, R., and R. Peto. 1976. Mortality in relation to smoking: 20 years' observations on male British doctors. *BMJ* 2:1525–36.

Doll, R., et al. 1994. Mortality in relation to consumption of alcohol: 13 years' observations on male British doctors. *BMJ* 309:911–18.

Doll, R., et al. 1997. Alcohol and coronary heart disease reduction among British doctors: Confounding or causality? *Eur. Heart J.* 18:23–25.

Edwards, G., et al. 1994. *Alcohol policy and the public good.* Oxford: Oxford Univ. Press.

Fisher, R. A. 1934. *Statistical methods for research workers.* 5th ed. Edinburgh: Oliver & Boyd.

Fisher, R. A. 1957a. Dangers of cigarette smoking. *BMJ* 2:43.

Fisher, R. A. 1957b. Dangers of cigarette smoking. *BMJ* 2:297–98.

Fisher, R. A. 1958a. Lung cancer and cigarettes? *Nature* 182:108.

Fisher, R. A. 1958b. Cancer and smoking. *Nature* 182:596.

Fisher, R. A. 1958c. Cigarettes, cancer and statistics. *Centennial Rev.* 2:151–66.

Fisher, R. A. 1959. *Smoking: The cancer controversy. Some attempts to assess the evidence.* Edinburgh: Oliver and Boyd.

Floderus, B., R. Cederlöf, and L. Friberg. 1988. Smoking and mortality: A 21-year follow-up based on the Swedish Twin Registry. *Int. J. Epidemiol.* 17:332–40.

Hall, E. M., A.Y. Olsen, and F. E. Davis. 1953. Clinical and pathologic review of 782 cases from 16,600 necropsies. *Am. J. Pathol.* 29:993–1027.

Hill, A. B. 1965. The environment and disease: Association or causation? *Proc. R. Soc. Med.* 58:295–300.

Hill, A. B., and E. L. Faning. 1948. Studies in the incidence of cancer in a factory handling inorganic compounds of arsenic. I. Mortality experience in the factory. *Brit. J. Industr. Med.* 5:1–7.

International Agency for Research on Cancer (IARC). 1980. *Some metals and metallic compounds. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans.* Vol. 25. Lyon: IARC.

IARC. 1981. *Wood, leather, and some associated industries. IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans.* Vol. 23. Lyon: IARC.

IARC. 1988. *Alcohol drinking. IARC monographs on the evaluation of carcinogenic risks to humans.* Vol. 44. Lyon: IARC.

Jarvis, J. 1994. A profile of tobacco smoking. *Addiction* 89:1371–76.

Kaprio, J., and M. Koskenvuo. 1985. Twins, smoking and mortality: A 12-year prospective study of smoking discordant pairs. *Soc. Sci. Med.* 29:1083–89.

Kennaway, E. 1924. On cancer producing tars and tar fractions. *J. Industr. Hyg.* 5:462–88.

Kilbourne, E. M., et al. 1988. Chemical correlates of pathogenicity of oils related to the toxic oil syndrome epidemic in Spain. *Am. J. Epidemiol.* 127:1210–27.

Klatsky, A. L., D. G. Friedman, and A. D. Siegelaub. 1974. Alcohol consumption before myocardial infarction. Results from the Kaiser-Permanente epidemiologic study of myocardial infarction. *Ann. Intern. Med.* 81:294–301.

Last, J. M., ed. 1994. *A dictionary of epidemiology.* New York: Oxford Univ. Press.

Marmot, M., and E. Brunner. 1991. Alcohol and cardiovascular disease: The status of the U-shaped curve. *BMJ* 303:565–68.

Medical Research Council. 1957. Tobacco smoking and cancer of the lung. *BMJ* 1:1523.

Moore, R. D., and T. A. Pearson. 1986. Moderate alcohol consumption and coronary heart disease. *Medicine* 65:242–67.

Pearl, R. 1926. *Alcohol and longevity.* New York: Knopf.

Poikolainen, K. 1995. Alcohol and mortality: A review. *J. Clin. Epidemiol.* 488:445–65.

Posada de la Paz, M., I. Abaitua Borda, and E. H. Kilbourne. 1987. Study of peripheral cases of toxic oil syndrome. Report presented to the meeting of the Liaison Group of the WHO Steering Committee for Toxic Oil Syndrome, Madrid, 27–28 Jan.

Posada de la Paz, M., et al. 1999. Epidemiologic evidence for a new class of compounds associated with Toxic Oil Syndrome. *Epidemiology* 10:130–34.

Rimm, E. B., et al. 1999. Moderate alcohol intake and lower risk of coronary heart disease: Meta-analysis of effects on lipids and haemostatic factors. *BMJ* 319:1523–28.

Roffo, A. H. 1940. Krebserzeugendes Benzpyren, gewonnen aus Tabakteer. *Z. Krebsforsch* 49:588–97.

Shaper, A. G. 1995. Alcohol and coronary heart disease. *Eur. Heart J.* 16:1760–64.

Spence, A. W., and E. F. Scowen. 1935. Use of gonadotropic hormones in treatment of imperfectly migrated testes. *Lancet* 2:1335–38.

Stampfer, M. J., et al. 1988. A prospective study of moderate alcohol consumption and the risk of coronary artery disease and stroke in women. *N. Engl. J. Med.* 319:267–73.

Stason, W. B., et al. 1976. Alcohol consumption and non-fatal myocardial infarction. *Am. J. Epidemiol.* 104:603–8.

Stolley, P. D. 1991. When genius errs: R. A. Fisher and the lung cancer controversy. *Am. J. Epidemiol.* 133:416–25.

Tabuenca, J. M. 1981. Toxic-allergic syndrome caused by ingestion of rapeseed oil denatured with aniline. *Lancet* 2:567–68.

Thun, M. J., et al. 1997. Alcohol consumption and mortality in middle-aged and elderly U.S. adults. *N. Engl. J. Med.* 337:1705–14.

Yates, F., and K. Mather. 1962. Ronald Aylmer Fisher, 1890–1963. *Biographical memoirs of fellows of the Royal Society of London* 9:91–120.