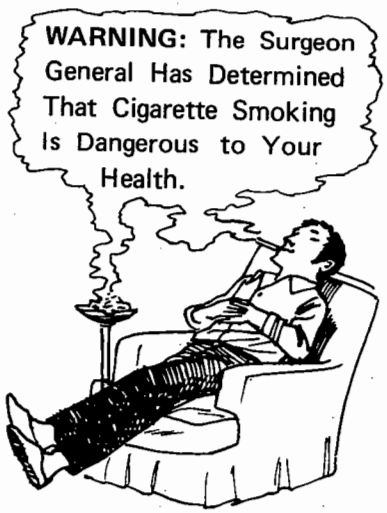

STATISTICS, SCIENTIFIC METHOD, AND SMOKING

B. W. Brown, Jr. *Stanford University School of Medicine*

AFTER HUNDREDS of years of tobacco use, smoking has been condemned on a scientific basis as a serious hazard to the health of the smoker. Future generations may regard the scientific indictment of smoking as a major contribution to preventive medicine and the health of the western world. Statisticians, statistical principles of scientific thought, and statistical methods of scientific study played essential parts in evaluating the effects of smoking.

Today a majority of health experts believe (though there are dissenters) that smoking is bad for the health—that it causes cancer and heart attacks and has other deleterious effects on the body. Governments have taken action to modify or suppress advertising by the tobacco industry, and to educate people about the hazards of smoking. Private health organizations, such as the American Cancer Society, actively propagandize against smoking. The effects of this new knowledge concerning smoking can readily be seen in surveys, which indicate that a growing number of people have quit smoking, that there is a decreasing proportion of smokers among young people, and that a decreasing number of cigarettes per capita are being sold. Over the last decade, the decrease in smoking at large parties, at meetings, and in public places especially among people in the health sciences, has been strikingly apparent to even casual observers.

How was it established scientifically that smoking is hazardous to the health? Why has this hazard been established only recently although smoking has been a widespread custom of the Western World for 300 years? When specific data on the question of hazard were published more than 40 years ago, why did the question spark such controversy, even among scientists most knowledgeable on the issues involved? And what role did statistics play in resolving the controversy?

EARLY VIEWS OF SMOKING AND HEALTH

Tobacco smoking is an ancient habit of man. Crude cigarettes have been found among the artifacts left by cave dwellers in Arizona; Columbus carried tobacco from the New World to the Old, with an endorsement by the Indians for its medicinal effects; Sir Walter Raleigh was a strong advocate of the use of tobacco; and many others have lauded its merits as cure and comfort for most of the diseases and distresses afflicting mankind. Cigarette smoking became so prevalent in the Western World that, according to a survey of the U.S., in the mid-sixties only 30% of males 17 years of age and older reported that they were not and had never been regular smokers. Fifty-one percent of the men and 34% of the women in the U.S. reported that they were currently regular smokers. Together, these smokers consumed more than a billion cigarettes each day in the year 1969.

Though tobacco has had its advocates from the beginning and has enjoyed an increasing, currently overwhelming popularity in the Western World, it has had opposition also, from the beginning. Swinburne said "James the First was a knave, a tyrant, a fool, a liar, a coward; but I love him, because he slit the throat of that blackguard Raleigh, who invented this filthy smoking." Some objected to the filth, while others felt the habit was sinful; still others objected that the habit was not good for the health. We can trace speculations concerning the ill effects of tobacco in records and writings as far back as three centuries, but these comments, at best, offered the authority of the experienced physician giving his impressions. They were not based on systematically gathered and evaluated scientific evidence. For example, a century ago Dr. Oliver Wendell Holmes, distinguished professor of the Harvard Medical School and father of Supreme Court Justice Holmes, said "I think tobacco often does a great deal of harm to the health. I myself

gave it up many years ago." But Dr. Holmes gave no evidence to justify his conclusion.

Papers on the effects of smoking that have appeared in medical journals over the last 100 years show the general tendency of medical scientists to collect their information more systematically and to evaluate it more carefully as scientific evidence. In 1927, Dr. F. E. Tylecote, an English physician, wrote that almost every lung-cancer patient he had known about had been a regular smoker, usually of cigarettes. In 1936, Drs. Arkin and Wagner reported more specifically that 90% of 135 men afflicted with lung cancer were "chronic smokers."

DIFFICULTIES OF RESEARCH

With the accumulation of such anecdotal reports by serious medical men it became clear that the question of the harmful effects of tobacco should be subjected to scientific study. But how could the question be examined scientifically? Although scientific method in the physical sciences had been thoughtfully discussed and developed over the past several centuries, the principles that had been evolved for the physics laboratory could not be adapted easily to scientific study in the biological and medical sciences. Specifically, the biological scientist fell far short of the physicist in his attempt to hold fixed all factors except those under investigation. Instead, he was faced with investigations involving many uncontrolled, or partially controlled, factors that caused unwanted variation in the data. Conclusions had to be made in the face of this biological variation in the study material.

The problem of drawing valid scientific conclusions in the life sciences attracted an assortment of extremely able English men of science; among them were Francis Galton, a genius and creative psychologist; Karl Pearson, trained as an engineer and ultimately an outstanding philosopher of science; and R. A. Fisher, a mathematician who laid down new principles of great importance to scientific study. It was Fisher, building on the work of Galton and Pearson, who suggested, in the twenties, an alternative to the impossible requirement that all factors be held constant except the factors under investigation.

Fisher suggested that if two fertilizers, for example, are to be compared, each should be allocated to a number of different plots of ground, and that the allocation be random, according to some sort of lottery or coin-tossing system. Fisher pointed out that such random allocation would tend to balance all differences between the plots receiving the two different fertilizers and thus would yield an unbiased experiment—that is, one fair to both fertilizers. He further pointed out that any differences in the results obtained with the two fertilizers could be evaluated, using the probability theory that had been developed for gambling games, to decide how likely it was that so large a

difference might have come about simply through a chance allocation of the more fertile plots to one of the treatments.

This new scientific methodology, employing randomization and probability theory, was a tremendous contribution to the young science of statistics. It spread quickly through the agricultural sciences and then to the medical sciences. Today a new drug will not be approved by the U.S. Food and Drug Administration for marketing until it has been compared with other treatments in what is called a *randomized clinical trial.* In the thirties, however, these rigorous standards for scientific study were new and scientists were just becoming familiar with their application in the medical sciences. How could the new methodology be applied to the smoking question?

RETROSPECTIVE STUDIES

Clearly it would be difficult, though perhaps not impossible, to conduct a randomized experiment on human beings to determine whether smoking is harmful. The first attempt at some sort of satisfactory substitute approach was to identify cases of lung cancer and, at the same time, to select some other persons comparable to the lung-cancer patients in age, sex, and other characteristics, and then to determine whether the persons with lung cancer smoked more heavily than the artificially constructed "control group." The first study of this kind was reported by Müller in 1939, who found much more smoking among the lung-cancer cases. Of course, it was clear to statisticians and others that this kind of data must be regarded with caution. The "control group" in such a study can only be selected somewhat arbitrarily and it is quite easy to imagine that the selection might tend to omit, or fail to recognize, smokers because of conscious or unconscious bias on the part of either the investigator or the persons responding to questions regarding their own smoking habits or those of their deceased relatives. Similarly, biases might exaggerate smoking experience among the lung-cancer cases.

Because the retrospective approach did not measure up to the standards of the randomized clinical experiment, other approaches were tried. Raymond Pearl (1938), an eminent medical statistician at Johns Hopkins University, had been keeping close records of health experiences of hundreds of families in the Baltimore area for some years. Pearl's work was going on at the time that serious study of the tobacco and health question began, and he decided to compile information about the smoking habits and longevity of all males in his files. Using the statistical methods common in computing life tables for life insurance purposes, Professor Pearl constructed a life table for nonsmokers, another for moderate smokers and a third for heavy smokers (see Figure 1). His data showed that 65% of the nonsmokers survive to age 60, whereas only 45% of the heavy smokers live that long. Indeed, Figure 1 shows that at every age between 30 and 90, proportionately more nonsmokers survive

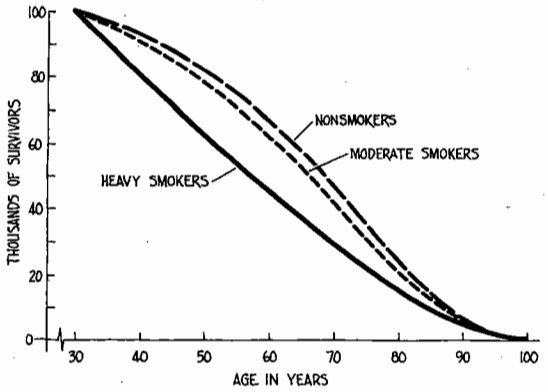

FIGURE 1

The survivorship lines of life tables for white males falling into three categories relative to their tobacco usage: (a) nonsmokers, (b) moderate smokers, smokers, and (c) heavy smokers. Source: Pearl (1938)

than moderate smokers, who, in turn, survive longer than heavy smokers. The impressive study hinted at smoking effects far beyond an increase in the relatively rare disease of lung cancer.

When Pearl's results were published, they added substantial weight to the evidence against smoking. But his claims were based on a special kind of study, so workers who did not have this sort of longitudinal data continued using the retrospective method of identifying lung-cancer patients and demonstrating that these patients were tobacco smokers to a much larger extent than any comparable control group that might be chosen. In the next decade many such retrospective studies were published, but the mounting number of scientific reports did not add conviction in proportion to their number. They all used the same method and they all might be drawing the same erroneous conclusion. Meanwhile, the consumption of tobacco per capita continued to increase, smoking among teenagers continued to rise, and the medical profession as a whole did not advise the public against the habit.

Medical scientists recognized the tremendous importance of the question. Tobacco was an intimate and welcome part of the lives of a substantial portion of the people. The tobacco industry was a vital part of the economy of a huge region of the U. S. and furnished the livelihood of additional thousands through related industries such as advertising, transportation, and sales. If tobacco were as harmful as Pearl's 1938 paper indicated, it would be the

obligation of the medical profession and government health agencies to help the public protect itself against this hazard. On the other hand, if the studies were erroneous, it would be extremely unfortunate if the medical profession or the government were to interfere in private lives and disrupt large segments of the economy on the basis of unsound scientific work. Thus medical scientists were deeply concerned about assessing the strength of the evidence against tobacco. Statisticians, especially in the biological and medical sciences, found themselves at the center of the dilemma facing the question. "Is the retrospective study a sound scientific approach?"

PROSPECTIVE STUDIES

It was felt that a new kind of study—more closely resembling the randomized clinical study—must be done. Pearl's choice of subjects and study were unsatisfactory in that his methods for choosing families and for choosing subjects within the families were never clearly set down. The chance for substantial biases in selection of the subjects and in the interviews for smoking experience seemed unacceptably large, and the report was not taken seriously. His prospective approach, however, in which smokers and nonsmokers are identified and then studied until death is an appealing one. It is a much closer approximation of a real experiment, in which a population would be split randomly into two groups; one group would be given one treatment (in this case, smoking) and the other group would receive an alternative treatment (no smoking). Both groups would be studied over time to determine the effects of the treatments. Even in the prospective study, the lack of random assignment of persons to be smokers and nonsmokers means that the conclusions are subject to the reservation that the smokers may be different from the nonsmokers in some systematic way that is entirely unrelated to their smoking habits, but that is causally related to some other factor that is deleterious to their health. For example, suppose that people who *like* smoking tend to get cancer and people who *don't like* it tend not to. Despite this inherent weakness in the prospective study approach, statisticians and other scientists called for this kind of study, and several were conducted in the fifties.

The first two prospective studies of the tobacco question were done by eminent statisticians in England and the U.S. In England, Dr. Richard Doll and Sir A. Bradford Hill, whose careful retrospective studies of the question had led them to conclude that smoking does cause lung cancer, were the first to report on a prospective study. They sent questionnaires to all members of the medical profession in the United Kingdom, roughly 60,000 men and women, and received about 40,000 replies. They asked about present smoking habits and a few other characteristics, such as age and sex. Then, through the Registrars-General in the United Kingdom, they determined the survivorship of these men and women over a period of several years. Their

study results were as anticipated: there were remarkably more lung-cancer deaths among the smokers than among the nonsmokers. Of course, this report, based on a different study approach in which the information on smoking habits was obtained first, rather than after the fact of death or disease, was regarded as extremely important additional evidence against smoking.

The prospective study offered new information that could not be obtained from the retrospective study. The retrospective study offered counts of smokers among lung-cancer cases. The prospective study offered counts of lung-cancer deaths among smokers and, thus, a direct measure of the death rate for lung cancer among smokers. The Doll and Hill data showed the rate for heavy smokers to be 1.66 per 1000 men per year, compared to 0.07 for nonsmokers; thus, heavy smokers had a lung-cancer death rate 24 times higher than nonsmokers. But even more important, the prospective study recorded all deaths and the recorded causes of these deaths, whereas the retrospective studies focused on the cause of death that was suspect. With information on all deaths that occurred in the group it was found that there was also a surplus of deaths due to heart attack among the smokers, and that this surplus was not readily explainable as chance variation in the data. Medical investigators had reported the effects of tobacco on the body (for example, the immediate cooling of the skin through contraction of the blood vessels when a person begins to smoke a cigarette), but the Doll and Hill study provided rather direct evidence that such effects on the heart and blood system could increase the risk of early death. The death rate for heart attacks among smokers was 5.99 per 1000 men per year for heavy smokers and 4.22 for nonsmokers—not a large difference, but extremely important, if real, because of the large number of deaths from this cause.

The second prospective study was reported by two American statisticians, Dr. E. Cuyler Hammond and Dr. Daniel Horn. Through the American Cancer Society, these men enlisted the aid of about 22,000 woman volunteers, each of whom was asked to select 10 healthy men between the ages of 50 and 69 and have each of them fill out a smoking questionnaire. Then these women reported on the health status of each of these men each year. Death certificates were obtained for each death reported. About 200,000 men were followed through a period of almost four years during which some 12,000 of them died. The Hammond and Horn report confirmed the findings of the Doll and Hill prospective study. Hammond and Horn reported a lung-cancer rate 23.4 times higher among heavy smokers (more than one pack a day) compared to nonsmokers and a death rate for heart and circulatory diseases 1.57 times as high among heavy smokers as among nonsmokers. Hammond and Horn emphasized the fact that although lung-cancer rates were strikingly high among smokers, the greater share of the excess deaths among smokers occurred in the heart and circulatory disease category, simply because this is a much more common cause of death.

NEW OBJECTIONS

Further prospective studies, involving different investigators and large groups of subjects chosen from various sources and followed for periods of up to a decade, were reported in the late fifties and early sixties. All of these studies substantiated the results of the first two studies. The evidence, which had been carefully gathered and evaluated by eminent men of medicine and statistics, seemed to condemn tobacco. But two factors were yet to be reckoned with: the impressive economic importance of tobacco as a prime industry and as contributor to other industries and the overwhelming strength of the smoking habit among the population. Again the statisticians played a major role in the drama, but this time, surprisingly, their influence was in defense of the "vile weed." One of the defenders was Sir Ronald Fisher, the man whose ideas were behind the randomized clinical trial and perhaps the greatest statistician who ever lived (Fisher died in the mid-sixties). Another statistician who came to the defense of tobacco was Dr. Joseph Berkson, Chief of Medical Statistics for the famed Mayo Clinic until he retired in the mid-sixties and one of the most creative people in medical statistics, having been trained in both medicine and medical statistics. Fisher and Berkson, both colorful and persuasive scientists, in the past had often found themselves on opposite sides of arguments concerning scientific methodology. This time they were on the same side, however, and Berkson was heard to say in jest that this was the only point that caused him serious doubt about his position on the tobacco-health question.

Although the positions taken by Fisher and Berkson gave aid and comfort to the habitual smoker, to the tobacco industry, and to the people who made their living through the tobacco industry, it must be stressed that these men had a larger and more important issue in mind, namely, the methodology of science itself. And, though the short-term effect of their writings was a delay in the scientific indictment of tobacco, the long-term effects will more than counterbalance these by tightening up scientific standards for medical studies in which appropriate randomized clinical experiments cannot be carried out because of possible danger to participants. In the late fifties, Fisher and Berkson both repeatedly pointed out that the evidence against tobacco as a hazard to the health was only circumstantial because clinical experiments could not be done. They pointed out, in detail, weaknesses in the evidence, even in the evidence from the prospective studies. They proposed other explanations that would account for the data if tobacco had no deleterious effects on the health.

Berkson's main point was that the apparent ill effects of smoking were too pervasive. The death rates seemed to be higher for too many different causes of death. Such a universally bad effect from tobacco was never anticipated, and there were no theories to explain how smoking could cause death through each of the multitude of diseases implicated by the reports from the prospective

studies. Berkson suggested that a more likely explanation was that there were biases in sample selection, or in the data collected, or both, and that these biases affected all of the prospective studies because they were inherent in the methodology of the prospective studies and they affected all causes of death. Otherwise, Berkson suggested, we must hypothesize some sort of general constitutional effect of smoking on the body, or some sort of accelerated aging caused by smoking; he was dissatisfied also by the absence of known mechanisms causing the disorders.

Fisher, who had been knighted for his work in genetics and who had always been concerned with the problems of interpreting nonexperimental (i.e., nonrandomized) studies, put these two interests together and argued that if persons with a hereditary tendency toward smoking also had a hereditary predilection for disease, the result would be exactly the kind that had been repeatedly demonstrated, both retrospectively and prospectively. Then, Fisher argued, if the explanation lay in such hereditary predilections, a smoker who quit could not change his genetic makeup or his peculiar higher risk of disease thereby, and smoking could not be regarded as harmful to health. Fisher proposed some (nonrandomized) twin studies that might shed light on his hypotheses, but little could be done because large groups of people were needed for definitive results. (See the essay by Reid, especially Table 2 and its discussion.)

Neither Fisher nor Berkson has been completely answered. In fact, the questions they raised cannot be answered definitively without randomized experiments. Of course, Fisher and Berkson knew this as well as anyone, but such conspicuous critiques, made by widely known men of high prestige, in the context of a scientific question of intense public interest, served to underline the dangers in drawing conclusions from nonrandomized studies.

FURTHER RESEARCH

The immediate effect of the Fisher and Berkson critiques was to spur on statisticians and medical scientists to an even more careful look at the tobacco-health question. Although the questions raised by Berkson and Fisher could never be answered directly, additional indirect evidence has been marshaled on many points. There is space here only to mention some of the approaches, without detailing their results.

Because experimental verification of smoking as a cause of ill effects in man did not seem practicable, scientists had attempted to investigate the effects of tobacco smoke on animals, where random assignment to smoking and nonsmoking groups was feasible. Although much experimental work had been done, until the early sixties the actual smoking experience of the human had not been adequately simulated in the laboratory and no direct positive evidence had been obtained. In the sixties, however, studies demonstrated that tobacco smoke and some of its constituents do cause lesions in animal

tissue that are similar to those seen in human lung cancer. When, finally, an experimental setup was attained that satisfactorily simulated the smoking habit in dogs, a few cases of lung cancer were induced in what seemed to be well-controlled, randomized experiments.

At the same time, in partial answer to Berkson's point that there was no theory or information that would lead one to suspect that smoking would affect the diversity of organs and systems cited in the prospective reports, much work in animals and humans has been carried out to determine what physiological effects tobacco has on the organ systems. One of the most important results has been the discovery that tobacco smoke does have deleterious effects on the blood vessels of animals and the verification of this same damage in the blood vessels of heavy smokers.

The evidence against tobacco has been strengthened by more detailed information and closer statistical evaluation in ongoing prospective studies. It has been found that smoking is not associated with all causes of death, but that for the causes implicated, the death rates increase regularly and convincingly with the amount smoked. When the details of smoking habits and smoking experience were analyzed, including especially the results for persons who had smoked but quit the habit for varying lengths of time, the death rates have been found to be closely consistent with what might be expected if smoking were indeed a causal factor in death.

THE SURGEON GENERAL'S REPORT

By the mid-sixties, scientific opinion had swung far toward the conclusion that smoking is harmful. Despite the realization that the evidence against smoking can only be indirect, scientists were convinced that some conclusion must be reached and acted on for the good of the public health. A special committee, appointed by the Surgeon General of the U.S. Public Health Service, made a long and thorough study of all aspects of the tobacco-health question. The committee of ten men included experts on chemistry, pharmacology, internal medicine, surgery, pathology, epidemiology, and statistics. The statistical representative was William G. Cochran, Professor of Statistics at Harvard University and a world-renowned expert in the application of statistical methodology to problems of the life sciences. In 1963 this Advisory Committee submitted to the Surgeon General a report that cited smoking as a cause of lung cancer and several other cancers and stated that the evidence pointing to smoking as a cause of death due to heart attack was strong enough to justify acting on this presumption.

LATER DEVELOPMENTS

In the sixties, independent health organizations and governmental health agencies in this country and throughout the world publicly announced their

conclusions that tobacco is a health hazard. Along with these indictments came a call for further action. Many health officials felt that it wasn't enough to inform the public through the usual scientific publications, carried along by brief news reports and by word of mouth. Rather, they felt that a public habituated to tobacco and under constant bombardment by advertisements that associated cigarettes with youth, health, attractiveness, and elegance must be protected. The issue became one of how far the responsibility and legal authority of the government can extend to protect the citizen against himself and against the influences of deleterious propaganda, especially when the evidence did not measure up to the standards of true experimental science.

By 1970, there were still scientists (a few of them statisticians) who found the evidence against smoking less than convincing and who were willing to so testify at congressional hearings. But, for the most part, statisticians will play a new part in the tobacco-health arena as the issue ceases to be a question of science and becomes one of public policy. In 1969, Bernard Greenberg, Professor of Biostatistics at the University of North Carolina and a noted public health statistician, published a paper discussing the strength of evidence against smoking and the extent to which such evidence might justify various levels of government action in protection of the individual. These possible actions ranged from rather modest propaganda campaigns to legal judgments against tobacco companies for causing untimely death and illness through the vending of their product. Professor Greenberg's paper was not a philosophical discussion, concerned with ethical, political, and legal principles; it was an attempt to use modern statistical theory—this time a branch of mathematical statistics called *decision theory*—to arrive at reasoned advice, based on the costs to the public of actions that might be taken and the gains and losses that might accrue to the public as the result of such actions. This new kind of thinking in public health requires a different kind of statistical theory than that for experiments.

SUMMARY

This essay has traced the role of some eminent statisticians and the statistical methods they used to determine that smoking is harmful to the public health. The work that has been done on this problem and the debates on scientific method that it has stimulated will have untold benefit both through the eventual elimination of smoking as a general habit of the people and in the development of better techniques and higher standards for the scientific study of other hazards to the public health in the future.

PROBLEMS

1. What did R. A. Fisher suggest to use as an alternative to holding all factors constant except the ones under investigation?

2. Explain what is meant by an "unbiased experiment."

3. Explain the terms "retrospective study" and "prospective study." Give the pros and cons.

4. Refer to Figure 1. Approximately what percentage of males survived up to age 50 in the three different categories?

5. Refer to Figure 1. What was the approximate median lifetime (i.e., the age up to which 50% of the people survived) in each of the three categories?

6. What were the objections of Berkson and Fisher to the results of both the prospective and retrospective studies?

7. If a prospective study shows that there is a higher incidence of lung cancer among smokers than among nonsmokers, can we conclude that smoking causes lung cancer? Why or why not?

REFERENCES

Joseph Berkson. 1958. "Smoking and Lung Cancer: Some Observations on Two Recent Reports." *Journal of the American Statistical Association* 53:28–38.

Harold Diehl. 1970. *Tobacco and Your Health: The Smoking Controversy.* New York: McGraw-Hill.

Alfred Dunhill. 1954. *The Gentle Art of Smoking.* New York: Putnam.

R. A. Fisher. 1959. *Smoking: The Cancer Controversy.* Edinburgh: Oliver & Boyd.

B. G. Greenberg. 1969. "Problems of Statistical Inference in Health with Special Reference to the Cigarette Smoking and Lung Cancer Controversy." *Journal of the American Statistical Association* 64:739–758.

A. E. Hamilton. 1927. *This Smoking World.* New York: Appleton–Century.

Raymond Pearl. 1938. "Tobacco Smoking and Longevity." *Science* 87:216–217.

U. S. Public Health Service. 1963. *Smoking and Health.* Report of the Advisory Committee to the Surgeon General.