Two approaches to etiology: the debate over smoking and lung cancer in the 1950s

Mark Parascandola

Statisticians R.A. Fisher and Joseph Berkson have become infamous for ending up on the ‘wrong’ side of the debate over the evidence linking smoking and lung cancer during the 1950s, and scholars have speculated about their personal motives in the controversy. But there were many senior biostatisticians and epidemiologists voicing similar concerns about the quality of the evidence at the time, albeit with less inflammatory rhetoric. This debate occurred during a time when epidemiological research methods commonly used today were understood by few and were only just beginning to work their way into public health and medicine. All of the participants in the debate over smoking and lung cancer saw the need for explicit and rigorous standards for evaluating etiological hypotheses, but they held conflicting views about what those standards should be. The differing opinions on the evidence reflected two different models of etiological research – controlled experiment as the crucial, objective test of a causal hypothesis versus inferential judgment based on a diverse body of evidence. This debate has relevance for current epidemiological practice, as tension between these two views still remains.

Epidemiologists have struggled to understand the scientific and personal factors that drove prominent biostatisticians Ronald Aylmer Fisher and Joseph Berkson to the ‘wrong’ side of the debate over smoking and lung cancer in the 1950s. They both maintained a strong skeptical stance towards the mounting evidence far beyond, some suggest, the point of reason. Viewing the debate retrospectively, it is all too easy to divide the participants into those who were right and those who were wrong, or those who were against cigarettes and those who were for them. But, of course, that is too simplistic. While Berkson and Fisher stand out in the debate because of their professional stature and their strong rhetoric, a substantial number of biostatisticians and epidemiologists shared their concerns. For these scientists, it was not merely tobacco that was at stake, but the role of biostatistics and epidemiology as scientific disciplines.

The debate over the epidemiological data linking smoking and lung cancer during the 1950s provides an excellent case study to understand the tensions in methodology that were emerging at this time. The epidemiological tools used in the early cigarette studies were relatively new and perhaps it is unsurprising that they were misunderstood and viewed suspiciously by pathologists, physicians and others lacking statistical training [1,2]. But the suspicions of statisticians and epidemiologists themselves cannot be put down to the ignorance of the naive laboratory researcher. ‘Now I should be the last person to attack evidence for being merely statistical,’ Fisher acknowledged [3]. The epidemiologists and biostatisticians who participated in this debate differed on what were sometimes subtle points of methodology, but these differences had substantial implications for characterizing the potential hazards of cigarette smoking. All of the participants saw the need for explicit and rigorous standards for evaluating etiological hypotheses, but they held conflicting views about what those standards should be. The differing opinions on the evidence reflect two different models of research – controlled experiment as the crucial, objective test of a causal hypothesis versus inferential judgment based on a diverse body of evidence.

Two modes of etiological research

The first half of the 20th century saw the development of two distinct but parallel approaches to etiological research – the randomized controlled experiment and the analytic epidemiological survey. The first approach came out of biostatistics, developed and vigorously promoted by Fisher, whose Statistical Methods for Research Workers [4] and The Design of Experiments [5] became the standard textbooks. Followers of Fisher, including Major Greenwood and Austin Bradford Hill, took these tools into the field of medical research and promoted what they called ‘experimental epidemiology’, which initially meant testing epidemiological hypotheses in animal experiments [6]. The experimental approach (albeit without true randomization) was also taken up by public-health researchers for evaluating the efficacy of new vaccines.

At the same time, some influential public-health investigators emphasized the role of the epidemiologist as one of assembling a diverse body of facts from different sources into a coherent explanation. American epidemiologist Wade Hampton Frost wrote that ‘epidemiology is something more than the total of its established facts. It includes their orderly arrangement into chains of inference which extend more or less beyond the bonds of direct observation.’ [7]. Instead of emphasizing a particular experimental design or statistical method, Frost stressed the importance of analytical skills and knowledge of a wide range of facts. These two approaches – experimental and inferential – are not mutually exclusive, but they do...
represent two distinct trends, and this subtle divide was crucial in the debate over cigarettes and lung cancer.

This debate began in 1950, with the publication of five case-control studies that compared the smoking habits of lung cancer hospital patients with non-lung cancer patients [8–12]. Case-control studies were especially useful at the early stages of the investigation, when several potential causes were being investigated. Yet the method clearly did not meet Fisher’s requirements for experimental design. A group of about a dozen biostatisticians in the Biometry Branch of the National Cancer Institute (NCI) played a leading role in advancing new methods for the analysis of case-control data and in building the case against cigarettes [13,14]. Large cohort studies, which compared cancer rates in smokers and non-smokers over time, also provided crucial epidemiological support, although some biostatisticians also questioned these more rigorous studies.

The crucial experiment

Joseph Berkson and R.A. Fisher have become infamous for their stubborn opposition to the emerging evidence about the dangers of cigarettes (Fig. 1). Both argued that biological knowledge was essential and statistical methods limited, but their aim was ultimately to defend a particular kind of statistical expertise – that of the biostatistician who designed and interpreted randomized controlled experiments. Berkson reinforced the central role of biological knowledge in guiding research: ‘if biologists permit statisticians to become arbiters of biological questions, scientific disaster is inevitable.’ However, knowledge of biological mechanisms was important not as an end in itself, but primarily because it suggested experiments. ‘The most important consideration with respect to a theory is not whether it appears plausible, but whether it suggests experiments, and what experiments are suggested.’ [15].

Fisher clung to the possibility that lung cancer and the smoking habit may both be produced by some common genetic trait. The statistical association between smoking and cancer, he insisted, did not imply that one must cause the other. Only a randomized controlled experiment could rule out the common cause hypothesis, he argued [3]. But Fisher was not alone in taking this hypothesis seriously. In fact, even Charles Cameron, president of the American Cancer Society, proposed that hormonal differences might explain both smoking habits and cancer susceptibility [16]. Moreover, studies conducted during the 1950s provided evidence that smokers did in fact differ in several ways from non-smokers, including personality, hospitalization rates, occupation, diet and physical characteristics [17,18]. Two of these studies were conducted by epidemiologists, one by Clark Heath, with tobacco industry funding, and one by Abraham Lilienfeld, then at Roswell Park Memorial Institute. But this debate rested on more than a pet hypothesis.

While Berkson and Fisher were especially outspoken and uncompromising in their views, their concerns were shared by a substantial number of senior biostatisticians at the time, including Donald Mainland at New York University, Antonio Ciocco at the University of Pittsburgh, and K.A. Brownlee at the University of Chicago. This debate was going on while the randomized controlled trial was a recent development in medical research and its place was still in doubt. Biostatisticians and clinical trialists were trying to persuade the medical establishment that their expertise was vital to research and that their scientific methods were comparable in validity to laboratory experiments [19,20]. Simple procedures like the test of statistical significance had worked their way into medical research, but statisticians objected that they were often naively applied by those without statistical training. Fisher himself probably contributed to this tendency by strongly emphasizing tests of statistical significance in his textbooks and showing an interest only in the interpretation of the results of single trials and not from a diverse body of evidence. In short, it was not just tobacco that was at stake in the debate, but the authority of the biostatistician in medical science.

Jacob Yerushalmy, a biostatistician at Berkeley, later took a similar critical stand in response to studies showing a link between maternal smoking and infant mortality. Here he argued that the data were equally consistent with competing interpretations, that either the smoking or the constitution of smokers could explain an association between maternal smoking and low birth weight. Non-randomized studies simply were helpless to resolve the conflict. Yerushalmy made it clear that the problem was not with the statistical form of the results. ‘The evidence may not be convincing, but not because it is ‘only statistical’, rather because the evidence is nonstatistical in the sense that the method of study which produced the evidence violates the basic principles for valid statistical inference.’ [21].

Statisticians also argued that smokers and non-smokers in epidemiological follow-up studies might differ in ways
other than in their smoking habits, because smokers (or non-smokers) might be more or less inclined to participate in a voluntary study, thus introducing selection bias [22]. Experienced public-health research methodologists were likely aware of earlier discussions over the problems of using volunteers in vaccine studies. Additionally, some statisticians criticized E. Cuyler Hammond and Daniel Horn’s use of 22,000 American Cancer Society volunteers to recruit and interview subjects for their smoking study. Mainland argued that researchers, particularly untrained volunteers, were likely to select a particular type of individual for the study; he conducted a survey of his students to demonstrate that they did indeed suffer from this procedural bias [23].

Most epidemiologists and statisticians, including those who defended the link between cigarettes and lung cancer, admitted that a controlled experiment, when it was available, offered the strongest evidence that any single study could offer [24–26]. Richard Doll, coauthor of one of the early case-control studies and also a clinical trialist, did conduct a randomized controlled trial of the effect of giving up smoking on the rate of healing of gastric ulcers. Doll and coauthors acknowledged that an association between smoking cessation and earlier healing could be due to a common cause. They concluded that while their earlier case-control study could not choose between two hypotheses, the controlled trial made the common-cause hypothesis less probable (in spite of the high noncompliance in their study) [27]. And Bradford Hill strongly advocated randomization and criticized excessive focus on tests of statistical significance, while also defending the observational approach [28].

However, several epidemiologists and biostatisticians who championed the cigarette–lung-cancer link, such as Hammond, Cornfield, and Lilienfeld, downplayed the role of experimentation. Even a randomized controlled experiment did not guarantee the validity of a causal inference, they argued, particularly when one wanted to generalize that inference outside the particular experiment. Cornfield and Hammond both explained that randomization by itself is insufficient for inference without additional judgments and assumptions on the part of the experimenter [29]. As Hammond explained in an experiment, ‘the cause of the effect on the dependent variable was the totality of conditions sufficient to produce it.’ [25]. The experimental intervention consisted not only of the experimental drug, say, but the manner in which it was delivered. Thus, drawing generalizations about the intervention’s effectiveness in other situations could be risky. These participants also maintained that the differences between different study designs, while real, had been overstated [30,31]. Cornfield acknowledged that while experimental studies offered greater control, ‘there are no such categories as first-class evidence and second-class evidence.’ [29]. The difference was one of degree rather than kind. There was no single, crucial experimental design that would trump all other forms of evidence. Instead, several participants emphasized the need to synthesize a diverse body of evidence and look at the interrelationships between findings. As Lilienfeld stated, ‘the plausibility of the causal hypothesis is assessed, not in terms of the results of one particular study or a ‘crucial experiment’, but in terms of the totality of available biological evidence.’ [32]. The consistency between case control and cohort findings seemed to further this argument.

**Conflicting criteria**

An influential 1959 paper by Yerushalmy and Carroll E. Palmer, a commissioned PHS officer with expertise in biostatistics, faulted epidemiological studies of cancer and urged that they should follow ‘the more rigorous methods long in use by bacteriologists.’ Bacteriologists had a set of rules – Koch’s postulates – for drawing etiological conclusions about infectious agents. But for chronic diseases there were only vectors, like cigarette smoke, that somewhere contained specific carcinogens, and statistical methods were particularly important in studying these hidden causes. Yerushalmy and Palmer, of course, saw tremendous value in statistical methods. But the problem in this case was that ‘conventional statistical techniques cannot be utilized without modification, because the fundamental requirement of group comparability, ordinarily achieved through randomization, is not satisfied.’ [33]. This paper was part of an ongoing debate during the 1950s and early 1960s about criteria for causal inference based on non-randomized epidemiological studies.

To help rule out non-causal spurious associations, Yerushalmy urged that investigators should test for the ‘specificity’ of an association. However, Yerushalmy’s interest in specificity was not driven by a naive monicausal theory. Instead, the rationale was based in statistics. Before reaching the stage of causal inference, Yerushalmy, writing with Herman Hilleboe of the New York State Department of Health, argued that researchers must determine whether the *association* observed is what it seems. They should ascertain whether ‘the association between two variables is in fact between the variables investigated and does not merely reflect relationships with a broader group, of which one or the other of the variables forms a part.’ [34]. Thus, is there really an association between cigarette smoking and lung cancer, or is that link merely a byproduct of a cluster of correlations, as poverty is associated with many health outcomes? Berkson similarly argued that most (88.5%) of the excess deaths among smokers in the Hammond and Horn study were not from lung cancer, but from various other causes, indicating some sort of selection bias [35]. Indeed, the fact that there was not a one-to-one deterministic relationship between smoking and lung cancer, that not all smokers developed lung cancer and some nonsmokers did, was simply not an important issue of debate among statisticians and epidemiologists. Those on both sides of the smoking debate explicitly defined a cause in probabilistic terms, as a factor that increases the probability of disease, rather than in terms of necessity and sufficiency [25,33,36].

Proponents of the cigarette–lung-cancer link proposed their own guidelines for drawing causal inferences from a body of epidemiological evidence, which differed significantly from those proposed by Yerushalmy and Palmer. For example, the strength of an association was held to be
important because a strong noncausal association required a strong confounder to explain it, and weak associations were more likely to be artifacts of selection bias. Such an obvious fact could hardly escape the attention of any conscientious investigator, the reasoning went. By contrast, Yerushalmy had rejected strength of association as a criterion for inference because its evaluation was necessarily subjective. ‘There is no rational way to decide how large a difference there must be before we accept it as indicating a cause–effect relationship,’ he wrote [37]. At the same time, the lack of specificity of smoking as a cause of lung cancer was not a source of worry, according to Lilienfeld, because the association between smoking and lung cancer in particular was so dramatic relative to other adverse effects [17] (Fig. 2).

Biological plausibility played a relatively minor role here, as only one piece in the larger puzzle. The need for understanding biological mechanisms was cited as important for suggesting avenues for prevention rather than as a crucial element in causal inference. Those who did mention biological plausibility did not require that biological mechanisms be worked out before claiming causation, only that findings should be consistent with current medical and biological knowledge. Additionally, because of the complex nature of cancer causation and the many differences between animal models and humans, these same scientists urged that negative animal experiments were of little relevance to human cancer etiology.

But most importantly, the scientists who supported the cigarette–lung-cancer hypothesis emphasized the overall consistency and coherence of the findings, rather than emphasizing the need for a crucial experiment as Yerushalmy and Berkson did. In short, a causal relationship should make sense in light of a broad range of biological, epidemiological and demographic facts. Thus, their reviews of the evidence focused on making sense of a variety of observations (such as the differences in smoking habits and lung-cancer rates by age and sex) as the primary method for choosing between rival causal hypotheses. Two workshops including key players such as Cornfield, William Haenszel, Hammond, Lilienfeld, Michael Shimkin, and Wynder, summarized the evidence from epidemiological, pathological and animal studies, explaining how it answered the particular concerns raised by skeptics like Fisher and Berkson. They concluded that the evidence linking cigarettes and lung cancer was ‘beyond dispute’ [38,39]. These same investigators also conducted studies that were solely directed at answering challenges to the overall coherence of the findings; for example, they compared urban versus rural smoking habits and studied the quality of cancer mortality survey data in the context of this debate [40,41].

It was this global approach to the evidence, both pragmatic and analytical, that was ultimately employed in the 1964 report of the Surgeon General’s Advisory Committee on Smoking and Health (Fig. 3). The Committee provided a set of criteria for judging a diverse body of evidence that incorporated key elements of the earlier debate: the criteria included consistency, strength, specificity, temporal relationship and coherence. They explained that, in the absence of experimentation, the ‘causal significance of an association is a matter of judgment.’ [42]. Today, the 1964 Surgeon General’s report and Hill’s subsequent causal criteria are routinely cited as authoritative statements of the proper method for assessing a body of etiological evidence [43,44].

Discussion
At least two distinct approaches to etiological research were in conflict in this debate, one rooted in experimental science and another rooted in a pragmatic need to make inferences from a diverse assortment of data (Table 1). The latter was more responsive to the concerns of public health practice. At the same time that the debate over smoking and lung cancer was being played out, epidemiologists were debating the methodological future of their discipline and its role in public health. In 1962, Milton and Terris urged that epidemiologists to look closely at Joseph Golberger’s pellagra experiments, rather than to John Snow’s observational cholera investigations, for their etiological research paradigm [45]. Morton Schweitzer at Columbia University replied that ‘the epidemiologist never exactly fulfills the requirements of statistical theory’ and, therefore, a formal experiment is not necessary for demonstrating causation [46]. Also at this time, planning was actively going on for the ambitious but ill-fated Diet–Heart study at the National Heart Institute; research advisors questioned whether a true randomized, double-blind trial of diet on heart disease was feasible [19].

These differences were also likely due, in part, to differences in professional training. Older biostatisticians like Fisher, Berkson and Yerushalmy had been trained in the analysis of experimental data and, as heirs to Karl Pearson, emphasized rigorous empirical methodology and scientific criticism, rather than inferential reasoning, as the foundation of objective science. Fisher was not disinterested in drawing conclusions about cause and

![Fig. 2. A lecture on the effects of smoking on the lungs. WHO photo by D. Henrioud.](www.sciencedirect.com)
effect, but he certainly held that such knowledge could only be had under very particular circumstances. Additionally, earlier Public Health Service epidemiologists like Palmer, Alexander Gilliam and Dorn had first-hand experience of the pitfalls of mortality data and voluntary vaccine trials, which likely made them more cognizant of the shortcomings of that data. By contrast, the epidemiologists and biostatisticians who were proponents of the smoking and lung-cancer hypothesis mostly belonged to a younger generation and were not steeped in Fisher’s experimentalism. Indeed, Cornfield, Nathan Mantel and William Haenszel had arrived at NCI as mathematical statisticians without experience in medical or biological research, and whilst there they aided laboratory scientists, clinical trialists and epidemiologists alike. Thus, it is not surprising that they would give more attention to devising ways to get more use out of existing data rather than to worrying about the origins of that data. Additionally, they were not reluctant to use those data to draw inferences.

A parallel methodological divide continues to create conflict even today. Over the past several years, a cohort of epidemiologists have waged a campaign against post-war academic epidemiology, arguing that it focuses too much on testing potential determinants of disease for the individual. Instead, they argue, the discipline should broaden its scope to consider social and ecologic causes and should focus more on developing theoretical models and less on conventional study design [47,48]. By contrast, epidemiologists who focus on methodology maintain that epidemiologists must strive for scientific objectivity, rather than allow themselves to be driven by specific public health goals [49]. Of course, both camps acknowledge the importance of traditional statistical methodology as well as pragmatic public-health concerns, but subtle differences in emphasis have created substantial divisions within the field. Methodological debate is healthy in any discipline and should not be discouraged. Yet epidemiologists today should be aware of how such subtle differences, particularly when powerful nonscientific interests are at stake, can create an atmosphere of uncertainty that persists for decades.

References
13 Cornfield, J. (1951) A method of estimating comparative rates from...
clinical data. Applications to cancer of the lung, breast, and cervix. J. National Cancer Institute 11, 1269–1275
27 Doll, R. et al. (1958) Effect of smoking on the production and maintenance of gastric and duodenal ulcers. Lancet 1, 657–662
39 Study Group on Smoking and Health (1957) Smoking and health. Science 125, 1129–1133

Do you want to reproduce material from a Trends journal?
This publication and the individual contributions within it are protected by the copyright of Elsevier. Except as outlined in the terms and conditions (see p. ii), no part of any Trends journal can be reproduced, either in print or electronic form, without written permission from Elsevier. Please address any permission requests to:

Rights and Permissions,
Elsevier Ltd,
PO Box 800, Oxford, UK OX5 1DX.
E-mail: permissions@elsevier.com

www.sciencedirect.com