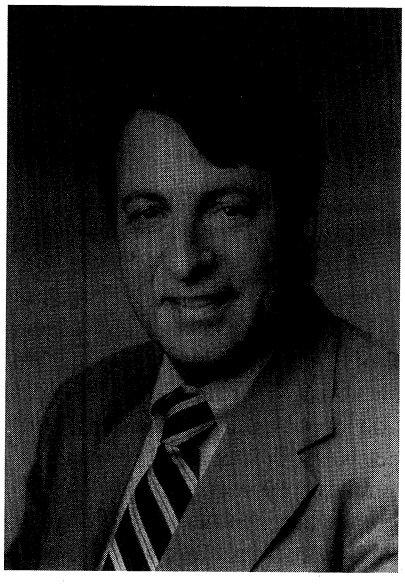
Statistics in Action

Mitchell H. Gail

Journal of the American Statistical Association, Vol. 91, No. 433 (Mar., 1996), 1-13.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199603%2991%3A433%3C1%3ASIA%3E2.0.CO%3B2-%23

Journal of the American Statistical Association is currently published by American Statistical Association.



Statistics in Action

Mitchell H. GAIL

1. INTRODUCTION

Presidents are sometimes asked to make predictions and to chart a course for the future of an association. I am reminded of the President of the Clairvoyance Society who brought his family to Orlando just last year so they could hear his Presidential Address. When he arrived at the Swan Hotel, he was astonished to discover that the Annual Meeting of the Clairvoyance Society had been cancelled—due to unforeseen circumstances. I took this lesson to heart and decided instead to illustrate my theme, "statistics in action," with two examples from the past: the advent of the randomized controlled clinical trial and the debate over whether smoking causes lung cancer. These two examples of statistics in action have increased the awareness of the importance of statistical thinking in medical and public health circles.

Sometimes the statistician's work is highly visible. For example, in 1964, U.S. Surgeon General Luther Terry issued a report, "Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service" (U.S. Department of Health, Education, and Welfare, Public Health Service 1964), that summarized 15 years of intense study and debate and set the stage for efforts to curb smoking in the United States. Statistical thinking, data collection, and analysis were crucial to understanding the strengths and potential weaknesses of the scientific evidence. This effort enhanced the stature and visibility of statisticians and the statistical method and gave rise to new methodologic insights and constructive debate on criteria needed to infer a causal relationship. These ideas form the foundation for much of current epidemiologic practice.

At other times, the role of the statistician is less apparent. Perhaps scientific collaborators will change their minds about the quality or strength of evidence or perhaps they will adopt a particular experimental design on the recommendation of a statistician. But even apparently small agreements that arise when the statistician finds an approach particularly suited to solving the issue at hand and convinces his co-workers of the appropriateness of that approach can have unforeseen and wide-ranging consequences. For example, in 1946, A. B. Hill persuaded members of the Whooping-Cough Immunization Committee of the Medical Research Council (MRC) to adopt a randomized trial design for studying the *H. pertussis* vaccine, and, a few months later, he persuaded members of the Streptomycin in Tuberculosis Trials Committee of the MRC to conduct a randomized trial comparing streptomycin and bed rest with bed rest alone in young patients with acute progressive bilateral pulmonary tuberculosis. These two trials set the pattern for the thousands of trials that followed and are ongoing in every area of preventive medicine and experimental therapeutics today.

I shall emphasize statistical applications in developing the theme of "statistics in action" by elaborating on these two examples; the advent of a new paradigm for randomized comparative clinical trials and the struggle to use observational data to infer that smoking causes lung cancer. By highlighting applications, I do not mean to slight the important role of statistical theory. Without the proper statistical tools and principles, and without a close familiarity with the scientific method and with the strengths and weaknesses

Mitchell H. Gail is Chief, Biostatistics Branch, National Cancer Institute, Bethesda, MD 20892. The author thanks Peter Armitage, Ruth Bittler Cornfield, Richard Doll, Joseph F. Fraumeni, Jr., Joseph L. Gastwirth, Edmund A. Gehan, Tricia Hartge, and Sholom Wacholder for helpful comments and Jennifer Donaldson for preparing the manuscript. The author thanks Lance Waller for finding the typewritten manuscript by H. S. Diehl in the Bio-Medical Library of the University of Minnesota that clarified the method of allocation of Diehl, Baker, and Cowan (1938). The author also thanks the Royal Statistical Society for providing the original print and giving permission to reprint Figure 1 (originally published in Armitage 1977), and Ruth Bittler Cornfield for providing the picture of Jerome Cornfield for Figure 2. This article was the basis for the Presidential Address to the American Statistical Association in Orlando, Florida, on August 15, 1995.

Journal of the American Statistical Association
March 1996, Vol. 91, No. 433, Presidential Address

of experimentation and observational study, the statistician is ill-equipped to contribute effectively to the solution of complex problems. But more is needed, including a serious effort to understand the problem itself, as seen by the scientist or administrator or businessperson who poses it. As my mentor at the National Cancer Institute, David Byar, often said, "Before consulting on a problem in medical statistics, I read 100 medical papers for every methodologic one." Such multifaceted preparations are needed if the applied statistician is to play an inspired role.

After reviewing these two seminal examples of "statistics in action," I will comment on some qualities and attitudes that typify the excellent applied statistician. I will also comment briefly on how we, as members of a diverse Association, can benefit from giving a prominent place to applications and to "statistics in action" in our meetings and publications and also in our dealings with those outside our profession.

2. CLINICAL TRIALS

2.1 Comparability and Randomization

Why did it take so long to introduce the randomized comparative trial as a decisive method for assessing medical treatments? It was not because physicians were unaware of the need for comparability when making treatment comparisons. Armitage (1983) reviewed the contributions of Pierre Charles Alexandre Louis, an astute clinician and pathologist, who introduced the "numerical method" for comparing treatments. Louis' idea was to keep careful records on the results of treatments and to compare treatments on groups of patients with similar degrees of disease or illness—or, as we would put it today, to compare "like with like." Using this method, Louis (1835, 1836) came to the surprising conclusion that "bloodletting has had but a very limited influence on the course of pneumonitis." And this was not because Louis and his colleagues were timid at bloodletting. They typically took about 12 ounces of blood in each of two lettings, or about $1\frac{1}{2}$ pints total. One patient died despite five repeated bloodlettings amounting to about $4\frac{3}{8}$ pints, or about 37% of the blood volume of a 70-kg man.

Louis (1837) explained his method as follows:

> I come now to therapeutics, and suppose that you have some doubt as to the efficacy of a particular remedy. How are you to proceed? ... You would take as many cases as possible, of as similar a description as you could find, and would count how many recovered under one mode of treatment, and how many under another; in how short a time they did so; and if the cases were in all respects alike, except in the treatment, you would have some confidence in your conclusions; and if you were fortunate enough to have a sufficient number of facts from which to deduce any general law, it would lead to your employment in practice of the method which you had seen oftenest successful.

Although the desirability of comparing patients "in all aspects alike, except in the treatment" was well accepted, it was not evident how to obtain such comparability. Nonetheless, the idea that randomization might be a useful tool for obtaining fair treatment comparisons is more than 300 years old. According to Armitage (1983), the Belgian pharmacist J. B. van Helmont offered to bet 300 florins that his curative powers were greater than those of academic physicians. He proposed in 1662:

> Let us take out of the hospitals, out of the Camps, or from elsewhere, 200, or 500 poor People that have Fevers, Pleurisies &c. Let us divide them into halfes, let us cast lots, that one half of them may fall to my share, and the others to yours; ... we shall see how many funerals both of us shall have.

Today we would call this a "group randomized trial" and not recommend it because only two groups were to be allocated. The academic physicians of the day rejected this design too, though possibly not on statistical grounds. Nonetheless the idea that randomization could yield a fair (unbiased) estimate of treatment effect is implicit in this proposal.

But even if one has an unbiased estimate of treatment effect, measured by the difference in mean treatment response between an experimental and control group, one still needs a valid estimate of the standard error of this difference to determine its importance. In the 1920s, Fisher described randomization as a method that could yield valid estimates of this standard error in agricultural field studies, because with randomization and under the null hypothesis of no treatment effect on any experimental unit, the variability of responses within treatment groups would be representative of the variability of responses overall and hence could be used to compute the standard error of the difference in mean responses (Fisher 1925, 1926; Fisher and MacKenzie 1923). Without randomization, complex trends in fertility patterns among plots could induce positive or negative correlations among responses within treatment groups and result in a biased estimate of standard error.

Fisher seems to have taken it as self-evident that randomization yields an unbiased estimate of treatment effect, but this idea is made explicit in *The Design of Experiments* (Fisher 1935), where he referred to an experiment involving 15 pairs of plants. A cross-fertilized plant was paired with a self-fertilized plant and grown in the same pot, with pairwise randomization used to allocate the cross-fertilized and self-fertilized plants either to the east or west side of the pot. Fisher commented on the role of pairwise randomization:

> This is to say much more than merely that the experiment is unbiased, for we might still call the experiment unbiased if the whole of the cross-fertilized plants had been assigned to the west side of the pots, and the self-fertilized plants to the east side, by a single toss of the coin ... Randomization properly carried out, in which each pair of plants are assigned their position independently at random, ensures that the estimates of error will take care of all such causes of different growth rates and relieves the experimenter from the anxiety of considering and estimating the magnitude of the innumerable causes by which his data may be disturbed.

2.2 Persuasion

Thus the necessity for comparing "like with like" had been established as a principle of medical therapeutics, and the method of randomization had been used in agricultural experimentation since the mid-1920s to produce unbiased estimates of treatment effect and reliable estimates of the corresponding standard error. Why, then, was this technique not being used in clinical therapeutics? Two hurdles had to be overcome. First, medical investigators had to be educated in the principles of experimentation and the need for

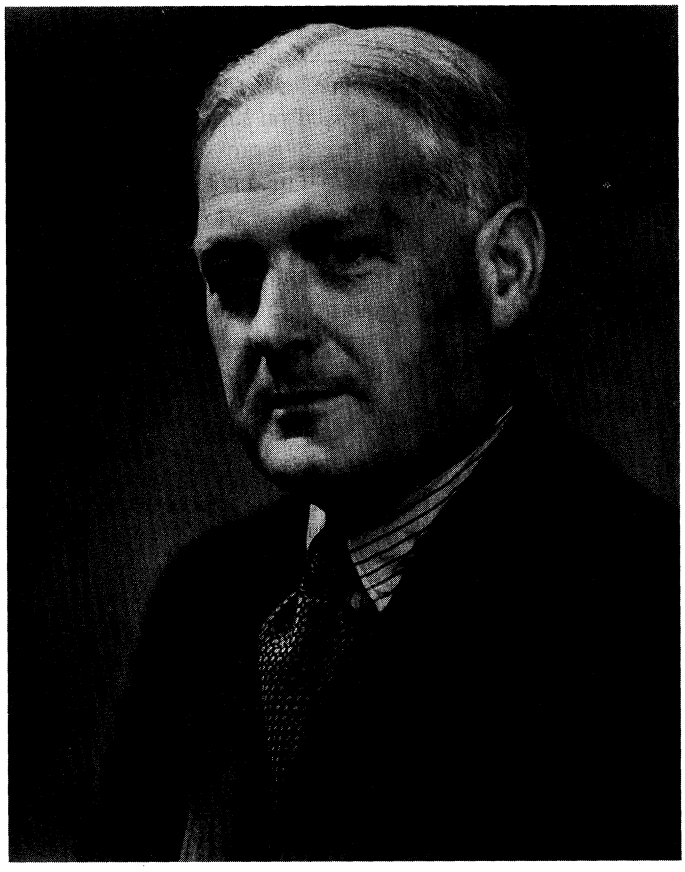


Figure 1. Austin Bradford Hill. (Reprinted from Armitage 1977 with permission from the Royal Statistical Society.)

random treatment allocation. Second, serious ethical concerns arise when an investigator controls the assignment of treatment to subjects in an experiment designed to obtain knowledge on treatment efficacy. Compared to an investigator who records and analyzes the outcomes of treatments assigned in the usual course of medical practice, an investigator who controls treatment assignments, whether through random allocation or otherwise, assumes an entirely different level of responsibility for the welfare of the study subjects. The former investigator is conducting an "observational study"; the latter is conducting an "experiment."

In 1937, the editor of the *Lancet* invited Austin Bradford Hill (Fig. 1), known to his colleagues as "Tony," to "prepare for our columns a series of short simple articles on such methods as his experience of medical statistics had shown him would be most useful in that field." These articles were republished as *Principles of Medical Statistics* in 1937 (see Hill 1971). In commenting on these articles and describing the "essentially innumerate medical profession in 1937," Richard Doll (1992) noted:

> It was Tony's genius that led him to respond to this situation not by pressing the need for deferring to a statistical consultant but by insisting that workers in medical problems, in clinical as well as preventive medicine, must themselves know something of statistical techniques . . . and learn to accept statisticians as partners in research, who on their side had to steep themselves in the realities of medical life.

For making treatment comparisons, Hill stressed the importance of comparing like with like and using concurrent rather than historical controls. In a reminiscence, Hill (1990) said:

> At the outset, I think I pleaded that trials should be made using alternate cases. I suspect if (and it's a very large IF) that, in fact, were done strictly they would be random. I deliberately left out the words 'randomization' and 'random sampling numbers' at that time, because I was trying to persuade doctors to come into trials in the very simplest form and I might have scared them off."

Through his writings and participation on advisory committees, Hill endeavored to impress upon medical investigators the need for appropriately controlled therapeutic comparisons. Meanwhile, Hill (1990) was waiting for an opportunity to introduce strict randomization. ("So I had been thinking about controlled trials for all of those 10 years and hoping the opportunity might arise.")

While Hill was waiting, a remarkable study, "Cold Vaccines: An Evaluation Based on a Controlled Trial," was published in the *Journal of the American Medical Association* (Diehl, Baker, and Cowan 1938). Students at the University of Minnesota were assigned to receive a cold vaccine or placebo by injection according to the following scheme:

> At the beginning of each year of the study, students were assigned at random and without selection to a control or to an experimental group. The students in the control groups were treated in exactly the same manner as those in the experimental groups but received placebos instead of vaccine. All students thought they were receiving vaccine and so had an unprejudiced attitude toward the study. Even the physicians who saw the students at the health service when they contracted colds during the period of study had no information as to which group they represented.

Today we would describe this as a randomized, placebo-controlled, double-blind study.

The number of colds decreased from 5.9 reported in the previous year to 1.6 in the year of study for those given vaccine, but similar decreases from (5.6 to 2.1) were noted among students treated by placebo (see Table 1). The authors attributed these decreases to subjects' inability to correctly remember the number of colds they had had in the year before the study. During the study period, the vaccinated group had .5 colds fewer per year than the control group. The authors concluded that "this difference in favor of the vaccinated group, although statistically significant, is too small to be of practical importance," a conclusion since confirmed by cold sufferers.

Taken at face value, this paper would seem to be the first published report of a randomized trial in human subjects (Lilienfeld 1982). But the paper included no description of a formal method of randomization. A typewritten manuscript by Diehl, "Adopted from the Sigma Xi Address, delivered in Northrup Auditorium, University of Minnesota, February 14, 1941," describes the method of allocation in the 1938

Table 1. Results of the Cold Vaccine Trial by Diehl et al. (1938)

	Vaccinated	*Unvaccinated*	*Difference*
Subjects who completed study	272	276	
Average number of colds in previous year	5.9	5.6	.3 ± .17
Average number of colds in study year	1.6	2.1	−.5 ± .08

*Table 2. Results from the Whooping-Cough Vaccination Trial**

	Vaccinated	*Unvaccinated*
Number allocated	4,515	4,412
Number who received three vaccinations	3,801	3,757
Cases of pertussis	149	687
Cases/1,000 child-months	1.45	6.72

* From the Whooping-Cough Immunization Committee of the Medical Research Council (1951).

paper as follows: "At the beginning of the study, students who volunteered to take these treatments were assigned alternately and without selection to control groups and to experimental groups." This is the same allocation method that Diehl had used previously in studies of cold medications (Diehl 1933). Thus it appears that strict randomization was not used in the 1938 study. In any case, this work set a high standard for clinical studies, including a concurrent placebo control group, a double-blind design, and careful attention to issues of endpoint measurement and interpretation.

Hill's first opportunity to use randomization was also in the "less emotive field of preventive medicine, e.g. a whooping cough vaccine or anticatarrhal was given to children in random order" (Hill 1990). Beginning in November 1946, children age 6–18 months were randomly allocated to receive pertussis vaccine (the "vaccinated" group) or an "anticatarrhal" vaccine that contained no *H. pertussis* (the "unvaccinated" group). Parents were informed that some children would receive "anticatarrhal" treatment instead of pertussis vaccine and that parents and investigators would not know until the end of the investigation whether a child was in the vaccinated or unvaccinated group. Only children who completed the required three vaccinations were included in the analysis. (Today, most analysts would also insist on a comparison involving all randomized children—an "intention-to-treat" analysis.) There were huge cyclical variations in whooping cough rates during the trial (Whooping-Cough Immunization Committee of the Medical Research Council 1951), but the randomized design led to convincing results nonetheless (see Table 2). Significance tests were not presented. Perhaps this is wise, because conventional statistical tests, including Fisher's randomization test, assume that responses are independent, which would not necessarily hold for an infectious disease like whooping cough. A. B. Hill was acknowledged with the comment that "the trials were planned with the advice of the Statistical Research Unit of the Medical Research Council and were arranged and supervised by Dr. W. C. Cockburn." Nowhere was Hill's name mentioned, but his randomized design had contributed enormously to the credibility of this study and had set an important precedent at the MRC.

It is one thing to test a vaccine on healthy children to prevent disease and quite another to conduct a therapeutic clinical experiment on deathly ill patients. For example, how could one withhold the new potentially beneficial antibiotic, streptomycin, from seriously ill patients with tuberculosis to obtain a convincing controlled proof of efficacy? Members of the Medical Research Council's Streptomycin in Tuberculosis Trials Committee began grappling with this issue in September 1946, and A. B. Hill, a member of that Committee, could argue the need for randomized, controlled experimentation based on his experience in designing the whooping-cough trial. In pressing for controlled experimentation, Hill had an able medical ally on the Committee, Dr. Philip D'Arcy Hart, who was an expert in the treatment of tuberculosis. Hart had just reviewed a century of studies on chemotherapy for tuberculosis (Hart 1946), and his frustration with uncontrolled experimentation is evident in an introductory paragraph describing the streptomycin trial (Streptomycin in Tuberculosis Trials Committee of the Medical Research Council 1948):

> The natural course of pulmonary tuberculosis is in fact so variable and unpredictable that evidence of improvement or cure following the use of a new drug in a few cases cannot be accepted as proof of the effect of that drug. The history of chemotherapeutic trials in tuberculosis is filled with errors due to empirical evaluation of drugs (Hart 1946); the exaggerated claims made for gold treatment, persisting over 15 years, provide a spectacular example. It had become obvious that, in the future, conclusions regarding the clinical effect of a new chemotherapeutic agent in tuberculosis could be considered valid only if based on adequately controlled trials."

According to Hill (1990), another crucial factor was the shortage of streptomycin. There was only enough streptomycin available in Britain for investigational purposes to treat about 50 patients. Hill (1990) wrote:

> This I think turned the scales: I could argue with the chairman, Sir Geoffrey Marshall (a good and sensible physician) and he would listen. I could argue that in this situation it would not be immoral to make a trial—it would be immoral not to make a trial since the opportunity would never rise again (streptomycin would be synthesized, there would be plenty of it, and so on). ... We were limited to this number, about 50 in the streptomycin arm of the trial, and I thought that was probably enough to get a reliable answer so long as it was strictly controlled and if streptomycin was really effective. And so it proved.

The trial included patients age 15–30 with acute progressive bilateral pulmonary tuberculosis, for whom the only available therapy was bed rest. Patient eligibility was determined, and then a randomly selected treatment (streptomycin and bed rest versus bed rest alone) was assigned at a central office, effectively preventing abuses of the allocation scheme. Patients were not told that they were participating in a controlled experiment. Radiologists and a clinician charged with evaluating the course of the illness over the next 6 months did not know which treatment the patients had received.

Results of the radiologic assessment at 6 months, excluding two randomized patients who died during an initial week of observation before treatment began, indicated an impressive benefit for streptomycin (see Table 3). The disproportion in deaths alone would be convincing ($p = .006$ by Fisher's exact test). This report did not include a significance test, however. The report carefully outlined the design and precautions taken to assure unbiased assessment of radiologic and clinical findings, and presented a full and meticulous description of the various effects of treatment on temperature, weight, erythrocyte sedimentation rate, radiologic changes, clinical changes, toxicity, and bacteriology. Despite the trial's impressive evidence of a benefit for

*Table 3. Assessment of Radiological Appearance at 6 Months as Compared to Appearance on Admission**

Radiological assessment	Streptomycin group		Control group	
Considerable improvement	28	51%	4	8%
Moderate or slight improvement	10	18%	13	25%
No material change	2	4%	3	6%
Moderate or slight deterioration	5	9%	12	23%
Considerable deterioration	6	11%	6	11%
Deaths	4	7%	14	27%
Total	55	100%	52	100%

* From the Streptomycin in Tuberculosis Trials Committee of the Medical Research Council (1948).

streptomycin, the investigators noted "first, that no clinical 'cures' were effected and that only 15% were bacteriologically negative (to direct examination and culture) at the end of six months; and, secondly, that this trial presents at the time of writing only a short-term evaluation (Streptomycin in Tuberculosis Trials Committee of the Medical Research Council 1948)." The report was comprehensive, descriptive, balanced, and, most important, convincing.

A. B. Hill was listed as a member of the Streptomycin in Tuberculosis Trials Committee in this paper, and the importance of concurrent randomized controls was reiterated in the Discussion: "The need of a control group in trials of a new drug for pulmonary tuberculosis is underlined by the finding that impressive clinical improvement was seen in some of the patients treated by bed rest alone (Streptomycin in Tuberculosis Trials Committee of the Medical Research Council 1948)." There seems to be agreement that this was the first randomized trial in clinical therapeutics. Armitage (1992) discussed many aspects of Hill's attitudes toward and advocacy of randomized trials.

Before leaving the topic of clinical trials, I would like to mention another statistician whose powers of persuasion contributed to the acceptance of clinical trial methodology: Jerome Cornfield (Fig. 2). Fred Ederer (1982) reviewed Cornfield's contributions to the conduct of clinical trials. He recounted Cornfield's efforts in 1957 to convince a group of radiologists to randomize individuals to receive supervoltage versus conventional radiation cancer treatment rather than to adopt a proposal of one radiologist to treat all eligible patients in one hospital with supervoltage radiation and all eligible patients in another hospital with conventional radiation. Cornfield (1958) took up the issue gingerly:

> I don't want to disagree with Dr. Fletcher, but statisticians are sometimes practical and my own answer to your question is that I would welcome, personally, any kind of information that can reasonably be gotten. If there is a reasonable possibility of collecting information on the results of therapy in different institutions under different sets of conditions, this is still information. If it is the best we can get, then let's get it and see what we can make of it. There are any number of medical problems in which controlled experimentation is just out of the question. We try to make progress in those situations despite that fact. But at the same time that we collect this kind of information, or consider an experiment in which one hospital uses one form of therapy and another hospital uses another form of therapy, we must be conscious of the fact that, although this may be the best we can do, there are enough uncontrolled variables so that we don't know what we have when we are through.

Cornfield then described an analogous study of medications for seasickness. Men on one ship would receive one kind of pill, and men on another ship would receive another type of pill. As it turned out, differences in ballast caused more turbulence on one ship than the other, so that it was "impossible to know whether the differences in the incidence of seasickness on the two ships could be attributed to the pills or to the ballast" (Cornfield 1958). Happily, the radiologists launched a series of multicenter clinical trials in which they randomized individual patients, not hospitals.

In later years, Cornfield was in constant demand as an advisor to groups wishing to begin clinical trials and as a collaborator on clinical trials. He chaired the Food and Drug Administration's Committee on Biometry and Epidemiology, which developed guidelines for the design and conduct of clinical trials. Ederer (1982) also alluded to Cornfield's concerns about the shortcomings of available statistical methodology for clinical trials. In particular, Cornfield disliked the rigidity of formal hypothesis testing and the limitations of this approach in monitoring interim results from clinical trials, in examining treatment effects in subgroups of patients, and in following leads from the data with further unplanned analyses. Reluctant to be constrained by rules designed to preserve the significance levels, Cornfield commented (Cutler, Greenhouse, Cornfield, and Schneiderman 1966) that "if maintenance of the significance level interferes with the release of interim results, all I can say is so much the worse for the significance level." This attitude partly reflected Cornfield's deep criticisms of significance tests and p values as measures of evidence and gave rise to his pathbreaking Bayesian efforts to confront these issues (Cornfield 1966a, 1966b, 1969, 1976).

Figure 2. Jerome Cornfield.

3. DOES SMOKING CAUSE LUNG CANCER?

3.1 The Problem

Today, the fact that smoking causes lung cancer and many other cancers is widely accepted, and cancer studies that fail to control for smoking status are open to criticism. In 1964, Surgeon General Luther Terry issued a report on "Smoking and Health" (U.S. Department of Health, Education, and Welfare, Public Health Service 1964), and the U.S. Public Health Service has promoted smoking cessation efforts ever since. As a member of the Public Health Service, I shall proudly refer to this report as the "Surgeon General's Report." Since the issuance of that report, smoking prevalence has decreased among males age 45–64 from 52% in 1965 to 29% in 1992 (National Center for Health Statistics 1995, p. 151). The age-adjusted male lung cancer incidence rate peaked between 1981 and 1987 and has been decreasing slowly since then (Ries et al. 1994). Women began smoking in large numbers after men did, and the prevalence of smoking among women age 45–64 has declined more slowly, from 32% in 1965 to 26% in 1992. Among women, age-adjusted lung cancer incidence rates continued to increase through 1991. Lung cancer has long been the most common cause of death from cancer among men and surpassed the previous leader, breast cancer, among women in 1987.

These facts are so well known and the causative role of smoking so widely accepted that it is hard to imagine the debates provoked by early observational studies demonstrating an association between lung cancer and cigarette smoking. These debates touched on such fundamental issues as:

- Can one infer a causal relationship from observational studies in which one does not control exposure assignment with a randomized experimental design?
- What biases afflict such studies?
- How can one interpret the findings of a case-control study in terms of a prospective measure of risk?

Many epidemiologists and other scientists contributed to the formulation and resolution of these issues, but none more decisively than Cornfield and Hill, with whom Richard Doll worked. It can be argued that the resolution of these issues laid the foundation for much of current epidemiologic practice. But the motivating problem was lung cancer.

Lung cancer was seldom reported before the 20th Century and was uncommon in the 1930s. But since 1930, age-adjusted U.S. lung cancer mortality rates (per 10^5 person years) had grown from 4 in 1930 to 35 in 1960 in men and from 2 to 5 in women. Cornfield et al. (1959) called the rising death rate from lung cancer "the most striking neoplastic phenomenon of this century."

3.2 The Association Between Smoking and Lung Cancer and Advances in Case-Control Methodology

In 1928, Lombard and Doering (1928) found an association between heavy smoking (mainly pipe smokers) and all types of cancer combined, and numerous earlier uncontrolled case series of patients with lung cancer had suggested an association with cigarette smoking (see references in Wynder and Graham 1950). Müller (1939) found that relatives of deceased lung cancer patients reported more tobacco smoking among these cases than was reported by healthy men of the same age. This study was the first of 30 case-control studies conducted from 1939 to 1963 and cited in the Surgeon General's Report.

In the United States, one of the most influential early studies (Wynder and Graham 1950) compared questionnaire responses of 605 male patients with histologically proven lung cancer to those of 780 hospitalized male patients with other diseases. Wynder and Graham found that 86.4% of cases were heavy smokers compared to 54.7% of controls. Similar results were obtained from a special substudy in which the interviewers had no knowledge of whether or not the patient being interviewed had cancer. The point of this substudy was to try to reduce the possibility of "recall bias" that might result when an interviewer who knows the diagnosis consciously or unconsciously encourages different responses from cases than from controls. Levin, Goldstein, and Gerhardt (1950) found that 66% of 236 lung cancer patients smoked cigarettes, compared to 44% of other patients at Roswell Park Memorial Institute in Buffalo, and the protection against "recall bias" in this study was that the smoking data were recorded as part of the routine medical history before a diagnosis was established.

Four months later, Doll and Hill (1950) published a study comparing 649 male and 60 female patients with lung cancer to equal numbers of age- and sex-matched controls from the same hospital. Patients and controls were recruited from 20 hospitals in London between April 1948 and October 1949. Doll and Hill demonstrated statistically significant differences in the proportions of nonsmokers between cases and controls and in the distributions of amount smoked among smoking patients with lung cancer and smoking control patients (see Table 4). A strong trend of increasing risk

Table 4. Most Recent Amount Smoked (Cigarettes per Day) in Men[a]

	Daily cigarette consumption						
	0	*1–4*	*5–14*	*15–24*	*25–49*	*50+*	*Total*
Lung cancer cases	2	33	250	196	136	32	649
Hospital controls	27	55	293	190	71	13	649
Odds ratio[b]	1.0	8.1	11.5	13.9	25.9	33.2	

[a] From Doll and Hill (1950). Pipe smokers are assigned an equivalent of four cigarettes per day for each ounce of tobacco smoked per week.

[b] Number of cases divided by number of controls for a given level of smoking divided by the same ratio (2/27) for nonsmokers.

with increasing dose is evident in these data (see the odds ratios in Table 4).

The paper by Doll and Hill (1950) was exemplary in pointing out the key issues for evaluating a case-control study. These authors argued that the lung cancer cases and controls were representative of all cases and controls seen at the 20 London hospitals. But this condition is not sufficient to guarantee that an association between disease status and smoking in this hospital population also exists in the general population. That would be true if cases and controls were representative of cases and noncases in the general population or if the selective pressures that determine who goes to a hospital do not depend on both disease and exposure status. (See Schlesselman 1982, pp. 128–131, for a precise statement of these conditions.) In trying to eliminate selection bias as an explanation of the association, Doll and Hill (1950) argued that hospitalized control patients had no lower prevalence of smoking than the general population. Indeed, in a second report, Doll and Hill (1952) showed that hospitalized control patients smoked somewhat more than a representative sample from the general population of Greater London, so that if anything, estimates of risk based on hospitalized controls would be too low.

Doll and Hill (1950) controlled for factors ("confounders") that were associated with both smoking and lung cancer risk and that could therefore potentially account for the association between smoking and lung cancer. They matched on gender and age. The only other potential confounder they discovered was that more lung cancer cases than controls tended to live in rural areas. This confounder would not explain the differences in smoking rates between cases and controls, however, because smoking was more prevalent in London than in rural areas.

Doll and Hill (1950) dealt with the issue of recall bias that might result from lung cancer patients' exaggerating the extent of their tobacco exposure or from the interviewers' knowing the patients' diagnoses. In fact, the subset of control patients who were incorrectly thought to have had lung cancer when originally admitted to the hospital and interviewed but who were later found to have other diseases also smoked much less than lung cancer patients.

One of the main difficulties in interpreting case-control studies is how to translate information on the exposure distributions among cases and controls into information on the probability of disease given exposure status. Doll (1993) later described Hill as "a master of the method by which arithmetic is made argumentative." (This had also been said of William Farr.) Some of this mastery is evident in Table 5, which contains the ratios of the number of lung cancer cases interviewed in the case-control study to the estimated size of the corresponding population in Greater London for various categories of age and smoking status. Doll and Hill (1950) coupled census data on the age distribution of the population in Greater London with data on the age and smoking status distribution among their control patients to estimate the sizes of the populations at risk for Greater London, writing that "if it can be assumed that the patients without carcinoma of the lung who lived in Greater London at the time of their interview are typical of the inhabitants of Greater London with regard to their smoking habits, then the number of people in London smoking different amounts of tobacco can be estimated." If, further, the case-control study had included all cases arising in a year in Greater London, then the ratios in Table 5 would be the actual rates of lung cancer per person-year—namely, the lung cancer risk. Doll and Hill (1950) knew that the 20 hospitals in their survey did not cover all lung cancer cases and that many cases registered at these hospitals were not interviewed. These facts explain their comment that "it must be stressed that the ratios shown in this table are not measures of the actual risks of developing carcinoma of the lung, but are put forward very tentatively as proportional to these risks." This proposition would indeed be true if the age-specific smoking distributions in controls were the same as in the population of Greater London, if the probability of inclusion as an interviewed case in the case-control study were independent of age and smoking status, and if all cases arose from the Greater London population. Provided that the elements in Table 5 are proportional to actual risks, ratios of elements correspond to relative risks of disease. Thus for people age 65–74, the risk of disease for a smoker of 25–49 cigarettes per day is $1{,}063/21 = 50.6$ times the risk of a nonsmoker, and we say the relative risk is 50.6. Combining data from age 45–74, Doll and Hill (1950) found the "relative risks become 6, 9, 26, 49 and 65 when the average number of cigarettes smoked a day are 3, 10, 20, 35 and, say, 60—that is, the midpoints of each smoking group." Thus by the magic of arithmetic and clear thinking, Doll and Hill had converted retrospective case-control data on smoking exposure into estimates of prospective relative risks.

Another remarkable aspect of Doll and Hill's (1950) paper was its effort to establish a causal interpretation of the association between lung cancer and smoking. Doll and Hill raised and discounted the possibility that lung cancer caused smoking, because "the habit of smoking was, however, invariably formed before the onset of the disease." They argued that smoking was specific for lung cancer and not associated with other respiratory diseases and cancer of other sites. (This argument from specificity was later used by Berkson [1955, 1958] and others to attack the causal hypothesis when it became clear that in fact smoking was associated with a great variety of diseases, including nonmalignant respiratory diseases and cardiovascular disease.)

Table 5. Ratios of Lung Cancer Patients Interviewed by Doll and Hill (1950) to the Corresponding Estimated Population Size in Greater London[a]

	Daily cigarette consumption					
Age	*0*	*1–4*	*5–14*	*15–24*	*25–49*	*50+*
25–34	0[b]	11[b]	2[b]	6[b]	28[b]	
35–44	2[b]	9[b]	43	41	67	77
45–54	12[b]	34	178	241	429	667
55–64	14	133	380	463	844	600
65–74	21	110	300	510	1,063	2,000

[a]Ratios are expressed as lung cancer cases per million population.
[b]Based on fewer than five cases.

Doll and Hill also pointed to the strong dose–response relationship between smoking and lung cancer risk as evidence of a causal relationship.

Cornfield was inspired by the work of Wynder and Graham (1950) not only to give up smoking (according to his wife, Ruth Bittler Cornfield, they each gave up a 2.5-pack-a-day habit), but also to think about how to interpret case-control data in terms of prospective risk of disease. Using Bayes's theorem and smoking as his motivating example, Cornfield (1951) showed that the probability of disease (D) given exposure (E) is

$$P(D|E) = P(D)P(E|D)/\{P(D)P(E|D) + P(\bar{D})P(E|\bar{D})\}, \tag{1}$$

where $\bar{D}$ denotes the absence of disease. Case-control studies allow one to estimate the conditional exposure probabilities $P(E|D)$ and $P(E|\bar{D})$. Thus if one also knows the rate of disease in the population, $P(D)$, then one can use case-control data to calculate the risk of disease.

Often one has no information on the overall disease rate in the population, $P(D)$. Even so, provided that the rate of disease is small, Cornfield showed that the relative risk $P(D|E)/P(D|\bar{E})$ is well approximated by the exposure odds ratio $\{P(E|D)/P(\bar{E}|D)\}/\{P(E|\bar{D})/P(\bar{E}|\bar{D})\}$, which is estimable from case-control data. Here $\bar{E}$ denotes absence of exposure.

Thus Cornfield had rigorously and independently proved that relative risks could be estimated from case-control data, as had been illustrated numerically by Doll and Hill (1950). Cornfield had also shown [see Eq. (1)] that exposure-specific absolute risks could be calculated from case-control data, provided that the overall rate of disease in the population was known. He illustrated such calculations for lung and cervical cancer. The next year, Doll and Hill (1952) published estimates of the rates of lung cancer according to daily cigarette consumption. These estimates were obtained by combining case-control data on smoking history with data on the overall lung cancer death rate in Greater London obtained from the Registrar-General. Thus the work of Cornfield and of Doll and Hill showed that case-control studies could be connected to the underlying base population (e.g., "Greater London") and could yield either relative risks alone or relative risks and exposure-specific risks, depending on whether or not the overall disease rate in the base population was known.

3.3 From Association to Causation

Despite this firm theoretical basis for case-control sampling, such studies were not regarded as entirely convincing because of difficulties remembering exposure history, because of potential recall bias, and because sampled cases and controls might not be representative of cases and noncases in the base population. For this reason, Doll and Hill (1956) sent a questionnaire to 59,600 members of the medical profession in the United Kingdom on October 31, 1951; they received 40,701 usable forms. These 40,701 respondents constituted a cohort that was followed prospectively until March 31, 1956, to determine lung cancer risk. The cohort study confirmed earlier case-control findings (Doll and Hill 1950, 1952) in several respects:

1. There was a striking confirmation of trends in relative risk with increasing tobacco consumption. The relative risks for men age 35–64 years for those smoking 1–14, 15–24, and 25 or more cigarettes daily (or its tobacco equivalent for pipe smokers) in the cohort study were 7, 12, and 24. Corresponding relative risks from men age 45–64 in the case-control study (Doll and Hill 1952) were 9, 12, and 22. (I have grouped finer consumption categories in the case-control data of Doll and Hill 1952 for this comparison by weighting risks in Table XII by proportions among controls in Table V.)
2. Cigarette smokers are at higher risk than pipe smokers.
3. Relative risk decreases with increasing time since giving up smoking.

An important feature of the paper by Doll and Hill (1956) is the discussion of possible biases and alternative explanations for the association between smoking and lung cancer. One possibility was that cancer was detected with greater probability among smokers than among nonsmokers; however, the same gradient of risk with smoking was seen for those with firm histologic evidence of cancer as for those diagnosed without histologic confirmation.

A second possibility, raised by Berkson (1955), was that healthy, nonsmoking members of the population joined the cohort study more readily than healthy smokers, introducing a differential bias in the composition of the cohort. But if this were true, then one would expect the greatest relative risks from smoking to occur at the beginning of the study, with smaller relative risks later as the initial cohort selection effects are attenuated. Doll and Hill (1956) pointed out that the relative risks remained constant.

The potential weaknesses of the cohort design are quite different from potential weaknesses of the case-control design, such as recall bias. The fact that very similar relative risks were obtained from these two different study designs indicates that the association found between smoking and lung cancer probably did not result from artifacts of measurement or selection.

The possibility of confounding by atmospheric pollution was dealt with by noting that gradients with smoking were seen in both urban and rural settings.

Berkson (1955, 1958) was suspicious of the observation that smoking was associated not only with increased risk of lung cancer, but also with cardiovascular disease and certain other causes of death. He thought it implausible that smoking tobacco could have such far flung consequences. Although the principle of "specificity" of effect has been mentioned as one indication of a causal relationship, it apparently does not apply to cigarettes. Cessation of smoking leads not only to reductions in the relative risk for lung cancer but also to rapid reductions in cardiovascular risk.

In June 1957, The Study Group on Smoking and Health, sponsored by the American Cancer Society, the American

Heart Association, the National Cancer Institute, and the National Heart Institute, summarized data from 14 case-control studies, 2 cohort studies (Doll and Hill 1954 and Hammond and Horn 1954), and laboratory studies and concluded:

> The sum total of scientific evidence establishes beyond reasonable doubt that cigarette smoking is a causative factor in the rapidly increasing incidence of human epidermoid carcinoma of the lung (Study Group on Smoking and Health 1957).

Later in June 1957, the Medical Research Council in Britain published a statement that concluded:

> In scientific work, as in the practical affairs of every day life, conclusions have often to be founded on the most reasonable and probable explanation of the observed facts, and so far no adequate explanation for the large increase in the incidence of lung cancer has been advanced save that cigarette smoking is indeed the principal factor in the causation of the disease. ... It is clearly impossible to add to the evidence by means of an experiment in man. The Council are, however, supporting a substantial amount of laboratory research which may throw more light on the mechanism by which tobacco smoke and other suspected causative factors exert their effect (Medical Research Council 1957).

There remained skeptics, such as R. A. Fisher, who stressed that associations that arise in observational studies need not reflect a causal relationship. Fisher (1957, 1958a) mentioned two alternatives: that lung cancer caused smoking, and that a third "constitutional" factor, possibly genetic, caused individuals both to become smokers and to develop lung cancer. Regarding the first possibility, Cornfield et al. (1959) wrote that "since we know of no evidence to support the view that the bronchogenic carcinoma diagnosed after age 50 began before age 18, the median age at which smokers begin smoking, we shall not discuss it further."

Then Cornfield and his coauthors, in a masterful review of the evidence and objections to the causal hypothesis, considered the alternative "constitutional" hypothesis. They noted that "nothing short of a series of independently conducted, controlled, experiments on human subjects, continued for 30 to 60 years, could provide a clear-cut and unequivocal choice between" the constitutional hypothesis and the hypothesis that smoking caused lung cancer (Cornfield et al. 1959). They adduced a series of facts that made the constitutional hypothesis less and less tenable:

1. Lung cancer mortality has increased continuously in the last 50 years and more so for men than women. The simple constitutional hypothesis would require some modifications, such as the hypothesis that a new environmental carcinogen is acting on the genetically susceptible subpopulation or that a new mutation is spreading rapidly in the population, to explain rapidly rising rates.
2. Tobacco smoke contains substances that cause cancer when applied to the skin of mice and rats. The constitutional hypothesis would require that such substances not be carcinogenic for human lungs.
3. Cigarettes cause mainly lung cancer, and pipes and cigars cause mainly mouth and throat cancers. There would need to be two constitutional makeups, one for cigarette smokers and one for pipe or cigar smokers.
4. Lung cancer mortality is lower in people who stop smoking than in those who continue to smoke. Thus the constitutional factor must decrease with age, allowing one to stop smoking and at the same time to experience decreasing risk from lung cancer unrelated to smoking.
5. Another counterargument (Cornfield et al. 1959) was based on the following theorem, proved by Cornfield:

> If an agent A, with no causal effect upon the risk of disease, nevertheless, because of a positive correlation with some other causal agent, B, shows an apparent risk, r, for those exposed to A, relative to those not so exposed, then the prevalence of B among those exposed to A, relative to the prevalence among those not so exposed, must be greater than r.

In other words, for a genetic "constitutional" confounder to explain the relative risk of lung cancer of 20 or more among heavy smokers, that constitutional factor would need to be at least 20 times more prevalent in heavy smokers than nonsmokers. No specific factors satisfying this condition had been identified or proposed.

Cornfield et al. (1959) concluded:

> No one of these considerations is perhaps sufficient by itself to counter the constitutional hypothesis *ad hoc* modification of which can accommodate each additional piece of evidence. A point is reached, however, when a continuously modified hypothesis becomes difficult to entertain seriously.

Cornfield and his coauthors argued against other objections to the causal hypothesis. For example, Fisher (1957) had pointed to the more rapid increases of lung cancer in men than women as inconsistent with "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." Cornfield et al. (1959) cited data on smoking in men and women that proved just the reverse and said that this preponderance of smoking in men constituted "in fact, one of the links in the chain of evidence implicating cigarettes." Cornfield and his coauthors argued methodically, fairly, and at times with irony, to meet the various objections of Fisher (1957, 1958a, 1958b, 1959), Berkson (1955, 1958), Neyman (1955), and many others.

The Surgeon General's Report appeared 5 years later, in 1964, and marked the beginning of a major campaign to promote smoking cessation in the United States. There had been an earlier pronouncement in the scientific literature by the U.S. Public Health Service that smoking was the "principal etiologic factor in the increased incidence of lung cancer" (Burney 1959), but the 387-page Surgeon General's Report was comprehensive, authoritative, and well publicized. It included topics ranging from the chemistry and pharmacology of tobacco products to psychosocial aspects of smoking. The Surgeon General's Advisory Committee on Smoking and Health consisted of 10 eminent members, including William G. Cochran, former President of ASA. This Advisory Committee obtained the "constant support" of nearly 200 individuals, groups, and institutions, among whom were Berkson, Cornfield, Doll, Dorn, Goldstein, Haenszel, Hammond, Horn, and Levin, whose work I have mentioned, and several other statisticians in the Public Health Service and at universities. The Surgeon General's Report emphasized the findings of 7 prospective cohort studies, but also considered the results of 30 case-control

studies, carcinogenicity experiments in laboratory animals, and the descriptive statistics of time trends in lung cancer incidence, in gender differences, and in patterns of tobacco consumption in reaching its conclusion that:

> Cigarette smoking is causally related to lung cancer in men; the magnitude of the effect of cigarette smoking far outweighs all other factors. The data for women, though less extensive, point in the same direction.

In reaching this conclusion, the Surgeon General's Report recognized that

> Statistical methods cannot establish proof of a causal relationship in an association. The causal significance of an association is a matter of judgement which goes beyond any statement of statistical probability. To judge or evaluate the causal significance of the association between the attribute or agent and the disease, or effect upon health, a number of criteria must be utilized, no one of which is an all-sufficient basis for judgement.

Among these criteria were the *consistency* of the association in study after study, whether prospective or retrospective, in various populations; the *strength* of the association, with relative risks above 10; the *temporal pattern* with exposure preceding disease; and the *coherence* of the causal hypothesis with a large and complicated body of evidence. These criteria form part of a mantra of causal criteria used by today's epidemiologists and enunciated by Hill (1965). This list also includes biologic gradient, typified by increasing lung cancer risk with increased smoking exposure; specificity of the effect of the agent in causing only one or a limited number of diseases; biologic plausibility; experiment (for example, in a setting where disease might be prevented); and analogy with the effects of another similar agent.

But A. B. Hill would not be satisfied with a checklist. His aim (Hill 1953) was no less than "the permeation of statistical research with the experimental spirit." If one cannot manipulate exposures directly, then one can at least "go seek more facts, paying less attention to the techniques of handling data and far more to the development and perfection of methods of obtaining them. In doing so one must have the experimental approach firmly in mind." Hill referred specifically to his work with Doll on lung cancer and wrote that "our aim was to make the field observations mirror an experimental design as nearly as possible." He illustrated these ideas with John Snow's triumphant work that led to the control of cholera even before the bacterium had been discovered; "Snow approached his problem not only as an incomparable master of logical deduction from observations but also, it should be noted, as the constructor of observations." For Snow, with one assistant, "tramped in the summer sun, learning for every cholera death the water supply of the household" and by that means demonstrated a death rate of 281 per 10,000 for houses served by the Southwark and Vauxhall Company, compared to only 5 per 10,000 for houses served by the Lambeth Company. And then he explained apparent anomalies by seeking further facts. For example, a 59-year-old widow living in Hampstead, where there had been no cholera, died of this disease on September 2, 1854. Snow determined that she had not visited Broad Street but that she had drunk from a bottle of water from the Broad Street pump on August 31, 1854. At Snow's recommendation, the handle of the Broad Street pump was removed, and the cholera epidemic, which had already begun to abate, ceased. This work

> shows—as many other examples have shown—that the highest returns can be reaped by imagination in combination with a logical and critical mind, a spice of ingenuity coupled with an eye for the simple and humdrum, and a width of vision in the pursuit of facts that is allied with an attention to detail that is almost nauseating (Hill 1953).

It is the complex skein of coherence, in the face of many "facts," that makes a strong case for a causal hypothesis from observational data. Of course, experimental scientists use much the same approach to test all the implications of a theory by gathering new and various facts. This is perhaps what Fisher had in mind in about 1945, when, according to Cochran (1965), he was asked what can be done in observational studies to clarify the step from association to causation, and he replied, "Make your theories elaborate."

Part of this process requires considering alternative explanations to the causal hypothesis and judging their plausibility in view of the "facts," as Doll and Hill (1952) had done in their discussion on "validity of the results." Cornfield (1954) insisted that one has an obligation to seek out and evaluate alternative explanations even when an association has been found in an experiment:

> To distinguish between statistical association on the one hand and relationships established by experimentation on the other, without any reference to alternative variables that are present in one case but not the other, seems to us to be neither good statistics, good science, nor good philosophy—though it may be good red herring.

Thus even well-conducted randomized controlled experiments require an appraisal of alternative explanations before drawing a causal inference.

4. IMPLICATIONS

A reader who has come this far might ask: "What does this have to do with me or with the American Statistical Association?" For my part, it was reward enough to review these masterful papers and their resolutions of problems that influence our daily practice. I recommend these articles to you, in the words of an admirer (Jean-Francois de La Harpe), of the fableist La Fontaine: "One should not praise La Fontaine, one should read him, reread him, and reread him again."

And yet there are some broader lessons. We can admire and hope to emulate in some degree the personal qualities of people like Hill and Cornfield. Both Hill and Cornfield were humorous and were wonderful speakers and writers. These skills made them persuasive in committee and as consultants and collaborators. But perhaps an even greater factor was their sympathy toward collaborators. Hill (1953) quoted Greenwood (1924) in describing his approach to collaboration and consultation:

> I used to see in the statistician the critic of the laboratory worker; it is a role which is gratifying to youthful vanity, for it is easy to cheat oneself into the belief that the critic has some intellectual superiority over the criticized. I do not think even now that statistical criticism of laboratory investigations is useless, but I attach enormously more value to direct collaboration, the making of statistical experiments, and the permeation of statistical research with the experimental spirit.

It was this sense of the value of "direct collaboration" that allowed Hill and Cornfield to gain the confidence and affection of colleagues in the fields of medicine, laboratory science and public health (Armitage 1977). Hill was an Honorary Fellow of the Royal College of Physicians and the Royal Society of Medicine (Armitage 1991); Cornfield was President of the American Epidemiological Society and Vice-Chairman of the Council of Epidemiology and Prevention of the American Heart Association (Schneiderman and Greenhouse 1980).

Cornfield was attracted to statistics as a field where diversity of interests could be advantageous (Cornfield 1975). Certainly Cornfield's interests were broad; he majored in history as an undergraduate and almost drifted into "the mental indolence of history" before his "random walk" ended in statistics. Cornfield made notable contributions to statistical theory, yet he had no advanced degree. His writings range from confidence intervals for the odds ratio to the philosophy and limitations of statistical methods (Cornfield 1976). Hill had planned to study medicine, but tuberculosis intervened, and after a 3-year convalescence, he studied economics in bed, going to London University only twice to take examinations (Doll 1992). Although he took courses in statistics at University College London, he "was more inspired by Karl Pearson's ideas, philosophy and enthusiasm than by the mathematical niceties of his statistical methods." In fact, "Bradford Hill never thought of himself as a statistician, but rather as an 'arithmetician' " (Doll 1993). He had, however, "an almost obsessive fascination for, and grasp of, numerical information, and it is difficult to recall any instance in which his rather intuitive approach failed to reveal any important aspects of the data" (Armitage 1991). Like Cornfield, Hill made great contributions to medical statistics without specialist qualifications in medicine, statistics, or laboratory science.

Perhaps this lack of highly specialized training, coupled with natural intelligence and broad interests, helped Hill and Cornfield to identify important problems, to approach them from first principles with an uncluttered mind, and to understand people and practicalities. Hill and Cornfield were also favored by chance and by the responsibilities of their positions at the Medical Research Council and the National Institutes of Health in being exposed to a range of exciting and potentially important problems. In any case, it was a willingness to take on important problems and, if necessary, to debate, that led to important practical results, to methodological advances, to a clarification of ideas for causal inference, and to high visibility and prestige for statistical thinking.

We can learn much from these examples. For despite important advances in statistical methodology over the past 25 years, the work of Doll and Hill and Cornfield reminds us that some of the most important elements of applied statistics do not require advanced statistical calculation. Seeking out important problems, working with colleagues in other fields to define critical issues and objectives, understanding the nuances of the consultees' problems before attempting a quantitative description, expressing objectives in measurable terms, developing an organized plan ("permeated with the experimental spirit") to gather needed data, paying special attention to possible sources of systematic error (such as "recall bias"), interpreting results in light of various alternative explanations, performing follow-up experiments to clarify special issues, communicating clearly with colleagues about the meaning of the data for their problem—these are critical elements we sometimes fail to emphasize. Often the real challenges concern bias and systematic error rather than random variation, and a preoccupation with p values is misplaced. This was certainly Hill's (1965) view when he wrote, "What is worse, the glitter of the t table diverts attention from the inadequacies of the fare."

But the most important lesson we can learn from Doll and Hill and Cornfield is to involve ourselves actively in the solution of real problems. In 1834, the Royal Statistical Society adopted as its motto "Aliis exterendum," which means "Let the others thrash it out." I suppose the idea was that statisticians should be dispassionate gatherers and distributors of facts, but that the interpretation and ultimate use of these facts was for "others" to decide (Hill 1984). This is surely not what Louis, Doll, Hill, and Cornfield have taught us, nor is it a motto that we can live by today. Our challenge is to seek important new problems or help our consultees deal effectively with old problems, by providing not only technical expertise but also context, perspective, problem definition, decisive observational or experimental plans for gathering facts, and informed interpretation and opinion, based on knowledge of the substantive issues and experimental constraints. This activity may evoke Hill's frenetic image in a discussion of Cochran's (1965) great paper on the planning of observational studies of human populations:

> I suspect that in our approach to observational studies of the human population, there is only one material difference between Professor Cochran and myself. He, as he points out in an early paragraph of his paper, has (in this situation) largely served as a referee, or, at the very least, as a linesman. Over the past forty years I have had to rush feverishly around the field of play, and, in this particular field, unfortunately, most of the missiles are aimed at the players; indeed it is not unknown for the referee to join in.

Of course we, as professional statisticians, face different times and different problems. In many ways, we have more to offer than our predecessors. Our statistical tool kit is fuller than those carried by practitioners even 25 years ago. We require more formal training to avoid injuring ourselves and others with these tools. Theoretical and applied methodological studies and advances in computing have put at our disposal a wealth of new statistical models, nonparametric methods for data exploration, graphical aids, resampling and advanced asymptotic methods for inference, approaches for handling missing or censored data, and other technical marvels. We also have access to a profusion of specialized journals and books, to relative ease of travel, and to electronic communications. Now, more than ever, we are in a position to learn from each other—from methodologists and from those interested in applications analogous to our own—and to bring the knowledge and experience of a diverse and talented association to bear on our problems.

Now, through electronic communication, the possibilities for collaboration and mutual support are nearly limitless—provided that one has the ASA Directory and a good sense of others' interests.

This, then, is an opportune time—and it is our time—to seek out new collaborations, to identify good problems, and to steep ourselves in their intricacies. It is our time to think broadly about the impediments to progress: Is it a technical issue? A question requiring persuasion? It is our time to share experiences and specialized technical skills by discussing real problems and giving each other access to ideas that have grown and evolved in the partial isolation of our separate fields. It is our time to return to colleagues in other fields with sympathy, with enthusiasm, with a deeper understanding, and with a broader view, and hope to hear them say: "That statistician really understands our problem and will help us solve it." That's statistics in action.

[Received September 1995. Revised September 1995.]

REFERENCES

Armitage, P. (1977), "A Tribute to Sir Austin Bradford Hill," *Journal of the Royal Statistical Society*, Ser. A, 140, 127–128.

——— (1983), "Trials and Errors: The Emergence of Clinical Statistics," *Journal of the Royal Statistical Society*, Ser. A, 146, 321–334.

——— (1991), "Obituary: Sir Austin Bradford Hill, 1897–1991," *Journal of the Royal Statistical Society*, Ser. A, 154, 482–485.

——— (1992), "Bradford Hill and the Randomized Controlled Trial," *Pharmaceutical Medicine*, 6, 23–37.

Berkson, J. (1955), "The Statistical Study of Association Between Smoking and Lung Cancer," *Proceedings of the Staff Meeting of the Mayo Clinic*, 30, 319–348.

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

Burney, L. E. (1959), "Smoking and Lung Cancer. A Statement of the Public Health Service," *Journal of the American Medical Association*, 171, 135–143.

Cochran, W. G. (1965), "The Planning of Observational Studies of Human Populations," *Journal of the Royal Statistical Society*, Ser. A, 128, 234–265.

Cornfield, J. (1951), "A Method of Estimating Comparative Rates From Clinical Data. Applications to Cancer of the Lung, Breast and Cervix," *Journal of the National Cancer Institute*, 11, 1269–1275.

——— (1954), "Statistical Relationships and Proof in Medicine," *The American Statistician*, 8, 19–21.

——— (1958), "Discussion," in *Research in Radiology*, ed. H. S. Kaplan, Nuclear Science Series, Report Number 22, Publication 571, Washington, DC: National Academy of Sciences—National Research Council, pp. 152–153.

——— (1966a), "Sequential Trials, Sequential Analysis and the Likelihood Principle," *The American Statistician*, 20, 18–23.

——— (1966b), "A Bayesian Test of Some Classical Hypotheses, With Applications to Sequential Clinical Trials," *Journal of the American Statistical Association*, 61, 577–594.

——— (1969), "The Bayesian Outlook and Its Applications," *Biometrics*, 24, 617–657.

——— (1975), "A Statisticians's Apology," *Journal of the American Statistical Association*, 70, 7–14.

——— (1976), "Recent Methodological Contributions to Clinical Trials," *American Journal of Epidemiology*, 104, 408–421.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959), "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions," *Journal of the National Cancer Institute*, 22, 173–203.

Cutler, S. J., Greenhouse, S. W., Cornfield, J., and Schneiderman, M. A. (1966), "The Role of Hypothesis Testing in Clinical Trials. Biometrics Seminar," *Journal of Chronic Diseases*, 19, 857–882.

Diehl, H. S. (1933), "Medicinal Treatment of the Common Cold," *Journal of the American Medical Association*, 101, 2042–2049.

Diehl, H. S., Baker, A. B., and Cowan, D. W. (1938), "Cold Vaccines: An Evaluation Based on a Controlled Study," *Journal of the American Medical Association*, 111, 1168–1173.

Doll, R. (1992), "Sir Austin Bradford Hill and the Progress of Medical Statistics," *British Medical Journal*, 305, 1521–1526.

——— (1993), "Sir Austin Bradford Hill, 1897–1991," *Statistics in Medicine*, 12, 795–808.

Doll, R., and Hill, A. B. (1950), "Smoking and Carcinoma of the Lung: Preliminary Report," *British Medical Journal*, 2, 739–748.

——— (1952), "A Study of the Aetiology of Carcinoma of the Lung," *British Medical Journal*, 2, 1271–1286.

——— (1954), "The Mortality of Doctors in Relation to Their Smoking Habits. A Preliminary Report," *British Medical Journal*, 1, 1451–1455.

——— (1956), "Lung Cancer and Other Causes of Death in Relation to Smoking. A Second Report on the Mortality of British Doctors," *British Medical Journal*, 2, 1071–1081.

Ederer, F. (1982), "Jerome Cornfield's Contributions to the Conduct of Clinical Trials," *Biometrics*, 38 (Supplement), 25–32.

Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

——— (1926), "The Arrangements of Field Experiments," *Journal of Ministry of Agriculture of Great Britain*, 33, 503–513.

——— (1935), *The Design of Experiments*, Edinburgh: Oliver and Boyd.

——— (1957), "Dangers of Cigarette Smoking," *British Medical Journal*, 2, 297–298.

——— (1958a), "Lung Cancer and Cigarettes?," *Nature*, 182, 108.

——— (1958b), "Cigarettes, Cancer and Statistics," *Centennial Review of Arts and Sciences*, 2, 151–166.

——— (1959), *Smoking, the Cancer Controversy. Some Attempts to Assess the Evidence*, London: Oliver and Boyd.

Fisher, R. A., and MacKenzie, W. A. (1923), "Studies in Crop Variation II. The Manurial Response of Different Potato Varieties," *Journal of Agricultural Science*, 13, 311–320.

Greenwood, M. (1924), "Is the Statistical Method of Any Value in Medical Research?," *Lancet*, 2, 153–158.

Hammond, E. C., and Horn, D. (1954), "The Relationship Between Human Smoking Habits and Death Rates: A Follow-up Study of 187,766 Men," *Journal of the American Medical Association*, 155, 1316–1328.

Hart, P. D. (1946), "Chemotherapy of Tuberculosis: Research During the Past 100 Years," *British Medical Journal*, 2, 805–810.

Hill, A. B. (1953), "Observation and Experiment," *New England Journal of Medicine*, 248, 995–1001.

——— (1965), "The Environment and Disease: Association or Causation," *Proceedings of the Royal Society of Medicine*, 58, 295–300.

——— (1971), *Principles of Medical Statistics*, 9th ed., New York: Oxford University Press.

——— (1990), "Memories of the British Streptomycin Trial in Tuberculosis: The First Randomized Clinical Trial," *Controlled Clinical Trials*, 11, 77–79.

Hill, I. D. (1984), "Statistical Society of London—Royal Statistical Society. The First 100 Years: 1834–1934," *Journal of the Royal Statistical Society*, Ser. A, 147, 130–139.

Levin, M. L., Goldstein, H., and Gerhardt, P. R. (1950), "Cancer and Tobacco Smoking: A Preliminary Report," *Journal of the American Medical Association*, 143, 336–338.

Lilienfeld, A. M. (1982), "Ceteris Paribus: The Evolution of the Clinical Trial," *Bulletin of the History of Medicine*, 56, 1–18.

Lombard, H. L., and Doering, C. R. (1928), "Cancer Studies in Massachusetts. 2. Habits, Characteristics and Environment of Individuals With and Without Cancer," *New England Journal of Medicine*, 198, 481–487.

Louis, P. C. A. (1835), *Recherches sur les Effets de la Saignée dans Quelques Maladies Inflammatoires*, Paris: Baillière.

——— (1836), *Researches on the Effects of Bloodletting in Some Inflammatory Diseases, and on the Influence of Tartarized Antimony and Vesi-*

cation in Pneumonitis (Translated by C. G. Putnam), Boston: Hilliard, Gray.

——— (1837), "The Applicability of Statistics to the Practice of Medicine," *London Medical Gazette*, 20, 488–491.

Medical Research Council (1957), "Cigarette Smoking and Cancer of the Lung. Statement by the Medical Research Council," *British Medical Journal*, 1, 1523–1524.

Müller, F. H. (1939), "Tabacmissbrauch und Lungencarcinom," *Zeitschrift Krebsforschung*, 49, 57–84.

National Center for Health Statistics (1995), *Health, United States*, 1994, National Health Interview Survey. Hyattsville, MD: Public Health Service.

Neyman, J. (1955), "Statistics—Servant of All Sciences," *Science*, 122, 401–406.

Ries, L. A. G., Miller, B. A., Hankey, B. F., Kosary, C. L., Harras, A., and Edwards, B. K. (1994). *SEER Cancer Statistics Review, 1973–1991: Tables and Graphs*, NIH Publication No. 94-2789, Bethesda, MD: National Cancer Institute.

Schlesselman, J. J. (1982), *Case-Control Studies: Design, Conduct, Analysis*, New York: Oxford University Press.

Schneiderman, M., and Greenhouse, S. (1980), "Jerome Cornfield, 1912–1979," *Journal of the National Cancer Institute*, 65, 1–2.

Streptomycin in Tuberculosis Trials Committee of the Medical Research Council (1948), "Streptomycin Treatment of Pulmonary Tuberculosis," *British Medical Journal*, 2, 769–782.

Study Group on Smoking and Health (1957), "Smoking and Health," *Science*, 125, 1129–1133.

U.S. Department of Health, Education and Welfare, Public Health Service (1964), *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*, Public Health Service Publication No. 1103, Washington, DC: U.S. Government Printing Office.

Whooping-Cough Immunization Committee of the Medical Research Council (1951), "The Prevention of Whooping-Cough by Vaccination," *British Medical Journal*, 1, 1463–1471.

Wynder, E. L., and Graham, E. A. (1950), "Tobacco Smoking as a Possible Etiologic Factor in Bronchogenic Carcinoma," *Journal of the American Medical Association*, 143, 329–336.