Biometrics 59, 1–8
March 2003

PRESIDENTIAL ADDRESS: XXI International Biometric Conference, Freiburg, Germany, July 2002

Are Statistical Contributions to Medicine Undervalued?

Norman E. Breslow

Department of Biostatistics, Box 357232, University of Washington,
Seattle, Washington 98195-7232, U.S.A.
email: norm@u.washington.edu

Summary. Econometricians Daniel McFadden and James Heckman won the 2000 Nobel Prize in economics for their work on discrete choice models and selection bias. Statisticians and epidemiologists have made similar contributions to medicine with their work on case-control studies, analysis of incomplete data, and causal inference. In spite of repeated nominations of such eminent figures as Bradford Hill and Richard Doll, however, the Nobel Prize in physiology and medicine has never been awarded for work in biostatistics or epidemiology. (The "exception who proves the rule" is Ronald Ross, who, in 1902, won the second medical Nobel for his discovery that the mosquito was the vector for malaria. Ross then went on to develop the mathematics of epidemic theory—which he considered his most important scientific contribution—and applied his insights to malaria control programs.) The low esteem accorded epidemiology and biostatistics in some medical circles, and increasingly among the public, correlates highly with the contradictory results from observational studies that are displayed so prominently in the lay press. In spite of its demonstrated efficacy in saving lives, the "black box" approach of risk factor epidemiology is not well respected. To correct these unfortunate perceptions, statisticians would do well to follow more closely their own teachings: conduct larger, fewer studies designed to test specific hypotheses, follow strict protocols for study design and analysis, better integrate statistical findings with those from the laboratory, and exercise greater caution in promoting apparently positive results.

Key words: Black box; Epidemiology; Nobel Prize; Observational studies; Ronald Ross.

The international community of statisticians was elated by news that the 2000 Nobel Prize in economics had been awarded to econometricians James Heckman and Daniel McFadden. Heckman was cited for the "development of theory and methods for analyzing selective samples," McFadden "for his development of theory and methods for analyzing discrete choice." The presentation speech by Swedish economist Professor Karl Gustav Jöreskog made clear that the selection of the 2000 Nobelists was based squarely on their contributions to the development of statistical methodology (`http://www.nobel.se/economics/laureates/2000/presentation-speech.html`):

> Their methods have become standard tools of microeconomic research in economics as well as in other social sciences, and have been applied to solving many important problems of society.
> . . .[They have] made it possible to study individual economic behavior in a statistically correct way.

The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel was established in 1968. Ever since then, the selection committee has broadly interpreted the domain of economic science, and has granted awards for methodology and interdisciplinary work. Previous economics prizes, for example, had gone to Frisch and Tinbergen in 1969 for development of "dynamic models for analysis of economic processes," to Arrow and Hicks in 1972 and Debreu in 1983 for "general economic equilibrium theory," to Solow in 1987 for mathematical models for macroeconomic growth, and to Haavelmo in 1988 for the "probability theory foundations of econometrics." The 2000 prize was a dramatic confirmation of the value and importance of statistical thinking in the social sciences.

McFadden's method for discrete choice analysis, developed to assist a graduate student to study freeway routing choices made by the California Department of Transportation, had its basis in formal theories of economic behavior. Under Marschak's random utility model, observed economic choices—such as whether to ride the bus or drive the car to work—result from multinomial sampling (Becker, DeGroot, and Marschak, 1963). Psychologist Duncan Luce (1959) added the rather strong condition that the ratio of odds for any two choices was independent of the set of other available choices.

These assumptions led McFadden (1973) to his "conditional logit model"

$$\Pr(x \mid s, B) = \frac{e^{\nu(s,x)}}{\sum_{y \in B} e^{\nu(s,y)}},$$

where x is the choice selected, B is the set of available choices, and s measures attributes of the individual decision maker. Corollary developments included methods for outcome dependent, "choice-based" sampling. Manski and Lerman (1977) proposed a weighted likelihood method with origins in sampling theory. Manski and McFadden (1981) developed a pseudolikelihood procedure to which they referred as conditional maximum likelihood or CML. Cosslett (1981) studied semiparametric efficient "maximum likelihood" estimation. McFadden and Reid (1975) realized the need to correct for attenuation when trying to forecast aggregate demand from disaggregated models, and proposed methods for the combination of data from individuals and aggregates.

As I pointed out in my 1995 Fisher Lecture to the Joint Statistical Meetings (Breslow, 1996), much of this work in econometrics was paralleled by biostatistical contributions to the analysis of data from case-control studies in epidemiology. These were stimulated by results of three case-control studies of lung cancer and smoking published in 1950, the most influential of which was that of Doll and Hill (1950). They included Cornfield's (1951) demonstration of the preservation of the odds ratio under case-control sampling, Mantel and Haenszel's (1959) summary relative risk estimator, and Cox's (1958b, 1966) semiparametric approach to logistic regression. Anderson (1972) and Prentice and Pyke (1979) sanctioned applications of "prospective" logistic regression to case-control samples, while Breslow and Day (1980) developed extensions for matched samples of cases and controls.

Heckman's work also has parallels in the (bio)statistical literature. His selection problem was based on the model

$$g(z) = \frac{\pi(z) f(z)}{\int \pi(u) f(u)\, du},$$

where g is the probability distribution observed for the selected (biased) sample, π is the selection probability, and f is the target (population) distribution. Solution of the problem stems from the realization that knowledge or estimation of the selection probability π allows recovery of f from g. The selection may be due to biased sampling—such as occurs with censored survival data, length-biased sampling in screening studies, or choice-based (case-control) sampling. Or it may be due to self-selection by study subjects, whether into the study in the first place, into the treatment group in nonrandomized trials, or into the group of dropouts in a longitudinal study.

Heckman's interest in correction for selection bias was stimulated by his realization that this was crucial for estimating the effects of interventions in structural econometric models. Heckman (1979), for example, noted:

> The wages of migrants do not, in general, afford a reliable estimate of what nonmigrants *would have earned* had they migrated.

This makes clear that the targets of inference are parameters in *counterfactual* models, a point brought out in the press release announcing the Nobel Prize (`http://www.nobel.se/economics/laureates/2000/press.html`, italics added): "[Heckman] ... has also proposed tools for solving closely related problems with *individual differences unobserved* by the researcher." His views on counterfactual causality were explicitly spelled out in a very readable review in the *Journal of Educational Statistics* (Heckman, 1989). Distilling the problem to its essence, he defined Y_1 to be the outcome if the subject was treated, Y_0 the outcome if it was not, and R to be a binary indicator of whether Y_1 ($R = 1$) or Y_0 was observed. The population of interest was the causal population specified by the joint probability distribution $F(y_1, y_0)$. What could be observed was the distribution $G(y, r)$ of the observed outcome and the treatment indicator. Citing Fisher (1951), Roy (1951), and Cox (1958a) as the source of these ideas, he noted that the fundamental problem of causal inference was the fact that

$$\beta = \mathrm{E}(Y_1 - Y_0) \neq \mathrm{E}(Y \mid R = 1) - \mathrm{E}(Y \mid R = 0).$$

One possible way of making progress was to identify covariates Z such that treatment assignment R was independent of treatment outcomes (Y_1, Y_0) conditional on Z. And he was very clear that this was always an assumption!

If this all sounds very familiar—it is! Precisely the same ideas involving counterfactual causality have been explored at length in the statistical literature. Key sources are Rubin (1974) on causal models, Little and Rubin (1987) on the analysis of incomplete data, Rosenbaum and Rubin (1984) on the propensity score, Robins (1986) on the G-computation algorithm, Robins, Hernan, and Brumback (2000) on the marginal structural model for longitudinal analysis, and Pearl (2000) on graphical causal models. I was most disappointed to learn that an invited paper session at this conference that was devoted to this important topic has had to be canceled.

The main point of all of this, of course, is that statisticians, biostatisticians, and epidemiologists have been just as active and successful as have been their counterparts in econometrics in developing statistical methods for causal inference, discrete outcomes, outcome-dependent sampling, and many other statistical problems. However, because they have chosen to apply their work in medicine and public health, rather than economics and social science, none is likely ever to win a Nobel Prize. I first became aware of this issue in 1972, when I went to work for the International Agency for Research on Cancer (IARC). I learned of the impact of the pioneering work of Doll and Hill—not only on reduction of deaths from lung cancer by identification of its major preventable cause, but also on the development of chronic disease epidemiology as a discipline. Richard Doll left mathematics for medicine at the university, but never abandoned his interest in statistics. He joined Bradford Hill in 1948 to work on their lung cancer study (Doll and Hill, 1950), which is still regarded as a model for the application of statistics in medicine. Still active in his nineties, Doll (2002) delivered the 23rd Fisher Memorial Lecture in Oxford last year on the topic "Proof of Causality: Deduction from Epidemiologic Observation." There, he reviewed the careful steps taken in 1948 to address questions of bias: exclusion of selection bias by showing that smoking rates in controls were the same as in patients with other cancers; exclusion of interview and recall bias by studying smoking rates in patients with a false positive

diagnosis of lung cancer. It would be 20 years later before he could publicly debate Ronald Fisher—whose *Statistical Methods for Research Workers* (Fisher, 1934) had stimulated his love for statistics—on the issue of confounding, and why it could not plausibly explain the data on smoking and cancer.

At IARC, I also became aware of the controversy surrounding the fact that this work had not yet been recognized by the selection committee for the Nobel Prize in physiology and medicine. It is now 30 years later, and they haven't gotten any wiser! Bradford Hill, revered by many as the major architect of the randomized clinical trial, is no longer with us. Although Nobel nominations are officially secret, I am told informally by persons outside the selection committee that Richard Doll has been put up at least 20 times, including in each of the last three years. An editorial in the *Lancet* (1999), noting that the pattern in Nobel awards is "one of an overwhelming preponderance of basic research," argued that "research more directly related to health should not be disregarded" and that "to honor [Nobel's] wish fully, clinical and epidemiological research must be rewarded as much as basic science."

The fact that neither statistical nor epidemiological approaches to medicine have been recognized by the most prestigious of awards is, of course, of little significance in its own right. It is primarily of interest as one indicator of the degree to which statistical contributions in medicine are valued—or not valued—by certain segments of the medical community. Fortunately, the founders and selection committees for other prestigious awards have taken a broader view. The General Motors Cancer Research Foundation awards three prizes annually (`http://www.gm.com/company/gmability/philanthropy/cancer_research/index.htm`): the Kettering Prize, "for the most outstanding recent contribution to the diagnosis or treatment of cancer," the Mott Prize, which similarly recognizes work "related to the cause or prevention of cancer," and the Sloan Prize, for "basic science related to cancer research." Richard Doll was the first recipient of the Mott Prize in 1979, being cited "for development of knowledge concerning the environmental causes of cancer in man," and for "definitive investigations of the roles of tobacco smoking and exposure to occupational and therapeutic x-rays." David Cox, in a wonderful recognition of the value of statistics in clinical medicine, received the 1990 Kettering Prize for "the development of the Proportional Hazards Regression Model."

Similarly, the Albert and Mary Lasker Awards (`http://www.laskerfoundation.org/awards/awards.html`) have separate categories for basic medical research, clinical medical research, and public service. Furthermore, they specifically focus on "diseases which are the main causes of disability and death." Many recipients in basic research have gone on to win the Nobel Prize in medicine. Of greater significance for us is the fact that epidemiological and statistical work has formed the basis for several recent Lasker awards. Alfred Sommer was cited in 1997, "for the understanding and demonstration that low dose vitamin A supplementation in millions of third world children can prevent death from infectious disease as well as blindness." Alfred Knudson received the Lasker in 1998 "for incisive studies in patient-oriented research that paved the way for identifying genetic alterations that cause cancer in humans." The paper that launched this research (Knudson, 1971) laid out the 2-stage mutational model of carcinogenesis, and bore the intriguing title "Mutation and Cancer: Statistical Study of Retinoblastoma." It is a remarkable example of how simple statistical observations—when cleverly and wisely interpreted—can foster basic science hypotheses that have now been validated at the molecular level. (This paper and its sequelae have stimulated much of my own statistical work with Wilms tumor of the kidney, another childhood neoplasm involving a paired organ.) The 2001 Lasker Award for public service went to William Foege "for his courageous leadership in improving worldwide public health, and his prominent role in the eradication of smallpox."

Finally, the 1997 European *prix Louis-Jeantet de médecine* was awarded to statistician and epidemiologist Richard Peto, a disciple and colleague of Richard Doll. (The Geneva foundation making the award stipulates that candidates be selected, not only for their past record, but also for future promise, and the prize money is to be used for their research.) Peto also has had enormous impact on tobacco sales and public policy, with his quantification of the health hazards and mortality risk of cigarette smoking. Some measure of this impact may be gleaned from the fact that he recently received the title "WHO Minister of Statistical Propaganda" from a U.S. libertarian prosmoking lobby. Meta-analyses conducted at the Oxford Clinical Trials Service Unit that he directs have helped set worldwide policy regarding the use of aspirin, tamoxifen, and other medicaments.

These examples demonstrate that contributions of statisticians and epidemiologists are indeed recognized and appreciated by an influential sector of the medical community. The failure of any of these contributions to earn recognition with a Nobel Prize can be viewed primarily as a reflection of the viewpoints of the individual members of the various Nobel selection committees. I was informed that no explicit policy rules out the selection of an epidemiologist, and that some of the Swedish medical faculty who constitute the selection committee are working hard to overcome the perceived basic science bias. In discussions with one of these faculty regarding the low esteem accorded epidemiology in some scientific circles, mention was made of a possible "malignant synergy between the charlatans and the media." Indeed, a good part of the problem that we statisticians and epidemiologists face in having our work valued as highly as it should be stems from our own failures. Many of these were discussed in the *Science* article by Gary Taubes (1995), "Epidemiology Faces its Limits," that began with the observation: "The news about health risks comes thick and fast these days, and it seems almost constitutionally contradictory." The statisticians and epidemiologists interviewed for the article were generally in agreement that there were serious problems. Harvard epidemiologist Dimitrios Trichopoulos, for example, opined: "People don't take us seriously anymore, and when they do take us seriously, we may unintentionally do more harm than good." Many of those interviewed suggested that results of individual observational studies—even those conducted according to the most stringent standards—should not be accepted as providing evidence of a causal assocation, unless the lower confidence bound on the relative risk exceeded 3 or 4. Yet, with the pressure on investigators to make a career through publication, few studies published in the literature and reported in the media meet this standard. Authors themselves are

often aware of the weaknesses, but fail to communicate these adequately—either in publication or in subsequent contact with representatives of the press.

The sometimes capricious nature of epidemiologic findings was brought home to me and my students in a dramatic way this past Autumn Quarter, when I taught a second-year class for master's level statisticians and doctoral-level epidemiologists that focused largely on logistic regression analysis of data from case-control studies. For use in the term project, my esteemed colleague, epidemiologist Noel Weiss, kindly provided data that had already been analyzed and published for the primary question, but not for a secondary question, of interest to the investigators. The document I handed out to the class described the study and the project goals as follows:

> Weiss and colleagues have conducted a series of case-control studies of endometrial cancer risk and hormonal replacement therapy in Western Washington... The study questionnaire contained sections on diet, smoking history and lifelong consumption of alcoholic beverages...The purpose of this project is to study the association between alcohol consumption and endometrial cancer using basic descriptive statistics and logistic regression modeling techniques... Some studies have suggested that alcohol in the form of a particular beverage is most protective...

A month later, after receiving 12 papers from teams of 2–3 students each, I edited and presented in class the summary conclusions from each team. Here is a selection of excerpts from those conclusions (italics added):

Team 10 After adjustment for confounding, alcohol consumption does not appear to affect the risk of developing endometrial cancer.

Team 4 A mild protective effect was observed for combined alcohol (OR = 0.73, 95%CI 0.58, 0.93).

Team 1 Alcohol intake can reduce the risk of endometrial cancer, and this protective effect only exists *for non-obese women.*

Team 8 Endometrial cancer risk is slightly, but not materially decreased by alcohol consumption. However, consumption of ethanol of 1.6–4.0 gm/day may reduce the risk of endometrial cancer *in obese women.*

Team 6 Alcohol consumption is not associated with significantly reduced risk for endometrial cancer after adjusting for known confounding characteristics. There may be some protective effect of wine consumption, but this needs further confirmation.

Team 3 Consumption of beer or wine was associated with a decreased risk of endometrial cancer, while consumption of liquor was associated with an increased risk... Beer, wine and liquor all showed a monotonic dose-response relationship...We were unable to demonstrate an interaction between alcohol consumption and *body mass index*... The current study goes a long way towards resolving the inconsistent epidemiological literature.

The first conclusion from these contradictory findings, of course, is that I must have failed as a teacher! When I looked closely at the reasons for the dramatically different conclusions, however, it seemed clear that they were the result of a chain of effectively arbitrary choices made with regard to scales of measurement, cut points for discretization, and variables selected as potential confounders. The most blatant contradiction, in the direction of the apparent interaction between body mass and alcohol, resulted from one team being so concerned about confounding between abstention and unmeasured risk factors that it dropped the abstainers from the analysis and sought a dose-response relationship among the drinkers alone. Standard texts, e.g., Clayton and Hills (1993, p. 256), advocate this strategy under certain circumstances.

This exercise involved teams of budding epidemiologists and statisticians who were given the same set of data, and who had all just had the same instruction in logistic regression modeling techniques. When I showed the results to Weiss, he had to admit that he wasn't really that surprised! We agreed that, had 12 teams been funded to go out and collect their own data in different parts of the world, the variability in the reported results would have been even greater.

What can be done to reduce this variability and, in so doing, improve the status of epidemiology as a scientific discipline? The few suggestions I want to mention have, for the most part, been repeatedly emphasized by others. Before I get to these, however, I would like to take a side trip and mention some thoughts on the subject from the writings of a most unlikely source: the winner of the second Nobel Prize in physiology and medicine. An exception who proves the rule, he was a brilliant man who first won the Nobel and then went on to become an epidemiologist and biostatistician.

I first heard of Ronald Ross (Figure 1) in 1967, while a postdoc at the London School of Hygiene and Tropical Medicine, which at that time boasted of "incorporating the Ross Institute." I was vaguely aware of his discovery linking mosquitoes and malaria, but it wasn't until reading the review by Dietz and Schenzle (1985) that I learned of his fundamental contributions to mathematical epidemiology. Ronald Ross was born in 1857 in the Himalayan hilltown of Almara, when his father was a captain with the 66th regiment of Gurkhas. He was sent to England for medical studies, receiving a licentiate in medicine in 1879, from St. Bartholomew's Hospital. Ross returned home as a military doctor with the Indian Medical Service, but apparently had ample free time for other pursuits. Between fishing, hunting, and tennis matches at the Madras Club, he studied mathematics on his own and wrote poetry and prose. During visits to London, he developed a close relationship with parasitologist Patrick Manson, of the London School of Hygiene and Tropical Medicine, who encouraged him to spend his free time working on the link between mosquitoes and malaria. In August 1897, Ross completed his work on the life cycle of the malaria parasite, showing its transmission from man to anopheline mosquito and back to man. (In fact, since human transmission was low in the region to which he was posted, Ross worked out the final step in the bird population.) The discovery was reported the following December in the *British Medical Journal*. Ross assessed the value of his achievement by writing a poem:

> This day relenting God
> Hath placed within my hand
> A wondrous thing; and God
> Be praised. At His command.
>
> Seeking His secret deed
> With tears and toiling breath,

Figure 1. Photo of Ross in Mauritius, 1908, courtesy of London School of Hygiene and Tropical Medicine.

> I find thy cunning seeds,
> O million-murdering Death.
>
> I know this little thing
> A myriad men will save.
> O Death, where is thy sting,
> Thy victory, O Grave!

Five years later, in 1902, Ross received the invitation to Stockholm for the second medical Nobel Prize. Professor Count Mörner, chairman of the Nobel Committee, summarized the significance of his accomplishment as follows (`http://www.nobel.se/medicine/laureates/1902/press.html`):

> By your discoveries you have revealed the mysteries of malaria. You have enriched science with facts of great biological interest and of the very greatest medical importance. You have founded the work of preventing malaria, this veritable scourge of many countries.

Ross turned away from biology and medicine, however, to continue his mathematical work, receiving in 1904 a doctor of science degree in mathematics from the University of Dublin. He took a position initially at the Liverpool School of Hygiene, but later moved to London, where he helped develop the institute that bore his name, dying there in 1932. Throughout this period, he devoted himself to mathematical epidemiology and its application to the prevention of malaria. His theory of "dependent happenings"—the paradigm "happening" being a case of malaria infection—first appeared in 1908, in a report on malaria control in Mauritius (Fine, 1975). The central equation was

$$N = p \cdot m \cdot i \cdot a \cdot b \cdot s \cdot f,$$

where

N = no. new infections per month
p = average population
m = proportion population infected
i = proportion infectious among infected
a = average no. mosquitoes/person
b = proportion of uninfected mosquitoes that feed on man
s = proportion of mosquitoes that survive through incubation period
f = proportion of infected mosquitoes that feed on man

Ross's own explanation for his terminology appeared in the second edition of his book *The Prevention of Malaria* (Ross, 1911):

> We shall deal with time-to-time variations not only of malaria, but of all disease, and not only of diseases of man, but those of any living organism. Still further, as infection is only one of many kinds of events which may happen to such organisms, we shall deal with *happenings* in general.

The major importance of this equation was that it led directly to what Ross called "the critical mosquito density," the key number for a malaria control program (Fine, 1975). Letting r denote the recovery rate, and assuming $b = f$, Ross set the number of recoveries per month $r \cdot m \cdot p$ equal to the number of new infections N, to arrive at the expression

$$a = \frac{r}{b^2} \cdot \frac{1}{i} \cdot \frac{1}{s} \approx \frac{0.2}{(0.25)^2(0.25)(0.33)} = 40.$$

Plugging in reasonable values for the recovery rate, the proportion of mosquitoes (infected or uninfected) that feed on man, etc., he arrived at the conclusion that malaria would die out if the mosquito density could be reduced to fewer than 40 mosquitoes per person in the population. This mathematical way of thinking contradicted the prevailing wisdom in medicine and public health, however, which held that only complete elimination of the mosquito would control malaria. The only way to be sure was to test the concept in practice. This Ross had done during a late 1902 consultancy with the Suez Canal Company in Ismailia. Figure 2, based on data from the published version of Ross's Nobel lecture (`http://www.nobel.se/medicine/laureates/1902/rosslecture.html`), dramatically illustrates the success of his ideas.

Ross's theory evolved from the product formula shown above for the number of new infections per month, to finite difference equations for the changes from one month to another, and finally to differential equations that express the law of mass action. With $x(t)$ denoting the number of susceptibles at time t, and $y(t)$ the number of infectives:

$$\frac{dx}{dt} = -\beta xy, \quad \frac{dy}{dt} = \beta xy - \gamma y,$$

where β is the transmission rate and γ the recovery rate. These equations, whose solution was approximated by Ross's

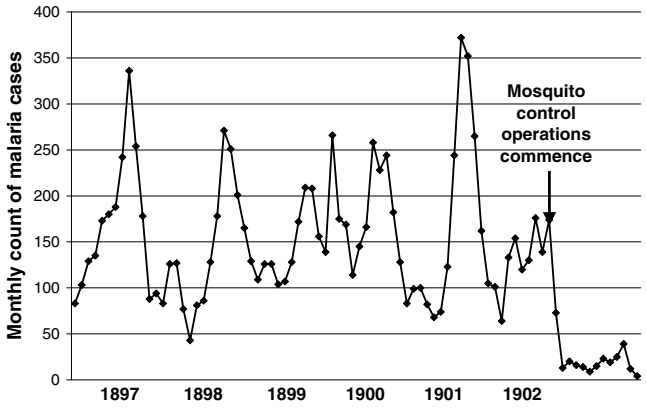

Figure 2. Monthly counts of malaria cases recorded by the Suez Canal Company over a seven-year period.

disciple McKendrick and a host of later workers, form the foundations for modern epidemic theory (Dietz and Schenzle, 1985). They lead, for example, to the concept of the basic reproductive number, today's equivalent of Ross's critical mosquito density. With the advent of the worldwide epidemic of HIV infection and renewed interest in tropical medicine, a new generation of research workers is learning about the fundamental contributions of Ronald Ross. Several sessions at this conference are devoted to this work and its extensions.

Although he won the Nobel Prize for his biology, in Ross's own opinion his "principal work has been to establish the general laws of epidemics." Ross's biographers Nye and Gibson (1997, p. 279) state: "If Ross had been born 100 years later he could have become an eminent biostatistician. His thinking in this area was well ahead of his time and perhaps did not achieve the recognition it deserved until much later." Ross (1916) described how his approach to medical statistics differed from that of others as follows:

> The whole subject is capable of study by two distinct methods which are used in other branches of science, which are complementary of each other, and which should converge towards the same results.

The approach that prevailed in Ross's day, and in our own, was termed by him the a posteriori method:

> . . . we commence with observed statistics, endeavor to fit analytical laws to them, and so work backwards to the underlying cause (as done in much statistical work of the day).

His clear preference, however, was for the a priori method:

> . . . we assume a knowledge of the causes, construct our differential equations on that supposition, follow-up the logical consequences, and finally test the calculated results by comparing them with the observed statistics.

Much the same distinctions between different epidemiologic approaches have been made in the review by Susser and Susser (1996) of the different eras in epidemiology and their distinctive paradigms—which they use as a basis for their vision of the future (Table 1). The Sussers define the "black box" approach as one that relates exposure to disease "without necessity for intervening factors or pathogenesis." Could one of the problems of modern epidemiology thus be that we have drifted back to a posteriori methods—fitting black box equations to data, rather than working out predictions from mathematical modeling of underlying processes? Klaus Dietz has been one of the foremost contributors to the a priori approach, yet his 1985 review with Schenzle stated

Table 1
The Sussers's paradigms in epidemiology

Period	Era	Paradigm
19th century	Sanitary statistics	Miasma
1st half 20th century	Infectious disease epidemiology	Germ theory (a priori)
2nd Half 20th century	Chronic disease epidemiology	Black box (a posteriori)

> By now about one thousand papers have contributed to mathematical epidemiology, but smallpox eradication was achieved without the use of any of them.

Dietz and Schenzle followed up, quoting this more detailed assessment by Bart et al. (1983):

> Mathematical formulations have been developed to describe outbreaks of infectious disease, to test concepts, and to provide insights into disease control and policy formulation. The resulting equations sometimes mirror the observed events, but to date have had little impact upon disease control or preventive practice. Instead they have been used more retrospectively to reassure rather than assist in the development of policy.

I am very glad to see that, in spite of this apparently pessimistic assessment, Dietz has not abandoned his work. He will be presenting an invited paper at this conference, and I look forward to discussing his current thinking with him.

Richard Peto (1985) has made much the same distinctions as do Ronald Ross and the Sussers. Writing about Doll and Hill's landmark work on smoking and lung cancer, he said:

> The importance of the story of this discovery is that it did not depend on any serious understanding of the mechanisms of carcinogenesis [or] of the . . . causative components of tobacco smoke. . . . This indicates that there are two complementary approaches. . .

One was the *mechanistic* approach, which suggests that one must first

> . . . understand the mechanisms of carcinogenesis and the preventive measures will follow.

Opposing this was the *black box* approach, which

> . . . seeks. . . correlates. . . of the risk. . . of cancer among people. [This] has yielded by far the. . . most important results. . .in the prevention of chronic disease.

Both Ross and Peto note, however, that the two approaches (i) mechanistic, or a priori, versus (ii) black box, or a

posteriori, are complementary to each other—which suggests that a synthesis is possible. Results of observational epidemiologic studies analyzed using black box techniques are most convincing when supplementary information on biological processes that support the observed association is available. This principle was well codified by Bradford Hill as the criterion of *biological plausibility* in his guidelines for causal inference.

In summary, statisticians and epidemiologists have contributed enormously to the prevention and treatment of disease. The low esteem accorded their work by many in the medical profession is due in part to the publication of too many contradictory findings from observational studies. Laboratory workers have the luxury of being able to quickly repeat their experiments; many are careful to do so, and thus replicate their results before submitting them for publication. Epidemiology also suffers from an inherent aversion to "black box" thinking, as the perjorative nature of this term implies. This has been a classic source of disagreement between public health officials who want to develop policy before mechanisms are fully understood and critics who insist on greater "proof" before implementing regulatory actions.

Specific recommendations for improvements in epidemiology—which hopefully will lead to a greater appreciation for the discipline—have been made by many researchers, including those mentioned in this talk. The same principles are enunciated in courses and textbooks, but too often seem to be forgotten—even by those who profess them. These include the call for fewer, larger studies that are designed to test specific *a priori* hypotheses. Epidemiologists would do well to emulate the clinical trials specialists, by developing strict protocols for both study design and subsequent analysis of the collected data. Doll and Hill worked from such a protocol, and had the added advantage that the punch card technology of their era made analyses so time-consuming that it prevented data dredging and the inclusion of all but the most essential variables in the analysis. We should pay more attention to biological mechanisms, and attempt wherever possible to integrate the epidemiologic findings with those emanating from the laboratory. The *a priori* and *a posteriori* approaches should indeed be viewed as complementary, rather than competing. We need more definitive reviews and syntheses of individual data records from different studies. Finally, greater caution is clearly needed in promoting positive results. While it may not be feasible to replicate a large study before publication, it could be prudent to wait until completion of another similar study, and then to publish the two together. Good science must always be the top priority.

Acknowledgements

I am grateful to the following individuals for conversation or correspondence related to the theme of this talk: Hans-Olov Adami, Richard Doll, Eric Feigl, Elizabeth Halloran, Michael Hills, Paul Fine, Alfred Knudson, Dimitrios Trichopoulos, and Noel Weiss. Special thanks go to the University of Washington students in Biostatistics 536, Autumn Quarter, 2001. This work was supported in part by grant 5-R01-CA40644 from the United States Public Health Service. Ross's three-stanza poem, from *Ronald Ross: Malariologist and Polymath* by Nye, E. R., and Gibson, M. E. (1997), is copyright © Nye, E. R., and Gibson, M. E. and is reprinted with permission of Palgrave Macmillan.

Résumé

Les économètres Daniel Mc Fadden et James Heckman ont obtenu le Prix Nobel d'Économie 2000 pour leurs travaux sur les choix de modèles discrets et les biais de sélection. Les statisticiens et les épidémiologues ont apporté des contributions équivalentes en médecine par des travaux sur les études cas-témoin, l'analyse des données incomplètes, et l'inférence causale. En dépit de nominations répétées aussi éminentes que Bradford Hill et Richard Doll, le Prix Nobel de Physiologie et de Médecine n'a jamais été attribué pour des travaux en biostatistique ou en épidémiologie (l'exception qui confirme la règle est Ronald Ross qui obtint en 1902 le second Prix Nobel de Médecine pour la découverte du vecteur le paludisme, le moustique. Ross développa alors la mathématique d'une théorie de l'épidémie, qu'il considéra comme sa plus importante contribution scientifique, et appliqua ses résultats à des programmes de lutte contre le paludisme). Le faible intérêt accordé aux biostatistiques et à l'épidémiologie dans certains cercles médicaux, et de plus en plus dans le public, est étroitement lié aux résultats contradictoires d'études observationnelles publiées si complaisamment dans la presse de vulgarisation. En dépit de son efficacité démontrée pour sauver des vies, l'approche "boite noire" épidémiologique des facteurs de risque n'est plus réellement respectée. Pour remédier à ces perceptions malheureuses, les statisticiens devraient suivre plus étroitement leurs propres principes: construire des études moins nombreuses mais plus larges pour tester des hypothèses spécifiques, suivre des protocoles stricts pour la définition de leurs études et leur analyse, mieux intégrer les résultats statistiques avec les faits de laboratoire, et prendre plus de précautions dans l'affichage de résultats apparemment positifs.

References

Anderson, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59,** 19–35.

Bart, K. J., Orenstein, W. A., Hinman, A. R., and Amler, R. W. (1983). Measles and models. *International Journal of Epidemiology* **2,** 263–266.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1963). Stochastic models of choice behavior. *Behavorial Science* **8,** 41–55.

Breslow, N. E. (1996). Statistics in epidemiology: the case-control study. *Journal of the American Statistical Association* **91,** 14–28.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research I: The Analysis of Case-Control Studies.* Lyon: International Agency for Research on Cancer.

Clayton, D. and Hills, M. (1993). *Statistical Models in Epidemiology.* Oxford: Oxford University Press.

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.

Cosslett, S. R. (1981). Maximum likelihood estimator for choice-based samples. *Econometrica* **49,** 1289–1316.

Cox, D. R. (1958a). *Planning of Experiments.* New York: Wiley.

Cox, D. R. (1958b). The regression analysis of binary sequences. *Journal of the Royal Statistical Society, Series B* **20,** 215–242.

Cox, D. R. (1966). Some procedures connected with the logistic qualitative response curve. In *Research Papers in Statistics: Festschrift for J. Neyman*, F. N. David (ed), 55–71. New York: Wiley.

Dietz, K. and Schenzle, K. (1985). Mathematical models for infectious disease statistics. In *A Celebration of Statistics: The ISI Centenary Volume*, A. C. Atkinson and S. E. Fienberg (eds), Chapter 8, 167–204. New York: Springer.

Doll, R. (2002). Proof of causality: deduction from epidemiological observation. *Perspectives in Biology and Medicine* **45,** 499–515.

Doll, R. and Hill, A. B. (1950). Smoking and carcinoma of the lung: Preliminary report. *British Medical Journal* **2,** 739–748.

Fine, P. E. M. (1975). Ross's *a priori* pathometry—a perspective. *Proceedings of the Royal Society of Medicine* **68,** 547–551.

Fisher, R. A. (1934). *Statistical Methods for Research Workers,* 5th edition. Edinburgh: Oliver and Boyd.

Fisher, R. A. (1951). *The Design of Experiments*. Edinburgh: Oliver and Boyd.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47,** 153–161.

Heckman, J. J. (1989). Causal inference and nonrandom samples. *Journal of Educational Statistics* **14,** 159–168.

Knudson, A. G. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* **68,** 820–823.

Lancet (1999). Narrowness of Nobel awards for physiology or medicine. *Lancet* **354,** 1399.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Luce, R. (1959). *Individual Choice Behavior*. New York: Wiley.

Manski, C. F. and Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica* **45,** 1977–1988.

Manski, C. F. and McFadden, D. (1981). Alternative estimators and sample designs for discrete choice analysis. In *Structural Analysis of Discrete Data with Econometric Applications*, C. F. Manski and D. McFadden (eds), Chapter 1, 2–50. Cambridge, Massachusetts: MIT Press.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22,** 719–748.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics,* P. Zarembka (ed), Chapter 4, 105–142. New York: Academic Press.

McFadden, D. and Reid, F. (1975). Aggregate travel demand forecasting from disaggregated behavioral models. *Transportation Research Record* **534,** 24–37.

Nye, E. R. and Gibson, M. E. (1997). *Ronald Ross: Malariologist and Polymath*. New York: St. Martin's Press.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Peto, R. (1985). The need for ignorance in cancer research. In *The Encyclopaedia of Medical Ignorance: Exploring the Frontiers of Medical Knowledge*, R. Duncan and W. S. Miranda (eds), 129–133. Oxford: Pergamon Press.

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66,** 403–411.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect. *Mathematical Modelling* **7,** 1393–1512.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11,** 550–560.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* **79,** 516–524.

Ross, R. (1911). *The Prevention of Malaria*. London: John Murray, 2nd edition.

Ross, R. (1916). An application of the theory of probabilities to the study of *a priori* pathometry, part 1. *Proceedings of the Royal Society, Series A* **92,** 204–230.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers* **3,** 135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* **66,** 688–701.

Susser, M. and Susser, E. (1996). Choosing a future for epidemiology: I. Eras and paradigms. *American Journal of Public Health* **86,** 668–673.

Taubes, G. (1995). Epidemiology faces its limits. *Science* **269,** 164–169.

Received October 2002. Revised October 2002.
Accepted October 2002.