EPILOGUE OF THE HEALTH-POLLUTION CONFERENCE

JERZY NEYMAN
STATISTICAL LABORATORY, UNIVERSITY OF CALIFORNIA, BERKELEY

1. *General remarks.* The papers published in this volume represent the result of the effort to compile a realistic cross section of the contemporary statistical thinking on the problems of pollution and health. The first seven papers stem from public institutions and, with unavoidable differences as to the amount of detail, reflect these institutions' interests. The next seven papers illustrate the sharp dispute about health effects of radioactive pollutants. The subjects of the remaining papers are varied, each representing a different "case history" connected with the problem of pollution. Thus far the problem of health and pollution has not attracted the attention of many mathematical statisticians and this volume contains just one paper, by Richard E. Barlow, that contains a theorem.

Four papers of the first group, one by Totter and the other three by Finklea, by Riggan, and by Nelson, with collaborators, describe in detail the very impressive programs of activities of the Biology-Medicine Division of the Atomic Energy Commission and of the Division of Health Research, Environmental Protection Agency. To a considerable extent, this includes not only the work of the two important agencies of the Federal Government, but also that conducted by the various contractors. As a result, these four articles do give a firsthand account of a large section of the contemporary statistical work on pollution. The paper by Sirken illustrates the commendable concern with the reliability of data collected by the National Center for Health Statistics.

While this coverage of institutional research is gratifyingly broad, it is regretted that the information in this volume on the impressive amount of work (with 468 papers published up to June 1969!) performed under the aegis of the National Academy of Sciences–National Research Council is only secondhand, being fragmentarily reported by several speakers. In particular this applies to the NAS-NRC biology-health studies of the atomic bomb casualties in Hiroshima and Nagasaki.

Even though the announced ultimate goal of the conference was to discuss and to plan a comprehensive statistical study of the relationship between human health and the various pollutants, and even though four skeletal plans have been submitted and published above, other material in the volume shows little enthusiasm for the project. In fact, Gofman and Tamplin are explicit in opposing

This investigation was partially supported by PHS Grant No. GM-10525-08, National Institute of Health, Public Health Service.

the idea. Most other authors showed indifference. Particularly this is the case with regard to the controversial pollutants such as radioactivity: a hot potato effect, perhaps. Yet, the papers presented do suggest that the current information on health effects of the various pollutants is fragmentary and that firm data needed for the formulation of national policy might be obtained only through a difficult, large, broadly multipollutant and multilocality statistical study. In particular, this applies to radiation because, apparently, the biological effects of radiation are not uniquely induced by radiation, but are also caused by other environmental agents. Among the other agents that may compete with radioactivity, various papers emphasize DDT, lead, mercury, and a defoliant 2,4,5-T. The programs of the Environmental Protection Agency and, particularly, of the Biology-Medicine Division of the Atomic Energy Commission are so extensive that the proposed comprehensive statistical health-pollution study might be included within the sphere of these institutions' activities. The drawback is that both the AEC and the EPA appear to be parties to the sharp controversy which the proposed comprehensive statistical study could help to resolve.

The purpose of the present *Epilogue* is to emphasize a few arguments in favor of a comprehensive study, perhaps on the lines of the "Skeletal Plan . . ." published earlier in this volume, and also to bring out some methodological-statistical difficulties which seem to have escaped the attention of a number of authors.

2. *Fragmentary character of information available on health effects of radioactivity.* Figure 2 in the paper by Totter refers to an experiment with mice irradiated with a high rate of X-ray. It contains a tentative extrapolation to humans. The figure seems to imply that, if a fetus is irradiated *in utero* at any time some two to five weeks after conception, then it is almost certain to develop abnormalities; also, a very substantial proportion of live births, perhaps 40 to 50%, are followed by neonatal deaths. All this applies to a high dose of X-ray radiation administered at a high rate, not in conditions of real life. Figure 4 in the same paper of Totter indicates that the effect of a fixed dose of gamma ray irradiation depends substantially on the rate at which this dose is administered. Specifically, the decrease of the rate from 6.7 rads/min to some 0.003 rads/min decreases the shortening of life of the experimental animals in the ratio of, roughly, 4 to 1.

All the above are results of experiments performed on animals and one is naturally interested in the effects on humans of such rates of radiation as one is likely to find in actual life. Here the paper by Vaughan provides some relevant information. This paper gives a description of an interesting study conducted in Alaska. The study is concerned with the food chain: fallout → lichens → caribou → Eskimos. The study seems to have covered the period 1963–1969 and involved direct observations and measurements of cesium 137 in the fallout, in lichens, in caribou, and in humans. The findings are summarized in Figure 6. Briefly and roughly, they are as follows. The fallout content of cesium 137 diminished steadily and reached essentially zero by the end of 1966. The lichens'

content of cesium 137 had a maximum in 1965. Then it showed a gentle decline, ending at a level about double that at the beginning of the study period. The curve representing the body-burden of cesium 137 in the Eskimos is somewhat complicated, showing a yearly cyclic variation. The yearly minima grow from about 160 nCi in 1964 to about 300 nCi in 1968. At the same time, the maxima increase, roughly, from 575 nCi to 700 nCi. No indications are given of the health effects that the accumulation of the radioactive cesium may have caused either to adult Eskimos or to their progeny. However, a detail in Figure 6 indicates a chromosomal study as the next step in continuing research. This may or may not mean an effort to answer the question suggested by Totter's Figure 2 about the effects on human fetuses of radiation administered at now observable levels.

It must be obvious that, while highly interesting, the study described by Vaughan can provide only a fragmentary answer to the general question of radioactive pollution and health. This will continue to be the case even when the indicated chromosomal study is completed and even if it is accompanied by similar studies of other pollutants mentioned by Vaughan: DDT, lead, mercury, and the defoliant 2,4,5-T. The reasons are that (a) malformations at birth and postnatal deaths need not be reflected in the observable changes in the chromosomes and (b) that whatever may be found for Eskimos in Alaska will refer to the whole local Alaskan pattern of pollutants and not to any one of them in particular.

Clearly, in order to be able to evaluate n unknowns (these would be separate effects on a particular health parameter of n different pollutants or their combinations), one needs to have n independent equations. Over a unit of time, each locality, such as Alaska, can produce only one such equation. To estimate n effects, one needs at least n localities with different patterns of pollutants; in order to allow for random variation, a multiple of n localities is needed.

One other paper in this volume is concerned with the body-burden of a radioactive chemical. This is the paper by Rosenthal giving measurements of strontium 90 in teeth and in bones of children, apparently accumulated through milk consumed by expectant mothers. The information given is obviously important but, by itself, it is only a fragment of the general picture that seems important to have.

As described by Totter, the effect of the accumulating strontium 90 body-burden in man, resulting from the world wide fallout from nuclear tests, is a continuing concern of the AEC. A number of experiments with various animals have been conducted and are still in progress. There seems to be no doubt that, at sufficient rates of intake, strontium 90 is deadly to animals through a number of forms of cancer. As far as humans are concerned, there is the omnipresent problem of extrapolation.

Figure 6 of Vaughan suggests the intriguing question about the source of cesium 137 in the fallout over Alaska, quite large in June 1963 and then declining to essentially zero in 1966. Here, several figures of Sternglass, beginning with

Figure 4, come to one's mind. Among other things, Figure 4 indicates that 1961–1963 were three years of the latest large H-bomb tests. These tests were conducted at distances from Alaska measured in thousands of miles and one is inclined to skepticism at the suggestion that just these tests could have resulted in measurable quantities of cesium 137 in the fallout, not to speak of the measurable body-burden of this chemical in the Eskimos that persisted at least up to 1968. Is there any other imaginable source of cesium 137?

The diagrams produced by Sternglass are intended to support his opinion that the observable deceleration in the decline of infant mortality is likely to have been caused by tests of nuclear weapons. While the covariation asserted by Sternglass seems to be there, the arguments for causality based on this covariation alone appear tenuous. However, if no imaginable source of cesium 137, in the fallout over Alaska, can be found other than the 1961–1963 H-bomb tests, then the findings described by Vaughan contribute to the credibility of the Sternglass hypothesis.

Remark. In what follows the words cause, causality, and the like are occasionally used. The reader will realize that these words are used without any metaphysical connotation. With reference to empirical facts, the statements to the effect that "A is caused by B" mean simply that in the past, the appearance of B was always followed by A. With reference to theoretical speculations "causality" means the hypothesis that not only in the past but also in the future the appearance of B will be followed by A, possibly, through a particular hypothetical mechanism.

No observational study can ever establish causality. The best one can hope for is that an observational study will suggest real causal relations. Thus, the proposed multipollutant and multilocality statistical health-pollution study cannot resolve all the controversies. However, if conducted with all due care, it can help to unify the existing fragmentary information to form a coherent general picture of actual happenings. Among other things, such a study could answer the question whether the body-burden of radioactive chemicals (cesium 137, strontium 90, etc.) in people living close to electric power generators using mineral fuels is less or is larger than among those living *in comparable conditions* next door to nuclear power plants. Another illustrative question that the proposed study might answer is that about the frequency of malformations at birth in *comparable localities* in which, however, the body-burdens of radioactive chemicals are different. Both these questions seem interesting and important. There are many others of the same kind.

The two questions just mentioned illustrate some of the difficulties of the proposed study. For one thing, the monitoring of malformations at birth must be made reliable. As illustrated by the conference organized last year by Professor Hook, this is a difficult problem. For another thing, the program of monitoring pollutants organized by the EPA should be enlarged to include a new kind of "environment"—the bodies of the people: the body-burdens of radioactive chemicals will be necessary for the study.

Whether all of these and a host of similar other problems can be authoritatively

solved, and by what uncommitted organization, is debatable. With reference to the defoliants, Professor Sterling mentions the American Statistical Association. An alternative possibility might be the International Association for Statistics in Physical Sciences which exhibited an independent interest in the matter. On the initiative of this organization an international symposium on pollution was recently held at Harvard, organized by Professor John W. Pratt.

The phrase "in comparable conditions" is easy to write. However, because of all the complexities of monitoring physical factors, and because of the relevance of race and of socioeconomic status, as discussed by Landau and by Winkelstein, the difficulties in reaching a reasonable level of comparability are enormous.

3. *Pitfalls of competing risks.* The much debated question of reliability or otherwise of the safety standards for man-made radioactivity hinges considerably on extrapolations of the results of studies conducted on several groups of individuals, who underwent heavy exposures to radiation administered at high rates. This includes the survivors of atomic bomb attacks on Hiroshima and Nagasaki. The studies concerned were conducted at least partly under the aegis of the National Academy of Sciences–National Research Council, in part through the Atomic Bomb Casualty Commission. No direct report on these studies is published in this volume, but a few fragmentary references to them indicate some of the difficulties involved. For example, the dosimetry proved difficult: on occasion it was difficult to estimate the exact dose of radiation to which a survivor at Hiroshima or Nagasaki was exposed. However, there was another great difficulty in the same studies which, with the apparently single exception, did not seem to attract the attention of the various authors. The difficulty in question is that of allowing for the presence of so-called "competing risks." [1]

The directly computable death rate from a cause C_1 is what is called the "crude rate," the value of which depends not only on the intensity of C_1 but also on that of all other competing causes, say C_2. The rate that characterizes C_1 alone, is the "net rate" which is the rate of interest. If C_2 is only mild, then the difference between the net and the crude rates of C_1 can be trivial. On the other hand, if the combination C_2 of all other causes competing with C_1 is intense, then the net rate of death from cause C_1 can be a large multiple of the crude rate.

It is quite plausible that the radiation induced deaths of infants in Hiroshima and Nagasaki had very intense competitors in, say, starvation, lack of maternal care, and the like. Again, with reference to some other studies in this country and abroad, of groups of individuals subjected to X-ray treatment against some disease D, deaths from that same disease, and from the various complications thereof, must have competed with deaths from the radiation induced cancer. How heavy were these cases of competition? What was done, and how, to elicit the all important estimates of net rates of deaths directly caused by radiation? The competence and the authority of the National Academy of Sciences are and should be great. Therefore, it is very regrettable that the methodology used to solve the problem of competing risks is not described in this volume.

Even though the term competing risk is not mentioned by Gofman and

Tamplin, their discussion indicates that they are fully aware of the issues involved, complete with the complication that the radiation induced leukemia tends to develop earlier than other radiation induced cancers.

4. *Pitfall of incomplete comprehensiveness of a multipollutant, multilocality study.* In a multipollutant, multilocality statistical investigation one of the threatening pitfalls is incompleteness of the set of pollutants studied: if the study involves a certain number s of pollutants, say $P_1, P_2, \cdots, P_s$, but neglects another pollutant P_0 that happens to be important, then the conclusions regarding $P_1, P_2, \cdots, P_s$, suggested by even very highly significant findings, may be completely misleading.

A case in point is a recently published multipollutant and multilocality study of a substantial number of pollutants, which omitted radioactivity. One of the findings was that an increase of copper content of the air diminishes significantly the experienced frequency of death from a certain disease D. The correct, but somewhat lengthy interpretation of the result is, roughly, as follows:

> Among the so many localities studied (and there were quire a few of them), the average frequency of deaths from disease D in localities with high copper pollution is less than in other localities in which copper pollution is low.

This interpretation is just a statement of facts and, apart from being cumbersome and somewhat incomplete, is not objectionable. However, after obtaining a result like this, one is tempted to go just a little farther and conclude that an increase in copper pollution of the air tends to diminish the frequency of deaths from D.

Radioactivity in the air, or in food or in water is frequently a suspected cause of premature death. Also, the levels of radioactive pollutants vary considerably from one locality to the next as, undoubtedly, do the levels of pollution with copper. *A priori* it seems possible that the dust of copper containing chemicals in the air is deleterious to health (contrary to conclusions suggested by the actual study) but that it is much less deleterious than radioactivity. Finally, it is possible that the localities with high air pollution levels of copper (possibly, localities with strip mining of copper ore) have relatively little pollution with radioactivity. In other words, among the localities studied in the particular investigation, there may have been a strong negative correlation between levels of radioactivity and of copper pollution. If this was so, then, on the average, the localities with plenty of copper in the air were also localities with little radioactivity and hence with relatively low mortality from D, even though the presence of copper in air caused some increases in deaths.

The reader will realize that all the above is purely hypothetical and that there is no intention to suggest that copper is deleterious or that the radioactivity is substantially more so. The purpose of the discussion is to indicate the danger of incomplete inclusiveness of pollutants in observational studies. Various considerations of convenience, and others, may (and do) suggest the formulation of the policy, such as:

> We are interested in health effects of pollutants $P_1, P_2, \cdots, P_s$ but, provisionally at least, not in others; therefore, even though we are aware of claims that some other pollutants $P_{s+1}, P_{s+2}, \cdots, P_n$ are deleterious to health, our own multipollutant and multilocality study shall be limited to $P_1, P_2, \cdots, P_s$.

This is a dangerous policy.

While practical considerations must impose limitations on the number of pollutants to be included in an investigation, it is important to be aware of what the omission of a particular pollutant may entail. In particular, the reader will have no difficulty in visualizing how the omission from the study of a particular pollutant, say P_{s+1}, may result in the appearance that the pollutant studied P_1 is deleterious to health while, in actual fact, it is beneficial.

5. *Pitfalls of "spurious correlations."* Spurious correlations have been ruining empirical statistical research from times immemorial. Apparently the first publicly discussed incident is recorded in three contributions published next to each other in 1897. The credit for identifying the noxious phenomenon belongs to Karl Pearson [2]. The victim whose spurious correlation mishap stimulated the discussion was W. F. R. Weldon [3]. The third contribution, intended to make Pearson's developments more clear intuitively, is due to Francis Galton [4].

Even when contrasted with "organic correlation," the term "spurious correlation" seems to be a misnomer. As used by Karl Pearson, the term refers to a very real and easily computable correlation, say R_1. What is spurious is the interpretation of R_1 as having something to do with another correlation, say R_2, termed "organic" which happens to be of primary interest but is not easy to compute. As rightly noted by Karl Pearson and fully understood by Galton, R_1 may have no relation to R_2.

Pearson's own awareness of difficulties connected with spurious correlations stemmed from studies of errors committed independently by several observers. Later he noticed similar difficulties in biology and economics. Weldon's studies were concerned with shrimp. In more modern times, spurious methods of studying correlations were involved in a great variety of empirical research: in astronomy, in farm economics, in biology, in the study of elasticity of demand, in the problems of drunkenness and crime, of railroad traffic, and of racial segregation. On occasion, they were used in arguments about public policy matters. This applies to the health-pollution literature, including some papers in this volume, which is the justification for the present somewhat long section of this article. In general terms referring to public policy matters, the situation is as follows.

Consider a not directly controllable phenomenon P as it develops in some units of observation U (perhaps different localities in a country in a given year, or over several consecutive years in the same locality). The phenomenon P manifests itself in some variable Y which is of public concern: the currently observed values of Y appear unacceptably high (or low). It is suspected that Y is somehow connected with another variable X which is subject to at least partial

control. (Y may be the number of deaths from a disease D in a specified section of the population, perhaps in a particular age group; X may stand for the level of a pollutant). A public measure, perhaps legislation, is contemplated to enforce a change in the values of X with the hope that this would result in desirable changes in Y, perhaps only on the average.

Authoritative information as to whether changes in X will cause changes in Y at least on the average, can be obtained through a well designed experiment. However, in the circumstances considered experiments are impossible and one is compelled to conduct an observational study. While no causal relations between X and Y can be expected from such investigation, it can reveal how the average values of Y in units of observation where X is large differ from those where X is small, which may be valuable information in deciding on the public measure contemplated.

The situation would be relatively simple if enough observational units could be found identical in all respects, except for values of X and Y. In real life such favorable conditions cannot be expected and one must be prepared to find that, generally, the observational units vary not only in values of X and Y but also in many other respects. In particular some variables $Z_1, Z_2, \cdots, Z_s$ come under consideration, the variation of which is likely to influence either X alone or Y alone or both. For example, in a health-pollution study with X and Y denoting, respectively, the level of a particular pollutant and Y the number of deaths from D, the several "nuisance variables" Z may be numbers of the exposed to risk from disease D who belong to particular racial and socioeconomic categories of the population. Again, some other nuisance Z may refer to pollutants other than that under study. Possibly Z_i may mean the average body-burden of strontium 90 and Z_j that of cesium 137 and the like. Obviously, the practical problem of estimating the changes in Y, that may result from the contemplated intentional changes in X, calls for the study of conditional regression, say $Y(x|z)$, of Y on X, with all the Z's maintaining some fixed values symbolized by the latter z. With a moderate number of the nuisance variables Z, and in favorable conditions of linearity of regression, etc., such an investigation is not very difficult. However, in actual life various complications occur that suggest looking for some shortcuts. One kind of shortcut may mean the replacement of several measures of qualitatively different radioactive pollutants by a single measure of all such pollutants combined. Another kind of shortcut may result from an effort to correct the number of deaths for the variation from one unit of observation to the next in the total number of exposed to risk and also in the numbers of those who belong to different racial and socioeconomic groups. Frequently, the temptation to take a shortcut is strong and one is involved in studying spurious correlations.

In the simplest case, for each unit U of observation one computes the supposedly corrected value, say V, of X and/or the supposedly corrected value W of Y which these variables X and Y would have had if all the nuisance Z did not vary from one unit of observation to the next, but maintained some typical

values. In effect V and W are certain functions f_1 and f_2 of the directly observed X, Y and Z,

$$V = f_1(X, Z), \qquad W = f_2(Y, Z). \tag{1}$$

Then the correlation between V and W is taken to represent that of X and Y, with the influence of the nuisance Z being eliminated.

Emphatically, the convincingness of the correcting functions f_1 and f_2 notwithstanding, THE CORRELATION OF V AND W NEED NOT BE INDICATIVE OF THE PARTIAL CORRELATION BETWEEN X AND Y WHEN THE NUISANCE VARIABLES Z ARE FIXED.

Some theoretical considerations relating to this problem will be found on pp. 143–154 of reference [5]. I am indebted to Robert Traxler for the following numerical example illustrating, with somewhat exaggerated precision, what may result from a "shortcut" in the case where, given the nuisance Z, the variables X and Y are strictly independent.

The example refers to the simplest case of spurious correlation studies, involving just one nuisance variable Z, which represents the number of individuals exposed to the risk of death from a disease D, and where only one of the variables studied, namely Y, is corrected for the variation of Z. Furthermore, the method of correcting is so convincing that it appears incredible that it may involve some pitfalls.

Traxler considers 54 localities with the necessary data given in Table 1. This table is divided into six panels, each panel referring to 9 localities, all characterized by the same number Z of exposed to risk measured in some convenient units, such as 10,000 or the like. In the first panel $Z = 5$, in the second $Z = 6$, etc., up to $Z = 10$ in the last panel.

While the value of Z in each panel is constant, there is a variation in X, the level of the pollutant studied. Measured in some units and from a conventional zero point, X has three different values in each of the six panels: 1, 2, 3 in the first, 2, 3, 4 in the second and so forth. This change in the values of X from one panel to the next reflects Traxler's idea that the increase in the number Z of the exposed to risk may mean an increase in the total population, with the more populated localities being more intensely polluted than those with smaller populations. Of course, this need not be always the case and the reader may find it interesting to investigate the situation in which the relation between X and Z is contrary to that of Traxler.

The detail of Table 1 which deserves particular attention is the arrangement of columns giving X and Y. In each panel, to each of the three different values of X there correspond three localities with varying numbers of deaths Y. The important point is that to all the three values of X there corresponds *the same* triplet of values of Y. In the first panel this triplet is 1070, 1100 and 1130 deaths from D. In the second panel, with somewhat more people at risk, the triplet of the numbers of deaths is 1270, 1300, and 1330, and so on.

TABLE I

SIX GROUPS OF LOCALITIES WITH POLLUTANT X, NO. OF DEATHS Y AND NO. AT RISK Z

Locality no.	No. at risk Z	Pollutant level X	No. of deaths Y	Locality no.	No. at risk Z	Pollutant level X	No. of deaths Y	Locality no.	No. at risk Z	Pollutant level X	No. of deaths Y
1			1070	19			1465	37			1860
2		1	1100	20		3	1500	38		5	1900
3			1130	21			1535	39			1940
4			1070	22			1465	40			1860
5	5	2	1100	23	7	4	1500	41	9	6	1900
6			1130	24			1535	42			1940
7			1070	25			1465	43			1860
8		3	1100	26		5	1500	44		7	1900
9			1130	27			1535	45			1940
10			1270	28			1665	46			2060
11		2	1300	29		4	1700	47		6	2100
12			1330	30			1735	48			2140
13			1270	31			1665	49			2060
14	6	3	1300	32	8	5	1700	50	10	7	2100
15			1330	33			1735	51			2140
16			1270	34			1665	52			2060
17		4	1300	35		6	1700	53		8	2100
18			1330	36			1735	54			2140

Table 1 produced by Traxler shows many facets of regularity which cannot be expected in any real study. Rather than have six sets of localities with exactly the same number Z of individuals at risk within each set, in a real study it would not be surprising to find that no two localities have the same value of Z. Here, then, there will be a real problem of dealing with the variation in Z and the method that the practicing statistician is likely to use is easy to perceive. The expected reasoning would be somewhat as follows. My primary objective is to find out the changes in the *frequency* of deaths from D that are to be expected if the level X of the pollutant is intentionally changed. Thus, the most direct way of studying the problem is by computing, for each locality, the death rate $W = Y/Z$, by classifying all the localities according to the level X of the pollutant and, finally, by computing the mean death rate $\overline{W}(X)$ that corresponds to any given value of X.

What such an analysis may lead to is illustrated in Table II. Table II is arranged so as to simplify the calculation of the $\overline{W}(X)$ somewhat. The simplification is based on the fact that all the 54 localities considered are divided into triplets, each triplet being characterized by a combination of values of X and Z. Thus, there is no point in calculating the death rate W separately for each locality. It is sufficient to compute the average rate per triplet, then to classify the triplets according to (X, Z) and to average over values of Z so as to obtain the desired $\overline{W}(X)$. Accordingly, the 8 lines of Table II correspond to the eight values of X and the six columns to the six values of Z. The last column gives the desired average death rates from D corresponding to increasing values of X.

TABLE II

ANALYSIS OF THE EFFECT OF POLLUTANT BASED ON DEATH RATES

Pollution level X	Mean death rates $\overline{W}(X, Z)$ in triplets of localities cross classified according to X and Z = no. at risk						Mean death rate $\overline{W}(X)$ in localities with pollution X
	$Z = 5$	6	7	8	9	10	
1	220.0						220.0
2	220.0	216.7					218.3
3	220.0	216.7	214.3				217.0
4		216.7	214.3	212.5			214.5
5			214.3	212.5	211.1		212.6
6				212.5	211.1	210.0	211.2
7					211.1	210.0	210.6
8						210.0	210.0

If any real study exhibited the correspondence between the average death rates $\overline{W}(X)$ and the level X of the pollutant studied, anywhere comparable to that in Table II (however, no comparable regularity can be expected!), the interpretation would be somewhat as follows:

(i) The pollutant studied *does* influence the death rates from D.

(ii) At least as far as the disease D is concerned, the pollutant studied is beneficial: if the level of the pollutant is low, say if it is lower than $X = 6$, then an increase in the level of the pollutant decreases noticeably the death rate from D.

(iii) For the above reasons, the adoption of a public measure should be considered to increase the level of the pollutant, perhaps by spraying the countryside, at least in those localities in which the current level of the pollutant is low.

In order to see what would be the result of adopting any such measure as suggested in (iii), we must return to Table I. Each of the six panels refers to localities with three different levels of the pollutant: 1, 2, 3, or 2, 3, 4, etc. It is seen that if, through spraying or otherwise, the two lower levels of the pollutant are replaced by the highest, the effect on the numbers of deaths would be exactly nothing. The numbers of deaths implied by the data would be unchanged. Thus, the contrary conclusion suggested by Table II is not inherent in the data. It is an artifact produced by dealing not with the triplets of values of (X, Y, Z) as given directly by the observations, but by values of X and $W = Y/Z$ computed for each locality. The anatomy of the phenomenon is interesting and the reader is urged to examine it both theoretically as in reference [5] and numerically. In particular, it may be interesting to see how Table II would be modified if the six values of Z in Table I ranged not from 5 to 10 but, say, from 4 to 9 or from 7 to 12, all other details of Table I remaining without change. While such changes in the range of Z may seem of little consequence, their effect on the appearance of Table II is likely to appear dramatic. Another question that may be interesting to answer is whether the range of values of Z can be so adjusted as to force Table II to yield an answer to the basic question which is at least approximately true.

Finally, the "anatomical" study of Tables I and II can help to meet the possible objection that, while Table I is unambiguous about the apparent effect of an increase in the pollution X in each of the six categories of localities by one or two units, it does not say anything about the possible effect of bringing X to its highest value $X = 8$, uniformly in all the 54 localities studied. Obviously, in order that Table I provides this kind of information, it must include data for more than 9 localities in each of its six panels, with the consequent increase in its complexity. However, it is likely to be interesting to consider how the objective could be attained.

Still another detail of Table II is worth noticing. This is that, while the computed $\overline{W}(X)$ decrease when X grows, the decrease is not linear.

The general principle that Traxler's example is intended to suggest may be heuristically formulated as follows. The object of the empirical study is to estimate the effect on a variable Y of an intentional change in the level of another variable X (or variables X). Here, the term "effect" is understood to refer to not any single unit of observation (locality) but to a population of such units. The information available for the study consists of values of not only X and Y but also of some s other variables $Z_1, Z_2, \cdots, Z_s$ which are suspected of

being somehow involved in the mechanism that connects X and Y. The safe method of studying the population effect on values of Y of an intentional change in X, while the values of the Z's are left to vary as they will, is through an investigation of the *joint* variability of all the $s + 2$ variables involved (X, Y, Z_1, Z_2, $\cdots$, Z_s). It is this simultaneous variation that characterizes the complex mechanism involved, of which we are interested in a single detail: what will happen to the values of Y (number of deaths) if the values of X are modified in a specified manner. Admittedly, the direct investigation of the variability of (X, Y, Z_1, Z_2, $\cdots$, Z_s) is cumbersome and the tendency to reduce the number of the nuisance variables Z is understandable. However, any such reduction is equivalent to the injection into the mechanism studied of some elements that are extraneous to it. Traxler's example illustrates the pernicious effect of replacing the triplet (X, Y, Z) by the pair (X, $W = Y/Z$) which looks very natural. If the reader investigates the suggested modifications in the range of Z, he will find that the effect of substituting the study of (X, W) for the study of the triplet (X, Y, Z) depends considerably on the properties of the joint distribution of the three variables, a detail of which is precisely the subject of investigation and *a priori* is unknown. The relationship between reality and the outcome of a real study in which one attempts, for example, to summarize in just one variable such directly observable quantities as, say, the body-burden of cesium-137, the body burden of strontium 90 and the radiation from walls of buildings, is a subject for speculation.

◇ ◇ ◇ ◇ ◇

I am grateful to Professor F. N. David for providing the references for Francis Galton and Karl Pearson.

REFERENCES

[1] C. L. Chiang, *Introduction to Stochastic Processes in Biostatistics*, New York, Wiley, 1968.

[2] Karl Pearson, "Mathematical contributions to the theory of evolution—On a form of spurious correlation which may arise when indices are used in the measurement of organs," *Proceedings, Royal Society of London*, Vol. 60 (1897), pp. 489–498.

[3] W. F. R. Weldon, "Note, January 13, 1897," *Proceedings, Royal Society of London*, Vol. 60 (1897), p. 498.

[4] Francis Galton, "Note on the memoir by Professor Karl Pearson, F.R.S., on spurious correlation," *Proceedings, Royal Society of London*, Vol. 60 (1897), pp. 498–502.

[5] J. Neyman, *Lectures and Conferences on Mathematical Statistics and Probability*, Graduate School, U.S. Department of Agriculture, Washington, D.C., 1952, pp. 143–154 (2nd ed.).