Environmental Health Perspectives
Vol. 32, pp. 241-254, 1979

Hazardous Substances, the Environment and Public Health: A Statistical Overview

by William G. Hunter* and John J. Crowley†

The purpose of this paper is to provide an overview of the statistical problems that exist and procedures that are available when attempts are made to assess the possible harm which has been or might be caused by substances in the environment. These issues bear directly on important decisions of public policy such as those related to the establishment and enforcement of regulations.

Some Complexities of the Problem

The central problem, which is illustrated in Figure 1, is to elucidate the relationship between human health and factors such as the food we eat, the air we breathe and our genetic make-up. This picture is deceptively simple. In fact, much careful, expensive and time-consuming detective work is necessary to unravel the complex mysteries of what factors or combination of factors cause what effects. The scientific goal is to develop a clear picture of this complicated reality.

Environment to most people means the air, water, and land around us, including animal and plant life and things we have made. In the public health field, environment tends to be defined as all external factors that act on an individual, which includes everything except genetics. The effect of exercise for example, is then classified as an environmental factor, as indicated in Table 1.

One major problem in trying to discover which environmental factors are harmful to human health is the great number of such factors. Some of them are listed in Table 1. There are two difficulties here: (1) It is virtually impossible to be sure that all the important factors are present in a list of this kind unless one goes to the extreme of simply listing all factors that exist, which makes the problem unmanageable given our current capabilities. (2) Data of sufficient quantity and quality do not exist on all factors of interest, especially when it is desired to estimate the effect of two or more factors acting together.

Besides the problem of the great number of factors, however, there are many others (see Table 2). Lurking variables may be present. A lurking variable is one that has an important effect and yet is not taken into account in the analysis because its existence is unknown or, if its existence is known, its influence is thought to be negligible or data on it are unavailable. Even if all the important factors are included in an analysis and data are available on all of

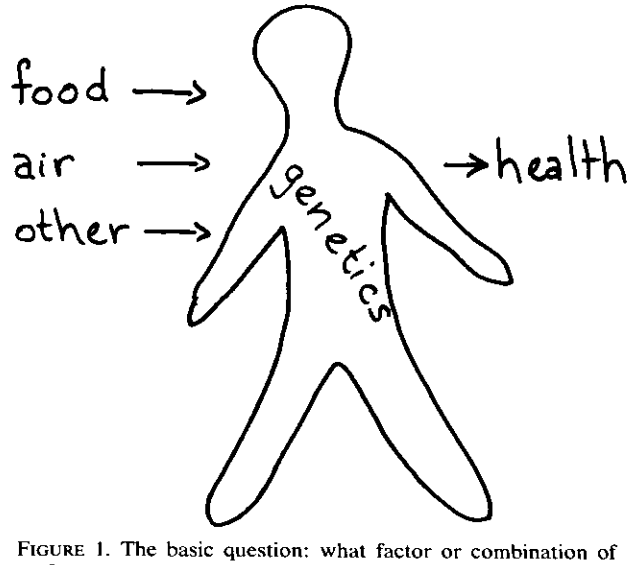

FIGURE 1. The basic question: what factor or combination of factors cause what effects on human health?

*Department of Statistics and Engineering Experiment Station, University of Wisconsin, Madison, Wisconsin 53706.
†Department of Statistics and Department of Human Oncology, University of Wisconsin, Madison, Wisconsin 53706.

Table 1. Some environmental factors that have an influence on public health.

"Food"	Air	Other
Asbestos	Cigarette smoke	Radiation
Nitrates	Carbon monoxide	Sunlight
Coloring	Hydrocarbons	Noise
Saccharin	Nitric oxides	Exercise
Pesticides	Photochemical oxidants	
Herbicides	Sulfur oxides	
Drugs	Particulates	
Arsenic	Asbestos	
Barium	Fluorides	
Cyanide	Acid mist	
Silver	Lead	
Selenium	Beryllium	
Lead	Mercury	
Mercury	Other trace elements	
Chromium		
Cadmium		
Other metals		

them, difficulties in interpretation can arise because of partial or complete confounding. Such problems are usually most severe in situations where the data have not been collected from an experimental design but rather where the data are historical or happenstance in nature, such as in epidemiological studies. For a further discussion of statistical aspects of lurking variables, confounding and other hazards of analyzing historical records, see Box, Hunter, and Hunter (*1*). Some of these issues are addressed for the particular case of beryllium by Wagoner, Infante, and Mancuso (*2*). The difficulties of interpretation when there are many lurking and confounding variables are illustrated by the situation in breast cancer, where there is the following constellation of correlated factors: height H, weight W, obesity (W/H^2, Quetelet's Index), and surface area ($H^{0.42}\,W^{0.51}$); also number of pregnancies, age at first pregnancy, lactation, early menarche, artificial menopause, and length of menopause (*3, 4, 27*). Considering also that many of these are related to nutrition, genetics, and socioeconomic status, the situation is difficult indeed.

Factors can interact with one another to produce a synergistic effect, that is, one that is more desirable or undesirable than would be expected on the basis of linearity and additivity from the results obtained with the factors individually. Such phenomena are sometimes called potentiation, promotion, or inhibition. For example, suppose there are two factors. In the absence of the second factor, suppose the first one at a certain level x_1 produces an effect y_1. In the absence of the first factor, suppose the second one at a certain level x_2 produces effect y_2. When they are both present in the amount $x_1 + x_2$, the combined effect may well be less than or greater than $y_1 + y_2$, and if this happens the factors are said to interact. Promoters are associated with results that are greater than $y_1 + y_2$ and inhibitors with results that are less than $y_1 + y_2$. Recent research, for example, has investigated the role of promoters and inhibitors for cancer. Andur (*5*) mentions some of these points about interactions, promoters, and inhibitors with reference to research on sulfur oxides. Speizer et al. (*6*) recount an incident in which high concentrations of sulfur dioxide (35 ppm) in a laboratory environment — filtered air was mixed with the chemical and breathed — produced dramatically less severe effects than lower concentrations (6 ppm) in a paper mill. It was hypothesized that sulfur dioxide became attached to particles present in the air in the mill thus permitting it to penetrate deeper into the tracheobronchial tree. Alternatively or in addition, because of interactions some of the sulfur dioxide in the mill may have been present in the form of acid droplets or, given the special circumstances that existed, may have been more readily converted into sulfuric acid. Nonlinearities were probably present as well. Fraumeni et al. (*7*) are the latest to report synergism between smoking and/or asbestos exposure in causing lung cancer. As a further example, recent labo-

Table 2. Some problems that complicate the interpretation of data on hazardous substances and public health and the establishment of appropriate regulatory mechanisms based on such interpretation.

No.	Problem
1	Many potentially important factors need to be assessed.
2	Lurking variables may be present.
3	Effects of factors may be partially or completely confounded, making the disentangling of individual effects difficult or impossible.
4	Relevant response functions may contain interactions and nonlinearities.
5	Some effects are only manifest after long latency periods.
6	Conducting experiments on humans is extremely complicated for ethical and scientific reasons.
7	Extrapolation of results from animals to humans is fraught with uncertainties.
8	Extrapolation of results from high to low dose is fraught with uncertainties.
9	Experiments at low dose are usually prohibitively expensive.
10	It is unclear whether thresholds exist, and, whether they do or not, it is also unclear what the ramifications should be as far as regulatory action is concerned.
11	Factors can act indirectly to produce effects on health.
12	The relationships between legal concepts such as endanger, harm, risk and equity, and scientific conclusions based on quantitative toxicological data are complex.
13	In establishing policy, potential benefit should be balanced against potential harm, and costs should be considered.
14	Money, time, and resources available for this work are limited.

ratory evidence indicates that saccharin, if anything, may be a promoter of carcinogenic activity but only a weak carcinogen (*8*). Nonlinearities in response relationships occur when, say, k times the basic level x_1 of a factor does not produce k times the basic effect y_1 of that factor.

Some factors only become apparent after long latency periods; for example, data suggest that it may take many years before the carcinogenic effect of exposure to asbestos manifests itself.

Conducting experiments on humans, of course, is extremely difficult because of ethical and scientific reasons (*9*). Thus tests are performed on animals. But at low levels of exposure corresponding to those encountered by humans (low-dose experiments), the resulting small effects are extremely hard to detect unless very large numbers of animals are used. Often this number is so big that it makes such testing prohibitively expensive. Consequently high doses are used, and results at relevant low-dose levels are obtained by extrapolation. As we shall discuss later, crucial decisions must be made about what models should be used and, given the model, what methods of inference should be employed. With a given set of data, wildly different answers are obtained depending primarily on what model is used. But, assuming that satisfactory answers exist regarding the extrapolation of animal experiments from high to low doses, the important question remains of what do the data from experiments on the effects of certain factors on the health of animals tell us about the possible effects of those factors on the health of humans. How best to make this extrapolation (or leap) from animals to humans is the subject, either directly or indirectly, of much current research work. References on extrapolation from high to low dose and extrapolation among species, have been compiled by Thomas and Blair (*10-25*).

The issue of thresholds is also controversial (*14*). But even if thresholds are shown to be present in isolated controlled experimental situations, there are those who argue that our bodies, which exist in a polluted environment, are "saturated" already and cannot tolerate any added toxic stress. That is, they argue that for toxic substances the controversy about thresholds is largely irrelevant (*26*).

Factors can act indirectly to produce effects on health; for example, there is the postulated link between presence of fluorocarbons and the depletion of stratospheric ozone and the higher incidence of skin cancer. As another example, consider the heating of a receiving water as a result of the operation of a power plant and its possible deleterious effect on plant and animal life there and, in combination with other actions of this kind, its possible ultimate effect on our well-being.

In the formulation of policy with regard to public health, the relationships between legal concepts such as endanger, harm, risk, and equity, and scientific conclusions based on quantitative toxicological data are not given due consideration, mainly because they are not well understood (*28-30*). In establishing policy, potential benefit should be balanced against potential harm, and costs should be considered (*31, 32*). A special issue discusses food additives, color additives, animal drugs, ritalin, and vinyl chloride; attention is given to the Toxic Substances Control Act and to problems of assessing environmental risk. An annotated source guide to information on toxic substances is given by Ross (*33*). The current situation with regard to the use of nitrites in meat illustrates the kind of considerations that must be weighed. Evidence suggests this additive is associated with cancer but prevents botulism, so reducing one of these risks increases the other (*34*). A similar situation exists with regard to drugs. In the Final Report of the Review Panel on New Drug Regulation (*35*), for example, it is stated that:

> "The legal requirement that new drugs be proven 'safe' and 'effective' is imprecise since no drug is absolutely safe or always effective. The statutory standard should be amended to reflect the fact that assessing the value and ultimate approvability of a new drug entails weighing its risks against its overall benefits."

The list in Table 2, of course, is not exhaustive. Another problem is that the responses (the development, say, of different cancers) are sometimes extremely difficult to determine and may be multiple in nature (different sites and cell types of cancer, for example). In addition, there may be statistical difficulties. Current methods of analyzing time-to-tumor data require the specification of tumors as either rapidly and uniformly fatal, so that deaths with tumor represent incidence data, or never fatal, so that such deaths represent prevalence data (*36, 37*), while the truth is usually somewhere in between. An approach to the intermediate situation has been given by Turnbull and Mitchell (*38*). Also needing attention is the role of competing causes of death (*39*).

Another complication in formulating sensible regulations is that high pollution levels measured in one location may actually originate hundreds of kilometers away. The Federal Standard of 0.08 ppm for ozone can be exceeded in places quite distant from the New York City metropolitan area because of pollution sources in that urban center; this phenomenon has also been observed in Los Angeles and elsewhere (*40*). There is also the difficulty of

properly taking into account data that are serially correlated in time (*41, 42*). The list in Table 2, however, does serve to illustrate some of the complications involved in trying to assess the influence on public health of important environmental factors and to implement policy to remedy the situation (*41, 43-46*).

Environmental Regulations, Scientific Data, and Assumptions

Figure 2 illustrates the way in which environmental regulations are related to scientific data and assumptions. Data are analyzed on the levels of exposure to certain potentially toxic or hazardous substances and the associated indicators of health. Based on these analyses, certain scientific conclusions are drawn regarding the possible harmful effect of selected substances. Making use of these conclusions and taking into account relevant economic, social, political, and technological factors, legislators establish environmental regulations governing the production, use, and disposal of these substances. The diagram in Figure 2 can be viewed as a structure resting on the twin foundations of data and assumptions. Data are generally preferable for this purpose because they offer a firmer basis on which to erect this regulatory edifice. Unfortunately, however, data are frequently unavailable and consequently, in order to proceed, it is necessary to make up for deficiencies in the data with assumptions. [Sometimes, even if data are available, they are of dubious value; for example, consider the experiments on the color additive FD&C Red No. 2 described by Boffey (*47*)].

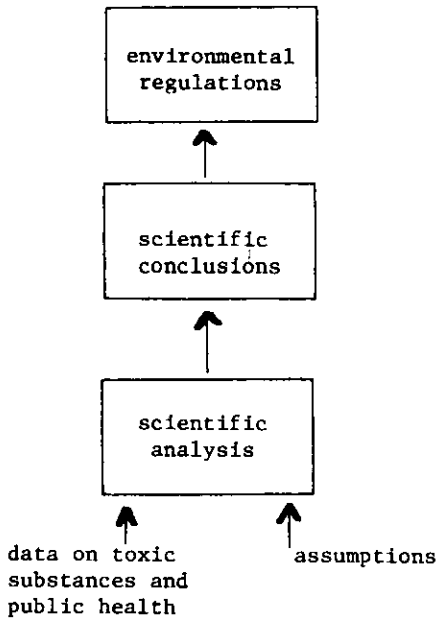

FIGURE 2. Regulatory structure and its twin scientific bases: data and assumptions.

Lawyers are involved in the establishment and enforcement of regulations and in disputes that sometimes arise in this context. In scientific studies, statisticians help plan the data collection process and analyze these data once they become available. We discuss this point in more detail later. Statisticians work with toxicologists, epidemiologists, and other investigators interested in data on both animals and humans.

Assumptions sometimes play an extremely important role. Obviously if the assumptions are incorrect, the regulations based on them may be inappropriate. In terms of Figure 2, if the assumptions are a major part of the foundation for the structure, then if they are shaky — or worse, flatly wrong — the structure may tilt, pointing therefore in some inappropriate direction, or it may even collapse. What are some of the most important assumptions that are injected into this process? The most important all involve extrapolations from the known to the unknown. We cite four examples.

First, there is the problem of extrapolating from high to low doses in animal experimentation. Assumptions are routinely made about which model to use (one-hit, two-hit, probit, and other empirical models); this involves making assumptions such as the existing or non-existence of thresholds and whether detoxification mechanisms work in the same manner at low doses as they do at high doses. Given a model, one must make assumptions to select the best statistical technique to use (Mantel-Bryan or some alternative).

Second, there is the problem of extrapolating from animals to humans. Assumptions play a very important role here, especially in determining on what basis this can best be done. By the Delaney Clause, any food additive that is shown to produce cancer in any animals or humans is assumed to be potentially harmful to humans and its use, therefore, must be banned. A tacit assumption here is that this risk, no matter how slight, automatically outweighs any benefits, no matter how great.

Third, there is the problem of extrapolating from controlled experiments to the real world. The main assumption made is that variables other than those studied have been properly taken into account. Controlled experiments, for example, have shown that fluorocarbons can react with ozone and that fluorocarbons are stable compounds. The assumption has been made that the chemical reactions ob-

served in the laboratory take place in the stratosphere, thus depleting the ozone layer there. As a consequence, regulations have been established in several countries limiting the use of fluorocarbons. The United States is one of these countries. In some countries, however, there are doubts about the basic assumptions necessary to reach these conclusions about ozone depletion. Consider one further example in which we have the opposite situation. Suppose controlled experiments (say, standard toxicology tests with mice) indicate that a certain substance produces no detectable ill effects. Does this mean this substance is safe? Before drawing that conclusion one must assume that promoters do not exist in the real world that act together with the tested substance (an interaction) to produce undesired effects. In a controlled experiment such a promoter may have been excluded so that the experimental observations show no effect of the substance even though in normal use its effect might be substantial.

Fourth, there is the problem of extrapolating from observational (for example, epidemiological) studies to conclusions about cause and effect. The main assumption made is that confounding variables have been taken into account in the analysis. Most of the effort involved in the interpretation of such data, which is extremely tricky business, is expended on trying to check on this assumption. Epidemiological studies often focus on the possible influence of a single factor on a particular disease. If the mechanism involves more than one factor, such a study might fail to yield reliable results. Consider the recent interest in the possible contribution of promoters in the development of cancer (*48*).

To recapitulate, there is a great deal of reliance on assumptions because, when data are lacking, assumptions are needed to take their place, and for many aspects of the general situation regarding the environment, hazardous substances, and public health, data are simply unavailable. A discussion of this point in a legal setting is given by Thomas (*29*), and the role of assumptions in modeling the dispersion of air pollutants is treated by Dinman (*49*).

When an assumption must be made, there is often a spectrum from which to choose and there is an understandable tendency to be conservative in this regard. Figure 3, in a nutshell, shows why this is so. Of the two mistakes that can be made — regulating a substance that is not hazardous and not regulating one that is — the more serious one is generally felt to be the latter.

With regard to cancer, heart, and lung disease, the First Annual Report to Congress by the Task Force on Environmental Cancer and Heart and Lung Disease (*50*) has a description of the government's progress and plans with regard to quantifying relationships between environmental pollution and these three diseases, developing strategies to reduce or eliminate the risks associated with these pollutants, and planning research to shed light on these problems. This report contains the conclusions that "[t]here is evidence that risk and occurrence of cancer, heart and lung disease increase with environmental pollution, broadly defined to include all environmental factors . . ." and that "[c]urrent preventive measures are believed to be inadequate to obtain desired reductions of risk and occurrence." It also states that increased knowledge of the pollution-disease relations are needed if intelligent policies are to be established. Abelson (*51*) and Hills (*52*) give some observations on the impact of regulations upon industry and public health.

		Is this substance hazardous? no	Is this substance hazardous? yes
Shall we ban or otherwise regulate this substance?	no	bravo!	mistake
	yes	mistake	bravo!

FIGURE 3. Two mistakes that can be made in deciding whether to regulate a particular substance.

Role of Statistics in Toxicological Studies

In science one often begins with a certain hypothesis (idea, conjecture, theory, or model) which strongly influences the choice of what data to collect. This hypothesis is then compared with the data. Assuming the data are reliable, if these two fail to match, the initial hypothesis is modified. An improved hypothesis therefore frequently points to the desirability of collecting new data, and so on. This sequence may therefore be repeated many times. Scientific work is characterized by this iterative pattern in which learning takes place gradually by a process of trial-and-error.

Toxicological studies typically consist of four stages: (1) an hypothesis about, say, the possible carcinogenicity of a particular substance; (2) the plan for what data are to be collected and how that is to be done; (3) the collection of the data, and (4) the analysis of that data. Statistics is a science that is particularly concerned with stages (2) through (4). The object of analysis is to attempt to extract all the useful information from a body of data. In planning

and monitoring the collection of data, one tries to ensure that the data will shed light on the questions at issue as brightly and inexpensively as possible. One important role of statistics, then, is to help scientists and others get reliable answers as economically as possible.

Analysis of data is dynamic, its aim being not only to reach answers to questions already posed but also to raise new questions, which can lead to new hypotheses that can provide better understanding of how our health depends on factors in the environment. Thus another important role of the statistician is to help conduct analyses of data in such a way as to catalyze the creation of new hypotheses.

Since analysis of data is an attempt to extract all the useful information they contain, it obviously requires not just statistical expertise. For example, Notices of Claimed Investigational Exemption for a New Drug and New Drug Applications received by the Office of New Drug Evaluation in the Federal Drug Administration are supposed to be studied by review teams consisting of medical officers, pharmocologists, chemists, and others.

Since raw data are never perfect, conclusions based on them necessarily contain uncertainty. Contributions of the statistician include helping to quantify uncertainties in such conclusions and to point out, where appropriate, other possible ways in which conclusions might be wrong.

Which statistical technique is the most appropriate to employ depends on the situation. One way to categorize situations is on the basis of the types of questions that are being asked. It may be that the question is: *Which* substances should be tested more thoroughly? Here the investigators may want to use efficient screening designs that will allow them to gather a modest amount of data on a great number of substances, that is, they will be performing screening tests. It may be, however, that a substance has been chosen for testing and the question is: *How* does this substance affect the responses of interest? The investigators here might be interested in developing an empirical model that will describe the dose-response relationship. Alternatively, it may be that the investigators want to probe deeper and the question might be: *What* can we learn about the basic mechanisms that operate in this situation? They would then be trying to understand *why* they observe certain effects. In some form they would be trying to develop a mechanistic model.

As illustrated in Table 3, many different types of investigations can be usefully viewed in terms of the progression of these three questions: Which? How? and Why? Cleveland et al. (*40*) and Phadke et al. (*53*), for instance, provide examples of environmental studies referred to in line 4 of Table 3. The better we are able to answer the question in the bottom right-hand corner of this table, the better we are able

Table 3. Some questions that arise in different types of studies.

Type of study	Screening questions	Empirical questions	Mechanistic questions
(1) Controlled experiments to measure effects of variables	Which variables are important?	How do these variables affect the responses?	Why do we observe these effects?
(2) Animal testing	Which substances are worth testing?	How do these substances affect animals?	Why do we observe these effects?
(3) Human testing	Which substances are worth testing?	How do these substances affect humans?	Why do we observe these effects?
(4) Environmental studies	Which variables should be considered?	How are these variables related to one another?	Why do we observe these inter-relationships?
(5) Epidemiological studies	Which variables should be considered?	How are these variables related to one another?	Why do we observe these inter-relationships?
(6) Comprehensive environmental studies, including animals and humans	Which variables and substances are important?	How do these variables and substances affect the environment?	Why do we observe these effects?

to develop effective policies for the preservation and improvement of our environment. Ultimately, therefore, scientists are striving to learn about the underlying mechanisms that produce disease. This work is often aided by trying to build models that will adequately represent phenomena of interest.

Figure 4 illustrates how this is done. The greater the quantity of reliable data that are available, the better the model-building process can go forward (box a). To create a tentative model, one might require a knowledge of biology, chemistry, mathematics, physics, or engineering (box b). Upon confronting the proposed model with the data, logically the first question one wants to answer is whether there is any evidence that the model is inadequate (box 1). If so, one must consider in what way(s) it is inadequate and return to the job of repairing it, or perhaps scrapping it altogether and building an entirely new one (feedback loop I). Also, if the model is inadequate, one may want to collect more data in order to gain more information about how best to modify the model (feedback loop II). If the model is found to be adequate, one then wants to obtain the best estimates of the parameters (constants) in that model (box 2). Then one should attempt to assess the precision with which these estimates have been obtained (box 3). If the precision is not high enough, one can return to the field or laboratory to obtain additional data, perhaps after improving the analytical equipment to permit more accurate data to be collected (feedback loop III). Note that, although logically question 1 precedes 2, in practice one must tentatively assume that the model is adequate and answer question 2 first and then return to consider question 1, by examining the residuals from the fitted model. (This way of looking at model-building was suggested to us by William Lawton of Eastman Kodak.)

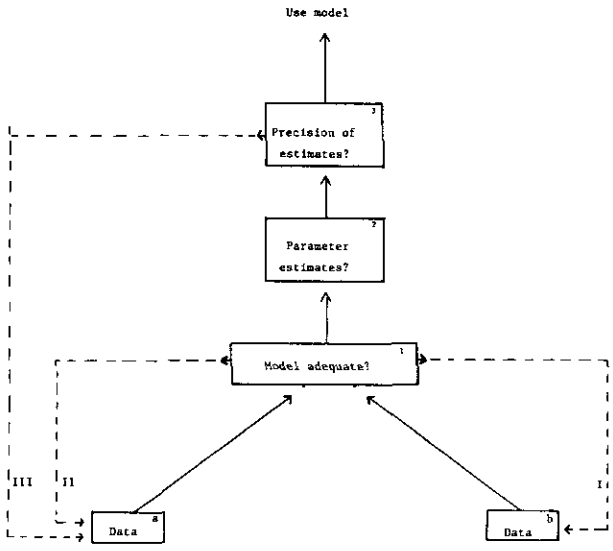

FIGURE 4. Schematic diagram of model-building process.

Viewed in this way, model building requires scientific, technological, and statistical expertise. Roughly speaking, it is required to have technology to collect the data (box a), science to construct the model (box b), and statistics to assess how well the model fits and to estimate the parameters in the model (questions 1, 2, and 3). Model-building projects, therefore, are typically multidisciplinary in nature.

Models are useful to the extent they help us answer questions such as the following: *Which* factors adversely affect health? *How*, empirically, do these factors affect health? *Why*, mechanistically, are these effects observed?

Next in this paper we will consider three sets of illustrations: the first concerning a mechanistic model-building project for some data on animals, the second concerning empirical model building for some hypothetical data on animals, and the third concerning the fundamental problem that plagues epidemiological studies, which, of course, involve data on humans.

Use of Mechanistic Models

Sauerhoff et al. (*54*) studied the dose-dependent pharmacokinetic profile of 2,4,5-tricholorophenoxy acetic acid (2,4,5-T), a plant growth regulator and herbicide, following intravenous administration to rats (Fig. 5). Concluding that the distribution and elimination of this compound are substantially different for low and high doses, they state:

> "There are direct toxicological implications for the dose-dependent elimination of 2, 4, 5-T from plasma as well as from the body. Systemic toxicity of a drug or foreign compound is often a function of the concentration and duration of that drug in plasma. If the drug or foreign compound is eliminated at a slower rate from the plasma, and/or different metabolites are formed at a high dose, then the ability of the drug and/or metabolite to induce toxicity at the high dose will be greater. Therefore, more severe toxicological manifestations than would have been predicted will often occur at the high dose level where nonlinear kinetics are operative. . . ."
>
> "The pharmacokinetic data presented in this report indicate that the statistical projection of results of experiments with large doses of 2,4,5-T to predict the hazard of exposure to small amounts is not justified because the capability of the body to handle the compound has been altered."

In their analysis these researchers used the Michaelis-Menton equation

$$-dC/dt = k_1C/(k_2 + C)$$

where C is the concentration of 2,4,5-T and t is the

time elapsed after administration of 2,4,5-T, and k_1 and k_2 are constants whose values are estimated from the data. This equation has some basis in theory and can therefore be regarded as different from purely empirical equations, such as straight lines and polynomials, which make no claim to explain what is happening on a mechanistic level. Accordingly, one might refer to the Michaelis-Menton equation as a mechanistic response function. Actually there is an entire spectrum stretching from the purely empirical at the one extreme to the purely mechanistic at the other, and it is doubtful that any model can be placed at either extreme. All models probably have at least some elements of both the empirical and mechanistic. Therefore the labels "empirical" and "mechanistic" merely indicate that a model is closer to one end of the spectrum than the other. A model can be termed "empirical-mechanistic" if it is somewhere in the middle of the spectrum. An example of such a model, in which Los Angeles air pollution data are analyzed, is given by Phadke et al. (*53*).

The following discussion of mechanistic models is adapted from Box, Hunter, and Hunter (*1*):

> A mechanistic model can contribute to scientific understanding, provide a basis for extrapolation, and provide a representation of the response function that is more parsimonious than one attainable empirically. For example, we might obtain a fair approximation locally to a response surface by fitting, say, a second-degree polynomial.
>
> If, however, a mechanistic model, believed to be supported by the basic biology of the system, could be verified, we would be in a much stronger position than would be attainable by mere empiricism. This is so because a well-tested mechanistic model does more than just graduate the data. It confirms that our scientific understanding of the system has been verified by the experiment. In addition, a polynomial equation, although it may be adequate to represent what is happening in the immediate region of study, provides only a very shaky basis for extrapolation. A mechanistic model, on the other hand, can suggest with greater certainty new sets of experimental conditions that are worthy of investigation. This better basis for extrapolation is provided because it is the mechanism, not a mere empirical curve, that is being supposed to apply more widely, and this mechanism is based on a partially verified understanding of the system itself. Of course, as we move in the space of the experimental variables, the mechanism may change or estimation errors may become serious, so unchecked extrapolation is never safe. Thus even a mechanistic model should preferably be used only to suggest regions where further experimentation might be fruitful.
>
> If the mechanistic model is well founded, we can expect it to give a closer representation of the response over a wider region than is possible with a purely empirical function. Estimation of the response will then be better, because lack of fit will tend to be less. In addition, since a mechanistic model is usually more parsimonious in the use of parameters, this causes less of the random error to be transmitted to the predicted values of the response.

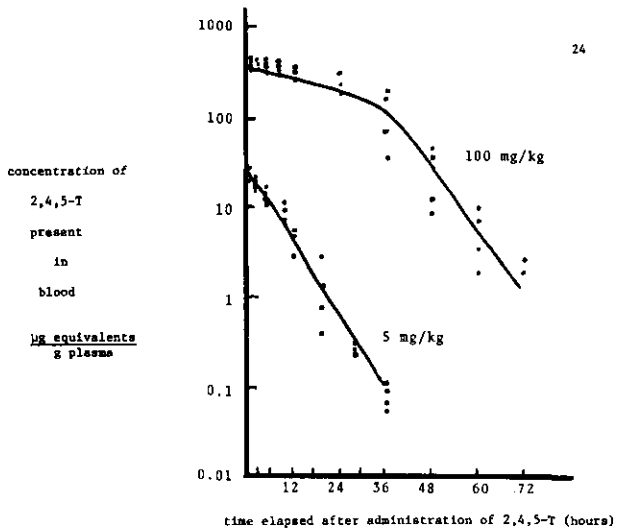

FIGURE 5. Data showing the amount of 2,4,5-T present in rats after single intravenous doses of 5 and 100 mg/kg. Data of Sauerhoff et al. (*54*).

Use of Empirical Models

Some hypothetical dose-response data are presented in Table 4. The doses, for example, might be the amounts of some chemical that have been administered to some mice after they have been divided randomly into three groups of equal size. The response might be the proportion of the mice that develop a certain type of cancer. For a given group of mice, let d represent the dose administered and p the proportion that respond, that is, develop cancer. Plotting the three data points on an ordinary piece of graph paper, as is done in Figure 6*a*, shows that they

Table 4. Hypothetical dose-response data.

Dose d	Response p
5	0.09
10	0.30
20	0.62

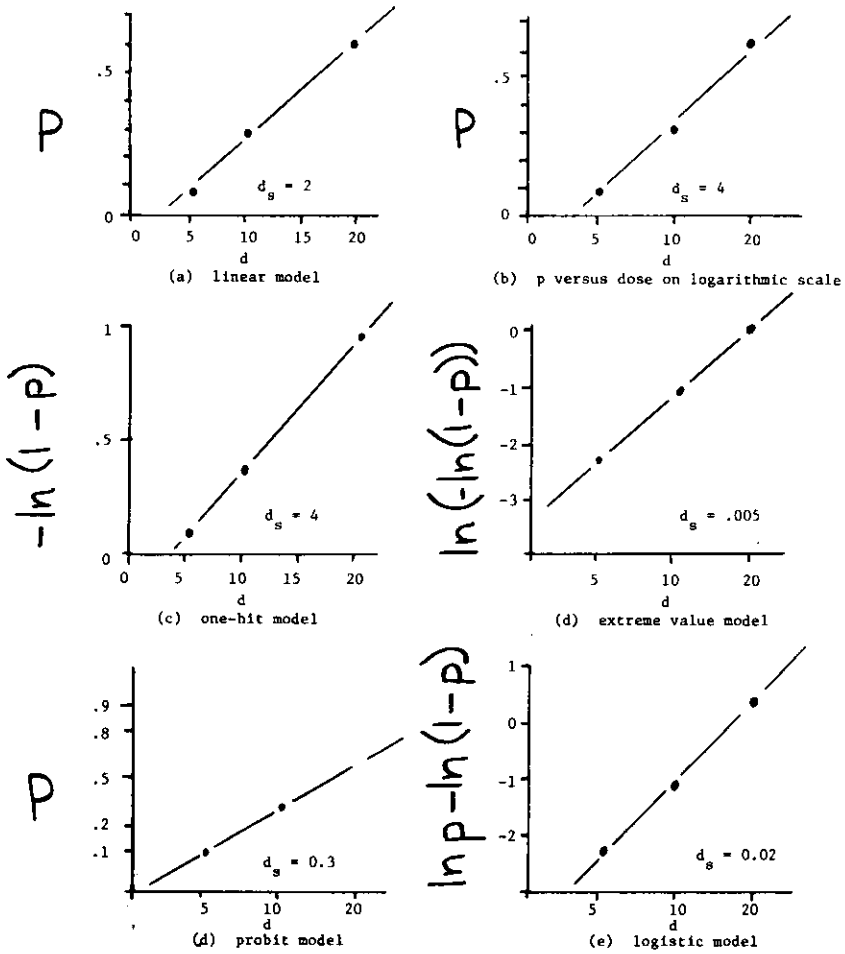

FIGURE 6. Hypothetical data from Table 4, where p = proportion of animals with cancer and d = dose administered, plotted six different ways. Values of "safe" dose d_s for the predicted value of p equal to 10^{-6} are indicated.

fall reasonably close to a straight line. The line shown in Figure 6*a* is the least-squares line. Extrapolating this line to the point where the predicted value of p is zero, one obtains a dose of $d = 2$. This might be regarded as a "safe" dose because, on the basis of the fitted straight line, it is the dose level that yields the value zero for the proportion of animals responding. There are, however, two major sources of uncertainty in this answer. The first is that, even if the true dose-response relationship were a straight line in these metrics, the three observed values of p are subject to observational or experimental error (that is, the observed values of p may differ from their true values). Since our predicted value $d_s = 2$ for the dose is derived from the fitted straight line which is derived, in turn, from the original data, uncertainty in the original data produces uncertainty in the predicted dose value. The second is that the true dose-response relationship may not be a straight line in these linear metrics.

Considering different metrics in which to fit a straight line produces dramatically different values for the "safe" dose, as is illustrated in Figure 6. In Figure 6*b* the dose d is plotted on a logarithmic scale. Fitting a straight line by the method of least squares and extrapolating to the point where the predicted value of p is zero gives a "safe" dose of 4, twice the value we obtained previously. In Figure 6*c* the response scale is modified but the dose scale remains linear, unchanged from Figure 6*a*. The quantity $-\ln(1-p)$ is plotted against d, and the "safe" dose is again 4. In Figure 6*d*, $\ln[-\ln(1-p)]$ is plotted against d on a logarithmic scale. [In these metrics it is impossible to locate the point where $p = 0$ because for this value $\ln[-\ln(1-p)]$ is infinite. What is customarily done in such situations is to define as

"safe" some sufficiently small value of p, say, 10^{-6} or 10^{-8}. We will proceed using the value of 10^{-6} and demonstrate that, in some circumstances, values for the "safe" dose are many orders of magnitude different from the original value of 2. Notice that the differences would be even more extreme if we had chosen the value 10^{-8}.] On this basis, then, in Figure 6*d* the "safe" dose is found to be 0.005. The dose is once again plotted on a logarithmic scale in Figures 6*e* and 6*f*; the response is plotted on a normal probability or probit scale in the first and on a $\ln p$-$\ln(1-p)$ scale in the second. These plots give "safe" doses of 0.3 and 0.02, respectively. These different ways of plotting the data that lead to such wildly discrepant estimates of the "safe" dose are based on several empirical models, as indicated in Figure 6. We note in passing that had we used 10^{-8} for the "safe" value of p, the corresponding "safe" dose levels would have been $d_s = 2, 4, 4, 0.0003, 0.1$, and 0.002 in the six cases. Note also that the curve fitting we have done is for illustration only, and that the techniques used make different assumptions about the existence of either a threshold or a background (zero dose) effect. We are not recommending this method in practice.

Discussions of some of the different dose-response models and statistical methods used in estimating parameters have been given by Mantel and Bryan (*22*), Mantel et al. (*21*), Hartley and Sielken (*20*), and Rai (*56*). Note that the plots shown in figures 6*a*, 6*c*, 6*d*, 6*e*, and 6*f* are associated with models with the following names: linear, one-hit, extreme value, pro bit and logistic, respectively.

Fundamental Problem in Epidemiological Studies

Figure 7 shows some real data on two factors, X and Y. If these data came from an epidemiological study where X was the level of some factor (for example, exposure to a certain chemical) and Y was the measure of incidence of some disease (for example, a certain type of cancer), there might be considerable interest in them. Perhaps the question would be raised about the possible desirability of promulgating regulations that would result in decreasing the levels of X in the environment. The implicit assumption in raising such a question would be that the data show a cause-and-effect relationship between X and Y.

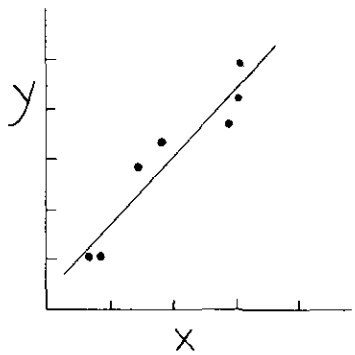

FIGURE 7. Observational data on two factors, X and Y.

But the data in Figure 7, which represent the number of storks X observed and the corresponding population Y in Oldenburg, certainly do not represent a cause-and-effect relationship. In interpreting epidemiological and other historical data of this kind, one must always bear in mind the possibility than an observed correlation between two quantities, no matter how striking (or "significant"), may not be the result of a cause-and-effect linkage between them but may be merely a correlation induced by the action of one or more lurking variables. This is the main difficulty in trying to draw valid conclusions from epidemiological studies and others like them that use happenstance data as raw material instead of data from carefully designed and controlled experiments, especially in which randomization has been employed.

One major distinction between the social sciences on the one hand and the physical and biological sciences on the other, for example, relates to the ability to conduct controlled experiments in which randomization is used. In the social sciences it is less often possible to perform such experiments, one notable exception being the study on negative income tax. Usually social scientists must rely on happenstance data. As a consequence it is typically a much more difficult task for them to correctly state their conclusions in terms of cause and effect. The crux of the matter is that correlation observed in happenstance data does not necessarily imply causation. The phenomenon exhibited in Figure 7 is sometimes called "nonsense correlation." Actually, there is nothing wrong or nonsensical about this correlation; the difficulty arises when one incorrectly imputes causation in circumstances of this kind. Therefore, perhaps a better name is "nonsense causation." But, by whatever name, examples of this phenomenon abound.

Suppose that a certain association or correlation is detected in a collection of historical data that suggests, for example, the possibility that a certain factor causes increased incidence of some disease. If sound research elucidates the basic mechanism by which this factor might be linked to the observed increased incidence, one would be more confident in asserting a casual link than if no such knowledge of

the underlying mechanism were available. Even if a plausible theory is available, one might still make mistakes in interpretation because of the presence of interactions, experimental errors, or lurking variables that have not been appropriately taken into account.

The major types of epidemiologic studies brought to bear on these problems (indirect, case-control, and occupational) all share these potential defects. Indirect studies are correlation studies of traits of geographic units, or studies of the same units over time (for example, colon cancer rates by country correlated with an estimate of the per capita consumption of animal fat). Problems peculiar to this type of study center on the use of population figures for individual exposures and the difficulty of obtaining adequate population measurements for potential confounders. Also complicating matters are latent periods (is the colon cancer rate to be correlated with current or past fat consumption?) and migration. These points are spelled out in detail by Breslow and Enstrom (*57*).

Case-control studies are retrospective comparisons of the characteristics of individuals with a given disease to a group of controls, often individually matched to reduce the effect of confounding variables. The most difficult design problem with this type of study is in the selection of cases and controls, and in obtaining comparable information on the two groups. For example, hospital-based studies may detect associations related to hospitalization rates and not the factor of interest. This general phenomenon is known as Berkson's fallacy and is the point at issue in the debated association between endometrial cancer and exogenous estrogens (*58, 70*).

Reliance on hospital charts in such studies may introduce biases because of selective questioning by the examining health care providers or selective abstracting by study personnel. Ziel and Finkle (*59*) present a strategy to avoid selective abstracting. Interviews and mail questionnaires are also subject to selective recall and nonresponse.

For these reasons one looks for replication of case-control studies, with different designs, and a synthesis of past results, before relying on the discovered associations or non-associations. For example, with the publication of the study by Williams et al. (*60*), there are now five studies purporting to show an association between reserpine and breast cancer, and four claiming no association. Williams et al. discuss the previous literature, give possible reasons for the discrepancies, and note that the nine studies are consistent in the direction (though not the statistical significance) of their results for one subgroup.

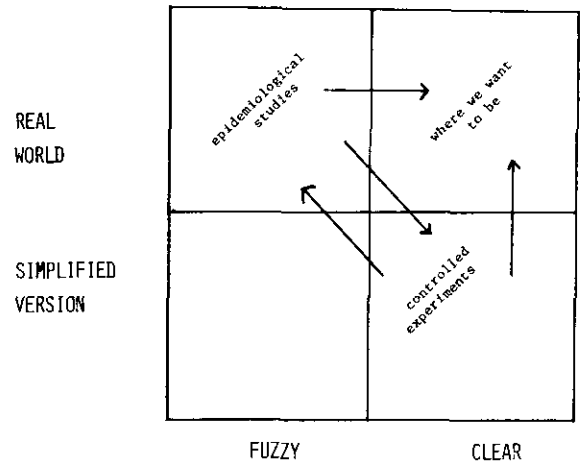

FIGURE 8. Controlled experiments and epidemiological studies.

Occupational studies are characterized by a description of the morbidity or mortality experience over time of defined groups of workers, often by relying on company or union records and death certificates. The workers' experience is then compared to that expected from the population at large, or internal comparisons are made (by job class, estimated exposure, etc). An example of current interest is the study of the effects of low-level radiation on the workers at the Hanford atomic plant (*61, 62*). Accurate followup and classification of cause of mortality are of course problems, as is the fact that workers are expected to be healthier than the rest of the population (the "healthy worker syndrome"). Further difficulties are associated with finding the proper way to measure and analyze exposure. Assessment of the effect of radiation in the Hanford study, for example, must take into account not cumulative exposure (which may be misleading) but exposure history (rate, pattern, duration, etc), as well as latency periods, age of the workers, and calendar period of exposure. Further discussion of these points is found in Breslow (*71*) and Pasternack and Shore (*63*).

For a more extensive discussion of using results from epidemiological studies to make inferences about cause and effect, see Lave and Seskin (*64*), Whittenberger (*65*), Tukey (*66*), Speizer et al. (*6*), Bloomfield (*42*), and Weiss (*67*) and the references listed therein.

Conclusion

We are not saying that controlled experiments are better than epidemological studies in all respects and that they should replace epidemiological studies. It is clear that we need both kinds of investigations, as is summarized in Figure 8. Epidemiological studies

may yield somewhat fuzzy pictures but at least they are pictures of the real world as it actually exists. Controlled experiments, while they may produce clearer pictures, do so only for simplified versions of the complex reality which is the world in which we live. Epidemiology presents us with landscape paintings and controlled experimentation gives us close-up photographs. What we want, of course, is a clear, detailed picture of the real world. Useful inferences about the environment and public health must be based on the analysis of both types of investigations and the interchange of insights thus provided (*65*).

We are definitely not advocating the attitude seen recently on a T-shirt in Madison: "I have given up my search for truth. I am now looking for a good fantasy." We are saying that the problems facing us are complex indeed, calling for multi-agency efforts as reflected in the saccharin question (*68*), and for multidisciplinary conferences such as this one.

As was mentioned in the Fourth Symposium on Statistics and the Environment, the essential nature of the problem was aptly summarized some years ago by Rachel Carson (*69*):

> "When one is concerned with the mysterious and wonderful functioning of the human body, cause and effect are seldom simple and easily demonstrated relationships. They may be widely separated both in space and time. To discover the agent of disease and death depends on a patient piecing together of many seemingly distinct and unrelated facts developed through a vast amount of research in widely separated fields."

REFERENCES

1. Box, G. E. P., Hunter, W. G., Hunter, J. S. Statistics for Experimenters: an introduction to design, data analysis, and model building. Wiley, New York, 1978, pp. 487, 498.
2. Wagoner, J. K., Infante, P. F., Mancuso, T. Beryllium: carcinogenicity studies. Science 201: 298 (1978).
3. MacMahon, B., Cole, P., Lin, T. M., Lowe, C. R., Mirra, A. P., Ravnihar, B. Salber, E. J. Valaoras, V. G., and Yuasa, S. Age at first birth and breast cancer risk. Bull. WHO 43: 209 (1970).
4. MacMahon, B., and Feinleib, M. Breast cancer in relation to nursing and menopausal history. J. Nat. Cancer Inst. 24: 733 (1960).
5. Andur, M. O. Toxicological guidelines for research on sulfur oxides and particulates. In: (Statistics and the Environment. Proceedings of the Fourth Symposium, March 3-5, 1976), American Statistical Association, Washington, D.C. 1977, pp. 48-55.
6. Speizer, F. E., Bishop, Y., and Ferris, B. G., Jr. An epidemiologic approach to the study of the health effects of air pollution. In: Statistics and the Environment. (Proceedings of the Fourth Symposium. March 3-5, 1976), American Statistical Association, Washington, D.C. 1977.
7. Fraumeni, J. F., Jr., Blot, W. J., Harrington, J. M., Toledo, A., Hoover, R., and Heath, C. W., Jr. Lung cancer after employment in shipyards during World War II. New Engl. J. Med. 299: 620 (September 21, 1978).
8. Hicks, R. M., and Chowaniec, J. The importance of synergy between weak carcinogens in the induction of bladder cancer in experimental animals. Cancer Res. 37: 2943 (1977).
9. Miké, V., and Good, R. A., eds. Medical research: statitics and ethics. Birnbaum Memorial Symposium. Science 198: 677 (1977).
10. Armitage, P., and Doll, R. Stochastic models for carcinogenesis. In: Proceedings of Fourth Berkeley Symposium on Mathematics and Statistics Probability. Vol. 4, Univ. California Press, Berkeley, 1961, pp. 19-38.
11. Bair, W. J., and Thomas, J. M. Prediction of the health effects of inhaled transuranium elements from a mathematical analysis of experimental animal data. Proc. IAEA Symposium on Transuranium Nuclides in the Environment, IAWA-SM-199/58, San Francisco, California, 1976, p. 569-585.
12. Chambers, E. A., and Cox, D. R. A comparison of models for determining safe levels of environmental agents. In: Reliability and Biometry: Statistical Analysis of Lifelength, F. Proschan, and R. J. Serfling, Eds., Society for Industrial and Applied Mathematics, Philadelphia, 1967, pp. 681-700.
13. Cochran, W. G. Problems arising in the analysis of a series of similar experiments. J. Roy. Statist. Soc. 4: 102 (1937).
14. Cornfield, J. Carcinogenic risk assessment. Science 198: 693 (1977).
15. Food and Drug Administration Advisory Committee on Protocols for Safety Evaluation, Panel on carcinogenesis. Report on cancer testing in the safety evaluation of food additives and pesticides. Toxicol. Appl. Pharmacol. 20: 419 (1971).
16. Finney, D. J. Probit Analysis, 3rd Ed., Cambridge University Press, Cambridge, England, 1971.
17. Gehan, E. A. Estimating survival function from the life table. J. Chron. Dis. 21: 629 (1969).
18. Guess, H. A., and Crump, K. S. Low-dose-extrapolation of data from animal carcinogenicity experiments — analysis of a new Statistical technique. Math. Biosc. 32: 15 (1976).
19. Guess, H. A., and Crump, K. S. Can we use animal data to estimate "safe" doses for chemical carcinogens? Proceedings of the Conference on Environmental Health, Alta, Utah, 1976, pp. 13-28.
20. Hartley, H. O., and Sielken, R. L. Jr. Estimation of "safe doses" in carcinogenic experiments. Biometrics 33: 1 (1977).
21. Mantel, N., Bohidar, N. R., Brown, C. C., Ciminera, J. L., and Tukey, J. L. An improved Mantel-Bryan procedure for "safety" testing of carcinogens. Cancer Res. 35: 865.
22. Mantel, N., and Bryan, W. R. "Safety" testing of carcinogenic agents. J. Nat. Cancer Inst. 27: 455 (1961).
23. Mayneord, W. V., and Clarke, R. A mathematical investigation into the carcinogenic risks associated with particulate sources of activity. Central Electricity Generating Board, Report No RD-B-N2872, London, 1974.
24. Mays, C. W. Estimated risk from ^{239}Pu to human bone, liver and lung. In: Biological and Environmental Effects of Low-Level Radiation, IAEA, Vienna, 1976, p. 373.
25. Morse, P. M. The combination of estimates from similar experiments, allowing for interexperiment variation. J. Am. Statist. Assoc. 62: 241 (1967).
26. Crump, K. S., Hoel, D. G., Langley, C. H., and Peto, R. Fundamental carcinogenic processes and their implication for low dose risk assessment. Cancer Res. 36: 2973 (1976).
27. Gehan, E. A., and George, S. L. Estimation of human body surface area from height and weight. Cancer Chemotherap. Repts. 54: 225 (1970).
28. Thomas, W. A., Ed. Scientists in the Legal System: Tolerated Meddlers or Essential Contributors? Ann Arbor Science Publishers, Ann Arbor, Mich., 1974.
29. Thomas, W. A. Judicial treatment of scientific uncertainty in the reserve mining case. In: Statistics and the Environment

(Proceedings of the Fourth Symposium. March 3-5, 1976), American Statistical Association, Washington, D.C. 1977.
30. Thomas, W. A. Equitable considerations in balancing energy equation. Environ. Health Perspect. 32: 253 (1979).
31. Portney, P. R., Ed. Current Issues in U.S. Environmental Policy. Johns Hopkins Univ. Press, Baltimore, 1978.
32. Ecology Law Quarterly. Hazardous Substances in the Environment. Law and Policy. 7 (2): (1978).
33. Ross, S. S. Toxic Substances Sourcebook. Environment Information Center, 1978; also Wrestling with TSCA? An annotated source guide for toxic substances. New Engineer, August/September, pp. 40-49, 1978.
34. Smith, R. J. Ever so cautiously, the FDA moves toward a ban on nitrites. Science 201: 887 and (1978).
35. Department of Health, Education and Welfare. Review Panel on New Drug Regulation, Final Report, May 1977. (For copies contact the Executive Secretarial (HF-1), Food and Drug Administration, Room 14-85, Parklawn Building, 5600 Fishers Lane, Rockville MD 20857).
36. Hoel, D. G., and Walburg, H. E. The analysis of survival experiments. J. Nat. Cancer Inst. 49: 361 (1972).
37. Peto, R. Guidelines on the analysis of tumor rates and death rates in experimental animals. Brit. J. Cancer 29: 101 (1974).
38. Turnbull, B. W., and Mitchell, T. J. Exploratory analysis of disease prevalence data from survival/sacrifice experiments. Biometrics 34: 555 (1978).
39. Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Fluornoy, N. Farewell, V. T., and Breslow, N. E. The analysis of failure times in the presence of competing risks. Biometrics 34: 541 (1978).
40. Cleveland, W. S., Kleiner, B., McRae, F. E., and Warner, J. L. The analysis of ground-level ozone data from New Jersey, New York, Connecticut, and Massachusetts: transport from the New York City metropolitan area. In: Statistics and the Environment. (Proceedings of the Fourth Symposium. March 3-5, 1976), American Statistical Association, Washington, D.C. 1977, pp. 70-91.
41. Finklea, J. F. Auto emissions and public health: questions, statistical problems, and case studies. J. Washington Acad. Sci. 64: 91 (1974).
42. Bloomfield, P. Spectrum analysis of epidemiological data. In: Statistics and the Environment. (Proceedings of the Fourth Symposium, March 3-5, 1976), American Statistical Association, Washington, D.C. 1977.
43. Schneiderman, M. A. Safe dose? Problem of the statistician in the world of trans-science. J. Washington Acad. Sci. 64: 68 (1974).
44. Hoel, D. G., Gaylor, D. W., Kirschstein, R. I., Saffiotti, U., and Schneiderman, M. A. Estimation of risks of irreversible, delayed toxicity. J. Toxicol. Environ. Health 1: 133 (1975).
45. Rall, D. P. Problems of low doses of carcinogens. J. Washington Acad. Sci. 64: 63 (1974).
46. Newill, V. A. Regulatory decision making: the scientist's role. J. Washington Acad. Sci. 64: 31 (1974).
47. Boffey, P. M. Color additives: botched experiment leads to banning of red dye no. 2. Science and 191: 450 (1976).
48. Marx, J. L. Tumor promoters: carcinogenesis gets more complicated. Science 201: 515 (August 11, 1978).
49. Dinman, B. D. The SO_2 ambient air quality standard, with special reference to the 24-hour standard — an inquiry into the health bases. In: Statistics and the Environment (Proceedings of the Fourth Symposium. March 3-5, 1976), American Statistical Association, Washington, D.C. 1977, pp. 45-46.
50. Task Force on Environmental Cancer and Heart and Lung Disease. First Annual Report to Congress, August 7, 1978. Environmental Protection Agency; National Cancer Institute; National Heart, Lung and Blood Institute; National Institute for Occupational Safety and Health; National Institute of Environmental Health Sciences. Printing Management Office, U.S. Environmental Protection Agency, Washington, D.C., 20460.
51. Abelson, P. H. Regulation of the chemical industry. Science, 202: 473 (1978).
52. Hills, J. P. Legal decisions and opinions in pollution cases. Environ. Sci. Technol. 10: 234 (1976).
53. Phadke, M. S., Box, G. E. P., Tiao, G. C. Empirical-mechanistic modeling of air pollution. In: Statistics and the Environment (Proceedings of the Fourth Symposium, March 3-5, 1976), American Statistical Association, Washington, D.C. 1977.
54. Sauerhoff, M. W., Braun, W. W., Blau, G. E. and Gehring, P. J. Toxicol. Appl. Pharm. 36: 491 (1976).
55. Hartley, H. O., and Sielken, R. L., Jr. Response to Professor Mantel's discussion of presentation by Hartley and Sielken on the estimation of 'safe doses' in carcinogenic experiments. J. Environ. Pathol. Toxicol. 1: 270 (1977).
56. Rai, K. On the estimation of "safe" level of carcinogenic agents. PhD dissertation, University of Wisconsin, Madison, Wisc., 1979.
57. Breslow, N. E., and Enstrom, J. E. Geographic Correlations Between Cancer Mortality Rates and Alcohol — Tobacco Consumption in the United States. Journal of the National Cancer Institute. 53: 631-639, 1974.
58. Horowitz, R. L., and Feinstein, A. R. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. New Engl. J. Med. 299: 1089 (1978).
59. Ziel, H. K., and Finkle, W. D. Increased risk of endometrial carcinoma among users of conjugated estrogens. New Engl. J. Med. 293: 1167 (1975).
60. Williams, R. R., Feinleib, M., Connor, R. J., and Stegens, N. L. Case-control study of antihypertensive and diuretic use by women with malignant and benign breast lesions detected in a mammograph screening program. J. Nat. Cancer Inst. 61: 327 (1978).
61. Marks, S., Gilbert, E. S., and Breitenstein, B. D. Cancer mortality in Hanford workers. Paper presented at the Symposium on the Late Biological Effects of Ionizing Radiation, International Atomic Energy Authority, Vienna, 1978.
62. Mancuso, T. F., Stewart, A. and Karale, G. Radiation exposure of Hanford workers dying from cancer and other causes. Health Phys. 33: 369 (1977).
63. Pasternack, B. S., and Shore, R. E. Statistical methods for assessing risk following exposure to environmental carcinogens. In: Environental Health: Quantitative Methods. A. Whittemore, Ed., SIAM, Philadelphia, 1977, p. 49-71.
64. Lave, L. B., and Seskin, E. P. Does air pollution cause mortality? In: Statistics and the Environment (Proceedings of the Fourth Symposium, March 3-5, 1976), American Statistical Association, Washington, D.C. 1977.
65. Whittenberger, J. L. Discussion of paper by Lave and Seskin. In: Statistics and the Environment (Proceedings of the Fourth Symposium, March 3-5, 1976), American Statistical Association, Washington, D.C. 1977, p. 35.
66. Tukey, J. W. Discussion of paper by Lave and Seskin. In: Statistics and the Environment (Proceedings of the Fourth Symposium, March 3-5, 1976), American Statistical Association, 1977.
67. Weiss, W. Lung cancer mortality and urban air pollution. Am. J. Publ. Health 68: 773 (1978).

68. Newell, G. R., Hoover, R. N., and Kolbye, A. C., Jr. Guest editorial: status report on saccharin in humans. J. Nat. Cancer Inst. 61: 275 (1978).
69. Carson, R. Silent Spring, Houghton-Mifflin, New York, p. 189. Quoted by M. L. Straf, Introduction to an overview of air pollution and human health. In: Statistics and the Environment (Proceedings of the Fourth Symposium. March 3-5, 1976), American Statistical Association, 1977, p. 24.
70. Berkson, J. Limitations of the application of fourfold table analysis to hospital data. Biometrika Bull. 1: 47 (1947).
71. Breslow, N. E. Some statistical models useful in the study of occupational mortality. In: Environental Health: Quantitative Methods. A. Whittemore, Ed., SIAM, Philadelphia, 1977, p. 88-103.