July 1, 1987

FROM MOUSE TO MAN: THE QUANTITATIVE ASSESSMENT OF CANCER RISKS

by

D A Freedman
Department of Statistics
University of California
Berkeley, Ca 94720
415-642-2781

H Zeisel
Law School
University of Chicago
Chicago, Ill 60637
312-702-9494

Technical Report No. 79
Department of Statistics
University of California
Berkeley, Ca 94720

To appear in
Statistical Science

Summary

Results from animal experiments are often used to assess cancer risks to humans from low doses of chemicals. This involves two extrapolations: from high dose to low dose, and from animals to humans. This paper will review the logic of both.

In general, absent other information, we think that a chemical which is carcinogenic in a well-run animal experiment should be viewed with some suspicion. However, there are real problems with most animal experiments as they are currently done, and there are serious inconsistencies in the results. One probable cause is poorly defined endpoints, and another is uncontrolled variation. A number of suggestions are made for improvement, including proper randomization, 'blinding' the necropsy work, and use of statistical techniques appropriate to multiple endpoints.

Numerical assessments of human risk, even if based on good animal data, seem well beyond the scope of the scientifically possible. There are substantial differences in sensitivity between species, strains, sexes, and individuals. Experimental work is needed, to quantify these differences and explore their biological bases.

The dose-response models now used in numerical extrapolation are quite far removed from the biology. At present there seems to be no sound way to choose a model on either biological or statistical grounds, and different models give substantially different risk estimates. On this score, there is little hope for progress until the biology of cancer is better understood.

The paper is organized as follows. The issues are set out in the first section. Then the one-hit model is introduced in the context of a stylized risk assessment for DDT. Next the main generalizations of the one-hit model are explained: the multi-hit, Weibull, and multi-stage. The biological foundations for these models are reviewed, and the impact of model selection on low-dose risk estimates is stressed. Dose scales and biological scaling factors are discussed, and then the conventional arguments for the mouse-to-man extrapolation. The DDT carcinogenesis literature is surveyed, to show the quality of animal experiments. Opinions by others are cited, and conclusions are drawn.

Authors' footnote. Some of the work discussed in this paper was done while the authors were consulting for the firm of Skadden, Arps, Slate, Meagher & Flom in a law suit about DDT contamination, ultimately settled out of court. We would like to thank the following individuals for useful discussions, without implying their agreement-- or disagreement: J Bailar, P Diaconis, L Gold, L LeCam, S Moolgavkar, L Moses, D Petitti, M Pike, J Robins, S Swan, A Tversky, A Whittemore.

Table of Contents

1. Introduction

New chemicals and some old ones must be tested for safety. From an abstract scientific viewpoint, the best data would come from controlled experiments. However, experiments on humans are ethically permissible only rarely, so other kinds of evidence must be brought to bear. In some cases, good epidemiological evidence is available, although such observational studies have weaknesses of their own.

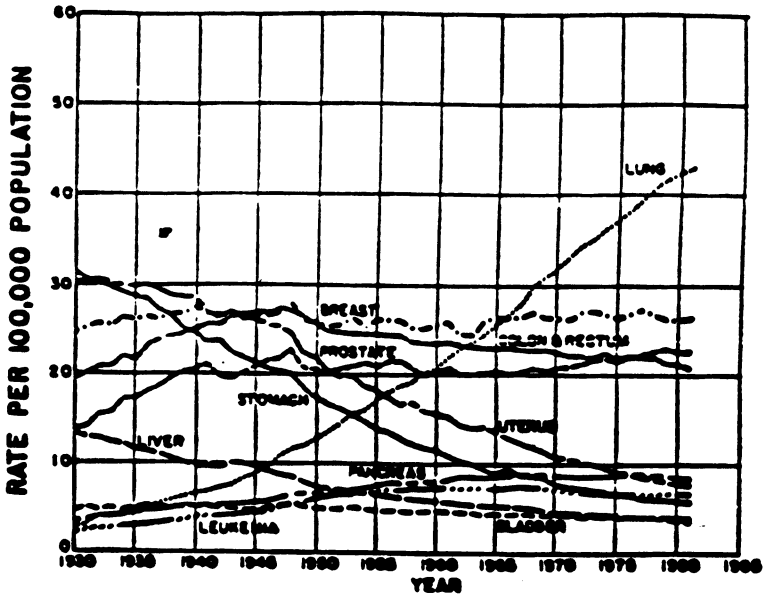
In most cases, no human data is available, and one turns to experiments on animals. Even if these are flawlessly done, two extrapolations are needed: from animals to humans, and from high doses in the experiments to the relatively low occupational or environmental exposures of interest. Our main topic is the reliability of these extrapolations. Before stating the scientific issues more sharply, we would like to explain a bit more of the practical background.

Two kinds of health effects must be considered. Some chemicals are acutely toxic in small doses. With others, exposure at ordinary levels does not cause any immediate harm, but chronic exposure at low levels may create a serious health hazard. In particular, chronic exposure to some chemicals in the workplace substantially increases the risk of cancer; asbestos and vinyl chloride are two prominent examples.

Cancer risks caused by chemicals are a matter of great public concern, because cancer is one of the most mysterious and frightening of modern diseases. In the US today it accounts for about one-fourth of all deaths. However, much controversy surrounds cancer statistics. Some commentators argue that there is an explosive cancer epidemic caused by exposure to chemicals; this view seems to be widely held by the public, although careful analysis of the available data does not lead to such alarming conclusions.

Crude cancer rates (not adjusted for age) have been going up, but this is mainly because of increased life expectancy: there are more old people at risk for the disease. To make sensible comparisons, it is necessary to standardize for age. On this basis, lung cancer rates have indeed been going up, following past increases in cigarette smoking. (But see Doll & Peto 1981 Appendix E or Franks & Teich 1986 p76 for evidence on a recent down-turn explained by changing smoking habits.)

Figure 1. Cancer death rates per 100,000, by site (age standardized to the 1970 population)



Source: ACS (1986). Rates are for both sexes combined except breast and uterus female population only and prostate male population only. Figure reproduced by courtesy of the American Cancer Society.

For other forms of cancer, the picture is quite mixed: for example, stomach cancer and liver cancer have been going down, leukemia up; the reasons are not well understood. On balance, except for the lung, cancer rates have been nearly constant (figure 1). For discussion, see eg Bailar & Smith (1986), Cairns (1985), Doll & Peto (1981), Higginson (1979), Peto (1980), National Academy of Sciences (1975, 1983b).

Cancer epidemiology depends on nonexperimental studies of human populations, with all the problems of confounding. The long periods of time between exposure and manifestation of cancer are a special complication: thus, asbestos workers from World War II are still developing mesothelioma in the 1980s. Furthermore, cancer seems to be inherently probabilistic: some non-smokers do get lung cancer while many smokers avoid this disease. Disentangling the causes of cancer in such circumstances is a very difficult exercise, but remarkable progress has been made: see eg Franks & Teich (1986) for a recent review.

Public concern, and the obstacles facing epidemiology, explain the heavy reliance on animal experiments. The Delaney amendment to the Food, Drug, and Cosmetic Act of 1954 outlawed residues in processed foods of chemicals which caused any risk of cancer to animals or humans. Such zero-risk requirements can be interpreted operationally as meaning "very low risk", where the risks are estimated from bioassays and extrapolated to humans. See, for example, National Academy of Sciences (1987).

Under prevailing standards, a good bioassay involves two species of test animals, typically rats and mice. Since cancer usually develops late in life, both for animals and man, the test species must have relatively short life-spans. Rats and mice live for about two years. They are small, cheap, and easy to maintain under lab conditions. Furthermore, experimentalists have much experience with rats and mice. There seem to be no other serious arguments for using these two test species in cancer testing. (Cf eg the Office of Technology Assessment 1981 p126.)

The basic axiom of toxicology is that the dose makes the poison: anything (even water) is harmful in large-enough quantities. Administering the test chemical at too high a dose kills the animals before they can develop cancer. Therefore, the experimental protocol requires the preliminary determination of the MTD, or maximally tolerated dose. By definition, above the MTD there are signs of acute toxicity (eg, stunting of growth, disorientation, etc).

According to standard protocols, in the main part of the experiment some animals are given the MTD, while others get specified fractions of the MTD. Some animals get no dose at all-- the control group. A control group is needed because the animals develop cancer spontaneously. Indeed, many strains of inbred lab animals seem particularly vulnerable to this disease. The bioassay is therefore intrinsically statistical. The whole idea is to compare the response of the test and control groups, to see if the incidence of cancer in the test group is above the chance level.

With three dose groups (eg, the MTD, half the MTD, zero), two sexes, and two test species, there are twelve groups of animals. It is conventional to start with 50 animals per group. So there will be 600 animals on test. This seems like a fairly modest experiment, but at the time of writing the cost is several hundred thousand dollars. Larger experiments have been done, like the 'mega-mouse experiment' with 24,000 animals on test, but these are clearly exceptional. (For some discussion of the mega-mouse experiment, see Staffa & Mehlman 1980.) The economics of bioassays dictate testing at the MTD. Indeed, with 600 animals, there is little chance of observing small effects at low doses, so the test would have negligible power at such doses.

Bioassays are used to make risk assessments, and then two extrapolations are needed:

i) Numeric extrapolation, from high dose to low dose.

ii) Species extrapolation, from the test animals to humans.

The extrapolation from high dose to low will be based on some type of mathematical model. The potential health hazards to humans usually result from doses which are 10 or 100 or 1000 times smaller than the experimental doses, so the extrapolation is over quite a range. How good are the dose-response models? What evidence validates them? These questions will be considered in sections 2 and 3.

While there are many similarities between mice and men, there are also many differences. What evidence shows the validity of the species extrapolation? Workers in the field call this issue the mouse-to-man problem. The statistical logic behind this extrapolation will be reviewed in sections 4 and 5.

The quality of the bioassays as experiments will also be considered. Does a positive finding mean that the chemical causes cancer in the test species, or is this likely to be an artifact of the experimental design and analysis? Section 6 reviews the DDT bioassays in an attempt to answer these questions.

Other literature is discussed in section 7, and conclusions are given in section 8. The balance of this section considers some of the public-policy issues, and some of the conventional responses to our sort of critique.

Cancer, and screening chemicals for carcinogenic hazards, is an explosive topic. In such a context, asking questions is seen as a political act-- especially if the questions turn out to have no satisfactory answers. (The interplay between the science and the politics is fascinating; see for example Epstein 1979 and Efron 1984.)

Some of the work on this essay was prompted by a consulting engagement with lawyers for a DDT manufacturer. The latter was sued by persons claiming damage from toxic wastes. The case was settled out of court, so there seems to be no reason to name the parties. Before working on the case, we felt-- along with every other educated person-- that DDT caused cancer. On review, the underlying evidence for this proposition turned out to be quite flimsy.

The contrast between the weakness of the evidence on DDT and the strength of the naive convictions is one motive for writing this essay, and DDT is used to illustrate the difficulties in risk assessment. The increasing use of risk models in the nation's law courts and government regulatory agencies, and our skepticism about the scientific foundations of the enterprise, explains the urgency we feel about the issue.

We do not wish to be understood as opposing government regulation of chemicals, or favoring uncontrolled pollution of the environment. Nor are we arguing against animal experiments. Great science has been done in the field of chemical carcinogenesis, and much animal work remains to be done if the biology of cancer is to be understood. However, routine bioassays have little to do with basic research, nor do they (in our opinion at least) contribute much to the scientific regulation of health or environmental hazards.

Any critique of the regulatory process generates two conventional responses. The first is that the human costs of introducing a carcinogen into the environment could be staggering, so any doubts should be resolved in favor of regulation. This may be right for food coloring, but the argument for a chemical like DDT is not so easy. Indeed, suppose the evidence for carcinogenicity of DDT in humans is weak, but DDT is a cheap, effective insecticide widely used in agriculture and for the control of diseases spread by insects (such as malaria). Finally, suppose that DDT is not especially toxic to humans, while the available replacements are not only more expensive, but also substantially more toxic.

Given these hypotheses, the balance of the costs in banning DDT is not so clear. On the one hand, DDT is clearly harmful to wild-life, and may pose some long-term hazard to humans. On the other hand, banning DDT may reduce the supply of food, increase the risks from malaria, and cause fatalities among insecticide workers: see Wald and Doll (1985, esp p119). An informed evaluation of the strength of the evidence for the carcinogenicity of DDT becomes crucial.

A second conventional response to our sort of critique: The existing technology for risk assessment may not be perfect, but it is better than nothing, and there is no replacement technology in sight. This argument is fine in some contexts. Engineers, for example, do know how to build roads. Therefore, someone who criticizes the plan for a road can quite reasonably be asked to produce a better plan.

With intellectual technologies like dose-response modeling (or the computer programming for Star Wars, to take an example with a different political flavor) the situation is quite different. Then the whole question is whether any technology can do the job. It may be worthwhile to face this question squarely. Indeed, the limits to knowledge may themselves be worth knowing; at least there is some precedent for taking such a position.

A final comment on the nothing-is-perfect argument. Dose-response models are imperfect. Nor was the maiden voyage of the Titanic a great success. Such understatements conceal more than they reveal. There are degrees of imperfection in theories ranging from quantum mechanics to astrology. The present essay attempts to locate risk assessment somewhare along this spectrum.

For a spirited defense of risk assessment (and only in part as a lesser evil), see eg Crouch & Wilson (1987), Lave (1987), or Russell & Gruber (1987).

2. An example: the one-hit model

A stylized account of a risk assessment involving DDT provides a useful starting point, and serves to introduce the one-hit model (the rationale for the name will be explained below). The object of the analysis was to estimate the risk of cancer caused by DDT contamination. Levels of contamination were estimated for individual plaintiffs in a lawsuit, and ranged from 1 to 30 parts per million (ppm) in the diet. This may seem low, but usual DDT levels run at parts per billion.

In the law case, the focus was on two metabolites of DDT, namely DDD and DDE. For now, the three substances can be considered together. To fix ideas, suppose an analyst wants to estimate the risk of cancer due to DDT at 20 ppm in the diet. A good data base for the purpose would show cancer rates for two similar human populations, one exposed at levels around 20 ppm and the other exposed at much lower levels. Any difference in the two cancer rates might be attributed to the difference in DDT exposures, subject to the usual arguments about interpreting observational data. For most risk assessments, including the one under discussion, such data do not exist.

At this juncture, risk assessment turns to animal data. To focus on essentials, suppose that lifetime exposure to DDT at 20 ppm in the diet causes an extra cancer risk of 10% in lab mice. This will be extrapolated to people, and background cancer rates must be considered. In round numbers, about 25% of the population of the US dies of cancer. Now a crucial step: if people are then exposed to 20 ppm of DDT for their lifetimes, their cancer rate are assumed to go up to

$$25\% + 10\% \text{ of } (100\% - 25\%) = 32.5\%$$

The first term represents the background cancer rate. The second term on the left represents the effect of DDT. The 10% has been extrapolated from mice to humans. That is the species extrapolation. (For official guidelines on this extrapolation, see eg US Environmental Protection Agency 1986; for a sympathetic presentation of examples, see eg Crouch & Wilson 1987.)

The procedure for combining the 25% background rate and the 10% additional risk from the exposure is called Abbott's formula. In effect, the 10% is the conditional chance of getting cancer from exposure to DDT at 20 ppm -- conditional on escaping cancer from all other causes. The basic assumption is the equality of this conditional chance for mice and men. On this hypothesis, Abbott's formula adjusts for the differences in background cancer rates between the lab mice and the human population. (The distinction between fatal and nonfatal cancer will be ignored for now; other conventional refinements and qualifications will be considered later; on Abbott's formula, see the Food Safety Council 1980 p716.)

Even for mice, the right comparative data on risks usually do not exist, because for reasons discussed earlier the dose levels in the bioassays are usually set much higher than the human exposures of interest. To fix ideas, suppose there is only one experiment to work with, in which the test animals were exposed to DDT at 250 ppm. There is a control group with no exposure. Assume the data turn out as shown in Table 1.

Table 1. Results from an animal experiment (Tomatis et al 1974)

dose (ppm)	number of mice	effective number	number with cancer	percent with cancer
0	190	188	34	18%
250	120	111	84	76%

Notes: Adapted from Tomatis et al (1974, Table 3). The dose is 125 ppm DDD plus 125 ppm DDE. The next column shows the number of mice initially assigned to the two groups; males and females are pooled. The 'effective number' is the number alive at the time of the first tumor, and the percents are relative to this number-- in effect, an adjustment for competing risks. The 'cancer' columns report on hepatomas, or liver tumors. The rationale for selecting this site is discussed later. Risk estimates based on these data are best thought of as applying to a 50-50 mix of DDD and DDE, since that was the substance on test.

Given that a mouse escapes cancer from other causes, its conditional chance of developing cancer from the DDT exposure is estimated (in effect, from Abbott's formula run backwards) as

$$(76 - 18)/(100 - 18) = 71\%$$

The exposure, however, is at 250 ppm; the analyst must extrapolate the risk down to 20 ppm, for the mice. This is the numeric extrapolation. After the risk at 20 ppm is estimated for the mice, the same estimate is used for people; this is the species extrapolation described above.

Numeric extrapolation involves a dose-response model which predicts response (chance of cancer) from dose (ppm in the diet). The formula will have one or more parameters which must be estimated from the data. There are many formulas to chose from, including the one-hit model and its generalizations like the multi-hit, the Weibull, and the multi-stage. These generalizations will be described below. For now, the focus is the one-hit model.

The basic equation in the one-hit model involves P(d), the total lifetime chance of getting cancer at dose level d. Thus, P(0) is the chance for the control animals, whose dose is zero. And P(250) is the chance for animals fed 250 ppm. These chances can be estimated by the fractions observed in the experiment.

The one-hit model involves the parameter k, which is called 'potency'. The basic equation can now be presented:

$$(1) \qquad P(d) = P(0) + [1 - P(0)] \times [1 - e^{-kd}]$$

The left side of equation represents the total chance of cancer, at dose d-- due to the exposure and to all other causes.

On the right side of equation (1):

P(0) represents the background chance of getting cancer due to all other causes-- at zero dose of the chemical on test

$1 - P(0)$ represents the chance of escaping cancer from all these other causes.

$1 - e^{-kd}$ gives the chance of getting cancer due to the exposure, at dose d-- conditional on escaping cancer from the other causes.

Technically, the one-hit model is the formula $1 - e^{-kd}$ for the conditional chance. In equation (1), this has been combined with the background chance P(0).

For fixed dose d, as k goes up the predicted chance of cancer goes up (hence the name, potency). Keeping the potency fixed, when the dose goes up the predicted chance of cancer goes up too, as is only reasonable. When the dose gets large, the chance of cancer approaches 1.0, or certainty. (This seems less reasonable, and with eg vinyl chloride or 2-AAF, the response rate in bioassays at high doses is substantially less than 100%.)

Now a minor technical fact: If kd-- the product of potency and dose-- is small, the equation is essentially

(2) $$P(d) = P(0) + [1 - P(0)] \times kd$$

Hence, the model is sometimes called 'linear'.

In Table 1, the number of mice at each dose level with cancer is considered to follow the binomial distribution, with probabilities governed by the model, so the parameter k can be estimated from the data eg by maximum likelihood. In round numbers, the estimated value for k is about .005, ie, each additional ppm of DDT in the diet causes an extra lifetime cancer risk of .005, and this completes the numeric extrapolation for the mice. (For more elaborate methods of estimating k, see the International Agency for Research on Cancer 1980 or Sawyer et al 1984.)

We are presenting a risk assessment used in a law case. To illustrate the mechanics of risk assessment based on the one-hit model, suppose that a plaintiff in the case has an estimated lifetime exposure of 20 ppm, and a background chance of .25 of getting cancer without the DDT exposure. The one-hit model can now be used to estimate that plaintiff's total lifetime chance of cancer, as follows. The product of potency and dose is .005 x 20 = .10. This is so small that the linear approximation (2) applies:

$$\begin{aligned} P(20) &= P(0) + [1 - P(0)] \times kd \\ &= .25 + [1 - .25] \times .10 \\ &= .25 + .075 = .325, \text{ or } 32.5\% \end{aligned}$$

In the model, the .25 is the background chance of getting cancer, without the DDT exposure; the .075 is the additional chance of getting cancer due to the DDT exposure at 20 ppm.

The basis for the risk assessment was Table 1, which reported liver tumors in mice, rather than tumors at any other site (lungs, bones, etc). This was a choice made by the analyst. As it turns out, such choices have substantial implications. The numerical results of a risk assessment depend on which experiment is used-- and which organ system.

To illustrate the point, Table 2 below presents additional data for the DDT experiment in question. There was an increased rate of liver tumors among the mice-- partly offset by decreases at other sites (eg, the bones). If the risk assessment had extrapolated from all sites in the mouse to all sites in humans, the estimated risk from DDT exposure would have been noticeably less, due to these offsets and to the high base rate of cancer among the controls.

Table 2. Results from Tomatis et al (1974). Male and female mice combined. Rates for all tumors and for liver tumors.

	All tumors	Liver tumors
Controls	89%	18%
250 ppm DDD	93%	27%
250 ppm DDE	91%	86%
125 ppm DDD + 125 ppm DDE	93%	76%

Note: The base of the percentage is the 'effective number', ie, the number of animals alive at the time of the first tumor observed. Table 1 reported the percentage of liver tumors for the controls and for the dose group 125 ppm DDD + 125 ppm DDE.

The estimated risk would also be much less if the extrapolation were from liver cancer in mice to liver cancer in people. Indeed, liver cancer is quite rare in the US. So P(0) in equation (1) for people would be much less. (Despite the increase in chemical pollution and the impact of chemicals on the mouse liver, the incidence rate of liver cancer in the US has been decreasing since the 1930s, as shown in Figure 1 above.)

The risk assessment also depends on assuming the formula (1) for mice and for humans, with the equality of k for the two species. The merits of all these assumptions will be considered below: but first, some of the main generalizations of the one-hit model for low-dose risk extrapolation will be discussed.

3. Other dose-response models

There are many dose-response models, ie, equations which predict response from dose and are used to extrapolate from high dose to low. The one-hit model and three generalizations will be reviewed, with some remarks on their biological foundations. (There are still other models which will not be reviewed, such as Cornfield's 1977 hockey-stick model, or the Moolgavkar-Day-Stevens 1980 two-stage model. Although more realistic on biological grounds, these models are seldom used in risk assessment. For a mathematical discussion of the various models, see Kalbfleisch-Krewski-van Ryzin 1983. Also see Moolgavkar 1986 for a review of the evidence on his model.)

As will be seen, the one-hit model does not fit typical data sets from animal experiments. The multi-hit, Weibull and multi-stage all tend to fit reasonably well, but lead to very different risk estimates at low doses. The biological foundations for all the models are quite weak, so there is no sound way to choose one rather than another, and no way to make reliable low-dose risk estimates.

The equations

First, the equations for the various models: let Q(d) be the chance of getting cancer at dose level d, due to the exposure, ie, conditional on escaping cancer from other causes. (Abbott's formula is used to bring in the latter.) As the equations show, the one-hit model is a special case of the multi-hit, Weibull, or multi-stage, since (4-6) specialize to (3) on setting m=1.

The one-hit, with parameter k:

$$Q(d) = 1 - \exp(-kd) \tag{3}$$

The multi-hit, with parameters k and m:

$$Q(d) = \int_0^{kd} t^{m-1} \exp(-t)\, dt/\Gamma(m) \tag{4}$$

The Weibull, with parameters k and m:

(5) $$Q(d) = 1 - \exp(-kd^m)$$

The multi-stage, with m stages sensitive and linear response at each stage:

(6) $$Q(d) = 1 - \exp(-\sum_{i=1}^{m} a_i d^i)$$

(This formula is a conventional approximation; the model will be explained in more detail below.)

Biological foundations for the multi-hit and Weibull

The multi-hit equation (4), for integer values of m, can be derived by assuming that 'hits' follow a Poisson process with parameter kd, and a cell becomes malignant when it suffers m hits. (The one-hit model requires only one hit, explaining the name.) However, these assumptions constitute a fable rather than a serious model, since there does not seem to be any precise biological definition for a 'hit', with some evidence that a specific number of hits causes cancer, or that hits follow a Poisson process.

The multi-hit (and Weibull) equations can also be derived by assuming that each individual in the population has a threshold, and gets cancer if the dose exceeds that threshold. Appropriate choice of the distribution for the thresholds leads to the equation of the model: gamma distributions give the multi-hit; extreme-value distributions, the Weibull. (In applications, m in the multi-hit model is often taken to be real rather than integer, so the 'hit' idea is not germane but the threshold idea still applies.)

Of course, the threshold hypothesis is open to some dispute. And there is no good reason why the distribution of thresholds should follow the extreme-value or gamma-- or any other textbook case.

The multi-stage model

The multi-stage model is more complicated and interesting-- and it will be discussed at some length here. The biological and statistical versions of the model will be distinguished; the statistical version turns out to be at some remove from biological reality.

In the biological model, a cell progresses through various stages until it becomes malignant. This seems reasonable, although the stages are seldom identified in any detail, or the process verified experimentally: for example, investigators cannot look at a cell and determine that it is in the 4th stage of a 5-stage progression. Indeed, it is very often impossible to decide whether a single cell in isolation is malignant or not. (However, recent progress in identifying DNA lesions must be cited, particularly for Wilm's tumor or retinoblastoma: see eg Franks & Teich 1986).

The statistical version of the multi-stage model involves a number of technical conditions which are usually not made explicit, and which are less reasonable:

> The order of progression through the stages is fixed and irreversible.
>
> The waiting times in the various stages are statistically independent, and follow the exponential distribution (in the case where exposure is constant).
>
> Cells go through the progression independently of one another.
>
> Independence of competing hazards.

In short, cancer is a Markov chain-- a pure birth process with absorption at the terminal state of 'malignancy'. A carcinogen is assumed to influence the rate of progression through the 'sensitive' stages. For example, an analyst might hypothesize that cancer is a 5-stage process, with DDT affecting the 1st and 4th stage. A carcinogen is assumed to act by increasing the rate at which the cell passes though each of the sensitive stages: this rate is assumed to be a linear function of dose, with different constants for each stage.

For reviews of the model, see Cairns (1981), Food Safety Council (1980), Kaldor & Day (1986), Lilienfeld & Lilienfeld (1980, especially p360), Peto (1977, especially pp1424ff). On the relationship between the Markov model and (6), see eg Freedman and Navidi (1987).

There are two kinds of evidence in favor of the model, human and animal. The main human evidence is as follows. For many kinds of cancer, the age-specific incidence is approximately a power of age: algebraically,

$$\text{incidence at age } t = \text{constant} \times t^p$$

This pattern is predicted by the model. More precisely, Armitage and Doll (1961) developed the model to explain this power law. In the model, the power p is related to the number of stages, which usually turns out to be between 4 and 6. See Peto (1977).

On the other hand, most cancers do not seem to follow the power law: see eg Cook-Doll-Fellingham (1969), Lilienfeld & Lilienfeld (1980, esp p360). Kaldor & Day (1986, sec III) discuss some of the difficulties in making such analyses. Pike (1983) gives an example of how anomalies might be resolved, at least for breast cancer; and Moolgavkar-Day-Stevens (1980) in effect give a counter-argument to that sort of resolution, by showing how their model, which is quite different, also fits that data.

Lung cancer may be the best studied, and is usually thought to follow the multi-stage model quite well: see Doll & Peto (1978). However, recent analyses of the data shows serious discrepancies even there: Freedman & Navidi (1987), also see Brown & Chu (1986). In brief, a variety of multi-stage models will fit the original Doll & Peto data for current smokers. No model fits the Dorn veterans cohort, or the American Cancer Society volunteers.

Turn now to the animal evidence on 'initiation' and 'promotion'. The idea is that an initiator causes a cell to change from its normal state to a pre-malignant state, in which it may remain indefinitely; a promoter causes an initiated cell to become cancerous. (Some writers consider a third stage of proliferation; others subdivide the promotion stage: see eg International Agency for Research on Cancer 1984.) Only one typical example need be given. DMBA (dimethylbenzanthracene) is considered an initiator, croton oil (more specifically, its phorbol ester constituent) a promoter. The reason: Applying the agents in the order

DMBA first, croton oil second

produces a large yield of tumors, mainly non-malignant papillomas. Applying them in the reverse order or separately gives a much smaller yield: see eg Boutwell (1964).

In the framework of the multi-stage model, an initiator is held to affect the rate of progression through an early stage; a promoter affects a late stage. Some 'complete' carcinogens are thought to be both promoters and initiators-- cigarette smoke, for example.

Now, some of the problems. These experiments relate to the progression of tumors-- colonies of cells: the mathematical model relates to the progression of an individual cell. Cells within a tumor become remarkably heterogeneous in their genetic makeup, so progression of the tumor is not good evidence about the progression of individual cells.

Even at the level of whole tumors, there are interesting new experiments which show that for some initiators and promoters the sequence

initiator, promoter, initiator

produces a much larger yield of malignancies than the sequence

initiator, promoter

Likewise, the order

promoter, initiator, promoter

increases the tumor yield. This is not easy to reconcile with the conventional view of initiation and promotion. See Hennings et al (1983), and for a review International Agency for Research on Ccancer (1984). The phenomenon was predicted on theoretical grounds by Moolgavkar & Knudson (1981).

Moreover, with typical initiator-promoter protocols, only one application of the initiator is needed. And the timing of the successive applications of the promoter is critical: if the applications are too far apart, or too close together, the effect disappears (Boutwell 1984). These facts cut against the model.

The idea of reversibility presents serious problems for the model, too. There is now much evidence for the reversibility of some lesions, including papillomas; the fixed order of progression through the stages of the model then comes into question. See eg Cohen et al (1984, esp p103), Slaga (1983); for reviews, Montesano-Bartsch-Tomatis (1980), Office of Science and Technology Policy (1985, chap 1, sec V), UK Department of Health and Social Security (1981).

The idea of 'epigenetic' cancer is also incompatible with the model: 'genetic' cancer is caused by damage to DNA, the genetic material of the cell; 'epigenetic' cancer is caused by some failure of surrounding tissue to regulate growth and differentiation, and this will affect all cells in the vicinity, contrary to the independence assumption in the model. For reviews, see Douglas (1984), Franks & Teich (1986), Rubin (1980), the UK Department of Health and Social Security (1981). For reports on experiments, see eg Stott et al (1981), Williams (1980, 1983).

Likewise, aging affects metabolic processes and may affect susceptibility to cancer; this would contradict the model's assumption of constant rates of progression. There is animal evidence (Peto et al 1975) to show that incidence of tumors depends on time since exposure rather than age, but there is also evidence going the other way. For reviews, see Likhachev-Anisimov-Montesano (1985); or Sohal-Birnbaum-Cutler (1985) on the molecular biology of aging.

Another difficulty: while some carcinogens act in synergy, there are antagonistic pairs. See eg Richardson-Stier-Borsos-Nachtnebel (1952), Miller-Miller-Brown (1952). Okey (1972) shows that DDT protects female rats against the induction of breast cancer by DMBA. Cohen et al (1979) demonstrates a protective effect from dioxin. A striking recent study shows that aspirin increases the effect of the carcinogen FANFT at one site but inhibits it at another-- Murasaki et al (1984). For a review of such interactions, see DiGiovanni et al (1980) or Shankel et al (1986). The phenomenon is well outside the scope of the model.

At this point, it may be useful to recall the distinction between the biological and statistical versions of the multi-stage model. In the biological version, a colony of cells progresses through stages on the way to cancer. In the statistical version of the model, an individual cell executes a Markov chain through a fixed order of states along the way to cancer, the transition rates being linear functions of dose: these are hypotheses largely about unobservable entities. The statistical model may lead to beautiful mathematics, and may have real heuristic power. But it is much more loosely coupled to reality than the biological model. The statistical model-- the relevant one for quantitative risk assessment-- is at a considerable distance from the realm of scientific fact.

Fitting the models to animal data

The biological foundations for all the models seem to be quite speculative, so there is no sound way to chose one over another on theoretical grounds. But the different models have very different implications for risk assessment. As will be seen, the one-hit model does not fit typical data sets from animal experiments: the multi-hit, Weibull and multi-stage all tend to fit reasonably well, but disagree by many orders of magnitude on the estimates of risk at low doses.

There is an excellent review of the models and data sets by The Scientific Committee of the Food Safety Council (1980). With 14 data sets, the one-hit model is rejected 6 times ($P < .05$ by chi-squared); the multi-stage model is rejected once; the Weibull and multi-hit fit all the data sets. In 10 out of 14 cases, the Weibull and multi-hit offer significant improvements over the one-hit. In this context, the chi-squared test does not have much power, so rejection is a strong signal. For a more recent review, with similar conclusions, see Hoel & Portier (1987).

In essence, the one-hit model is linear at low dose, and this linearity is often contra-indicated by the data. The other models are sufficiently flexible to fit typical dose-response data. Since there are at most 6 dose groups in the Food Safety Council data sets, this is perhaps not such a strict test. Few animal data sets have as many as 6 dose groups, so power to differentiate among the models is low. With time-to-tumor data, the multi-hit model may not fare so well. Also see Carlborg (1981ab), who argues for the Weibull over the multi-stage in the mega-mouse experiment.

For the purposes of risk assessment, it is a crucial point that many models will fit most of the data, while the choice of model has a profound impact on the estimated risks at the low doses of interest. In general, the one-hit model gives the highest risk estimates, and the multi-hit gives the lowest-- by quite large factors. A few examples may be of interest: see Table 3. For eg aflatoxin, the one-hit model gives 30 times the risk estimated from the multi-stage, 1000 times the risk from the Weibull, and 40,000 times the risk from the multi-hit. The results in Table 3 are not unusual. See, for example, Hoel-Kaplan-Anderson (1983), Krewski & van Ryzin (1981), Rai & van Ryzin (1979, 1981).

Table 3. The impact of the model on low-dose risk estimates

substance	one-hit	multi-stage	Weibull	multi-hit
Aflatoxin	1	30	1000	40,000
Dioxin	1	400	400	800
DMNA	1	700	700	2,700
Dieldrin	1	3	200	1,000
DDT	1	2	70	200

Notes. From Food Safety Council (1980, Table 4). The 'Virtually Safe Dose', or VSD, is estimated from each of the four models, as that dose giving a risk of one in a million. The column for the multi-stage model shows the ratio of its estimated VSD to the VSD estimated from the one-hit, for each of the 5 substances. Likewise for the Weibull and the multi-hit.

Saccharin is another example of some interest. Published risk estimates, starting from the same animal data but using various models, differ by factors of over 5,000,000. See the National Academy of Sciences (1978, p3-72 for the data and pp3-61ff for discussion).

A final example is Haseman & Hoel's (1979) study of risk estimates derived from animal experiments on DDT. In all cases, the multi-stage model was used. With 8 studies and two sexes, there were 16 sub-experiments. For lung tumors there were 11 cases where the risk estimate was zero: in the remaining 5 cases, the risk estimates varied by factors up to 1000. For liver tumors, as Haseman & Hoel remark, "the agreement was better:" there was only one case where the risk estimate was zero, and in the remaining cases the variation was only by a factor of 250.

The artificiality of the models, and the sensitivity of the results to the modeling assumptions, show how far removed risk assessment is from an objective science. Indeed, the Food Safety Council (1980, p718) quotes the Commissioner of the Food and Drug Administration as follows:

> The Commissioner has extensively reviewed the known procedures that may be used to derive an operational definition of the non-residue standard of the act from animal carcinogenesis data. This review has persuaded him that the same scientific and technological limitations are common to all. Specifically, because the mechanism of chemical carcinogenesis is not understood, none of these procedures has a fully adequate biological rationale. All require extrapolation of risk-level relations from responses in the observable range to that area of the dose-response curve where the responses are not observable. Matters are further complicated by the fact that the risk-level relations adopted by the various procedures are practically indistinguishable in the observable range of risk (5 percent to 95 percent) but diverge substantially in their projection of risks in the non-observable range.

Why is low-dose extrapolation so difficult? The Commissioner explained the answer quite clearly: Not enough is known about the biological mechanisms of cancer. In fact, there are some 200 different kinds of cancer, classified by site and tissue, with many different biological mechanisms. Although much has been learned about the biology in the past few decades, many crucial details remain to be elucidated. In this light, any attempt to develop one simple mathematical formula to describe cancer risks seems naive.

Some of the biological complexities in low-dose risk extrapolation should be mentioned explicitly-- eg, the role of metabolic pathways, genetics, repair mechanisms. For example, high doses may overwhelm repair mechanisms or metabolic pathways leading to detoxification: see Hoel-Kaplan-Anderson (1983) for the impact on risk modeling, or Whittemore-Grosser-Silvers (1986); and for a review, Office of Science and Technology Policy (1985, esp sec 3IIB).

Another example. Repeated injury to body tissue may increase the risk of cancer: the cells proliferate to repair the injury, and if the insult continues, this could increase the chance of mistakes in DNA replication, leading in the end to heritable mutations. For some of the relevant animal experiments, see Mirsalis et al (1985), Moore et al (1982), Stott et al (1981). For reviews and discussion of the implications for models of carcinogenesis, see Ames-Magaw-Gold (1987, pp275-6), Farber (1984) or Iversen & Astrup (1984). For a different opinion, see Ward (1984).

The human liver is quite vulnerable to repeated insults: see eg Bloom & Fawcett (1962, pp600ff) or Weinbren (1978, esp pp1207 & 1243-1262). Consider alcohol: at high doses, it causes cirrhosis of the liver by cell-killing and subsequent proliferation; at low or moderate doses, this does not occur. Likewise for acetaminophen, the active ingredient in many pain-killers. For these substances, low-dose extrapolation on cirrhosis would be a scientific blunder, and we are not aware of attempts in that direction.

On the other hand, in bioassays many animal carcinogens like DDT seem to affect the mouse liver through cell-killing at high doses. And there certainly are attempts at low-dose risk assessment for such substances, even though the cell-killing mechanism is unlikely to operate at low doses.

(At sites other than the liver, acetaminophen seems to be weakly carcinogenic by a different mechanism, while alcohol has a potent synergistic effect with tobacco. Risk assessment at these sites would run into serious problems too: see Doll & Peto, 1981, p1225 on alcohol; and International Agency for Research on Cancer 1982 p17 on phenacetin, of which acetaminophen is the active metabolite.)

To sum up, the choice of models has a decisive impact on low-dose risk estimates, and in the present state of knowledge there is no sound way to pick one model rather than another. All except the one-hit will fit typical dose-response data sets, and none have adequate biological foundations. That is why reliable estimates of risks at low dose cannot be made on the basis of present knowledge. This completes the discussion of extrapolation from high dose to low; turn now to the species extrapolation.

4. Dose scales and the species extrapolation

What is the basis for the species extrapolation? First, the definition of dose must be considered in more detail: Indeed, even granting that a man is just a big mouse, one milligram of DDT cannot mean the same thing for both of them, due to the difference in size. However, there turn out to be many different ways to measure this difference. For example, a man weighs 2800 times as much as a mouse, eats 300 times as much per day, and lives 40 times as long (Table 4). Which factor should be used to rescale the dose?

Table 4. Comparative size factors on 4 species.

	Weight kg	Food g/d	Lifetime years
mouse	.025	5	1.75
rat	.25	15	2
dog	10	250	10
man	70	1500	70

Source: Crouch & Wilson (1979, p1110). Also see Gold et al (1984, p13).

The stylized risk assessment in section 2 measured dose in parts per million in the diet. On that scale, it was assumed that men and mice would react similarly to similar doses. Other standard dose scales include mg of intake per day per kg of bodyweight, and mg of lifetime intake per kg of bodyweight. Some authors recommend adjusting by surface area rather than bodyweight, surface area being estimated as a power of body weight.

The choice of dose scale can itself affect the risk estimates by a factor of 50 or more. Given the dose scale, a risk model can be sophisticated by the inclusion of a scaling factor to represent species sensitivity. However, at present there is no real basis for choosing the dose scale, or estimating a scaling factor.

The crucial biological problem in choosing a dose scale or a scaling factor has already been mentioned, in connection with the numeric extrapolation-- metabolic pathways and rates. In more detail: Miller (1970) and Miller & Miller (1977) suggested, and it is now widely believed, that few substances are carcinogenic in their original form: highly reactive and unstable metabolites produced by enzymatic breakdown are the proximate carcinogens. For example, the enzyme 'Cytochrome P-450' is implicated in a number of cases. Genetics must therefore play a crucial role and current research on oncogenes reinforces this view.

For discussion, see eg Bishop (1987), Franks & Teich (1986), Gibson (1971), Montesano-Bartsch-Tomatis (1980), International Agency for Research on Cancer (1984), National Academy of Sciences (1978, Chap 3), Office of Science and Technology Policy (1985).

Metabolic pathways and rates play a major role in carcinogenesis. That is one basis for individual, inter-strain or inter-species differences in susceptibility. For such reasons, men cannot be expected to react to DDT the way mice do. After all, if a man had exactly the same metabolism as a mouse, he would <u>be</u> a mouse.

With some notable exceptions, pathways and rates are not known in detail, so pharmacokinetic models for the activation of carcinogens cannot at present be developed and tested. For some detailed argument on these topics from a variety of perspectives, see Calabrese (1984), Clayson (1985, 1986), Clayson-Krewski-Munro (1985, chaps 1-5), Office of Science and Technology Policy (1985, chap 1), Smith (1986); also see the proceedings of the symposium on estimating human risk from animal data (J Toxicol Pathol 1985 vol 13 no 2).

For all these reasons, a scientific basis for choosing the dose scale and biological scaling factor is not presently available. That is one way to state the fundamental difficulty in the species extrapolation.

5. The qualitative extrapolation

The main focus so far has been the quantitative extrapolation from animal experiments to human populations. This section considers the qualitative extrapolation-- the idea that if a substance causes cancer in animal experiments, it is likely to be a human carcinogen. The idea has intuitive appeal, but the evidence for it is far from solid. The main arguments for the validity of the qualitative extrapolation will be reviewed, and then some evidence from epidemiology will be considered.

The mammalian argument

One oft-recited argument is that humans and mice are both mammalian species. This verges on sentimentality. If the test species of choice were trout, we would all be vertebrates together.

The mouse-to-rat argument

A more substantive argument is that results in the mouse are predictive for the rat, and so by extension for humans. This argument has been made, for example, by Tomatis-Partensky-Montesano (1973).

Table I in Tomatis-Partensky-Montesano (1973) lists the chemicals considered at that time to induce tumors in mice. Were these chemicals carcinogenic for rats or hamsters? There were 58 chemicals, and 11 were classified as negative for the rat, while another 7 had not been tested: 6 were negative for the hamster, 29 had not been tested. The error rate for rats was 11/51; for hamsters, 6/29.

These seem quite low, but depend on the list of chemicals used as the test set. To illustrate the point, take chlorinated hydrocarbon pesticides-- the class which contains DDT. We could identify 9 in the Tomatis list and all were reported as carcinogenic in mice. Table 5 below shows what happens when these compounds were tested on rats or hamsters. With respect to chlorinated hydrocarbon pesticides, the mouse results do not seem so predictive for other rodents.

Table 5. From Tomatis-Partensky-Montesano (1973): Nine chlorinated hydrocarbon pesticides which are carcinogenic in mice, classified according to carcinogenicity in rats and hamsters.

	positive	negative	not tested
rats	2	6	1
hamsters	0	1	8

There are a number of other surveys on the reliability of the mouse-to-rat extrapolation. Haseman et al (1984) review the National Toxicology Program bioassays on mice and rats. (These bioassays are all designed with a common protocol, which is as good as any in widespread use.) Of the 86 compounds on test, 43 were carcinogenic in at least one of the two test species. Of the 43 carcinogens:

17 were positive in mice and rats both.
14 were positive in mice only
12 were positive in rats only.

These figures include three compounds that were tested in mice only and two in rats only. Of the carcinogens, then, only 17/43 = 40% were positive in both test species. (See p634.)

There is a similar review of the National Cancer Institute bioassays-- predecessors to the National Toxicology Program-- by Griesemer & Cueto (1980), also see Office of Technology Assessment (1981, p126). The number of chemicals tested was 190, of which

64 were non-carcinogenic in both species
28 were equivocal
98 were carcinogenic in at least one of the two test species.

Of the 98 carcinogens:

44 were positive in mice and rats both
54 were positive in only one species.

Again, of the carcinogens, only 44/98 = 45% were positive in both test species.

Di Carlo (1984) gives a similar picture. Ward-Griesemer-Weisburger (1979) conclude there is a reasonable correlation between bioassay results for rats and mice; so does Purchase (1980).

The lack of concordance between rodents and monkeys should also be mentioned. For example, 5 out of 6 'model rodent carcinogens' are negative in the monkey: Adamson & Sieber (1983). Results on 2-AAF, an intensively studied animal carcinogen, are worth noting too. This substance tests as carcinogenic in the cat, chicken, dog, guppy, hamster, mouse, newt, rabbit, and rat: not in the cotton rat, guinea pig, monkey, x/gf mouse, rainbow trout, or steppe lemming. The tally is 9 to 6. See Weisburger (1981, esp p3); Weisburger (1983, p23) comments on difficulties in metabolic interpretations. Here as elsewhere, some of the 'negative' findings may be due to low power, just as some of the positive findings may be artifactual.

There is no clear bottom line to report. Taking all the experiments at face value, there is some measure of agreement between the results for rats and mice, and some measure of disagreement. Now rats and mice are much more similar to each other than either is to humans. The validity of the mouse-to-man extrapolation seems hard to argue on the basis of these data.

Crouch & Wilson (1979) is often cited to show good inter-species correlations of carcinogenic potency. However, Bernstein et al (1985) suggest that Crouch & Wilson may have been misled by a statistical artifact of bioassay design. Zeise-Wilson-Crouch (1984) reports a correlation between toxicity and carcinogenic potency. If this is real rather than another artifact, it may be evidence for the cell-killing mechanism of carcinogenesis: see Bernstein et al (1985, p86). Zeise-Wilson-Crouch propose using the correlation in quantitative risk assessment, relying on the one-hit model: this ignores much evidence against the model. Crouch-Wilson-Zeise (1987) attempt to refute Bernstein et al (1985), but their statistical argument seems inappropriate.

The man-to-mouse argument

Another argument for the qualitative extrapolation is quoted in Tomatis (1979):

> The difficulties in assessing the significance of experimental [animal] results for predicting similar hazards in humans are both qualitative and quantitative and can be summarized in the following questions.
> 1. Are chemicals that have been shown to be carcinogenic to experimental animals also carcinogenic to humans?
> 2. Do experimental animals (rodents, in particular) and humans have similar susceptibility to the carcinogenic effect of chemicals, or are rodents incomparably more susceptible than humans?
> A partial answer to the first question is usually given by reversing the terms of the question: Most of the chemicals that are carcinogenic to humans are carcinogenic to at least one, and in most cases to more than one, animal species.

The question at issue is this: will most animal carcinogens turn out to be human carcinogens? The argument given is that most human carcinogens turn out to be animal carcinogens. This blurs together two conditional probabilities: P(A|B) can be quite small, while P(B|A) is quite large. Here, A is the set of animal carcinogens; B, the human carcinogens. So, as Tomatis acknowledges, even if most human carcinogens are animal carcinogens, the converse implication does not really follow.

Nor does the factual base of the argument seem right, as will now be explained. The test data will be drawn from the IARC, which publishes periodic reviews of the evidence for carcinogenicity of suspect chemicals, compiled by working groups of experts.

(The IARC is the International Agency for Research on Cancer, based in Lyon. It is one of the major research agencies in chemical carcinogenesis. Tomatis is a leading experimentalist and at the time of writing, the director of the IARC.)

At the time of writing, the most recent review on carcinogenicity was the IARC (1982), with one minor and one major revision reported in press: IARC (1987, 1988). Plainly, the classification of suspect chemicals is a moving target, but the data base for Tables 6 and 7 below is defined as the IARC (1982). That list of 155 suspect chemicals shows 30 'proven' carcinogens in humans (some well-known carcinogens, like tobacco, had not yet been reviewed by the IARC).

The list also includes information on the animal evidence. The IARC grades the evidence as 'sufficient', 'limited', or 'inadequate'. For animals, 'sufficient' evidence means that the chemical causes tumors in two strains or species, or unusually severe tumors in one. 'Limited' evidence includes one positive experiment. Negative or inconsistent results may be set aside.

The animal data is summarized in Table 6 below and is not in such good agreement with the human data after all. As it turns out, there is 'sufficient' proof of carcinogenicity for animals in only 13/22 = 59% of the human carcinogens.

Table 6. The IARC (1982) list of proven human carcinogens, classified by degree of evidence for carcinogenicity in animals.

sufficient	13		
limited	6		
inadequate	2		
no data	1		
		22	
animal data irrelevant		8	
			30

Consider next the data proposed in Tomatis (1979). His Table 1 lists 26 chemicals, groups of chemicals, and processes which are 'associated or strongly suspected of being associated with cancer induction in humans'. (Of these, 18 are considered by the IARC to be proven human carcinogens. For the other 8, the IARC does not consider the evidence sufficient, eg, isopropyl oils appear in Tomatis' Table 1, and in the IARC group 3 of things which 'cannot be classified as to [their] carcinogenicity in humans.')

Of Tomatis' 26 human carcinogens, 17 are carcinogenic in the mouse, 15 in the rat, and 6 in the hamster: 65%, 58% and 23% respectively. Thus, there is a fair amount of discordance among rodent species, as well as a significant discrepancy between the animal data and the epidemiology-- which is the next topic.

Consistency with epidemiology

How consistent is animal data with epidemiology? This question seems straightforward, but is full of complexities. There are relatively few chemicals which have been carefully evaluated by both methods. Nor does that set constitute a representative sample from the universe of all chemicals. Indeed, the chemical carcinogenesis community sets so much store by the man-to-mouse argument that enormous efforts are made to demonstrate the carcinogenicity in mice of likely human carcinogens; see Wald & Doll (1985, p4).

As before, data from the IARC (1982) will be used. Their Table 1 reports on 155 chemicals, groups of chemicals (eg, soots, tars and oils) and processes (eg, hardwood furniture manufacture) tested for carcinogenicity. With respect to 19-- including eg hardwood furniture manufacture, 'certain combined chemotherapy for lymphoma', and various forms of oral contraceptives-- the IARC judges that animal experiments are irrelevant. In some cases, they are clearly right and in others they may be wrong, but for present purposes we accept their judgment. In 3 cases there are no animal data. Eliminating the 19 irrelevant cases and the 3 without data leaves a test set of 133 chemicals and groups of chemicals.

The IARC considers three types of evidence: epidemiological studies, animal bioassays, and short-term tests (for mutagenicity in vitro). The grades of evidence were discussed above: for humans, 'sufficient' evidence means good epidemiology.

Table 7 below classifies the test set by grade of evidence for carcinogenicity in humans and animals (as determined by the IARC). There is a fair amount of discord in Table 7: with respect to only 21% of the animal carcinogens is there 'sufficient' evidence for human carcinogenicity.

The 'insufficient' category in the table combines IARC grades of limited or inadequate evidence. This may be unconventional, but seems fair, given the IARC definitions. Indeed, for reasons to be given in the next section, even 'sufficient' animal evidence may not be compelling.

Table 7. The test set of 133 relevant chemicals and groups of chemicals reviewed by the IARC, for which data is available, classified by grade of evidence for carcinogenicity in animals and humans.

	grade of evidence for carcinogenicity in animals	
in humans	sufficient	insufficient
sufficient	21%	11%
insufficient	79%	89%
	100%	100%
number	61	72

In principle, the evidence in Table 7 is decisive: carcinogenicity in lab animals is poor evidence for an effect in humans. Questions about the representativeness of the test set and doubts about the quality of the underlying studies (both positive and negative) weaken this conclusion appreciably. We do not take up such questions because we made no systematic review of the underlying studies, and only report the classifications reported by the IARC.

The overall conclusion from Table 7: the research reports of the cancer community (even taken at face value) do not sustain the conventional arguments for the validity of the qualitative extrapolation. For a more detailed discussion of inconsistencies between animal evidence and epidemiology, see Wald & Doll (1985). For an establishment view of the evidence, see Wilbourn et al (1986); the correlation in their Table III reflects only compounds which are positive in humans.

We remain sympathetic to the idea that animal data have some predictive value for carcinogenicity in humans, at least qualitatively; and perhaps even to establish rankings of potential hazards as suggested by Doll & Peto (1981, pp1215ff). But the evidence for such propositions is surprisingly weak.

Experimental studies to quantify inter-species differences in sensitivity would clearly be very useful, if expensive. Research to determine the biological bases for these differences would be even more useful.

6. Review of carcinogenesis experiments

Some general questions will be raised about the quality of animal experiments on carcinogenicity, and then the DDT literature will be reviewed, to illustrate the points. There turn out to be substantial inconsistencies in the experimental data, perhaps attributable to the multiplicity of endpoints and uncontrolled variation. Proposals are made for improving the experiments.

Reproducibility of results seems to be a crucial issue, and a preliminary remark on definitions is in order. As noted above, cancer is not a unitary disease. In animal experiments, there are some 25 major organ systems which are checked by autopsy for tumors of various types. Even the type of lesion which will be taken as evidence for carcinogenicity may only be decided during the course of the experiment.

There are marked differences in carcinogenicity across sexes, strains and species. Often, the same chemical will cause one kind of cancer in one experiment, and another kind in another experiment (but see Gold et al 1986b). Indeed, the most hard-bitten advocates of animal experiments do not claim to be able to predict which organ will be affected in humans by a chemical which is carcinogenic in animals (see eg Wilbourn et al 1986, esp Table II).

Some of the differences in carcinogenic response must be due to differences in the biology, and some to uncontrolled variation in the experimental design. What are the likely sources of such variation? For one, animals may not be properly randomized to the various treatment groups; and there may be strong litter effects, especially in multi-generation studies (Grice-Munro-Krewski 1981, Turusov et al 1973). Likewise, animals are seldom randomized to cages; but position in the rack seems to be a risk factor for cancer (Lagakos & Mosteller 1981).

Indeed, many other apparently extraneous factors substantially change the incidence of tumors. These include stress, calorie restriction, and viral infection. See eg Clayson (1975,1978), Gellatly (1975), Jose (1979), Peto (1980), Roe (1981), Tannenbaum (1940-2), National Academy of Sciences (1983b).

A final example of a design problem: the pathologists who identify the tumors often know the treatment status of the animals, and this leaves room for bias in the diagnostics. Pathologists see themselves as professionals exempt from bias and resist suggestions for blinding, as in Weinberger (1973, 1979): despite the author's intentions, these papers vividly show how knowledge of dose status can influence diagnostic results.

The magnitude of this bias is not easy to document from the medical literature. For some evidence in the setting of clinical oncology, see McFarlane-Feinstein-Wells (1986); for evidence on bias in reading echocardiograms, see Tape & Panzer (1986); on X-rays, see Reger-Butcher-Morgan (1973) or Reger-Petersen-Morgan (1974). For evidence on the variability in reading pathology slides, see eg Siegler (1956) or Metter et al (1985); and for a recent review, Swan & Petitti (1982).

The high spontaneous tumor rates in the experiments contribute to the difficulty: Multiple endpoints matter, because there are many types of tumors and many sites. Then artifacts of chance or poor design create the likelihood that in one experiment there will be a high cancer rate at one site, and in another experiment, the excess will be observed at another site, even if there is no real carcinogenic effect.

For a general discussion of excess variation, see Haseman (1983) who reviews 25 National Toxicology Program bioassays on various chemicals and shows that increases in cancer at one site are matched by decreases at another site. Such decreases are usually explained away by asserting that the animals in the treatment groups do not live long enough to develop tumors. Haseman rejects this explanation because the animals in the treatment groups live a bit longer than the controls: the difference is small, but statistically significant. (In all cases, the test species was the Fischer 344 rat; the increased tumor rates were mainly in the liver; the decreases, lymphomas and leukemias.)

These points will be illustrated using long-term animal experiments where DDT, DDD and DDE was fed to mice, rats or hamsters. To minimize selection bias, we took only papers referenced in IARC (1974, 1979, 1982). This screened out some bad studies, and some good ones. Too, we may have missed a few papers referenced in IARC (1974) but not summarized there. The sample is listed in Table 8 below, with comments.

Table 8. The sample of papers.

The mouse (11 papers, 9 studies).

Innes et al (1969). JNCI 42 1101.
Two strains of mice, X = (C37BL/6 x C3H/Anf)F1 and Y = (C37BL/6 x AKR)F1. Tested 120 compounds, with about 20,000 mice; found 11 carcinogenic, including DDT at 140 ppm; found DDD and 19 other compounds 'require further evaluation', but did not report data. Common control group. Survival is to term, and the denominator for cancer incidence is the number sent to necropsy.

Kashyap et al (1977). Int J Cancer 19 725.
Pure Swiss inbred mice. Reports multiple experiments; we analyze only the feeding experiment on DDT at 0 or 100 ppm. Survival is to 80 weeks.

Shabad et al (1973). Int J Cancer 11 688.
Multi-generation study on A-strain mice. The design is not easy to discern from the paper: compare IARC (1974, p98). Table 11 reports on DDT at 0 or 10 ppm, pools sex and generation, and gives six-month survival. The tumors are lung adenomas, and according to the authors, 'No other tumors were observed in the treated animals'.

Tarjan & Kemeny (1969). Fd Cosmet Toxicol 7 215.
Multi-generation study on BALB/c mice, DDT at 0 or 3 ppm. Denominators shown in Table 3 for males and females combined; we elected to pool F1-5.

Terracini et al (1973a). Int J Cancer 11 747.
Terracini et al (1973b). In WB Deichmann, ed. Proceedings of the 8th Inter-American Conference on Toxicology: Pesticides and the Environment, A Continuing Controversy.
Multi-generation study on BALB/c mice; DDT at 0, 2, 20, 250 ppm. Survival among the males was poor, in part due to fighting; so results are given only for females. Results in the second paper were not in usable format for present purposes. Data are from Table III in the first paper; denominators are initial number of mice; liver cysts not counted.

Thorpe & Walker (1973). Fd Cosmet Toxicol 11 433.
CF1 mice. We analyze only the DDT results, at 0 or 100 ppm. Data from Table 2. Liver tumors (a+b) taken relative to effective number; at other sites, relative to initial number of animals. Survival at 21 months (Table 1).

Tomatis et al (1972). Int J Cancer 10 489.
Turusov et al (1973). JNCI 51 983.
Multi-generation study on CF1 mice. DDT at 0,2,10,50,250 ppm. Hemangioendotheliomas not counted.
Data in the second paper not in usable format.

Table 8. The sample of papers, continued.
The mouse, continued

Tomatis et al (1974). JNCI 52 883.
CF1 mice. Common control group. Three treatment groups:
i) 250 ppm DDD ii) 250 ppm DDE iii) 125 ppm DDD + 125 ppm DDE.

Walker et al (1973). Fd Cosmet Toxicol 11 415.
CF1 mice. Reports multiple experiments. We analyze only the DDT results, at 0, 50, 100 ppm. Data from Table 4. Liver tumors of type a & b are pooled, as are adenomas and carcinomas of the lung. Incidence rates are relative to the initial number of animals.

The rat (9 papers)

Cabral et al (1982). Tumori 68 11.
MRC Porton rats; DDT at 0,125,250,500 ppm. 80 week survival. Incidence rates relative to initial numbers.

Deichmann et al (1967). Toxicol Appl Pharmacol 11 88.
Osborne-Mendel rats; synergy experiment; we analyse only data on DDT, at 0 or 200 ppm. Survival at 24 months. Incidence rates relative to initial numbers.

Fitzhugh & Nelson (1947). J Pharmacol & Exp Ther 89 18.
Insufficient data for tabulation.

Lacassagne & Hurst (1965). Bull Cancer 52 89. No control group.

Nishizumi (1979). Gann 70 835.
Synergy experiment, reports only on DDT in conjunction with other carcinogens.

Radomski et al (1965). Toxicol Appl Pharmacol 7 652.
Osborne-Mendel rats; synergy experiment; we analyze only data on DDT, at 0 or 80 ppm. Incidence rates relative to initial numbers; benign and malignant tumors pooled.

Rossi et al (1977). Int J Cancer 19 179.
Wistar rats; DDT at 0 or 500 ppm; survival at 100 weeks.

Treon & Cleveland (1955). J Agric Fd Chem 3 402. No data.

Weisburger & Weisburger (1968). Fd Cosmet Toxicol 6 235. No data.

The hamster (2 papers)

Agthe et al (1970). Proc Soc Exp Med NY 134 113.
Reports only a small number of tumors, and not by site.

Cabral et al (1982). Tumori 68 5.
Syrian golden hamster; DDT at 0,125,250,500 ppm; survival at 50 weeks.

Most of the authors did address the issue of comparability in husbandry among the various test groups, but not in convincing detail. No paper discussed the issues created by multiple endpoints, or 'open' reading of slides. By contrast, much space is routinely spent describing comparative pathology of tumors, with illustrations-- clearly the topic of interest.

No experiment in Table 8 had two test species, although one did have two strains of mice. Only two of the papers summarized in Table 8, both from the same laboratory, explicitly mention randomization of animals to treatment. Since there are a variety of standard randomization schemes, we lean to the view that the other authors did not, in fact, randomize the animals to the various dose groups. (We can also report that in one major institution, 'randomization' means that a technician takes animals by hand out of a cage.)

Table 9 below attempts to analyze the sample of papers in a unified way. It is based on the chi-squared test for trend, as in Armitage (1955). In effect, the test regresses the site-specific incidence of cancer on the dose, weighting by the number of animals at risk, and divides the slope by its standard error, which is estimated on the hypothesis of binomial variation. If there are only two dose groups, the test coincides with the usual one for equality of two binomial probabilities.

Epidemiologists routinely use this procedure to see whether a response goes up with dose, or down, or sideways. Simplicity is its virtue, but it does not distinguish between linear or curvilinear responses. On the other hand, with only a limited number of dose groups, such distinctions may not be feasible.

The table reports the ratio Z of the estimated slope to its standard error. If Z is positive, the rate tends to go up with the dose, and DDT is harmful; if negative, the rate goes down, and DDT is protective. If Z is bigger in absolute value than 2, the effect is 'statistically significant'.

(In many cases, the sample size is so small that the asymptotics are only a rough guide to the significance level; Fisher's exact test was feasible, but seemed unnecessarily complicated for our purposes, which are largely descriptive; likewise for maximum likelihood estimates of potency.)

Table 9. A study of studies: the impact of DDT and its metabolites on mice, rats and hamsters. Z-tests for dose-response in death rates and tumor incidence rates by site.

Study		Deaths	Liver	Lungs	Lymphoma Leukemia	Osteoma	Kidneys	Testes Ovaries	Mammaries	Pituitary	Adrenals	Thyroid
						MICE						
Innes												
X	M	-1.4	5	.7	-1.1	?	?	?	?	?	?	?
X	F	4	4	-.8	.2	?	?	?	?	?	?	?
Y	M	-.4	4	0.0	1.3	?	?	?	?	?	?	?
Y	F	1.1	1.2	-.8	4	?	?	?	?	?	?	?
Kashyap												
	M	-.5	1.4	1.2	2.0	?	?	?	?	?	?	?
	F	0.0	1.7	1.0	2.0	?	?	?	?	?	?	?
Shabad		-.7	0.0	5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tarjan		?	?	?	6	?	?	?	?	?	?	?
Terracini 73a												
P	F	-.9	9	-1.6	-2.8	?	?	?	?	?	?	?
F1	F	1.5	14	.9	-4	?	?	?	?	?	?	?
Thorpe												
	M	.8	5	-1.9	-2.1	?	-.6	-.8	?	?	?	?
	F	2.0	6	-2.6	-1.5	?	-.6	.9	?	?	?	?
Tomatis 72												
P	M	3	5	-2.0	-2.2	-.1	-1.3	?	0.0	?	?	?
P	F	1.9	10	2.1	-1.2	-.8	-.5	-1.2	-.6	?	?	?
F1	M	4	6	-.2	-1.4	-1.6	-1.0	?	0.0	?	?	?
F1	F	2.0	10	-2.1	-.7	.7	0.0	-1.7	-.6	?	?	?
Tomatis 74												
DDE	M	7	5	-2.1	-4	.6	?	?	?	?	?	?
	F	8	12	-3	-5	-2.3	?	?	?	?	?	?
DDD	M	-1.3	2.3	4	-.7	-.3	?	?	?	?	?	?
	F	.2	.3	4	.6	-1.1	?	?	?	?	?	?
Mix	M	.7	5	-.3	-3	-.8	?	?	?	?	?	?
	F	1.7	10	-3	-.4	-.6	?	?	?	?	?	?
Walker												
	M	1.7	4	1.0	-.01	?	.05	2.0	?	?	?	?
	F	-.2	5	-2.0	-2.0	?	-1.3	-1.2	?	?	?	?

Table 9. Continued. A study of studies: the impact of DDT and its metabolites on mice, rats and hamsters. Z-tests for dose-response in death rates and tumor incidence rates by site.

Study		Deaths	Liver	Lungs	Lymphoma Leukemia	Osteoma	Kidneys	Testes Ovaries	Mammaries	Pituitaries	Adrenals	Thyroid
						RATS						
Cabral												
	M	1.2	.9	?	?	?	?	-.9	1.4	1.1	?	1.9
	F	.9	2.8	?	?	?	?	1.4	-2.9	-1.2	?	0.0
Deichmann												
	M	-.8	0.0	0.0	0.0	?	?	?	-1.0	?	?	?
	F	-2.7	0.0	0.0	-1.0	?	?	?	-1.7	?	?	?
Radomski												
	M	?	0.0	1.0	0.0	?	?	0.0	0.0	0.0	1.0	?
	F	?	0.0	1.7	0.0	?	?	0.0	-1.0	1.0	-1.0	?
Rossi												
	M	1.1	4	-.9	.1	?	?	-1.6	.2	?	-2.3	.3
	F	.3	5	.7	.1	?	?	.1	-.6	?	-1.4	1.1
						HAMSTERS						
Cabral												
	M	-.4	.3	?	?	?	?	?	?	0.0	1.6	1.4
	F	-.5	0.0	?	?	?	?	?	?	.2	2.2	.6
						SUMMARY						
Z-values												
+2.0 or more		7	21	5	4	0	0	1	0	0	1	0
+0.1 to +1.9		12	6	7	5	2	1	3	2	3	2	5
0.0 exactly		1	6	3	4	1	2	3	4	3	1	2
-0.1 to -1.9		10	0	7	9	7	6	6	6	1	2	0
-2.0 or less		1	0	7	8	1	0	0	1	0	1	0
???		3	1	5	4	23	25	21	21	27	27	27

Authors were not uniform in reporting survival data; often a table was provided, sometimes only a graph-- which we did our best to read. Where possible, 90-week survival was tested in Table 9; sometimes, another period had to be substituted.

There is a preference in the field for reporting tumor incidence by site and sex, so we followed suit. Authors were not at all uniform in choice of sites to report; a question mark in the table means no report for that site. Since authors will report on the sites with many tumors, a question mark suggests the lack of any carcinogenic response.

Some authors failed to report the sexes separately, and then results are given for all animals pooled. With one exception, we reported as separate experiments the separate generations in multi-generation studies: P is the parental generation, F1 the first generation of offsprings, etc.

Many authors-- but by no means all-- report the 'effective number', ie, the number of animals alive in each group at the time of the appearance of the first tumor in that group. This represents a partial adjustment for differential mortality in the test groups, especially due to the toxicity of the test substance. If the effective number is given, incidence rates are computed relative to it. Otherwise, the denominator is taken eg as the number of animals sent to necropsy, or the number of animals initially assigned to the group. (See Table 8 above for details.)

Tallies are shown at the bottom of Table 9, and are collected in Table 10 for all sites other than the liver, and all experiments. (In the last line of Table 10, there are 180 combinations of sites and sub-experiments where no tumor incidence rates were reported.)

Table 10. Z-statistics for dose-response in tumor incidence rates by site, other than the liver. Mice, rats, hamsters; DDT and metabolites.

+2.0 or more	11
+0.1 to +1.9	30
0.0 exactly	23
-0.1 to -1.9	44
-2.0 or less	18
no data	180

The binomial model behind the Z-test represents some idealization of the experimental results. It assumes randomization of animals to treatment and conditions of husbandry, and no observer bias. Of course, even if the chemical has no effect, the actual variation may be appreciably larger than binomial, for reasons indicated above.

Insofar as the binomial model has any validity, with respect to mice the data suggest that DDT shortens the lifespan and causes liver tumors. At other sites, and for other species, the picture can only be described as mixed. Indeed, taking the results of the bioassays at face value, DDT seems on balance to inhibit tumor development. (But see Rossi et al 1983 on DDE and hamsters; also Cabral 1985 on DDE and rats.) Other evidence on protective effects has already been presented (Haseman 1983); and for DDT itself, Okey (1972).

With respect to evidence for the carcinogenicity of DDT, much rides on the interpretation of liver tumors in mice. And for workers in the field, this is something of a controversial area. As the Office of Science and Technology Policy (1985) says:

> Critical to decisions about carcinogens is the biological significance and human relevance of certain types of tumors, particularly the liver tumors in the mouse. This matter has been the subject of heated debate for the past 15 years....

The discordance of the results in Table 9 is its message. Of course, even with a good protocol, maintaining quality control is difficult, and the difficulty increases with the number of animals; this may put an upper limit on the power of a biossay. For other discussions of these issues, see Gart et al (1985), IARC (1980), the National Toxicology Program (1984), the UK Department of Health and Social Security (1982).

None of the NTP or NCI bioassays turned up in the test set; the protocol for those bioassays seems much better than the standard in Table 8, especially with respect to multiple endpoints. However, for a review of the NCI bioassays along present lines, see Salsburg (1983). For a useful summary of a large set of bioassay results in standard format, see Gold et al (1984).

Much more care is needed in defining endpoints before the experiment starts. We also recommend using statistical analyses which recognize the multiple-endpoint problem explicitly. It may be helpful to pool results across sites and even sexes, testing whether the percentage of tumor-bearing animals in the different treatment groups is dose-related. In this regime, correction for other causes of mortality would be quite important.

Pooling is contrary to standard practice in the field, which calls for separate analysis by site and sex. See Haseman et al (1986), who review the impact of pooling on NTP test results; of course, a difference in results does not show that one rule or another is superior. The authors argue that stratification should increase statistical power; and that combining tumors of different types will not lead to biologically meaningful results.

Their first set of arguments may be dominated by the multiple endpoint problem, which compromises all bioassay results whether apparently significant or apparently insignificant. On the other hand, as they point out, background tumor rates are so high in rats (97% for males, 83% for females) that increases would be very hard to detect. The possibility of analysis by major tumor types might be worth exploring; equally, the value of testing on a strain with such a high spontaneous tumor rate would be worth reconsidering-- even for strong proponents of bioassays.

Haseman et al's biology argument provides an effective critique of extrapolations from the most sensitive site in test animals to all tumors in humans. That sort of worst-case analysis is common practice in risk assessment: indeed, as noted above, animal experiments do not predict the sites that will be affected in humans (Wilbourn et al 1986). If it makes no sense to pool results for the animals, it makes no sense to pool predictions for the humans. Also see US Environmental Protection Agency (1986).

Multiple-comparison techniques might be a useful supplement to pooled analyses, as heuristic aids in identifying the sites responsible for a dose-response relationship. (The NCI/NTP bioassay protocol already uses such methods.) Anomalies such as the appearance of rare tumors would also have suggestive value. However, given the complexities of real bioassays, statistical analysis cannot by itself screen out the artifacts. Replication is crucial.

To summarize our other recommendations: the necropsy work should be done 'blind' so far as possible. Strict attention should be paid to randomization, using computer-generated random numbers or the like to make assignments to the different cages and treatment groups.

This completes the discussion of the animal experiments; the next topic is the usual justifications for risk assessment.

7. What do others say?

There is a large scientific bureaucracy, in Washington and elsewhere, concerned with the regulation of chemicals on the basis of animal studies. What do they say in defense of the activity? The most interesting documents seem to be Environmental Protection Agency (1975), Food Safety Council (1980), IARC (1979, 1980, 1982, 1983), National Academy of Sciences (1975, 1978, 1980), Office of Science and Technology Policy (1985), Office of Technology Assessment (1977, 1981).

The Food Safety Council

The Food Safety Council (now defunct) was jointly funded by industry, consumer groups, and government, to review dose response modeling in the setting of safety standards. The Scientific Committee included Jerome Cornfield and John van Ryzin. Earlier, we reviewed some of the discussion in the report (Food Safety Council 1980). From pp711, 718, 730.

> Human risk assessment is a very inexact exercise, based largely upon theoretical assumptions concerning interspecies extrapolations. The uncertainties involved should be fully recognized by the scientific community and society.
>
> Regulatory decisions, however, must be made even in the absence of complete knowledge. Decisions based on informed scientific judgment, moreover, may be more easily criticized than those based on the systematic application of an objective set of decision making criteria which provide insofar as possible for the biological and statistical uncertainties involved. [emphasis in original]
>
> ....the low-dose extrapolated risk estimates are highly model dependent. Because of this inexactness [sic] of the behavior of the models in the low-dose range, plus the fact that they cannot be firmly justified on either statistical (goodness-of-fit, say) or biological grounds, the choice of how one does the extrapolation is primarily a matter of judgment.

In cruder terms, the argument comes down to this: the regulatory process must proceed, whether or not there is a suitable scientific basis for it. And it is better to avoid the appearance of subjectivity by deriving the risk estimates from a model, even though choosing the model is itself a critical step in the process, and a highly subjective one-- but not so visible.

The IARC

The IARC research program is well respected, and draws working groups of scientists from all over the world. However, cancer is more than a scientific puzzle, and the working groups seem to walk a fine line. The IARC (1982, p13) was fairly blunt about risk assessment:

> At present, no objective criteria exist to interpret data from studies in experimental animals or from short-term tests directly in terms of human risk.

Here they are, a year later (IARC, 1983, p18):

> In the absence of adequate data in humans it is reasonable, for practical purposes, to regard chemicals for which there is sufficient evidence of carcinogenicity in animals as if they presented a carcinogenic risk for humans. The use of the expressions 'for practical purposes' and 'as if they presented a carcinogenic risk' indicates that at the present time a correlation between carcinogenicity in animals and possible human risk cannot be made on a purely scientific basis, but only pragmatically. Such a pragmatical correlation may be useful to regulatory agencies in making decisions related to the primary prevention of cancer. [emphasis in original.]
>
> In the present state of knowledge, it would be difficult to define a predictable relationship between the dose (mg/kg bw/day) of a particular chemical required to produce cancer in test animals and the dose which produce a similar incidence of cancer in humans. Some data, however, suggest that such a relationship may exist [Rall, 1977, National Academy of Sciences 1975], at least for certain classes of carcinogenic chemicals, but no acceptable methods are currently available for quantifying the possible errors that may be involved in such an extrapolation procedure.

The first paragraph still makes our point, if diplomatically. The second paragraph only says that while it may be possible to extrapolate from the mouse to man, at present it is impossible to estimate the resulting errors.

In the passage quoted above, as in many other such official documents, the NAS (National Academy of Sciences, 1975) is cited to support extrapolations from animal data. This NAS study is based on a review of risk assessments for six chemicals. The observed differences ranged up to a factor of 500. That still understates the problem, since the human risk 'data' are themselves obtained by modeling, which appears to force some agreement on the estimates. For other studies by the NAS, see below. Rall (1977), also cited by the IARC, is not persuasive either. Indeed, Rall presents only two systematic surveys: one is the NAS (1975), and the other is about toxicity not carcinogenicity.

The OSTP

The Office of Science and Technology Policy (OSTP) is the President's advisory group. They reviewed cancer risk assessment technologies in 1985 (also see US Interagency Staff Group on Carcinogens 1986). Here is what they have to say:

> Many components of the risk assessment process lack definitive scientific basis....
>
> Many scientists would agree that, while there is a significant amount of evidence to support qualitative animal to human extrapolation for carcinogenesis, the evidence falls short of establishing this proposition as a scientific fact (when determining the response of different species to chemicals, many chemicals appear to be carcinogenic in one species or strain and not in another, even when only rodents are being compared). Nonetheless, this principle has been accepted by all health and regulatory agencies and is regarded widely by scientists in industry and academia as a justifiable and necessary inference.

At first reading, the OSTP seems to be contradicting itself. However, 'justifiable and necessary' only means that they can't do risk assessment otherwise.

The OTA

The Office of Technology Assessment (OTA) is the advisory group for Congress. They have reviewed the cancer risk assessment technology at least twice: in 1977 when the saccharin controversy was raging, and again in 1981. In 1977, they were quite optimistic about the technology; even so, here they are on p82:

> It is generally accepted that an animal carcinogen is also a human carcinogen. Extrapolation between cancer incidence in animals and expected incidence in humans is necessary to quantify the risk for human populations from exposure to a chemical.

'Necessary' is to be distinguished from 'scientifically defensible'.

By 1981, the OTA was much more critical of the technology--although still for it. From the OTA (1981, pp12-13, 113, 122):

> Opinions differ about whether and how extrapolation methods should be used in estimating the amount of human cancer that might be caused by exposure to a carcinogen....The disagreements among the groups who hold different opinions about use of extrapolation methods are vocal and current.
>
> The fact that some regulations are based on nonhuman test systems shows that proof that a chemical is a human carcinogen is not demanded. This illustrates that prevention of cancer is seen as so important that it is appropriate to make decisions to restrict exposures before human damage is observed.
>
> A substantial body of experimentally derived knowledge and the preponderance of expert opinion support the conclusion that testing of chemicals in laboratory animals provides reliable information about carcinogenicity.

With respect to the last paragraph, the 'preponderance of expert opinion' is hard to assess. But the 'substantial body of experimentally derived knowledge' seems quite ambiguous on the crucial question of extrapolating risk estimates from laboratory animals to people.

The EPA

The Environmental Protection Agency (EPA) is one of the lead agencies in regulating carcinogenic hazards. In 1972, they banned DDT; and in 1975, they wrote a white paper justifying the decision. From pp87 and 252:

> Although the target tissue may be different, the mouse can, in specific cases, serve as a reliable and proven indicator of the carcinogenicity of a chemical in other species including man. However, although carcinogenic effects in mice are valid when dealing with certain chemicals, the results can vary greatly depending on the compound tested and may not always be a reliable basis for extrapolation to other species.
>
> Proponents argue that DDT has a good human health record and that alternatives to DDT are more hazardous to the user and more costly. Opponents to DDT, admitting that there may be little evidence of direct harm to man, emphasize other hazards connected with its use.

In sum, the mouse is a 'reliable and proven indicator' of carcinogenicity, which works for some compounds but not others, perhaps not even for DDT.

The Saffiotti report

An ad hoc committee of fairly determined bioassay proponents expressed its views in a 1970 report to the Surgeon General. The chairman was Umberto Saffiotti, then Associate Scientific Director for Carcinogenesis at the National Cancer Institute. (This passage is quoted in Epstein 1979.)

> Evidence of negative results, under the conditions of the test used, should be considered superseded by positive findings in other tests. Evidence of positive results should remain definitive, unless and until new evidence conclusively proves that the prior results were not causally related to the exposure.
>
> In order to evaluate the hazard of a chemical for man, one must extrapolate from the animal evidence. It is essential to recognize that no level of exposure to a carcinogenic substance, however low it may be, can be established to be a 'safe level' for man.
>
> In recognizing a chemical as a carcinogen, the limiting factor is the sensitivity and specificity of the bioassay system used. A bioassay system designed to detect tumor induction only at or above a given level under the conditions of the test (eg, a 25 percent incidence of a specific tumor type) will fail to reveal carcinogenicity below that level. Compounds whose carcinogenic effects fall below specific bioassay detection limits must not be considered innocuous.
>
> Chemicals should be subjected to scientific scrutiny rather than given individual 'rights'; they must be considered potentially guilty unless and until proven innocent.

The view seems to be that chemicals are carcinogenic until proven otherwise, and proof of innocence is almost impossible. The dictum that positive evidence supersedes negative is hard to justify, given the degree of inconsistency within the animal experiments, or the conflict between animal data and epidemiology.

The NAS

The National Academy of Sciences (NAS) conducts studies for government agencies, using ad hoc panels. They reviewed cancer testing technology in 1975, 1978, 1980, and 1983. The 1975 report was fairly positive; the 1978 and 1983 reports, quite guarded. The 1980 report (especially chap 4 and appendix A) was extremely critical; eg, from pp81-83:

> ....current understanding of carcinogenesis and related pathologies is not adequate to permit reliable extrapolations from animal experimentation and simpler assay systems to actual quantified hazards to human health....at least two extrapolations of inadequately tested reliability must generally be applied to bioassay data to derive estimates of human cancer incidence.... inferences drawn by means of current extrapolation methods lack scientific justification....the provision of a sophisticated quantitative estimate of human cancers provides a high potential for misinterpretation because the estimates may be used without the required attention to the inherent constraints....users are so hungry for numbers that quantitative estimates, once presented, take on a life and authority of their own, despite all the reservations that [the analyst] may attach to them....

Other authors

Ames

Ames (1983) gives a reductio-ad-absurdum argument against quantitative risk assessment. In brief, every day people ingest 'natural' carcinogens and mutagens: eg, M-IQ in broiled hamburger, aflatoxins in peanut butter, phorbol esters in herb teas, theobromine in cocoa, safrole and piperine in sassafras or black pepper, hydrazines in mushrooms, furocoumarins in celery.

The methodology of quantitative risk assessment-- extrapolation from bioassays and short-term tests-- shows that the risks from the natural carcinogens dominate the risks from environmental contamination by chemicals. (Cigarette smoking, some drugs, and some occupational exposures do present extraordinary hazards.) Also see Ames-Magaw-Gold (1987), Felton & Hatch (1986), Knudsen (1986), and the symposium report in the August, 1986 issue of Environmental Health Perspectives.

Doll & Peto (1981, pp1215-16)

> ....animal feeding studies have great value in certain circumstances but may not offer an uncomplicated and straightforward means of discovering preventable causes for the majority of human cancers, and at the very least it certainly does not seem likely that they can offer a reliable means of estimating quantitative human hazards.
>
> If our perspective on both short-term and animal tests is accepted, then quantitative human 'risk assessment', as currently practiced, is so unreliable, suffering not only from random but also probably from large systematic errors of unknown direction and magnitude, that it should definitely be given another name: 'Priority setting' might perhaps be a more honest, although less saleable, name.

Tomatis (1977, pp1349 & 1352)

We all agree that the mouse should be discarded as a testing tool if a better experimental model can be found. At present, however, it seems that it is no worse qualified than any other species for detecting the carcinogenicity of environmental chemicals and for predicting a possible human hazard.

It is clear that, at present, there is no general consensus on the validity of using experimental results to predict human hazards.

Tomatis-Breslow-Bartsch (1980)

The good empiric correlation between human and experimental animal data for the limited number of chemicals for which both human and experimental data are available indicates, as shown previously, that experimental animal data may predict a qualitatively similar response in humans, although the validity of this empiric correlation cannot be extended to predict possible quantitative variations of that response in different species. The data obtained in animal tests may, however, represent different degrees of evidence of a carcinogenic effect, this drawback being due mainly to our insufficient knowledge of the mechanisms of carcinogenesis.

The fact that the appropriateness of the various models is a matter of considerable debate within the scientific community, and that few relevant data are available to the proponents of either side, serves to emphasize the uncertainties inherent in the process of extrapolation.

8. Conclusions

Bioassays

The IARC (1980) guidelines on the conduct of bioassays seem quite sound, and we wish they were more often followed in practice. Since bioassays are inherently statistical, randomization is critical-- to treatment groups and to conditions of husbandry. 'Blinding' the necropsy work would also be a valuable precaution.

The multiple endpoint problem is quite serious. A possible solution would be to pool results across sites and even sexes, and test whether the percentage of tumor-bearing animals in the different treatment groups is dose-related. In this regime, correction for other causes of mortality would be quite important.

If the idea of pooling is accepted, multiple-comparison techniques might be a useful supplement, at least as heuristic aids in identifying the sites responsible for a dose-response relationship. Anomalies such as the appearance of rare tumors would also have suggestive value. However, given the complexities of real bioassays, statistical analysis cannot by itself screen out the artifacts. Replication is crucial.

Qualitative extrapolation

If a substance is carcinogenic in a bioassay, we think that is some evidence for carcinogenic potential in humans. If the bioassay was well run, the evidence is stronger. Replicability across experiments and across species makes the case even stronger. Conversely, flaws in the experiment or failure to replicate weaken the argument. As Wald & Doll (1985, p225) say:

> Only one rule is absolute: that all the available evidence must always be taken into account.

The validity of the qualitative extrapolation seems to be a topic on which much useful research could be done. A feasible-- but expensive-- program would call for a series of good bioassays on a representative sample of agents and species, with the object of measuring inter-species differences in sensitivity. Further research into the biological parameters which determine species susceptibility would be even more useful.

Quantitative extrapolation

In the present state of the art, making quantitative assessments of human risk from animal experiments has little scientific merit. Valid extrapolations would be possible only on the basis of mathematical models grounded in biological reality, and carefully tested against empirical data.

As presently formulated, public policy depends on quantitative risk assessment. This guarantees a steady supply of such assessments; and an equally steady supply of apologetics written by scientific oversight committees who ask, "What better technology is there?" A wiser course might be to reformulate the policies so that regulation could be accomplished on the basis of what regulators actually knew, rather than what they wished they knew.

Policy implications

We find that the models now used in risk assessment do not have much by way of scientific foundation, yet we do not propose new models. This position cannot be agreeable to workers in the field, or to anyone wanting statistics used in settling public-policy questions. On the other hand, a disagreeable position may be right.

We sympathize with the goal of bringing some statistical order into cancer prevention, and regulating chemicals on the basis of extrapolation from animal experiments. We also like the idea of advancing biological knowledge through statistical models. Our concern is the feasibility of such enterprises, given the present limits to knowledge in biology and statistics.

At one time, the multi-stage model seemed like a promising avenue to explore, and it has lead to some good research. But in the end, the scientific claims made for that model (and for others like it) must be judged by ordinary scientific standards. If we are right about the technical issues, quantitative risk assessment cannot be justified on those standards. It may be advisable to give up the pretense of a scientific foundation where none exists.

An objective procedure for licensing chemicals may be needed, and some well-defined version of risk assessment may be the answer. In the long run, as biological understanding develops, better models may become available. In the short term, the arbitrariness in the modeling approach can be reduced only by administrative fiat. We tend to be suspicious of that sort of science-by-decree, because it leads to a spurious sense of precision.

Governments have to make many crucial decisions in a rough and ready way, including public-health decisions. We see no evidence that regulatory modeling leads to better decisions than informal argument, and find the latter more appealing because it brings the uncertainties into the open. The factual basis for decision-making could be improved by putting more resources into epidemiology, or basic research on the causes of cancer and the origins and magnitudes of species differences. Either way, obscuring the scientific uncertainties cannot be good public policy.

References

RH Adamson & SM Sieber (1983). Chemical carcinogenesis studies in nonhuman primates. In R Langenbach, S Nesnow & JM Rice, eds. Organ & Species Specificity in Chemical Carcinogenesis. Plenum, NY

American Cancer Society (1986). Cancer facts and figures.

BN Ames (1983). Dietary carcinogens and anticarcinogens. Science, 221, 1256-1264. With a reply by Epstein & Swartz, and a rejoinder by Ames.

BN Ames, R Magaw & LS Gold (1987). Ranking possible carcinogenic hazards. Science 236 271-279

P Armitage (1955). Tests for linear trends in proportions and frequencies. Biometrics 11 375-86

P Armitage & R Doll (1961). Stochastic models for carcinogenesis. In L LeCam & J Neyman, eds. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. UC Press, Berkeley.

MS Arnott, T Yamauchi & DA Johnston (1979). Aryl hydrocarbon hydroxylase in normal and cancer populations. In AC Griffin & CR Shaw, eds. Carcinogens: Identification and Mechanisms of Action. Raven, NY.

JC Bailar III & EM Smith (1986). Progress Against Cancer? New England J Med 314 1226-32

L Bernstein, LS Gold, BN Ames, MC Pike & DG Hoel (1985). Some tautologous aspects of the comparison of carcinogenic potency in rats and mice. Fund Appl Toxicol 5 79-86

JM Bishop (1987). The molecular genetics of cancer. Science 235 305-311

W Bloom & DW Fawcett (1962). A Textbook of Histology. Saunders.

RK Boutwell (1964). Some biological aspects of skin carcinogenesis. Progr Exp Tumor Res 4 20750

RK Boutwell (1984). On the role of tumour promotion in chemical carcinogenesis. In Models, Mechanisms and Etiology of Tumour Promotion. IARC Sci Publ No 56. Lyon.

CC Brown & K Chu (1986). Use of multistage models to infer stage affected by carcinogenic exposure: example of lung cancer and cigarette smoking. To appear in J Chron Dis.

JRP Cabral (1985). DDT: laboratory evidence. In NJ Wald and R Doll, eds. Interpretation of Negative Epidemiological Evidence for Carcinogenicity. IARC Sci Publ No 65. Lyon.

J Cairns (1979). Cancer: Science and Society. Freeman, San Francisco.

J Cairns (1981). The origins of human cancers. Nature 289 353.

J Cairns (1985). The treatment of diseases and the war against cancer. Scientific American 253 51-59.

EJ Calabrese (1984). Suitability of animal models for predictive toxicology: theoretical and practical considerations. Drug Metabolism Reviews 15 505-23.

FW Carlborg (1981a). Multi-stage dose response models in carcinogenesis. Fd Cosmet Toxicol 19 361

FW Carlborg (1981b). 2-acetylaminofluorene and the Weibull model. Fd Cosmet Toxicol 19 367

DB Clayson (1975). Nutrition and experimental carcinogenesis: a review. Cancer Research 35 3292-3300

DB Clayson (1978). Overview, fact, myth and speculation. J Environ Path Toxicol 2 1-8

DB Clayson (1985). Dose response relationships in experimental carcinogenesis: dependence on multiple factors including biotransformation. J Toxicol Pathol 13 119-127.

DB Clayson (1986). Needs for biological risk assessment ~~in~~ interspecies extrapolation. Unpublished technical report. Health and Welfare Canada, Tunney's Pasture, Ottawa

DB Clayson, D Krewski & I Munro, eds (1985). Toxicological Risk Assessment. two vols. CRC Publishing, Florida.

GM Cohen et al (1979). Anticarcinogenic effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin on benzo(a)pyrene and 7,12-dimethylbenz(a)anthracene tumor initiation and its relationship to DNA binding. Cancer Res 39 4027-4033

SM Cohen et al (1984). Urinary bladder carcinogenesis. In Models, Mechanisms and Etiology of Tumour Promotion. IARC Sci Publ No 56. Lyon.

JM Concon, TW Swerczeck & DS Newburg (1979). Black pepper (piper nigrum): evidence of carcinogenicity. Nutrition and Cancer 1 22-6

PJ Cook, R Doll, SA Fellingham (1969). A mathematical model for the age distribution of cancer in man. Int J Cancer 4 93-112

J Cornfield (1977). Carcinogenic risk assessment. Science 198 693

E Crouch & R Wilson (1979). Interspecies comparison of carcinogenic potency. J Toxicol Environ Health 5 1095-1118

E Crouch, R Wilson & L Zeise (1987). Tautology or not tautology? J Toxicol Environ Health 20 1-10

K Crump & RB Howe (1984). The multistage model with a time-dependent dose pattern: applications to carcinogenic risk assessment. J Risk Analysis 4 163-176

NE Day (1984). Epidemiological data and multistage carcinogenesis. In Models, Mechanisms and Etiology of Tumour Promotion. IARC Sci Publ No 56. Lyon.

FJ DiCarlo (1984). Carcinogenesis bioassay data: correlation by species and sex. Drug Metabolism Reviews 15 409-413

J DiGiovanni, TJ Slaga, DL Berry, MR Juchau (1980). Inhibitory effects of environmental chemicals on polycyclical aromatic hydrocarbon carcinogenesis. In TJ Slaga, ed. Modifiers of chemical carcinogenesis. Raven, NY, 145-168.

R Doll & R Peto (1978). Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidem and Comm Health 32 303-313

R Doll & R Peto (1981). The Causes of Cancer. JNCI 66 1191-1309 Reprinted as an Oxford University Press paperback.

JF Douglas (1984). Carcinogenesis and mutagenesis testing. Humana Press, Clifton, NJ.

WH DuMouchel and JE Harris (1983). Bayes methods for combining the results of cancer studies in humans and other species. JASA 78 293-315

E Efron (1984). The Apocalyptics. Simon & Schuster, New York.

Environmental Protection Agency (1975). DDT: A Review of Scientific and Economic Aspects of the Decision to Ban its Use as a Pesticide. Washington, DC.

SS Epstein (1979). The Politics of Cancer. Anchor Press.

E Farber (1984). Cellular biochemistry of the stepwise development of cancer with chemicals. Cancer Research 44 5463-5474

JS Felton & FT Hatch (1986). Mutagens in Cooked Food, Energy & Technology Review, Lawrence Livermore National Laboratory, Calif.

Food Safety Council (1980). Quantitative risk assessment: Report of the Scientific Committee. Fd Cosmet Toxicol 18 711-34

LM Franks & N Teich (1986). Introduction to the Cellular and Molecular Biology of Cancer. Oxford University Press.

D Freedman & W Navidi (1987). The multi-stage model for cancer. Manuscript in preparation. Department of Statistics, University of California, Berkeley.

JJ Gart et al (1985). The Design and Analysis of Long-term Animal Experiments. Draft report. IARC, Lyon.

JBM Gellatly (1975). The natural history of hepatic parenchymal nodule formation in a colony of C57BL mice with reference to the effect of diet. In WH Butler & PM Newberne, eds. Mouse Hepatic Neoplasia. Elsevier.

JB Gibson (1971). Parasites, liver disease, and liver cancer. In Liver Cancer. IARC Sci Publ No 1. Lyon.

LS Gold et al (1984). A carcinogenic potency database of the standardized results of animal bioassays. Environ Health 58 9.

LS Gold et al (1986). An empirical comparison of methods used to estimate carcinogenic potency in long-term animal bioassays: lifetable vs sunmmary incidence data. Fund Appl Toxicol 6 263-269

LS Gold et al (1986). Reproducibility of results in "near-replicate" carcinogenesis bioassays. To appear in JNCI.

P Grasso & RF Crompton (1972). The value of the mouse in carcinogenicity testing. Fd Cosmet Toxicol 10 418

HC Grice, IC Munro & DR Krewski (1981). In utero exposure in chronic toxicity/carcinogenicity studies. Fd Cosmet Toxicol 19 373-9.

RA Griesemer & C Cueto (1980). Toward a classification scheme for degrees of experimental evidence for the carcinogenicity of chemicals for animals. In R Montesano, H Bartsch & L Tomatis, eds. Molecular and Cellular Aspects of Carcinogen Screening Tests. IARC, Lyon.

JK Haseman (1983). Patterns of tumor incidence in two-year cancer bioassay feeding studies in Fischer 344 rats. Fund Appl Toxicol 3 1-9

JK Haseman & DG Hoel (1979). Statistical design of toxicity assays: role of genetic structure of test animal population. J Toxicol Environ Health 5 89-101

JK Haseman et al (1984). Results from 86 two-year carcinogenicity studies conducted by the National Toxicology Program. J Toxicol Environ Health 14 621-39

JK Haseman, EC Tharrington, JE Huff, & EE McConnell (1986). Comparison of site-specific and overall tumor incidence analyses for 81 recent National Toxicology Program carcinogenicity studies. Regulatory Toxicology and Pharmacology 6 155-170

E Hecker (1984). Cocarcinogens of the tumour promoter type as potential risk factors of cancer in man. In Models, Mechanisms and Etiology of Tumour Promotion. IARC Sci Publ No 56. Lyon.

H Hennings et al (1983). Malignant conversion of mouse skin tumors is increased by tumour initiators and unaffected by tumour promoters. Nature 304 67-68

J Higginson (1979). Perspectives and future developments in research on environmental carcinogenesis. In AC Griffin & CR Shaw, eds. Carcinogens: Identification and Mechanisms of Action. Raven Press, NY.

DG Hoel, NL Kaplan & MW Anderson (1983). Implication of nonlinear kinetics on risk estimation in carcinogenesis. Science 219 1032-37.

DG Hoel & CJ Portier (1987). Nonlinearity of carcinogenesis dose-response functions. To appear in Science.

IARC (1974). On the Evaluation of Carcinogenic Risk of Chemicals to Man: Some Organochlorine Pesticides. Mon 4. Lyon

IARC (1979). Chemicals and Industrial Processes Associated with Cancer in Humans. Mon Suppl 1. Lyon.

IARC (1980). Long-Term and Short-Term Screening Assays for Carcinogens: A Critical Appraisal. Mon Suppl 2. Lyon.

IARC (1982). On the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Mon Suppl 4. Lyon.

IARC (1983). Miscellaneous Pesticides. Mon No 30. Lyon.

IARC (1984). Models, Mechanisms and Etiology of Tumour Promotion. Sci Pub No 56, esp the articles by Cohen et al, Hecker, Scherer et al and Yuspa.

IARC (1987). On the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Mon Suppl 6. Lyon. In press.

IARC (1988). On the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Mon Suppl 7. Lyon. In press.

OH Iversen and EG Astrup (1984). The paradigm of two-stage carcinogenesis: a critical attitude. Cancer Investigation 2 51-60

DG Jose (1979). Dietary deficiency of protein, amino acids and total calories on development and growth of cancer. Nutrition and Cancer 1 58-63

JD Kalbfleisch, DR Krewski & J van Ryzin (1983). Dose-response models for time-to-response toxicity data. Canadian J Statistics 11 25-45

JM Kaldor & NE Day (1986). Interpretation of epidemiological studies in the context of the multi-stage model of carcinogenesis. Draft technical report, IARC, Lyon.

Ib Knudsen ed (1986). Genetic Toxicology of the Diet. Liss. Esp the article by T Sugimura et al at p85, Overview: mutagens and carcinogens in cooked food.

D Krewski & J van Ryzin (1981). Dose response models for quantal response toxicity data. In M Csorgo et al, eds. Statistics and Related Topics. North Holland.

S Lagakos & F Mosteller (1981). A case study of statistics in the regulatory process: the FD&C Red No 40 experiments JNCI 66 197.

LB Lave (1987). Health and safety risk analyses: information for better decisions. Science 236 291-295

A Likhachev, V Anisimov, R Montesano, eds (1985). Age-related Factors in Carcinogenesis. IARC Sci Publ No 58, Lyon.

AM & DE Lilienfeld (1980). Foundations of Epidemiology. Oxford University Press.

B MacMahon (1979). Strengths and limitations of epidemiology. In National Research Council in 1979: Current Issues and Studies.

MJ McFarlane, AR Feinstein & CK Wells (1986). Necropsy evidence of detection bias in the diagnosis of lung cancer. Arch Intern Med 146 1695-8.

GE Metter et al (1985). Morphological subclassification of follicular lymphoma: variability of diagnoses among hematopathologists. J Clin Oncology 3 25-38.

EC Miller, JA Miller & RR Brown (1952). On the inhibitory action of certain polycyclic hydrocarbons on azo dye carcinogenesis. Cancer Res 12 282-283

JA Miller (1970). Carcinogenesis by chemicals: an overview. Cancer Research, 30, 559

JA & EC Miller (1977). Ultimate chemical carcinogens as reactive mutagenic electrophiles. In HH Hiatt, JD Watson, & JA Winsten, eds. Cold Spring Harbor Conferences on Cell Proliferation, Vol 4 pp605-27

JC Mirsalis et al (1985). Induction of hepatic cell proliferation and unscheduled DNA synthesis in mouse hepatocytes following in vivo treatment. Carcinogenesis, 6, 1521-24

R Montesano, H Bartsch & L Tomatis, eds (1980). Molecular and cellular aspects of carcinogen screening tests. IARC, Lyon.

SH Moolgavkar (1986). Carcinogenesis modeling: from molecular biology to epidemiology. Ann Rev Public Health 7 151-169.

SH Moolgavkar, NE Day & RG Stevens (1980). Two-stage models for carcinogenesis: epidemiology of breast cancer in females. JNCI 65 559-69

SH Moolgavkar & AG Knudson (1981). Mutation and cancer: a model for human carcinogenesis. JNCI 66 1037-52

DH Moore et al (1982). The effect of dose and vehicle on early tissue damage and regenerative activity after chloroform administration to mice. Fd Chem Toxicol 20 951-4.

G Murasaki et al (1984). Inhibition by aspirin of N-[4-(5-nitro-2-furyl)-2-thiazolyl]formamide-induced bladder carcinogenesis and enhancement of forestomach carcinogenesis. Carcinogenesis 5 53-5

National Academy of Sciences (1975). Contemporary pest control practices and prospects: the report of the executive committee. Washington, DC.

National Academy of Sciences (1978). Saccharin: Technical Assessment of Risks and Benefits. National Research Council. Committee for Study on Saccharin and Food Safety Policy. PB 292 695. Washington, DC.

National Academy of Sciences (1980). Regulating Pesticides: Report of the Committee on Prototype Explicit Analyses for Pesticides. Washington, DC.

National Academy of Sciences (1983a). Managing the process: Risk Assessment in the Federal Government. Washington, DC.

National Academy of Sciences (1983b). Diet, Nutrition, and Cancer. Washington, DC.

National Academy of Sciences (1984). Toxicity Testing. Washington, DC.

National Academy of Sciences (1987). Regulating Pesticides in Foods: The Delaney Paradox. Washington, DC.

National Cancer Institute (1978). Bioassays of DDT, TDE, and p,p'-DDE for Possible Carcinogenicity. Technical Report 131. Washington, DC.

National Toxicology Program Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation (1984). Washington, DC.

LB Novey et al (1986). Causation and Financial Compensation. Conference Report. Institute for Health Policy Analysis, Georgetown University Medical Center.

Office of Science and Technology Policy (1985). Chemical Carcinogens: A Review of the Science and its Associated Principles. Federal Register, Thursday, March 14, Part II, pp 10371-10442.

Office of Technology Assessment (1977). Cancer Testing Technology and Saccharin. OTA-H-55. Washington.

Office of Technology Assessment (1981). Assessment of the Technologies for Determining Cancer Risks From the Environment. OTA-H-138. Washington.

AB Okey (1972). Dimethylbenzanthracene-induced mammary tumors in rats: inhibition by DDT. Life Sciences 11 833-843.

J Peto (1984). Early- and late-stage carcinogenesis in mouse skin and man. In Models, Mechanisms and Etiology of Tumour Promotion. IARC Sci Publ No 56. Lyon.

R Peto (1977). Epidemiology, multistage models, and short-term mutagenicity tests. In HH Hiatt, JD Watson, & JA Winsten, eds. Origins of Human Cancer, Book C, p1403. Cold Spring Harbor.

R Peto (1980). Distorting the epidemiology of cancer: the need for a more balanced overview. Nature 284 297. With a reply by SS Epstein & JB Swartz (1981). Fallacies of lifestyle cancer theories. Nature 289 127.

R Peto, FJC Roe, PN Lee, L Levy & J Clack (1975). Cancer and ageing in mice and men. Br J Cancer 32 411-426.

MC Pike (1982). Some general remarks on cancer epidemiology with special reference to environmental causes. In JP Middaugh, ed. Energy, Health, and the Environment, p72-83; answers the article by Epstein, p36-51

MC Pike (1983). Hormonal risk factors, breast tissue age and the age-incidence of breast cancer. Nature 303 767

IF Purchase (1980). Inter-species comparisons of carcinogenicity. Br J Cancer 41 454-468

K Rai & J van Ryzin (1979). Risk assessment of toxic environmental substances using a generalized multi-hit dose response model. In N E Breslow & A S Whittemore, eds. Energy and Health, SIAM, Philadelphia.

K Rai & J van Ryzin (1981). A generalized multi-hit dose-response model for low-dose extrapolation. Biometrics 37 341-52

DP Rall (1977). Species differences in carcinogenesis testing. In HH Hiatt, JD Watson, & JA Winsten, eds. Origins of Human Cancer, Book C, Cold Spring Harbor, NY.

RB Reger, DF Butcher & WKC Morgan (1973). Assessing change in the pneumoconioses using serial radiographs: sources and quantification of bias Am J Epid 98 243-254

RB Reger, MR Petersen & WKC Morgan (1974). Variation in the interpretation of radio- graphic change in pulmonary disease. Lancet i 111-113

HL Richardson, AR Stier & E Borsos-Nachtnebel (1952). Liver tumor inhibition and adrenal histologic responses in rats to which 3'-methyl-4-dimethyl- aminoazobenzene and 20-methylcholanthrene were simultaneously administered. Cancer Res 12 356-361.

FJC Roe (1981). Testing in vivo for general chronic toxicity and carcinogenicity. In JW Gorrod, ed. Testing for Toxicity. Taylor & Francis, London pp29-43

L Rossi et al (1983). Carcinogenicity study with technical-grade [DDT] and [DDE] in hamsters. Cancer Research 43 776-81

H Rubin (1980). Is somatic mutation the major mechanism of malignant transformation? JNCI 64 995

M Russell & M Gruber (1987). Risk assessment in environmental policy-making. Science 236 286-290

U Saffiotti et al (1970). Ad Hoc Committee on the Evaluation of Low Levels of Environmental Chemical Carcinogens. Report to the Surgeon General, USPHS, Washington, DC.

DS Salsburg (1983). The lifetime feeding study in mice and rats-- an examination of its validity as a bioassay for human carcinogenesis. Fund Appl Toxicol 3 63-67.

C Sawyer, R Peto, L Bernstein, MC Pike (1984). Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. Biometrics 40 27-40.

DM Shankel, P Hartman, T Kada & A Hollaender, eds. (1986). Antimutagens and Anticarcinogens: Mechanisms. Plenum, NY.

P Shubik & DB Clayson (1976). Application of the results of carcinogen bioassays to man. Inserm 52 241-252.

EE Siegler (1956). Microdiagnosis of the carcinoma in situ of the uterine cervix. Cancer 9 463-469.

CC Smith (1986). Animal studies: how well do they predict xenobiotic metabolism in humans? In WE Lloyd, ed. Safety Evaluation of Drugs and Chemicals. Hemisphere, Washington.

RS Sohal, LS Birnbaum & RG Cutler, eds (1985). Molecular Biology of Aging. Raven Press, NY

TJ Slaga (1983). Overview of tumor promotion in animals. Environ Health 50 3-14

JA Staffa & MA Mehlman, eds (1980). Innovations in Cancer Risk Assessment (ED01 study). J Environ Pathol Toxicol 3 1-246.

WT Stott et al (1981). Genetic and nongenetic events in neoplasia. Fd Cosmet Toxicol 19 567-576.

SH Swan & DB Petitti (1982). A review of problems of bias and confounding in epidemiological studies of cervical neoplasia and oral contraceptive use. Am J Epid 115 10-18

A Tannenbaum (1940). A relationship of body weight to cancer incidence. Arch Pathol 30 509.

A Tannenbaum (1942a). Initiation and growth of tumors. I. Effects of underfeeding. Am J Cancer 38 335

A Tannenbaum (1942b). The genesis and growth of tumors. II. Effects of caloric restriction per se. Cancer Research 2 460.

A Tannenbaum (1942c). The genesis and growth of tumors. III. Effects of a high-fat diet. Cancer Research 2 468.

TG Tape & RJ Panzer (1986). Clinical information biases the interpretation of imaging tests. J Internal Medicine, 1 300-304.

L Tomatis (1977). The value of long-term testing for the implementation of primary prevention. In HH Hiatt, JD Watson, & JA Winsten, eds. The Origins of Human Cancer, Cold Spring Harbor, NY pp1339-57

L Tomatis (1979). The predictive value of rodent carcinogenicity in the evaluation of human risks. Ann Rev Pharmacol Toxicol 19· 511-30

L Tomatis, C Partensky & R Montesano (1973). The predictive value of mouse liver tumour induction in carcinogenicity testing-- a literature survey. Int J Cancer 12 1-20

L Tomatis et al (1974). The effect of long-term exposure to [DDE and DDD], and the two chemicals combined on CF1 mice. JNCI 52 883

L Tomatis, NE Breslow & H Bartsch (1980). Experimental studies in the assessment of human risk. In D Schottenfeld & JF Fraumeni, eds. Cancer Epidemiology & Prevention. Saunders. pp43-73

FH Tschirely (1986). Dioxin. Scientific American, February. pp29-35

VS Turusov et al (1973). Tumors in CF-1 mice exposed for six consecutive generations to DDT. JNCI 51 983-997.

UK Department of Health and Social Security (1981). Report of the British Committee on Mutagenicity of Chemicals in Food, Consumer Products and the Environment. Guidelines for the Testing of Chemicals for Mutagenicity. HMSO, London.

UK Department of Health and Social Security (1982). Report of the British Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment. Guidelines for the Testing of Chemicals for Carcinogenicity. HMSO, London.

US Department of Commerce (1986). Bureau of the Census. Money Income of Households, Families, and Persons in the United States: 1984. Current Population Reports, Consumer Income, Series P-60, No. 151. Washington, DC.

US Environmental Protection Agency (1986). Guidelines for Carcinogenic Risk Assessment. Federal Register 51 3392

US Interagency Staff Group on Carcinogens (1986). Chemical carcinogens: a review of the science and its associated principles. Environ Health 67 201-282.

NJ Wald and R Doll, eds (1985). Interpretation of negative Epidemiological Evidence for Carcinogenicity. IARC Sci Publ No 65, Lyon.

JM Ward (1984). Pathology of toxic, preneoplastic, and neoplastic lesions. In JF Douglas, ed. Carcinogenesis and mutagenesis testing. Humana Press, Clifton, NJ.

JM Ward, RA Griesemer & EK Weisburger (1979). The mouse liver tumor as an endpoint in carcinogenesis tests. J Toxicol Appl Pharmacol 51 389-97

WM Wardell (1973). Fluoroxene and the penicillin lesson. Anesthesiology 38 309-12

MA Weinberger (1973). The 'blind' technique. Letter to the editor, Science, 181, 219-20

MA Weinberger (1979). How valuable is blind evaluation in histopathological examinations with animal toxicity studies. Toxicol Pathol 7 14-17

K Weinbren (1978). The Liver. Chapter 21 in Vol III of Systematic Pathology, by Thirty-Eight Authors. Churchill Livingstone.

EK Weisburger (1981). N-substituted aryl compounds in carcinogenesis and mutagenesis. NCI Monograph 58.

EK Weisburger (1983). Species differences in response to aromatic amines. In R Langenbach, S Nesnow & JM Rice, eds. Organ & Species Specificity in Chemical Carcinogenesis. Plenum, NY

AS Whittemore, SC Grosser & A Silvers (1986). Pharmacokinetics in low dose extrapolation using animal cancer data. Fund Appl Toxicol 7 183-190.

J Wilbourn et al (1986). Response of experimental animals to human carcinogens: an analysis based upon the IARC Monographs programme. Carcinogenesis 7 1853-1863

GM Williams (1980). Classification of genotoxic and epigenetic hepatocarcinogens using liver culture assays. Ann NY Acad Sci 349 273-82

GM Williams (1983). Epigenetic effect of liver tumor promoters and implications for health effects. Environ Health Perspectives 50 177-183.

R Wilson & EAC Crouch (1987). Risk assessment and comparisons: an introduction. Science 236 267-270

L Zeise, R Wilson & E Crouch (1984). Use of acute toxicity to estimate carcinogenic risk. J Risk Analysis 4 187-99

TECHNICAL REPORTS

Statistics Department

University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.

2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosciences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.

3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.

4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11. No. 1, 185-214.

5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.

6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.

7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.

8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.

9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.

10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.

11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.

12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.

14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.

15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.

16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.

17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review.

18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.

19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).

20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.

21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.

22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.

23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984.

25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.

26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.

27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.

28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimates. Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.

29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.

30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES, 1985 Vol 3 pp. 1-13.

31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.

32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.

33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.

34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.

35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science, Feb 1986, Vol. 1, No. 1, 3-39.

36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.

37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly, Feb 1986, Vol. 93, No. 2, 123-125.

38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.

39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.

40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.

41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.

42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.

43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?

44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.

45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.

46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.

47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data.

48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.

49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift. 1986. D. Reidel.

50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada. January, 1986.

51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.

53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.

54. BLACKWELL, D. (November 1985). Approximate normality of large products.

55. FREEDMAN, D. A. (December 1985, revised Dec. 1986). As others see us: A case study in path analysis. Prepared for the Journal of Educational Statistics.

56. LE CAM, L. and YANG, G. L. (January 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies.

57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.

58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.

59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.

60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.

61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.

62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.

63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.

64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.

65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.

66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.

67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data.

68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).

69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.

71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.

72. LEHMANN, E.L. (July 1986). Statistics - an overview.

73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.

74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomials, and some continued fractions of Gauss.

75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.

76. O'SULLIVAN, F. (September 1986). Relative risk estimation.

77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.

78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.

79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks.

80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.

81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.

82. DOKSUM, K.J. and LO, A.Y. (November 1986). Consistent and robust Bayes Procedures for Location based on Partial Information.

83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.

85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory.

86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.

87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.

88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.

89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987). Estimating and testing in a two sample generalized odds rate model.

90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.

91. DIACONIS, P and FREEDMAN, D.A. (March 1987). A finite version of de Finetti's theorem for exponential families, with uniform asymptotic estimates.

92. DABROWSKA, D.M. (April 1987). Kaplan-Meier estimate on the plane.

92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.

93. DABROWSKA, D.M. (June 1987). Nonparametric quantile regression with censored data.

94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.

95. RIZZARDI, F. (Aug 1987). Two-Sample t-tests where one population SD is known.

96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in *Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.*

97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for cancer.

98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.

99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.

100. LE CAM, L. (July 1987). On the standard asymptotic confidence ellipsoids of Wald.

101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators.

102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.

103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.

104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer.

105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.

106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.

107. CHENG, C-S. (August 1987). Some orthogonal main-effect plans for asymmetrical factorials.

108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.

109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \le k \le n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.

110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals.

111. BICKEL, P.J. and GHOSH, J.K. (August 1987). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.

113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.