Carcinogenic Potency: A Progress Report

BRUCE N. AMES, KIM HOOPER, CHARLES B. SAWYER, ALAN D. FRIEDMAN, RICHARD PETO,* WILLIAM HAVENDER, LOIS S. GOLD, THOMAS HAGGIN, ROBERT H. HARRIS, AND MARGARET ROSENFELD

Department of Biochemistry
University of California
Berkeley, California 94720

*Radcliffe Infirmary
University of Oxford
Oxford OX2 6HE, England

Chemicals in the environment, both natural and man-made, are now recognized as increasing the risk of human cancer. Current regulatory policy and scientific research have focused mainly on determining whether a chemical is capable of inducing cancer, but little systematic effort has been directed at quantifying how much hazard a given substance poses. It has become increasingly evident, however, that quantifying the intrinsic carcinogenic power of different chemicals is crucial to developing a sensible policy response to these hazards. This realization has come about because the relative differences in carcinogenic potency that may even now be inferred from animal experiments are enormous. For example, a daily dose of saccharin that would give cancer to 50% of exposed rats would be more than 10^6-fold larger (mg per kg body weight basis) than the dose of aflatoxin that would yield the same incidence of tumors. Regulatory decisions aimed at avoiding the largest number of cancer deaths should take into account the extent and level of human exposure to chemicals and the vast differences in intrinsic carcinogenicity.

To improve risk assessments, risk-benefit judgments, and regulatory policy, an index number of carcinogenic strength (or potency) for each chemical is desirable. Ideally, we would like to have quantitative information on the capacity of various chemicals to cause cancer in man, but with rare exceptions this is not available. An alternative source of information is animal bioassays. In our laboratory, we have been engaged for several years in creating a comprehensive data source incorporating the animal bioassays reported in the world's literature that are suitable for determining a potency value. This paper details our progress in attaining the following project objectives:

1. The calculation of a quantitative measure of potency.
2. A thorough search of the world's literature to identify tests that are sufficiently complete to allow potency estimates to be made and, from this raw data, development of a computerized data base that enables convenient storage and manipulation of the information.
3. Analysis of the sources of variation in the data.

DEVELOPMENT OF A POTENCY INDEX

Attempts to quantify estimates of potency began as early as 1930 (Twort and Twort 1930, 1933; Iball 1939; Bryan and Shimkin 1943; Druckrey 1967; Meselson and Russell 1977). The most generally satisfactory of these indices was devised by Meselson and Russell (1977), where the potency, k, was given by

$$k = \frac{-\ln(1-I)}{D \times t^{n+1}} \quad (1)$$

and where I is the cumulative single-risk incidence observed at time t (expressed as the fraction of a normal lifetime), D is the administered dose rate (expressed as mg per kg body weight per day), and n was taken as 3 (selected as the best estimate of the dependence of tumor appearance upon duration of exposure).

This index has two limitations. First, it does not take into account the incidence of spontaneous tumors in control animals, which in practice can vary widely and has a substantial effect on the magnitude of the calculated potency. Second, it does not take into account the progressively smaller number of animals at risk because they have died from causes other than tumors during the course of the experiment. (The effect of failing to account for this mortality is to underestimate the true potency.) Richard Peto (Oxford University) together with Charles Sawyer and Alan Friedman of our group have developed the theory of calculating carcinogenic potency from animal bioassays and converted this to computer programs. Our index, the TD_{50} (tumorigenic dose$_{50}$), is the daily dose rate required to decrease by half the probability of an animal remaining tumor-free at the end of a standard lifetime (taken as 104 weeks for rats and mice). The calculation of this index takes into account whatever spontaneous tumor incidence occurs in control animals and, where life-table data is available, corrects for intercurrent mortality. It has the additional merit that the dose rate to be estimated (that which gives cancer to half of otherwise tumor-free animals) is usually not far from an actual dose used in an experiment that yields statistically positive results; therefore, only a small extrapolation from experimental observation is necessary. This means that the choice of a particular dose-response function (ranging from Mantel-Bryan to linear) will not greatly affect the estimate of TD_{50}.

Computer programs have been developed to estimate a TD_{50} together with its confidence limits; this program also analyzes the shape of the dose-response relation and the probability that the TD_{50} is significant. It is now a routine task to determine a TD_{50} from any suitable set of data. Separate programs have been developed to calculate life-table TD_{50} values (where complete data are available giving the time of death and tumor occurrence for each animal) and summary TD_{50} values (where tumor data have been reported only as a summary for a group of animals).

We have also developed a measure of the sensitivity of negative bioassays, which can differ enormously. The sensitivity depends on the dose levels used and on the experimental design. A negative test is described as excluding TD_{50} values

below a certain limit rather than simply as "negative." Some experiments have such small numbers of test animals and use such low doses that they could not have detected any but the most potent carcinogens. The comparison of research designs and dose levels will sometimes make it possible to reconcile positive and negative results with the same compound: if two such tests examined different regions of the TD_{50} range, they need not be contradictory.

THE CREATION OF A DATA BASE

Whereas there is a paucity of data on human carcinogens, there is an abundance of research reports of animal bioassays on hundreds of chemicals. We have conducted a painstaking search of the world's literature to collect and evaluate all tests that would be suitable for the calculation of a TD_{50}. In addition to exhaustively scanning the major cancer journals and *Current Contents*, we have consulted several major bibliographies of cancer tests, including the monographs on chemical carcinogens prepared by the International Agency for Research on Cancer (1972-1979) and the PHS survey of carcinogens (Shubik and Hartwell 1948-1973). In addition, we have obtained all the bioassays carried out by the National Cancer Institute (NCI) that have been released.

Most of these tests have utilized quite diverse and unsystematized protocols; this makes direct comparisons difficult. We have tried to cope with this problem by selecting only those tests in which:

1. exposure occurred chronically over at least one-half the animal's normal lifespan,
2. the route of exposure was by diet, gavage, water, or inhalation (i.e., analogous to the major human exposures),
3. the whole body was exposed rather than only a specific site, as with subcutaneous injection or skin painting, and
4. there was a control group.

We now have analyzed over 1500 experiments that meet these criteria, chiefly rat and mouse-feeding studies, which have sufficient data for calculations. We define an experiment as the control group and the various dose groups for one chemical in one species, strain, and sex from one research report.

Most of the papers we have collected report only the cumulative number of tumor-bearing animals seen over the course of the experiment and the number at risk at the start. TD_{50} values calculated from such summary data are more subject to bias than TD_{50} values calculated from life-table data. Fortunately, the NCI has recently completed tests on nearly 200 chemicals using comparable protocols, and they have supplied us with the full life table from each experiment. Here it has been possible to estimate unbiased TD_{50} values using the Peto-Cox theory.

This vast sum of data has been stored in a computerized data base, which facilitates rapid retrieval and manipulation of information. Thus, it is possible

to make comparisons in potency between sexes, strains, and species, as well as to carry out mathematical analyses.

This phase of the project is nearly complete and currently includes information on over 600 chemicals. Many of these include mulitple tests that use different strains and species. The output includes the estimated TD_{50}, confidence limits, tumor type and site, and information on the dose-response relationship. We have partially completed our error check on the data base.

ANALYSIS OF THE DATA

A goal of this analysis is to see how well results in one species of animal predict those in another and to examine the reasons for aberrations. Questions of interest that can readily be approached with the computerized data base include:

1. How similar are the TD_{50} values calculated from independent tests on the same compound?
2. How well do males and females compare within a strain in a single bioassay?
3. How well do strains within a species correlate with each other?
4. How well do rats and mice compare in overall sensitivity as well as in the preferred target organ? On the basis of preliminary results, we anticipate that interspecies extrapolations, at least among rodents, will usually agree within a factor of ten. Seen against the possible variation of potency of some 10^7-fold, such results will be clearly useful. Chemicals that deviate far from the usual correlations between sexes and species can alert us to special circumstances—unusual pharmacokinetics, or a peculiar metabolic route—that could be investigated. This would result in a better understanding of the conditions under which useful predictions can be made between laboratory animals and humans.
5. How well do rodents compare with other species? An area of particular theoretical interest will be the comparison of long-lived and short-lived species; the very fact that long-lived animals are long-lived, that is, they do not succumb to cancer after only 2 years, as do rats, points to the existence of mechanisms of cellular and tissue control that are quantitatively, and perhaps qualitatively, different from those of short-lived mammals. This quantitative difference is, in fact, much larger than the simple ratio of lifetimes would suggest because of the fourth- to fifth-order dependence of cancer incidence upon age within a species (Peto 1977, 1979). This comparison bears directly on the problem of extrapolating risk estimates from rodents to man. Therefore, a major and important focus of this project will be to analyze the potency of chemicals that have been tested in both rodents and monkeys. We are collaborating with R. Adamson (NCI), who is currently conducting extensive cancer tests with monkeys. He will give us life-table data on the seven compounds that he finds positive in monkeys and data on 19 test compounds that have either failed to induce

tumors or have not been under test a sufficient length of time. These TD_{50} values will provide us with an important set of reference points for making rodent-primate risk extrapolations. We also intend to explore the available human data on carcinogenic potency following up previous work (Meselson et al. 1975).

6. One would like to know which environmental chemicals pose the greatest hazard to humans, though this is a very difficult and complicated subject. In the meantime, we will examine the TD_{50} values of the chemicals that have already aroused public concern because of widespread human exposure (such as DDT, dioxin, benzene, saccharin, benzo[*a*]pyrene, vinyl chloride, ethylene dibromide, and ethylene dichloride).
7. How do TD_{50} values compare when the same chemical is administered by different routes (e.g., inhalation versus diet) or different dosing schedules?

ACKNOWLEDGMENTS

This work was supported by Department of Energy contract DE-AM-03-76SF00034 PA156, by a California Policy Seminar grant to B. N. A., and by National Institute of Environmental Health Sciences Center grant ES-01896. We are indebted to Jade Goldstein and Elizabeth Higgins for help with the data base and to Ken Chu and Jerrold Ward of NCI for much help with the NCI bioassays.

REFERENCES

Bryan, W. R. and M. B. Shimkin. 1943. Quantitative analysis of dose response data obtained with three carcinogenic hydrocarbons in strain C3H male mice. *J. Natl. Cancer Inst.* 3:503.

Druckrey, H. 1967. Quantitative aspects in chemical carcinogenesis. *UICC Monogr. Ser.* 7:60.

Iball, J. 1939. The relative potency of carcinogenic compounds. *Am. J. Cancer* 35:188.

International Agency for Research on Cancer. 1972–1979. The evaluation of carcinogenic risk of chemicals to man. *IARC Monogr.* 1-19.

Meselson, M. and K. Russell. 1977. Comparisons of carcinogenic and mutagenic potency. *Cold Spring Harbor Conf. Cell Proliferation* 4:1473.

Meselson, M. S., B. N. Ames, P. M. Dolinger, V. H. Freed, J. Kolojeski, R. L. Metcalf, M. A. Schneiderman, E. K. Weisburger, C. F. Wurster, and R. Hansberry. 1975. Health hazards of chemical pesticides. In *Pest control: An assessment of present and alternative technologies, Vol. 1. Contemporary pest control practices and prospects: The report of the executive committee*, pp. 4 and 54. National Research Council, National Academy of Sciences, Washington, D.C.

Peto, R. 1977. Epidemiology, multistage models and short-term mutagenicity tests. *Cold Spring Harbor Conf. Cell Proliferation* 4:1403.

Peto, R. 1979. Detection of risk of cancer to man. In *Proc. R. Soc. Lond. B Biol. Sci.* 205:111.

Shubik, P. and J. L. Hartwell, ed. 1948-1973. *Survey of compounds which have been tested for carcinogenic activity.* Public Health Service Publication number 149. Government Printing Office, Washington, D.C.

Twort, C. C. and J. M. Twort. 1930. The relative potency of carcinogenic tars and oils. *J. Hyg.* 29:373.

———. 1933. Suggested methods for the standardisation of the carcinogenic activity of different agents for the skin of mice. *Am. J. Cancer* 17:293.

COMMENTS

KARY: Dr. Ames, where do you differentiate between routes of exposure?

AMES: We put down the route of exposure so that it is in the computer print-out. For both an inhalation study and with a feeding study we calculate the dose. Whether inhalation is going to be exactly the same as feeding, we don't know. We can ask the computer to compare all the inhalations and feedings. We'll start to do that kind of thing soon. Getting this computer print-out has been an enormous amount of work, and there are still many errors that we have to weed out. As a result, we really haven't had time to sit down and try and analyze it in any great detail.

MALTONI: I want to compliment Dr. Ames on this tremendous amount of work that he is doing. I think everybody is aware of how much we need a quantitative approach for this. If a substance is a carcinogen, we must know if it is more or less carcinogenic than a possible subacute compound.

Our laboratory is working on this problem in a different way from Dr. Ames. We are not working so much on world-wide data, but rather working in our own system, always using the same type of animals with some 60 different compounds and keeping them entirely and strictly homogeneous and serialized to the following extent. We would like to be able to record all the subtle types of differences within the doses, the routes, the schedule of treatment, the role of species, of sex, age, etc., for a series of compounds that may have a similar type of structure. We also would like to assess what effect molecular structure will have—what will be the bearer of the slightest changes and to find out, really, what is the practical value in selecting one compound over another.

An important element in these assessments is the weight of the lab variability and the experimental situation, which may affect your type of quantitative monitoring. An aspect really bound to the monitoring you have studied is the biological model that is engaged. Time of survival of this animal, when it has been examined, and where it has been kept, observed, and interpreted may bring enough information so that when you go to a difference from a picogram down to a gram, data are coming up pretty well, but when you are working in a range of the milligram you will have extremely wide sources of data.

AMES: I know. There are 2000 labs using our *Salmonella* test, and I know there are some people messing it up completely, even though we write the directions down very precisely. Some people do mess up experiments, and animal cancer tests are more complicated than *Salmonella*, so all we can do here is put down what people report.

But I am surprised, so far, on how much labs agree, given that we're not very concerned about a factor of ten in all of this. That will all come out in the analysis. It may be that for discrepancies some expert will have to look at that paper and say "garbage." We don't know enough now to say that any paper is garbage. Experts will have to start analyzing, using our data base as a guide, and say, "Here's a discrepancy; what do we make of this paper or that one?"

MALTONI: You know that just 2 years ago we published the results of tests on benzene—claimed to be a negative compound for some 30 years. All of the previous experiments in their entirety were too little, too inadequate, too short, too impure. But just by performing a very small experiment on 200 animals kept in a very controlled way, you can pick up very quickly that benzene is a carcinogen (Maltoni and Scarnato 1977). And so, to pick up the false negative is quite a good job.

AMES: I'd like to make one more point. One of the reasons we got into this is that Meselson and Russell (1977) had published in the *Origins of Human Cancer* an analysis of our data on mutagenic potency in our test versus carcinogenic potency (which they calculated) for about a dozen chemicals. That was brave. Clearly, ground up rat liver is only some crude first approximation of a rat. Can ground up rat liver plus *Salmonella* tell you anything about potency in an animal?

Meselson found a rough correspondence between mutagenic and carcinogenic potencies for most, but not all, of those chemicals. There was a million-fold spread in the potency range for both.

We became interested in the mutagenic and carcinogenic potency. We soon found that carcinogenic potency needed our full attention, and we have been so busy thinking about carcinogenic potency we haven't really thought about mutagenic potency. But I think out of this will come calibration points for calibrating all of the short-term tests. Scientists are finding new mutagens by the bucketful—there are so many mutagens out there. Cancer tests are never going to catch up; only a few hundred cancer tests are being done in the world in a year. We need some way of determining if a powerful mutagen is more likely to be a greater human hazard than a weak mutagen as determined by a battery of short-term tests. We don't know, but at least I think we will have the calibration points. (Joyce McCann will look into all of this.) One can use human liver in some of the short-term tests, and compare human liver with rat liver. If there's a species difference between rats and mice, one can put in mouse liver and rat liver and perhaps explain a difference and then try human autopsy liver. We hope that with these calibration points we can see how good or bad short-term tests are quantitatively.

References

Maltoni, C. and C. Scarnato. 1977. Le prime prove sperimentali dell'-azione cancerogena del benzene. *Gliospedali della Vita* 4:111.

Meselson, M. and K. Russell. 1977. Comparisons of carcinogenic and mutagenic potency. *Cold Spring Harbor Conf. Cell Proliferation* 4: 1473.

Banbury Report

5

ETHYLENE DICHLORIDE: A Potential Health Risk?

Edited by

BRUCE AMES
University of California

PETER INFANTE
OSHA

RICHARD REITZ
Dow Chemical Company

COLD SPRING HARBOR LABORATORY
1980