# RISK ASSESSMENT IN REGULATION OF CHEMICALS

P.J. Gehring
The Dow Chemical Company, PO Box 1706
Midland, Michigan U.S.A. 48640

*Summary.* Extrapolation of animal toxicity data to predict the risk of humans exposed to low doses of chemicals is being increasingly employed to regulate chemicals. This is particularly true for chemicals shown to cause cancer in animal studies. Although the use of such risk assessment techniques has positive attributes, the methodology as currently performed has great potential to overestimate risk and lead to unjustified regulation. Unless risk assessment is scientifically sound and creditable, using it for regulatory purpose will ultimately be a disservice. To be creditable the database utilized must be critically assessed and mechanistic, and pharmacokinetic parameters must be considered in selecting models for estimating the risk. Subsequently, the estimated risks must be consistent with experience.

## INTRODUCTION

Risk assessment for regulation of exposure to chemicals is growing in popularity and will likely be adopted. Recently, the National Research Council in the United States proposed that a *de minimus* risk concept replace the Delaney Clause in regulating carcinogenic pesticides in the U.S. food supply (11).

The use of risk assessment to regulate exposure to pesticides and other chemicals is conceptually more scientifically sound than past procedures and worthy of support. However, adoption of risk assessment for regulations of chemicals necessitates an understanding of the validity of the data utilized for the assessment, the *a priori* assumptions inherent to the risk assessment process and ultimately the reliability of the estimated risk.

If attention to these items is not heeded, risk assessment will provide a fallacious foundation for regulation and destroy creditability of all concerned - the industry, the regulators, the scientists.

The purpose of this presentation is to provide some insights of the potential problems and fallacies in risk assessments for chronic toxicity, in particular carcinogenicity.

Currently, the fulcrum of risk assessment is data generated from the chronic bioassay. Only in rare incidences have data generated from other scientific studies such as epidemiology triggered risk assessment and subsequent regulatory restriction of chemical exposure. In this light, it is appropriate to examine the adequacy of this fulcrum *per se*. More critical and essential is to define what type of lever should this fulcrum be used to support, and how much leverage should it be given.

In assessing the adequacy of the chronic bioassay as a fulcrum, the first question to ask is whether it is science. To address this question I quote Lord Kelvin: "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be".

In this context, the chronic bioassay is science, the data are derived from measurement and they are expressed in numbers. Some may quibble about accuracy of some of the measurements used; for example, whether a liver nodule is a hepatoma or a hepatocellular carcinoma. I concede that the science of measurement and number generation may still be improved but risk assessment will be benefited little by such improvements.

Lord Kelvin provided an excellent starting point for evaluating the bioassay as a science. Additional criteria, however, must be considered. Economists and sociologists measure and count; they rely heavily on case specific studies, short-term peaks at the future anticipation of discontinuities. Charitably these professions are referred to as soft science. Not until these professions can utilize their measurements and numbers to accurately and creditably predict future outcomes will they receive the dignity to the called science. In this same context, the chronic bioassay definitely cannot be called science.

The quality of a scientific study is judged on the basis of three criteria: (i) the adequacy and appropriateness of the experimental design, (ii) the competancy and completeness with which the study was carried out and reported, and (iii) the evaluation and interpretation of the results. Each of the criteria as they pertain to the bioassay are discussed below.

## THE COMPETANCY AND COMPLETENESS WITH WHICH THE BIOASSAY IS CARRIED OUT AND REPORTED

Undoubtedly, this criterion has received the most criticism over the past decade. Conduct of a chronic bioassay is the easiest to creditably criticize. Thus, allegations such as mix-up of dose groups, poor animal health, losing animals or tissues, reincarnation of animals, etc. have been made in an attempt to negate the results of bioassays. Unfortunately outright fraud has been alleged and occasionally proven. The creditability of the entire science is obviously harmed by such allegations. Within reason competency must always be expected and assured.

However, it is to no ones advantage to expect that absolute perfection, no mistakes whatsoever, is essential for a bioassay to be meaningful. Most chronic bioassays require approximately one-half million measurements or observations. In addition, identifiable and unidentifiable variables may contribute to the final outcome. It is not feasible to perform such a task with absolute perfection. Adherence to "Good Laboratory Practices" together with reasonable care and common sense will negate performance as a significant flaw in the bioassay.

When identifiable or unidentifiable variables are likely to have influenced the results, they need to be evaluated to assess their applicability. To assure usefulness of the results, precautions beyond those presently defined and in place in most laboratories conducting chronic bioassays are not needed. However, I believe strongly that the same precautions are needed whether the study is conducted within industrial laboratories, or within governmental or academic laboratories; it is ludicrous for anyone to advocate a double standard in this regard.

## THE ADEQUACY AND APPROPRIATENESS OF THE EXPERIMENTAL DESIGN

Experimental design is almost inextricably associated with the third criterion, that of evaluation and interpretation of the results. For this reason, experimental design must be tailored to satisfy the needs for

evaluation and interpretation.

Originally, the objective of the chronic bioassay was to determine whether a chemical could cause an oncogenic response when given at a maximum tolerated dose, five days per week for a lifetime. However, the large amount of data generated and the problems of false positives, both statistical and biological, have made the data very difficult to interpret. This led Bill Ruckelshaus to state that: "The results of a two year rodent carcinogenicity study are like a captured spy in that, if you torture them hard enough, they will tell you anything you want to know".

As may have been anticipated, the captive provided a high percentage of positive responses. One tabulation by Purchase (5) found that of 250 compounds 98 (38%) were negative, 109 (44%) were positive in both rats and mice, 21 (8%) were positive only in mice, 17 (7%) in rats only and 5 (2%) had results differing from some other species. Hottendorf and Pachter (4) summarized data from 200 bioassays in The National Cancer Institute program and found similar results.

Considering the objective as well as state of our knowledge at the time, I do not fault these studies or the results. It was necessary for us to determine whether carcinogenic activity was a unique property of only a few chemicals. We now know that carcinogenicity is not unique to a single or likely even a few unique biological mechanisms. Furthermore, the accumulating data indicate that many if not the majority of chemicals absorbed into the body, man-made and natural occurring, produce or are likely to produce an oncogenic response if given at doses sufficient to cause frank cellular and tissue damage. Ames, *et al.* (1) show that pesticide residues and water pollution are likely a minimal carcinogenic hazard relative to the background levels of natural occurring substances.

Nonetheless, the concept of using the maximum tolerated dose remains with us. True, we have refined it somewhat in that we have reduced the degree of pathology required to constitute an MTD, to that compatible with longevity. However, still needed for selection of the doses to be used in the chronic bioassay is more attention to biochemical and/or physiological changes not necessarily accompanied by frank toxicity; for example non-linear pharmacokinetics.

A design based on the MTD is not wrong if the objective is to provide quantitative data to asses whether a chemical will cause cancer within a dose range causing, or near to that causing, observable signs of toxicity. For some chemicals, this may be a worthy goal! However, carcinogenicity at such doses is a qualitative, not a quantitative result with respect to assessing risks of much lower doses. In the absence of additional information and insight, it is naive, if not senseless, to extrapolate such information to predict creditably, the oncogenic response of much lower doses.

As scientists, we are only justified in advocating the adequacy of the scientific quality of the chronic bioassay for the objective it is designed to address. Without additional information to justify the use of data from chronic bioassay studies for assessing the risk of lower doses, risk assessment is by interpolation not extrapolation. Such efforts deserve no more dignity, perhaps even less, than those of economists and sociologists.

The adequacy of the species utilized for the chronic bioassay is frequently an issue. No one is likely to deem the rat or mouse a totally suitable surrogate for the human species in evaluating the oncogenicity potential of chemicals

for the latter. Nonetheless, most will agree that these species and perhaps the hamster, provide the best practical test systems.

A number of chemicals have been found to induce only hepatomas in mice (10). Most of these are weakly or non mutagenic to bacteria, and there are no clear chemical characteristics common to all, albeit many are halogenated compounds. Thus, the relevancy of the induction of hepatomas in mice is highly questionable with respect to risk assessment.

Using both rats and mice for each chemical has a very questionable return on investment. Allowing a more complete toxicodynamic evaluation of a chemical in one or more species will provide a far better return on investment than conducting lifetime bioassays using both rats and mice for each chemical; particularly in the fact that studies using mice provide little information of value beyond that provided by studies using rats (8).

Other issues of "experimental design" warranting attention are gavage as a route of administration, using corn oil as a vehicle, non-oncogenic pathology, the number of tissues requiring microscopic evaluation, need for microscopic evaluation of the tissues from all dose groups, etc. These and other issues of experimental design require continued critical reassessment.

## EVALUATION AND INTERPRETATION OF THE RESULTS

As indicated previously, the experimental design predetermines the ultimate evaluation and interpretation of the results. For the most part, the usefulness of the chronic bioassay will be little served by better evaluation and interpretation of the data derived from such studies *per se*. Generally, the result of the study will indicate whether the test chemical did or did not cause an oncogenic response under the conditions of the test. Of course there will be questions; for example, whether or not a specific oncogenic response is a random variation or attributable to the chemical under test. I do not wsih to pass such assesments off as trivial; they are critical and require training and experience. If data are available to determine whether the response is reproducible, the job is easier; reproducibility is a basic tenet of the scientific method. Unfortunately, many within the profession of toxicology as well as outside are ill prepared to address these issues. Nonetheless, these issues should not constitute major weaknesses in risk assessment.

At this point you may have concluded that all is well with chronic bioassay data for risk assessment. Wrong! The chronic bioassay provides data to assess the potential oncogenicity of chemicals given at or near toxic dose levels of the chemical to the species utilized in the study. Risk assessment must strive for creditable projections, or extrapolations, below these dose ranges and to another species, namely man. Currently, this process is somewhere between an art form and a science, just as is the use of macroeconomic data by a stock broker to advise you what purchases and sales will rended you wealthy. None of you is likely to use a broker who doesn't use such information, however, none will believe or use one who claims 100% or even 50% accuracy using solely macroeconomic measurements.

The current risk assessment procedures that rely solely on screening bioassay results from the most sensitive species are not scientific, they are not creditable or reliable and as such they are often in direct conflict with the incidence of cancer in populations of people in the work environment. Risk assessments derived from extrapolation of chronic bioassay data have uniformly over predicted the incidence of cancer in exposed populations.

Shown in Table 1 are some risk assessments for a few compounds, using the linearized multistage model. For perchloroethylene, trichloroethylene, acrylonitrile, and butadiene, exposure to 1/2 the threshold limit value (TLV) for 20 years is predicted to cause 8 to 26% increase in cancer. Although definitive epidemiological studies are not available, it is highly unlikely that such increases have occurred without recognition. For perchloroethylene and trichloroethylene, at least 8-10 small studies have been published which indicate no increase in cancer of the liver. While not one of these studies is in itself indicative, their consistency is an indication that the incidence of liver cancer is well below that predicted.

Table 1. Risk assessment based on extrapolation of animal data[a]

| Chemical | Exposure | Risk |
|---|---|---|
| Perchloroethylene | 60 ppm for 20 years | 0.23 |
| Trichloroethylene | 60 ppm for 20 years | 0.08 |
| Acrylonitrile | 10 ppm for 20 years | 0.13 |
| Butadiene | 500 ppm for 20 years | 0.26 |
| Vinyl chloride | 200 ppm for 20 years | 0.16 |
| Bischlormethylether | 0.01 ppm for | |
| | 10 years | 1.00 |
| | 1 year | 0.45 |
| Ethylene dibromide | 3 ppm for | |
| | 20 years | 0.65 |
| | 10 years | 0.41 |
| | 4.2 years | 0.20 |
| Aflatoxin | Average U.S. consumption | 789/100,000 |

[a]Risk calculated from cancer assessment group potency estimates published in methylene chloride health assessment document EPA/600/8-82/004F, Feb. 1985.

Vinyl chloride has been predicted to produce a 0.16 risk of cancer (3). However, two large inter-industry studies indicate no increase in cancer in vinyl chloride monomer or PVC polymer workers other than angiosarcoma of the liver (2, 9). The worldwide incidence of angiosarcoma, 108 cases from 1955-1983 (Forman, *et al.*, unpublished data, 1985), is far less than predicted by the risk models, considering that tens of thousands of workers have been exposed and that, historically, exposure to several hundred ppm were common.

Exposure to concentrations of bischlormethylether reportedly sufficient to cause marked respiratory irritation (1 ppm and greater) caused cases of respiratory tract cancer; however, the incidence was not 100% or even 45% as predicted from linear models.

For ethylene dibromide, no increase in cancer was seen in 156 employees exposed to an average of 2 ppm for 4.2 years, while the predicted incidence is 20% (6).

The incidence of mortality from liver cancer in the U.S. is 161/100,000, yet the risk predicted from the average exposure to aflatoxin alone is 789/100,000.

The lack, or inadequacy of epidemiological evaluation weakens the foregoing superficial analysis. However, it is not likely that the magnitude and consistency of over prediction will be discounted by more intense evaluation. More rigorous comparison of predicted risk with epidemiological results is needed. Such comparisons are needed to define the boundaries for the preconceptions inherent to risk assessment models.

In addition to using risk assessment models to predict the incidence of cancer in man from chronic bioassay data, similar predictions of the incidence in the same, or other species of animals exposed to lower doses in other experiments, need to be made when feasible. For chloroform, such as assessment shows the risk assessment models fail miserably (7). At the very least, it is reasonable to expect such predictions to be more accurate than predictions for man.

Rather than being in a "have model will asses risk" mode, we need to be in a "have model will asses hypothesis" mode.

Current modeling techniques relying on the linearized multistage model are intended to predict only an upper bound to risk. This is distressing for two reasons: (i) the whole process of communication and utilization of the results of risk assessments usually ignores the fact that the estimates are upper bounds; and (ii) no serious attempt is made to arrive at reasonable estimates of actual risk. As a scientist, I find the second point appalling. Risk estimates need to reasonably reflect actual risks if risk assessment is to move from the realm of an art and qualify as a science. Judgments must be exercised to produce reasonable estimates of actual risk consistent with human experience and knowledge of biological science.

## DEPLOYMENT OF FUTURE RESOURCES

To enhance the usefulness of chronic bioassay data in risk assessment significant resources should <u>not</u> be devoted to the following:

1. Additional good laboratory practices.
2. Increasing the number of animals used.
3. More mathematical risk modeling in the absence of biological rationale.
4. More pathology.
5. More species, or changing species.
6. Repeats of well conducted studies.
7. DNA alkylation *per se* in the absence of an established association with a carcinogenic response in the target tissue.

In contrast, resources and efforts <u>should be</u> directed to elucidate the mechanisms of carcinogenesis and extrapolation rationale. Such efforts include:

1. Role of proliferative stimuli.
   a. Hormonal
   b. Cellular damage including oxidative, *a la* superoxide
   c. Growth promoters
2. Cellular differentiation and the role of regulators such as Vitamin A.
3. Role of anti-carcinogens/mutagens.
4. Role of oncogenes.

5. Intercellular communications (membrane effects).
6. Role of the immune system.
7. Species differences in the ease of inducing immortal cells (mouse > rat > hamster >> man).
8. DNA repair processes *in vivo*.
9. Understanding the mechanisms (limitations) of *in vitro* tests used to augment bioassay data (pH, osmolarity, toxicity, etc.).
10. Physiological pharmacokinetic modeling for more accurate internal dose estimations between species and routes of exposure.
11. Non-linear pharmacokinetics.
12. Better understanding of spontaneous genetic interactions (food, lifestyle, etc.) for better perspective of low level impact of chemicals on the genome.
13. Comparisons of cancer incidence in exposed populations to predicted risks.
14. The relationship of cellular, organ, and whole body functions altered by toxic doses of a chemical to those in the "normal" state.

In particular, a concentration of effort is needed to determine the relationship between the biochemical or physiological status of animals and cells exposed to toxic amounts of a chemical and their status in the normal state. If changes induced by toxicity can be elucidated together with their relevancy to the "normal" state, the usefulness and reliability of chronic bioassay data in risk assessment will be greatly enhanced.

Acquisition of the aforementioned information must not be viewed as a panacea for solving risk assessment issues. When such information has been available, the demand for absolute proof of the nature of its association to the carcinogenic process has been demanded.

For example, if it is demonstrated that cancer only occurs when detoxification pathways are saturated leading to toxicity including DNA damage, it is posed that one cannot assure no DNA damage whatsoever at doses far below those needed to saturate detoxification processes. Others indicate that the human population is very heterogenous and someone, somewhere may not detoxify even low doses. In absence of such absolute proof, risk management has been based on the most conservative assessment rather than the most biologically rational assessment.

In science we have many theories, a lot of hypotheses, an abundance of indications, and plenty of probables - but there are few laws and few statements we can make with 100% certainty. While we wait for generation of laws to practice risk assessment, a second "dark ages" will occur if risk management is based on ultraconservative, non-scientifically based criteria.

It should not be anticipated that acquisition of new knowledge will provide definitive estimates of risk at $1x10^{-6}$ or even at $1x10^{-5}$, anyone advocating reliability of such estimates is doing so as an advocate not a scientist. I do believe reliable prediction of $1x10^{-3}$ risk may be a realizable goal, perhaps even $1x10^{-4}$. In spite of the amount of new knowledge acquired, scientifically based judgment will remain the most critical material in the fulcrum utilized for risk assessment. Absolute, unequivocal science is unattainable for assisting the risk assessment process.

If risk assessment is to become a creditable regulatory tool, the reliability of the risk estimate must be elucidated and communicated. Subsequently, comparisons with known risks are needed to assess whether the risk in question is acceptable. It will be disastrous to insist that as risk of less than

1/1,000,000 (*de minimus* risk) is required for registration of a pesticide if the estimate is grossly overestimated. This is even more ludicrous considering the natural background risk is 250,000 times higher.

In summary, the chronic bioassay for cancer, although misleading, is not the primary source of difficulty in risk assessment. Rather, it is the interpretation and subsequent extrapolation, or better described as interpolation, that is unscientific. To improve risk assessment will require acquisition and use of more pharmacokinetic and mechanistic data. Even with incorporation of such data, rational scientific assessment of low level risk will not be feasible if absolute proof of applicability is required. Almost anyone can make a wise decision when all the data and information are available. Unfortunately, the need for more experiments and data is never satiated in science. Ultimately, the best possible scientific judgment is the most important constituent of creditable risk assessment.

## REFERENCES


1. Ames, B.N., Magaw, R. and Gold, L.S. 1987. Sci. 236, 271-280.
2. Fox, A.J. and Collier, P.F. 1977. Brit. J. Indust. Med. 34, 1-10.
3. Gehring, P.J., Watanabe, P.G. and Park, C.H. 1979. Tox. Appl. Pharm. 49, 15-23.
4. Hottendorf, G.H. and Pachter, I.J. 1982. Tox. Pathol., 10, 22-26.
5. Purchase, I.F.H. 1980. Brit. J. Cancer. 41, 454-468.
6. Ramsey, J.C., Park, C.N. Ott, M.G. and Gehring, P.J. 1978. Tox. Appl. Pharm. 47, 411-414.
7. Reitz, R.H., Quast, J.F. Stott, W.T., Watanabe, P.G. and Gehring, P.J. 1980. Water Chlorination: Environmental Impact and Health Effect. Eds. R.L. Jolley, W.A. Brunge and R.B. Cummings (Ann Arbor Press: Ann Arbor, MI) Vol. 2. 48106, pp. 983-993.
8. Schach von Wittenau, M. and Estes, P.C. 1983. Fundamental and Applied Toxicology. 3, 631- 639.
9. Tabershaw, I.R. and Gaffey, W.R. 1974. J. Occupat. Med. 16, 509-518.
10. The Nutrition Foundation 1983. The Relevance of Mouse Liver Hepatoma to Human Carcinogenic Risk. A Report of the International Expert Advisory Committee to the Nutrition Foundation.
11. National Research Council, Board on Agriculture, Committee on Scientific and Regulatory Issues Underlying Pesticide Use Patterns and Agricultural Innovation. 1987. Regulating Pesticides in Food. The Delaney Paradox. (National Academy Press, Washington, D.C.)