



RAYTHEON COMPANY
ENVIRONMENTAL & OCEANOGRAPHIC SERVICES

P.O. BOX 350
PORTSMOUTH, RI 02871

TEL: 401-847-8000
TWX: 710-502-5923
TELEX: 092-7787

EVALUATION OF STATISTICAL METHODOLOGY USED IN
DERIVING EPA WATER QUALITY CRITERIA

(as stated in the May 18, 1978 Federal Register,
Vol. 43, No. 97, pp. 21506-21518)

Prepared for:

THE AMERICAN PETROLEUM INSTITUTE

Priya Wickramaratne
Priya J. Wickramaratne,
Principal Investigator

July 25, 1978

CMA 049450

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
A	Introduction	1
B	Summary of Results	2
C	Discussion of Results	6
D	References	24

A. INTRODUCTION

This report documents an evaluation of the statistical methodology used in deriving EPA water quality criteria, as stated in the May 18, 1978 Federal Register, Vol. 43, No. 97, pp. 21506 - 21518.

B. SUMMARY OF RESULTS

A summary of significant conclusions reached in our evaluation are itemized below:

1. Definition of Correction Factor - Preliminary analysis of data shows that use of a ratio as a description of the functional relationship between standardized LC_{50} and existing LC_{50} values seems to be an adequate description for the majority of cases. However, analysis of data from Table 3, indicates that the relationship may not be valid for the flow through/static correction factor for all compounds. We suggest that a more statistically rigorous method such as regression analysis be used to derive correction factors.
2. Definition of the mean ratio as the geometric mean of the individual ratios is valid under the justifiable assumption that the concentrations of the compounds are log-normally distributed.
3. Estimation of the ratios in Table 1 as the "averaged measured concentration divided by the average nominal concentration in each toxicity test" is not appropriate because it is generally assumed that measured concentration is log-normally distributed.

4. The use of one correction factor for all compounds is justified only if it can be correctly assumed that the correction factors are independently distributed random variables with a common mean and variance, i.e. the correction factors are not compound dependent. If this is not the case, then the estimates derived from these data bases could be highly biased. Similarly the "pooling" of data from different tests, procedures, waters, species, etc. without any compensating weighting factors is only justifiable under the assumption that the correction factors are independent of these variables. Neither of these assumptions, to our knowledge, have been validated or even tested. We suggest that the standard statistical methods of analysis of variance and covariance be used to investigate these aspects of data variability.
5. The estimation of flow through/static correction factors in Table 3, using only 12 compounds, the majority of which are pesticides, will probably result in a biased estimate, i.e. not be representative of other, non-pesticide compounds. We recommend that a larger and more representative data base be used to derive these correction factors.

6. Implicit in the use of one sensitivity factor to represent all compounds for various species of fish, (Table 4) is the assumption that the variance of LC_{50} values over different species of fish is independent of compound. This hypothesis should be investigated using standard statistical methods for investigating homogeneity of variance before making this assumption.

Since it is intended that the sensitivity factor be applied to standardized LC_{50} values, we suggest that the LC_{50} values used in computing the sensitivity factor be standardized LC_{50} values, i.e., LC_{50} values observed at 96 hrs, measured concentrations using flow through methods.

7. Comments in Section C above, also apply to Table 5, Table 8, and Table 9.
8. No attempt has been made by EPA to estimate the prediction error associated with each correction factor, and the propagation of errors using several correction factors.

Data in Tables 1, 2 and 3 were used to estimate the prediction error generated by standardizing an LC_{50} value measured at 24 hours, using static methods at nominal concentration to an LC_{50} value measured at 96 hours using flow-through method methods and measured concentration. The minimum average error in prediction, under these conditions, was found to be 90%.

Data in Tables 6 and 7 were used to estimate the prediction error generated by standardizing an LC_{50} value measured at 24 hours using static methods, to an LC_{50} value measured at 96 hours using flow-through methods (for invertebrates). The minimum average error in prediction under these conditions was found to be 161%.

9. Data in Table 5 were reported in mg/l and μ g/l without units. Calculations were run as if they were the same units which will result in incorrect values of the average of the logarithms of the variance.

For example, MATC values for Endrin, Chlordane and Methyl mercury are given in milligrams. While the MATC values for other compounds are given in μ grams.

10. In the estimation of a final fish chronic value, the use of equation $X\sqrt{YZ}$ (Section II.C.c) to compute a calculated fish chronic value is not justified under the following conditions:

Z = Final fish acute value from I.G.2., where the LC_{50} value obtained from I.G.2. is associated with a species whose MATC value is known.

C. DISCUSSION OF RESULTS

This section gives a more detailed discussion of the items listed in the summary. Items 7, 9 and 10 have not been included in this section, since we feel that they are self-explanatory.

1. Definition of correction factor:

Let y_i be the i^{th} standardized concentration and
 let x_i be the i^{th} existing concentration, where y_i represents
 either

$$y_i \text{ (measured concentration)} - x_i \text{ (normal concentration)}$$

$$y_i \left(\begin{array}{l} \text{concentration measured} \\ \text{under flow through} \\ \text{conditions} \end{array} \right) - x_i \left(\begin{array}{l} \text{concentration measured} \\ \text{under static conditions} \end{array} \right)$$

$$y_i \text{ (LC}_{50} \text{ measured at 96 hrs)} - x_i \text{ (LC}_{50} \text{ measured at 24, 48, 72 hrs)}$$

If we assume [1] that the y_i are random variables
 with a log-normal distribution, then the $\ln y_i$ are random
 variables with a normal distribution.

$$\text{Let } \ln y_i = f(\ln x_i)$$

and assume that a linear relationship exists between $\ln y_i$
 and $\ln x_i$, i.e.,

$$\ln y_i = \beta \ln x_i + \alpha + e_i$$

where e_i are independent, normally distributed random
 variables.

The standard statistical method of estimating β and α
 is by regression analysis, using the method of least squares
 [2]. This method gives estimates of β and α defined by:

$$\hat{\beta} = \frac{\sum (\ln x_i - \overline{\ln x}) (\ln y_i - \overline{\ln y})}{\sum (\ln x_i - \overline{\ln x})^2}$$

$$\text{and } \hat{\alpha} = \overline{\ln y} - \hat{\beta} \overline{\ln x}$$

On the other hand, if the correction factor is defined as the ratio:

$$\frac{y_i}{x_i} = k$$

where k is a constant

Then, taking logs: $\ln y_i = \ln x_i + \ln k$.

Comparing this formulation with that above for the regression analysis, the two formulations coincide if $\hat{\beta}=1$. Hence, we can conclude that the definition of the correction factor as a ratio is appropriate if $\hat{\beta}=1$.

To see if $\hat{\beta}=1$ for toxicants of interest, consider Table 1a which gives a list of toxicants with their corresponding regression equations. The coefficients ($\alpha, \hat{\beta}$) were derived from analyses done on data taken from available references in Table 1 of the EPA document.

Table 1b gives a list of toxicants with the corresponding regression equations derived from analysis done on data from available reference for Table 3 of the EPA document.

It will be noted the values of $\hat{\beta}$ in Table 1b are different, indicating that the ratio relationship may not be valid for some of the compounds in Table 3.

Table 1c gives a comparison of the equations defining the relationship between measured and nominal concentrations: a) using the EPA definition of the correction factor as a ratio, and b) using the correction factor derived from regression analysis. The comparisons are made for the toxicants listed in Table 1a (y is defined as the measured concentration, and x as the nominal concentration).

Table 1.a. Regression Equations for Selected Toxicants
From Table 1 of EPA Document

<u>Toxicant</u>	<u>Regression Equation</u>
Malathion	$\ln y = 1.01345 \ln x - 0.27785$
Hexachlorobenzene	$\ln y = 0.97739 \ln x - 0.65373$
Chlordane	$\ln y = 0.88925 \ln x - 0.25437$
Cadmium	$\ln y = 0.97519 \ln x - 0.14637$
Kepone	$\ln y = 1.07580 \ln x - 0.48856$

Table 1.b. Regression Equations for Selected Toxicants
From Table 3 of EPA Document

<u>Toxicant</u>	<u>Regression Equation</u>
Pyrethrum	$\ln y = 1.25106 \ln x - 0.90202$
RU - 11679	$\ln y = 0.67105 \ln x - 0.81792$
SBP - 13842	$\ln y = 0.48466 \ln x - 0.91404$
TFM	$\ln y = 0.99301 \ln x + 0.18491$

Table 1.c. Comparison of correction factor derived by EPA with correction factor derived using regression equations for toxicants listed in Table 1.a.

Toxicant	Correction Factor Derived by EPA	Correction Factor Derived Using Regression Equation
Malathion	$y = 0.79x$	$y = 0.76 x$ 1.01
Hexachlorobenzene	$y = 0.61x$	$y = 0.52 x$ 0.98
Chlordane	$y = 0.56x$	$y = 0.78 x$ 0.89
Cadmium	$y = 0.90x$	$y = 1.16 x$ 0.98
Kepone	$y = 0.83x$	$y = 0.61 x$ 1.08

2. Definition of Mean Ratio as Geometric Mean

Let the relationship between y_i and x_i be defined as

$$\ln y_i = \ln x_i + \ln k \quad i = 1, \dots, N$$

where

y_i is the standardized concentration,

x_i is the existing concentration

and k is the ratio y_i/x_i

If it is assumed that the individual concentration of the toxicants are log-normally distributed, then the minimum variance unbiased estimate of the mean ratio, $\ln \hat{k}$, over the total number of N observations is given by [2]:

$$\begin{aligned} \ln k &= \frac{1}{N} \left(\sum_{i=1}^N \ln y_i - \sum_{i=1}^N \ln x_i \right) \\ &= \frac{1}{N} \left(\sum_{i=1}^N (\ln y_i - \ln x_i) \right) \\ &= \frac{1}{N} \left(\sum_{i=1}^N \ln \left(\frac{y_i}{x_i} \right) \right) \\ &= \frac{1}{N} \ln \prod_{i=1}^N \frac{y_i}{x_i} \\ &= \ln \sqrt[N]{\prod_{i=1}^N \frac{y_i}{x_i}} \\ &= e^{\ln \left(\sqrt[N]{\prod_{i=1}^N (y_i/x_i)} \right)} \end{aligned}$$

This implies that the minimum variance, unbiased estimate at k is

$$\hat{k} = \sqrt[N]{\prod_{i=1}^N (y_i/x_i)}$$

= the geometric mean of the individual ratios

It is noted that if the individual concentrations are not log-normally distributed, then $\hat{k}$ = the geometric mean of the individual ratios is not necessarily the best estimate of k . Thus, implicit in the use of the geometric mean as an estimate of k , is the assumption that the individual concentrations are log-normally distributed (See Section 3).

3. Estimation of Ratios in Table 1

Estimation of an average value for the ratio in each toxicity test in Table 1 (of the EPA document) as the "averaged measured concentration divided by the average nominal concentration in each toxicity test" is not appropriate under the condition that each measured concentration is log-normally distributed.

The appropriate estimate of the mean ratio for each toxicity test is the geometric mean of the individual ratios as derived in section 2 above.

It should be noted that the "averaged measured concentration divided by the average nominal concentration in each toxicity test" does not have a log-normal distribution, under the condition that each individual measured concentration is log-normally distributed.

4. Use of One Correction Factor for All Toxicants

The use of one correction factor for all toxicants is justified only if it can be correctly assumed that the correction factors are independently distributed random variables with a common mean and variance, i.e., the correction factors are not toxicant dependent.

Similarly, the "pooling" of the data from different tests, procedures, water, species, etc., without compensating weighting factors is only justifiable under the assumption that the correction factors are independent of these variables.

The objective of this analysis as we see it is not to estimate a "grand" mean correction factor that is representative of all toxicants under a wide variety of conditions. This, however, appears to be what EPA has taken as its objective. Rather, the objective should be to estimate a correction factor, or series of factors, which will predict the desired concentration with a minimum of error.

Let us assume that the correction factors corresponding to the individual toxicants are random variables, $y_1, y_2, y_3, \dots, y_n$.

Let

$$\begin{aligned}\ln y_1 &\in N(\mu_1, \sigma^2) \\ \ln y_2 &\in N(\mu_2, \sigma^2) \\ \ln y_3 &\in N(\mu_3, \sigma^2) \\ \ln y_n &\in N(\mu_n, \sigma^2)\end{aligned}$$

i.e., $\ln y_i$ are normally distributed, with the same variance but different means.

It is desired to find estimates of $\mu_1, \mu_2, \dots, \mu_n$.

If $\mu_1 = \mu_2 = \dots = \mu_n = \mu_0$

then the best estimate of μ_0 is

$$\frac{\hat{\mu}_1 + \hat{\mu}_2 + \dots + \hat{\mu}_n}{n}$$

But if $\mu_1 \neq \mu_2 \neq \dots \neq \mu_n$

then $\frac{\hat{\mu}_1 + \hat{\mu}_2 + \dots + \hat{\mu}_n}{n}$

will give a biased estimate of $\mu_1, \mu_2, \dots$ and μ_n .

To illustrate this point we selected 4 toxicants at random from Table 1:

<u>Toxicant</u>	<u>Ratio</u>
Acrolein	$0.76 = \hat{\mu}_1$
Carbaryl	$0.99 = \hat{\mu}_2$
Carbofuran	$0.20 = \hat{\mu}_3$
Chlordane	$0.56 = \hat{\mu}_4$

Under the assumptions described above (i.e., the y toxicants have different expected values) using the geometric mean of these 4 values to derive one correction factor for all four toxicants will give biased results. The geometric

mean of the four ratios was computed to be 0.538. Using this value as the correction factor for all four toxicants will result in consistently under-estimating the measured concentration of Acrolein and Carbaryl, and consistently over-estimating the measured concentration of Carbofuran.

Under these circumstances a far more accurate method of predicting the measured concentration would be to use the 4 individual ratios as estimates of the correction factors for the 4 toxicants.

If, however, we can show that all 4 correction factors do indeed come from the same population, then the estimate $(\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4) / 4$ is the best estimate of μ (the common expected value) and it would be appropriate to use a common correction factor.

In the light of these observations, we strongly recommend that the standard analysis of variance methods be used to test the hypothesis that the mean correction factors for all toxicants are equal. If this hypothesis is rejected, then we can use various other statistical analyses to identify which toxicants or group of toxicants have significantly different correction factors. A series of correction factors can then be derived to predict the standardized concentrations.

The same methods can also be used to investigate the variation of correction factors with parameters other than toxicants, such as hard and soft water, methods of measuring LC_{50} , etc.

5. Estimation of Flow Through/Static Correction Factors

We feel that the data contained in Table 3 is an inadequate data base from which to estimate flow through/static correction factors, for the following reasons:

- 1) Only 12 toxicants are represented, the majority of which are pesticides. Thus, the correction factor derived from these data, may not be representative of other non-pesticide compounds.
- 2) There is a wide variation in correction factor estimates even within the toxicants listed in Table 3, reflected by the difference in functional form of the associated regression equations.

We suggest that a larger and more representative data base be used to derive these correction factors.

6. Sensitivity Factors

The use of one sensitivity factor to represent all compounds for various species of fish, in Tables 4, and 8 is justified only under the assumption that the variance of LC_{50} values, and MATC values over different species of fish is independent of toxicant.

We suggest that this hypothesis be investigated using standard statistical methods for investigating homogeneity of variance, before making this assumption. If this hypothesis is not rejected, the best estimate of the common logarithmic variance is [3]. •

$$\frac{\sum_{i=1}^k \left[\sum_{j=1}^{n_i} \left(\ln x_{ij} - \overline{\ln x_j} \right)^2 \right]}{N - k}$$

Where x_{ij} are the individual observations for a given toxicant, $\overline{\ln x_j}$ is the sum of the logarithmic values over a given toxicant, k , is the number of individual toxicants listed in a given Table, and N is the total number of observations. We suggest that under these circumstances, the above estimate be used in place of the "Average logarithmic variance" proposed by EPA in their computation of the sensitivity factor.

If this hypothesis is rejected then we suggest using a series of sensitivity factors associated with groups of toxicants of similar variance.

These comments also apply to the sensitivity factors derived in Tables 5 and 9, for invertebrates.

8. Prediction Error and Propagation of Error

No attempt has been made by EPA to estimate the prediction error associated with each correction factor, and the propagation of errors using several correction factors.

Table 8 gives the logarithmic standard deviation of the correction factors from Tables 1, 2, 3, 6, 7 and 10. It also lists the geometric mean and the 95% confidence limits for each correction factor.

The last column in Table 8 gives the average percentage error of prediction in using a given correction factor. This value is computed in the following manner [4]:

$$\text{Let } x = \hat{LC}_{50} \quad (1)$$

where LC_{50} is the standardized LC_{50} values using the estimated correction factor from the appropriate table.

$$\text{Let } u = \ln \hat{LC}_{50} \quad (2)$$

If the error in prediction, using the value $\ln \hat{LC}_{50}$ to estimate the actual value of $\ln LC_{50}$ is denoted by Δx , then:

$$\Delta x = \ln LC_{50} - \ln \hat{LC}_{50} \quad (3)$$

From (1) and (2) we have:

$$\begin{aligned} x &= e^u \\ \text{and } \Delta x &= x \Delta u \end{aligned} \quad (4)$$

Defining σ_x and σ_u as estimates of Δx , and Δu we obtain from (4) the relationship:

$$\sigma_x = x \sigma_u$$

$$\text{And } \frac{\sigma_x}{x} = \sigma_u$$

Defining error in x (LC_{50}) which results from using the correction factor as σ_x , the percentage error in LC_{50} is σ_x/x which is equal to σ_u .

It should be noted that the estimate of the logarithmic variance given in the above table is the minimum estimate of this variance, under the assumption that each individual rat in the original tables has zero error. Thus, the average percentage prediction error shown is the minimum average error.

Prediction Errors Associated With Correction Factors

Table # of EPA Docu- ment.	Correction factor	ln Stan- dard deviation	Geometric mean	95% Con- fidence limits	% error in prediction
1	Measured/ Nominal	0.3236	0.77	(0.40, 1.45)	32.36
2 (Fish)	$LC_{50}^{96} / LC_{50}^{24}$	0.4953	0.66	(0.25, 1.74)	49.53
	$LC_{50}^{96} / LC_{50}^{48}$	0.2465	0.81	(0.50, 1.31)	24.65
	$LC_{50}^{96} / LC_{50}^{72}$	0.1152	0.92	(0.73, 1.15)	11.52
3 (Fish)	Flow Through/ Static	0.6701	0.71	(0.19, 2.64)	67.01
6 (Inverte- brates)	$LC_{50}^{96} / LC_{50}^{24}$	0.8270	0.26	(0.05, 1.32)	82.70
	$LC_{50}^{96} / LC_{50}^{48}$	0.6168	0.43	(0.13, 1.44)	61.68
	$LC_{50}^{96} / LC_{50}^{72}$	0.7122	0.55	(0.14, 2.22)	71.22
7 (Inverte- brates)	Flow Through/ Static	1.3766	1.1	(0.07, 16.34)	137.66
10	HC/LC ₅₀	0.6891	0.44	(0.11, 1.69)	68.61

The error generated by using more than one correction factor is derived by using the following relationship:

Let $LC_{50}^{(e)}$ be the existing LC_{50} value and
let $LC_{50}^{(s)}$ be the standardized LC_{50} value obtained
by multiplying $LC_{50}^{(e)}$ by n correction factors $\overline{R}_1$,
 $\overline{R}_2, \dots, \overline{R}_n$.

$$\begin{aligned} \text{Then } LC_{50}^{(s)} &= \overline{R}_1 \cdot \overline{R}_2 \cdot \dots \cdot \overline{R}_n \cdot LC_{50}^{(e)} \\ \ln LC_{50}^{(s)} &= \ln LC_{50}^{(e)} + \ln \overline{R}_1 + \ln \overline{R}_2 + \\ &\quad \dots + \ln \overline{R}_n \end{aligned}$$

Define the variance of $\ln R_1$ to be σ_1^2 , $\dots$ $\ln R_n$ to be σ_n^2 ,
and assume that $\ln R_1, \dots, \ln R_n$ are independent random variables
with expected value $\ln \overline{R}_1, \ln \overline{R}_2, \dots, \ln \overline{R}_n$.

$$\text{Then var. } \ln LC_{50}^{(s)} = \sigma_1^2 + \dots + \sigma_n^2 \quad (2)$$

Data in Tables 1, 2 and 3 were used to estimate the prediction error generated by standardizing an LC_{50} value measured at 24 hours, using static methods at nominal concentration to an LC_{50} value measured at 96 hours using flow-through methods and measured concentration.

In this example, we can see from Table 8, that:

$$\sigma_1 = 0.3236$$

$$\sigma_2 = 0.4953$$

$$\sigma_3 = 0.6701$$

Substituting these values of σ_i in equations (1) and (2), we find that the minimum average error in prediction under these conditions was found to be 90%.

Data in Tables 6 and 7 were used to estimate the prediction error generated by standardizing an LC_{50} value measured at 24 hours using static methods, to an LC_{50} value measured at 96 hours using flow-through methods (for invertebrates). The minimum average error in prediction under these conditions was found to be 161%.

REFERENCE

1. Finney, D.J., 1964. Statistical Method in Biological Assay. Second Edition. (Charles Griffin & Co. Ltd., London) pp. 37-43
2. Draper, N.R., and Smith, H. 1966. Applied Regression Analysis. (John Wiley & Sons, Inc. New York) pp. 1-35
3. Zar, Jerrold H., 1974. Biostatistical Analysis. (Prentice-Hall, Inc.) pp. 135-136
4. Bevington, Philip R. 1969. Data Reduction and Error Analysis for the Physical Sciences. (McGraw Hill Book Co., New York) pp. 56-65

CMA EXHIBIT B

COMMENTS ON WATER QUALITY CRITERIA
EPA APPENDIX C
Federal Register March 15, 1979
Pages 15974-15980
and
ATTACHMENTS IV - VI

CMA 049476

The comments to the Health Guidelines will be broken into five sections.

- 1.0 A Proposal for Assessing Risk
- 2.0 The Extrapolation Models
- 3.0 Specific Comments
- 4.0 Conclusion
- 5.0 Literature Cited

1.0 A PROPOSAL FOR ASSESSING RISK

We do not agree with the policy statement that only the one hit-model should be used for risk assessment of carcinogens. Such a rigid policy precludes the use of alternate models which may best describe the experimental data.

We would like to propose a general framework for the problem of making risk assessments of carcinogens.

1. Qualitative Assessment. This may be made along the lines suggested under the heading of Effects, p 15976.
2. Quantitative Risk Assessment. Once the qualitative assessment has been made, an initial effort at quantitation is performed. The basic premise is that risk assessment is largely a judgmental process made by scientific experts and will have to be made on a case-by-case basis. For example, if the experimental dose response data is more appropriately fitted by another model then such a model should be used for extrapolation. Epidemiological data should also be considered in assigning the final risk. This will allow the use of all the experimental data in making the judgment as to risk. To do otherwise would be to ignore the data for the sake of the simplicity of using the one-hit model.
 - a. Simplistic Approach. The simple one-hit model as advanced by the Agency has utility in information gathering and for prioritizing chemicals as to which materials need additional data. If we confine our attention of the one-hit model to prioritizing, then it does have utility in the ranking of materials.
 - b. Once further data are obtained, a choice of extrapolation methods should be guided by individual facts from all experimental data relevant to carcinogenicity. Such models as the

multi-hit model, the linear extrapolation from the most reasonable data set, the pharmacokinetic model, and any other model which has relevancy to the experimental data or emerging theoretical considerations should be used.

- c. Once the risk in animals has been defined, the considerations to be used in translating to man include the following:
 - i. Metabolic similarities
 - ii. Epidemiological data
- d. In order to assess the risk in humans, a program should be designed to estimate exposure of the population at risk.

With these considerations in mind, the mathematical description of the one-hit model as advanced by the Agency appears to be adequate. This simple model should be used for prioritization of possible carcinogens and used for risk assessment only when no other data supports a less conservative model. As an example, the comments on the tetrachloroethylene document should be examined. In the comments by Dr. Hartung it is clearly indicated that the one-hit model produces a risk assessment which is contrary to what has actually been observed. In such a case the one-hit model is inadequate and a model which more closely reflects reality should be used for risk assessment.

2.0 THE EXTRAPOLATION MODEL

Historically the evaluation of safety of non-carcinogens has consisted of the application of a safety factor or uncertainty factor to an experimentally determined "no observed effect level" (NOEL). While a safety factor approach (1:5000) applied to a NOEL has been proposed as an adequate control for carcinogens (Weil, 1972), this approach has not received wide-spread acceptance.

Instead, there has been a significant trend towards the use of various extrapolation models. While there have been extensive arguments pro and con regarding the existence of threshold doses to produce carcinogens, nearly all of the presently proposed extrapolation models assume the absence of a threshold. The first extrapolation leading to a standard can be found in the BEIR Report (NAS, 1972) which used a linear model to estimate the effects of ionizing radiation on man. Most of the data in the report were derived from acute exposures of A-bomb victims or from medical x-ray exposures. In many instances several different extrapolation models could have been accommodated by the spread in the data of the BEIR Report. The linear model was chosen because it was the most conservative, i.e. it gave the highest estimates of risk

for any given dose level.

However, in the case of experimental data in animals, pure linear models of the type $F_t = bd$ (where F_t is the fractional tumor incidence, d is the daily dose, and b is the slope of the line) do not provide a good fit of the experimental data. Such data have traditionally been analyzed by probit analysis (Finney, 1971), or logistic analysis (Berkson, 1944). Both of these models are non-threshold models which fit the dose-response data after transformation.

For the probit analysis the fractional tumor incidence F_t is converted to probit (P) by the relationship:

$$F_t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{P-5} e^{-\frac{1}{2}u^2} du \quad (\text{See Table 1})$$

The dose (d) is transformed to $x = \log d$. After maximum likelihood solution a probit regression line in the form $P = a + bx$ is developed which can be extended outside the experimental range with the usual caveats.

For the logistic analysis the fractional tumor incidence is converted to logits (L) by:

$$L = \ln \frac{F_t}{1-F_t} \quad (\text{See Table 2})$$

The dose (d) is transformed to $x = \log d$. After a minimum chi-square solution a logistic regression line in the form $L = a + bx$ is developed which can be extended outside the experimental range with the usual caveats.

In general, both probit or logistic analyses provide a good fit for experimental data within the actual dose ranges of the experiment. However, when these two models are extrapolated outside the experimental range, they start to diverge significantly (see Table 3).

Mantel and Bryan (1961) developed an extrapolation method based on the probit method that provides more conservative estimates of doses required to produce risks of less than specified values. In that method they select the upper 99th percent confidence value of an experimental point and extrapolate downward from that point with a slope of 1 probit/log cycle. This is illustrated in detail in Figure 1 from Mantel and Bryan's publication, copied here without their permission. The method is applied to the data in Table 3 for comparative purposes, with

Table 1
Percent to Probit Conversion Table

	0	1	2	3	4	5	6	7	8	9
0		2.673	2.946	3.119	3.249	3.355	3.445	3.524	3.595	3.659
10	3.718	3.773	3.825	3.874	3.920	3.964	4.006	4.046	4.085	4.122
20	4.159	4.194	4.228	4.261	4.294	4.326	4.357	4.388	4.418	4.447
30	4.476	4.505	4.533	4.561	4.588	4.615	4.642	4.669	4.695	4.721
40	4.747	4.773	4.798	4.824	4.849	4.875	4.900	4.925	4.950	4.975
50	5.000	5.025	5.050	5.075	5.100	5.125	5.151	5.176	5.202	5.227
60	5.253	5.279	5.305	5.331	5.358	5.385	5.412	5.439	5.469	5.495
70	5.524	5.553	5.582	5.612	5.643	5.674	5.706	5.739	5.772	5.806
80	5.841	5.878	5.915	5.954	5.994	6.036	6.080	6.126	6.175	6.227
90	6.282	6.341	6.405	6.476	6.555	6.645	6.751	6.881	7.054	7.327

Extreme Values	
Risk	Approx. Probit
10^{-3}	1.909
10^{-4}	1.281
10^{-5}	0.735
10^{-6}	0.247
10^{-7}	-0.199
10^{-8}	-0.612

1/Calculated according to polynomial approximations of Abramowitz and Stegun(1966)
Literature Section - Abram and Stegun Handbook of Mathematical Functions(National
Bureau of Standards, 1968).

$$L = \ln \frac{P}{1-P}$$

Table 2
Percent to Logit Conversion Table

	0	1	2	3	4	5	6	7	8	9
0		-4.595	-3.892	-3.476	-3.178	-2.944	-2.752	-2.587	-2.442	-2.314
10	-2.197	-2.091	-1.992	-1.901	-1.815	-1.735	-1.658	-1.586	-1.516	-1.450
20	-1.386	-1.325	-1.266	-1.208	-1.153	-1.099	-1.046	-0.995	-0.944	-0.895
30	-0.847	-0.800	-0.754	-0.708	-0.663	-0.619	-0.575	-0.532	-0.490	-0.447
40	-0.405	-0.364	-0.323	-0.282	-0.241	-0.201	-0.160	-0.120	-0.080	-0.040
50	0.000	0.040	0.080	0.120	0.160	0.201	0.241	0.282	0.323	0.364
60	0.405	0.447	0.490	0.532	0.575	0.619	0.663	0.708	0.754	0.800
70	0.847	0.895	0.944	0.995	1.046	1.099	1.153	1.208	1.266	1.325
80	1.386	1.450	1.516	1.586	1.658	1.735	1.815	1.901	1.992	2.091
90	2.197	2.314	2.442	2.587	2.752	2.944	3.178	3.476	3.892	4.595

Extreme Values

Risk	Logit
10^{-3}	-6.907
10^{-4}	-9.210
10^{-5}	-11.513
10^{-6}	-13.816
10^{-7}	-16.118
10^{-8}	-18.421

CMA 049481

Table 3

ILLUSTRATION OF VARIOUS BASIC PROJECTION MODELS.

Assume: a dose (d) of 10mg/kg/day of compound A produces a 50% tumor incidence in man,
and a dose of 5mg/kg/day produces a tumor incidence of 25%;

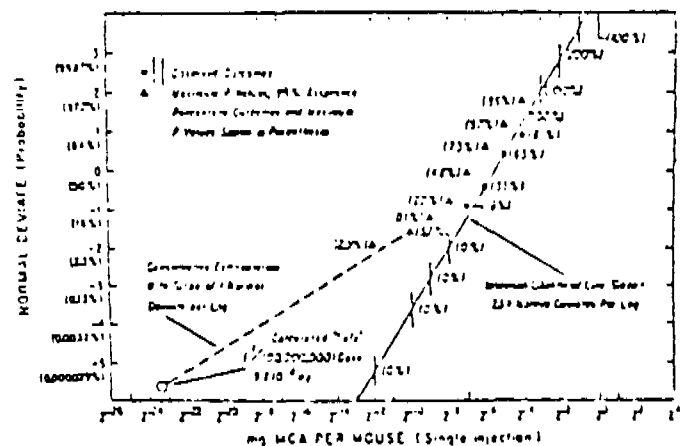
Then: the probit regression model is $P=2.761 + 2.239x$; the logistic regression model
is $L=-3.651 + 3.651x$, Where P =probit, L =logit, and x =log dose.

Fractional Tumor Incidence	Probit Projection	Logistic Projection	Projected Doses mg/kg/day		Single Hit Projections	
			Mantel Bryan * Projections d=10	d=5	d=10	d=5
0.50	10.0	10.0	10.0		10.0	12.0
0.45	8.79	8.81	7.50		8.62	10.4
0.40	7.71	7.75	5.58		7.37	8.88
0.35	6.73	6.77	4.12		6.22	7.49
0.30	5.83	5.86	2.99		5.15	6.20
0.25	5.00	5.00	2.12	5.00	4.15	5.00
0.20	4.21	4.17	1.44	3.40	3.22	3.88
0.15	3.45	3.35	0.92	2.17	2.34	2.82
0.10	2.68	2.50	0.52	1.23	1.52	1.83
0.05	1.84	1.56	0.23	0.53	0.74	0.89
0.01	0.91	0.55	0.047	0.11	0.14	0.17
10^{-3}	0.416	0.128	8.1×10^{-3}	1.9×10^{-2}	1.4×10^{-2}	1.7×10^{-2}
10^{-4}	0.218	3.0×10^{-2}	1.9×10^{-3}	4.5×10^{-3}	1.4×10^{-3}	1.7×10^{-3}
10^{-5}	0.124	7.0×10^{-3}	5.4×10^{-4}	1.3×10^{-3}	1.4×10^{-4}	1.7×10^{-4}
10^{-6}	7.5×10^{-2}	1.6×10^{-3}	1.8×10^{-4}	4.2×10^{-4}	1.4×10^{-5}	1.7×10^{-5}
10^{-7}	4.8×10^{-2}	3.8×10^{-4}	6.3×10^{-5}	1.5×10^{-4}	1.4×10^{-6}	1.7×10^{-6}
10^{-8}	3.1×10^{-2}	9.0×10^{-5}	2.4×10^{-5}	5.8×10^{-4}	1.4×10^{-7}	1.7×10^{-7}

* These Mantel-Bryan projections originate directly from the dose-response curve.
The use of a 99% confidence level would significantly decrease the dose estimates
at these series in this table.

data. Above the first 5 data points the triangles shown correspond to the maximum P values of table 1. Extrapolation, with the slope of 1 normal deviate per common log to the over-all calculated "safe" rate is indicated by a broken line. All triangles, other than the one from w.

...as to make impractical
by humans. But we are
and if any compounds are
zero, this is one of them.
...analysis of the experi-
on on the vertical scale,
...is shown as negative
...present the outcomes at
...shown by arrow. The
...profit has fitted to the



Text-figure 1—Estimation of the "safe" dose from test results with a carcinogen, methylcholanthrene, at several dose levels. At each test level both the observed percentage response and an upper limit, 99 percent assurance, based on combi data are shown. Solid line is the maximum likelihood probit line fitted to data. The "safe" level of 9×10^{-4} mg per mouse is in this instance estimated by extrapolation with the conservative slope of 1 normal deviate per log from upper limit on P at the second dose level.

the difference that the extrapolation is made from the central value rather than an upper 99th percent confidence value. The method combines two "safety factors" by extrapolating at a very shallow slope which is lower than normally encountered in empirical data, and by extrapolating from an upper 99% confidence value. The Mantel-Bryan extrapolation model has received only limited acceptance, possibly due to the unknowns associated with the assumptions used in the extrapolation process.

More recently, several variants of the "single hit" model have been used as the basis for extrapolation models. The basis of this model is the theory that a tumor can be produced by the interaction of a single molecule with a susceptible target molecule in the organism. In its simplest terms this model can be presented as (Hoel et al., 1975):

$$F_t = 1 - e^{-Rd}$$

where R has been termed potency and the other terms are as defined previously. Then, for very low doses this relationship becomes linear, so that

$$F_t = Rd$$

A variant of this model has been adopted by the U.S.E.P.A. to extrapolate carcinogenic data to man. The model attempts to introduce various correction factors for the duration of the test, the size of the test animal, and the incidence of tumors in control animals. It should be noted that the single hit model often does not provide as good a fit of the experimental data as the probit or logistic models (see example in Table 3).

The National Academy of Sciences Safe Drinking Water Committee proposed the use of a multi-step model for use in extrapolating carcinogenicity data (NAS, 1977).

This model was developed by Armitage and Doll (1961) and is presented by:

$$F_t = 1 - e^{-(\lambda_0 + \lambda_1 d + \lambda_2 d^2 + \dots + \lambda_k d^k)}$$

It is clear that there are many extrapolation models, and that these models result in significantly different extrapolations (Table 3). The extrapolations at risk levels of $1:10^8$ are grossly different and exceed any observed differences between species and also exceed any differences observed for any experimental designs.

It should be noted that each one of these extrapolation models is conceptually based on very simplistic foundations. None of the models considers the multiplicity of causative and repair factors that influence the final occurrence of a hit. A valid extrapolation model should consider at a minimum the causative factors, metabolism,

immunology, repair and individual variability. Since the balance between these factors may differ drastically for different chemicals it is that a single extrapolation model cannot be effective to cover all chemicals. If there were such a theoretical model, then it would certainly be much more complex than the single-hit model.

3.0 SPECIFIC COMMENTS

Page 15974. Under the section Objective (I) it is stated that ideally one would want to derive criteria for absolute safety. However, the determination of absolute safety is in itself a logical impossibility. The Agency makes this same point in the preamble, consequently the statement on p 15974 should be deleted as an objective. While some of the aspects of this problem are covered in the discussion under objectives, they are not sufficiently clear.

The difficulties in establishing safe or absolutely safe levels are inherent in the process of safety evaluations. Most scientific investigations seek to determine that specific findings have really occurred and that they are not due to the vagaries of chance. While this is the exact approach taken during the evaluation of environmental and toxicological processes, the application of data gathered in these investigations to safety evaluations require significant change in the handling of the data. In addition to data which demonstrates statistically significant effects, it is necessary to attempt evaluation of data which demonstrate no effect. To demonstrate that a "no-measured effect level" is real and not due to statistical variability is difficult. The difficulties are escalated when one considers that all studies are conducted in model test systems. The results from such studies are used to estimate the chances that no untoward effects will be seen in different species, under different circumstances and from much larger populations. Under such circumstances, it becomes clearly impossible to prove the negative, that is to prove absolute safety. Thus, safety evaluations are inherently judgmental and only the relative level of safety can be described, but absolute safety cannot be guaranteed.

Page 15975. Types of Criteria (II)

The use of the terms "stochastic effects" versus "non-stochastic effects" is misleading, especially since they are defined in terms of the occurrence of a threshold and the absence of a threshold. Such distinctions should be made strictly in terms of threshold and non-threshold models rather than confusing the issue with the terms of stochastic versus non-stochastic, which in themselves do not necessarily define the occurrence or absence

of a threshold. At the present time, the concept of threshold versus non-threshold effects relate largely to the mathematical models applied to the experimental data rather than to the empirical data themselves. To clarify, it is common to find dose levels at which experimentally no effects are found even for events such as carcinogenicity. To clearly establish whether a threshold does or does not exist on an experimental basis is difficult. Laboratory data which demonstrate no carcinogenic effect at a certain low dose level, when such an effect is seen at a high dose level, are always open to the conjectural criticism that the effect might have been seen if more animals had been exposed. Epidemiological data are also limited in their applicability because of background rates of incidence of tumors in populations which have apparently had no exposure to a particular chemical. Since the empirical data are not in themselves conclusive in establishing which events are threshold events and which ones are non-threshold events, the selection of those effects which were to be treated by non-threshold models has been limited to one category of effects and that is carcinogenicity. Since various defense and repair mechanisms have been identified even in the induction of cancers by chemicals, it has not been clearly established that the same non-threshold model can be applied to the evaluation of very low dose level of all chemical carcinogens.

The question of the possible occurrence of thresholds cannot be clearly answered at the present time and should thus be left open. A better understanding of metabolic effects, immunological responses and repair mechanisms will undoubtedly suggest practical threshold phenomena which may relate to key steps of the carcinogenic process on a case-by-case basis.

The statement on organoleptic effects should be re-worded to read: "It is recognized that criteria based solely on organoleptic effects do not represent satisfactory approximations of low risk levels."

The Approach, p 15975

We approve of the Approach in that the four indicated topics need to be addressed, namely, exposure, pharmacokinetics, toxicity and criterion formulations. We especially commend the Agency for their recognition of the importance of pharmacokinetics as a method for assessing the species to species extrapolation and in characterizing the mode of toxic action. Clearly, the more we know about the mechanism the more reliable will be the resulting criterion.

Page 15976. The refinements of the use of the bioconcentration factor appear to be very useful. And the distinction between those substances which are lipid soluble and those which are non-lipid soluble is very necessary.

The reference to the EPA Interagency Cancer Guidelines should indicate that this document is under intensive peer review and the conclusions could change once the review is completed. We agree with the Agency on their recognition that a chemical which has not induced a significant cancer response in humans or experimental animals is not identified as a possible carcinogen. In this connection the Agency cites that there are cases where weak carcinogenic effects are observed but the data cannot be used to establish criteria. We would add that bioassay systems that are conducted at doses greater than the maximum tolerated dose or by routes other than oral may suggest that the chemical has potential for human cancer but should not be used to establish numerical Water Quality Criteria.

Pages 15977-15978. The assumption that those chemicals which induce mutations are also likely to induce cancer in animals and therefore presumably in people is only a theory. While the majority of materials which have been shown to be carcinogenic in animals are also capable of inducing mutations, it has not been demonstrated what proportion of chemicals that are found to be mutagenic will also be found to be carcinogenic in animal tests. In addition, epidemiological evidence for man has failed to provide evidence of carcinogenicity for many materials which have been found to be carcinogenic in animal tests. In addition, it must be noted that there are a few chemicals which appear to be carcinogenic in man by epidemiological evaluations, which to date have not been shown to be carcinogenic in laboratory animals. Therefore, it is clear that the extrapolation of data gathered in bacterial systems to experimental animals to humans is fraught with great difficulty. At the present time, the finding of mutagenic effects can only be utilized as a screening test and for the establishment of priorities for further testing before materials come into contact with man.

To date, the selection of animal models for carcinogenicity testing has been relatively non-critical and has been largely restricted to the selection of animal strains which have demonstrated high sensitivity to chemical carcinogenesis. In the face of all these uncertainties, even the animal data should be examined critically on a case-by-case basis.

In the discussion section on stochastic effects and the justification for the selection of the single-hit model, it is stated "The possibility of threshold mechanisms has given support to other dose response models (Guess et al., 1977) for extrapolation of cancer risks, such as log-probit, logit and multi-stage models. These models are nonlinear and result in lower risks

than the one-hit model for a given low exposure." This statement is at least partially in error and the error is significant. All of the extrapolation models cited above are non-threshold models and, in addition, all of these models have the characteristic that they tend to fit the experimental data better than the single-hit model does. At extremely low dose levels these models are close enough to linearity that the selection or rejection of one model over the other cannot be made on the basis of experimental data even for such artificial models as mutagenesis in bacterial systems induced by radiation or chemical. Similarly, all of these models can be readily applied to the epidemiological studies done in man and will fit well within the margins of error provided by those studies. Therefore, most of the criteria used for the selection of the single-hit model are also applicable to the log-probit, logit and multi-stage models. Thus, the only reason remaining for the selection of the model is one of policy and the fact that the single-hit model provides greater risk estimates than other models.

The selection of the extrapolation model to be used requires a great deal more attention and justification than has been given in this criterion document. The reasons for that will be obvious when one examines in detail the consequences of selection of various extrapolation models.

4.0 CONCLUSION

In summary, we feel the alternative approach suggested in our Specific Comments is a more reasoned approach to setting health effects criteria for potential carcinogens. The Agency's use of the one-hit model precludes the use of any other model which may better fit the data. CMA urges strong consideration of our alternative approach.

5.0 LITERATURE CITED

- Armitage, P. and R. Doll. 1961. Stochastic models for carcinogenesis. Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability. 4:19-38 Univ. of California Press.
- Berkson, J. 1944. Application of the logistic function to bioassay. J. Am. Stat. Assoc. 39:357-365.
- Doll, R. 1971. The age distribution of cancer: implications for models of carcinogenesis. J. Prog. Stat. Soc. 13:133-166.
- Finney, D.Z. 1971. Probit Analysis. 3rd ed. Cambridge U. Press, 333 pp.
- Hoel, D.G. et al. 1975. Estimation of risks of irreversible, delayed toxicity. J. Toxicol. & Environ. Health 1:133-151.
- Mantel, N. and W.R. Bryan. 1961. "Safety" testing of carcinogenic agents. J. National Cancer Inst. 27:455-470.
- NAS. 1977. Drinking Water and Health. Safe Drinking Water Committee. National Academy of Sciences, Washington, D.C. 939 pp.
- NAS/NRC. 1972. The effects on populations of exposure to low levels of ionizing radiation. (BEIR Report). Div. of Med. Sci. NAS/NRC, Washington, D.C. 217 pp.
- Ramsey, J.C. et al. 1979. Carcinogenic risk assessment: ethylene dibromide. Toxicol. & Appl. Pharmacol. 47:411-414.
- Weil, C.S. 1972. Statistics vs. safety factors and scientific judgment in the evaluation of safety for man. Toxicol. & Appl. Pharmacol. 21:454-463.