research for
man and environment

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEUHYGIENE
NATIONAL INSTITUTE OF PUBLIC HEALTH AND ENVIRONMENTAL PROTECTION

IMPORTANT-READ CAREFULLY BEFORE OPENING SOFTWARE PACKAGE.
By opening the sealed package containing the software, you indicate your acceptance of the following ETX License Agreement. If the terms of this License Agreement are not acceptable, the unopened software should be returned promptly to the place of purchase for a full refund.

ETX License Agreement

This is the ETX License Agreement between you (either an individual or an entity) and RIVM (the National Institute of Public Health and Environmental Protection, Bilthoven, the Netherlands).

By this License Agreement, you are granted a license to use ETX subject to the following terms, restrictions and limitations:

1. The license to use this software is non-exclusive. RIVM retains title and all ownership rights to the software.
2. Publications and reports, containing results obtained by using ETX, have to refer to RIVM as the copyright owner of the software.
3. This software may be installed on, but not used on, more than one computer at a time. You may, however, make copies of ETX for achival or back-up purposes only.
4. You are not allowed to transfer the software and/or accompanying written materials to another party.
5. Neither RIVM, nor the programmer, Tom Aldenberg, will be liable for any direct, indirect, special, incidental, or consequential damages arising out of the use of, or inability to use, this software or documentation. In particular, neither RIVM, nor Tom Aldenberg, is responsible for any costs including, but not limited to, those incurred as a result of lost profits or revenue, loss of use of software, loss of data, the costs of recovery of such software or data, the costs of substitute software or data, claims by third parties or for other similar costs.
6. RIVM will not provide any after sales services. However, notification of errors, or any other remarks, by the user will be acknowledged.
7. RIVM will inform you about new versions, or extensions, of the software and about other related programs. You will have the opportunity to obtain new versions at prices to be fixed by RIVM.
8. This Agreement is effective upon acceptance by you and shall remain in effect perpetually, except if amended in writing and signed by both parties.
9. Any dispute arising out of this Agreement shall be settled by the competent Court of The Hague.
10. This Agreement will be governed by the laws of the Netherlands.

Antonie van Leeuwenhoeklaan 9, P.O. Box 1, 3720 BA BILTHOVEN, The Netherlands. Telephone: ..(31) 30 - 74 91 11, Telex: 47215 rivm nl, Telefax: ..(31) 30 - 74 29 71

Confidence Limits for Hazardous Concentrations Based on Logistically Distributed NOEC Toxicity Data

TOM ALDENBERG AND WOUT SLOB

National Institute of Public Health and Environmental Protection (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Reprinted from ECOTOXICOLOGY AND ENVIRONMENTAL SAFETY, Volume 25, No. 1, February 1993

ECOTOXICOLOGY AND ENVIRONMENTAL SAFETY 25, 48–63 (1993)

Confidence Limits for Hazardous Concentrations Based on Logistically Distributed NOEC Toxicity Data

TOM ALDENBERG AND WOUT SLOB

National Institute of Public Health and Environmental Protection (RIVM), P.O. Box 1, 3720 BA Bilthoven, The Netherlands

Received May 15, 1992

This paper deals with the calculation of Hazardous Concentrations of toxic substances from small sets of laboratory toxicity data, e.g., NOECs. A procedure due to Van Straalen and Denneman, as adapted from Kooijman (case $n = 1$), in which one seeks a concentration that protects 95% of the biological species is modified to account for the uncertainty in the estimates. New constants are obtained by simulation. These allow the calculation of the one-sided 95% left confidence limit of the Hazardous Concentration, from the mean and standard deviation of a sample of (laboratory) toxicity data. This 95% confidence limit is always lower than the 95% certainty value calculated with the Kooijman ($n = 1$)/Van Straalen method. The authors also derive constants to calculate a one-sided 50% confidence value, that overpredicts as often as it underpredicts. This value may be used as a median guess of the Hazardous Concentration. It will always be higher than the 95% certainty value of the Kooijman ($n = 1$)/Van Straalen method. However, by using the 50% value, one runs the risk of protecting substantially less than 95% of the biological species.

INTRODUCTION

This paper estimates safety factors for the extrapolation of laboratory toxicity data to allowable toxic substance concentrations in the field, using statistical methodology. Species differ in their sensitivity to a toxic substance. The statistical approach focuses on some presupposed distribution of these species sensitivities for a particular substance. In fact, this article treats some essential modifications to earlier procedures; hence, for a motivational introduction we refer to the original articles: Kooijman (1987) and Van Straalen and Denneman (1989).

In Kooijman (1987), a hazardous concentration for sensitive species (HCS) is defined, and an algorithm is given for its computation from a sample of LC_{50} values of different test species on the basis of the logistic distribution. Several, more or less independent, components in his theory are the choice of input data (LC_{50}'s), the type of statistical distribution employed (logistic), the definition of hazardous concentration, and the statistical methodology, i.e., the algorithm to calculate hazardous concentrations from small samples of toxicity data.

This study essentially follows a modification of Kooijman's theory by Van Straalen and Denneman (1989). Whereas Kooijman considered the probability of harming the most sensitive of a number of species, e.g., 1000, Van Straalen and Denneman will be followed in considering the probability of just harming species. This is the current approach in the Netherlands (Health Council, 1989, DGEP, 1988–1989). The authors also follow Van Straalen and Denneman in their choice and motivation with regard to the input data, NOEC toxicity data, instead of LC_{50} data in Kooijman, and adhere to the choice of the logistic distribution as well.

0147-6513/93 $5.00

However, an alternative approach has been developed to the statistical methodology in calculating the agreed upon hazardous concentration levels, and this is the main concern of this paper. Hence, the presentation is statistically oriented. Of course, different calculation methodologies lead to different outcomes as regards to what seems a justifiable safety factor, or acceptable concentration, and this is where the environmental implications cannot be easily overestimated. However, these implications are discussed elsewhere.

According to Van Straalen and Denneman (1989), a concentration of a certain compound is considered hazardous for $p\%$ of the species, if the probability of selecting a species with a NOEC smaller than this concentration is equal to $p\%$. In other words, above this concentration, called HC_p, $(100 - p)\%$ of the species is relatively safe, while $p\%$ of the species may not function properly or even worse. The general approach is to strive for 95% species protection, i.e., $p = 5$.

Figure 1 shows the logistic probability density function against the logarithmic NOEC concentration. The logistic distribution is very much like the well-known normal distribution. The logistic has more extended tails and therefore can be regarded as a more conservative assumption in comparison to the normal distribution. It, furthermore, has some nice mathematical features that make certain calculations relatively easy. (Most of the technical aspects have been provided in the Appendix.) The base of the logarithm by which the raw NOEC data are transformed does not matter, as long as the back-transformation of the results to concentrations is done with respect to the same base. Hence, the generic term "log" that may either stand for natural logarithms or for logs to the base 10, or otherwise, has been used. Also indicated in Fig. 1 is "log HC_5," the logarithm of HC_5, below which 5% of the species is in danger (shown shaded). In fact, one is looking for the fifth percentile of the distribution of (laboratory) species NOEC toxicity data. The difficulty is how to account for uncertainty in trying to estimate this percentile from a limited data set.

In this paper, improved extrapolation constants that allow straightforward calculation of estimates of HC_5 from mean and standard deviation of a sample of NOEC data are presented. The procedure is essentially identical to the one of Van Straalen and Denneman (1989), but the focus is on meeting the required confidence level exactly, in order to protect against overprediction. The previous extrapolation constants are shown to lead to unacceptably high percentages of overprediction of the true HC_5, and therefore do not meet their confidence level. Furthermore, constants that can be considered as a best guess and that overpredict as often as not are obtained to calculate estimates of HC_5. As an example, the cadmium data from Van Straalen and Denneman has been recalculated.

ESTIMATING HAZARDOUS CONCENTRATIONS

In order to estimate the agreed upon hazardous concentration (95% species protection) from a usually small number of toxicity data, a statistical procedure must be developed to correct for uncertainty due to small sample size. Hence, there is a need to quantify the uncertainty of the estimates, and one certainly does not want to overestimate too often. Therefore, a confidence approach seems natural.

Suppose one knew the mean, μ, and the standard deviation, σ, of the presupposed logistic distribution of log NOEC data of test species, as the one depicted in Fig. 1. Then the log Hazardous Concentration for 5% of the species is easily calculated as (cf. Appendix)

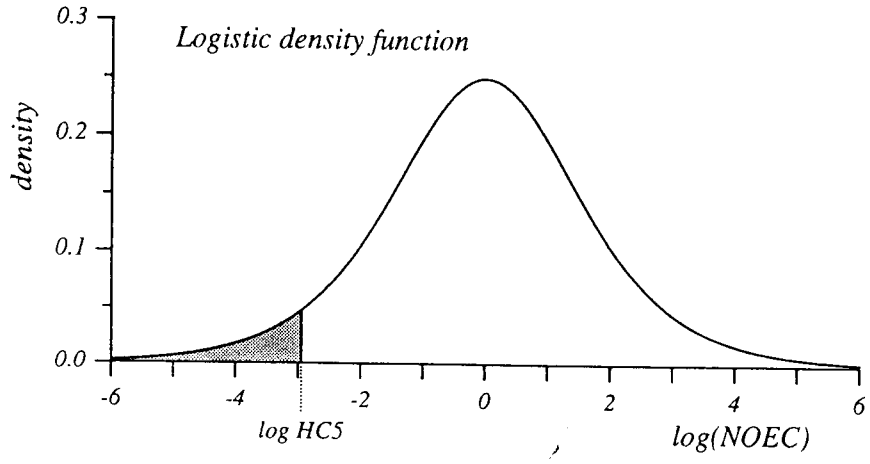

FIG. 1. The standard logistic distribution of log(NOEC) values. Log $HC_5 = -2.94$ is the log Hazardous Concentration to be estimated. The fraction of the species harmed is shown shaded.

$$\log HC_5 = \mu - 1.62 \cdot \sigma.$$

One can estimate mean and standard deviation from the usual sample mean, $\bar{x}_m$, and sample standard deviation, s_m, of m test species and estimate the log Hazardous Concentration straightforwardly, i.e., by substituting the sample statistics for the population statistics:

$$Z = \bar{x}_m - k_Z \cdot s_m.$$

With $k_Z = 1.62$, one acts as if mean and standard deviation did not come from a sample, but were the true ones, but this would suffer from frequent overprediction.

Figure 2 shows sampling distributions of Z for sample sizes $m = 2$, 5, 10, and 20. These sampling densities are simulated through Monte Carlo sampling (cf. Appendix for details). The respective percentages of overprediction are estimated to be 67, 61, 57, and 55%. Note that all of them overestimate by more than 50%. If Z in a particular sample would come out higher than log HC_5, then obviously more than 5% of the species may be affected. In fact, a recipe is wanted that overestimates log HC_5 in a minority of samples only, so that with large *confidence* one can say that no more than 5% of the species is affected.

Kooijman/Van Straalen Extrapolation Constants

The reason for reconsidering this estimation question is that Kooijman (1987) does not intend to construct an estimate with this confidence property—in fact, his Equation (16) and subsequent derivations cannot be motivated from a confidence point of view—while Van Straalen and Denneman (1989) do interpret the results that way.

The final expression (Kooijman 1987, Eq. (24); Van Straalen and Denneman 1989, Eq. (6)), which is called K here, looks very similar to Z,

$$K = \bar{x}_m - k_K \cdot s_m,$$

only with a different k-value, here called k_K. For an estimate based on a sample, this constant depends on the sample size. The original expression for k_K is given in the References and repeated in the Appendix.

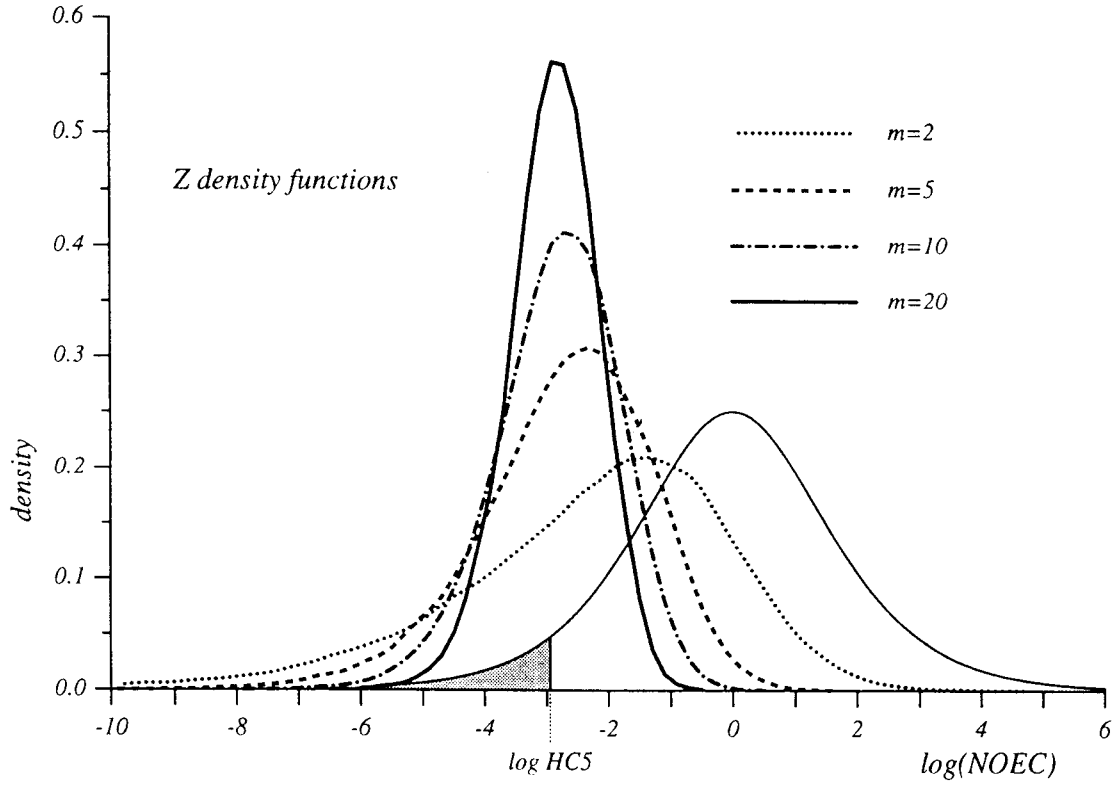

FIG. 2. Simulated variation of the answers that result from using Z to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $Z = \bar{x}_m - 1.62 \cdot s_m$. Z is the formula for estimating the hazardous concentration, as if we had complete information about the logistic curve, i.e., without adjusting for uncertainty, due to a limited sample.

Table 1 lists various values of k_K for two certainty levels: 95 and 50%, calculated from Kooijman (1987). The term "certainty" is ours to distinguish these constants from those to be given in the next paragraph. The reason to consider 50% certainty or confidence will be discussed later. The columns in Table 1 correspond to $\delta_2 = 0.05$ and 0.5 in Kooijman (1987, Table 1), respectively.

Note that the asymptotic value corresponds to the value 1.62 under complete knowledge about the mean and standard deviation of the distribution at hand. But, if interpreted as constants to calculate the log HC_5 with a certain level of *confidence*, Table 1 is suspect for two reasons. First, for 95% certainty, the constants do not seem to "blow up" enough for decreasing sample sizes, e.g., $m = 4, 3, 2$. This effect is well-known for confidence limits of the mean in normal distribution theory, Student's t values, and one expects it to be even worse for confidence limits of a tail value, what log HC_5 essentially is. But, second, the 50% column, if interpreted as confidence factors, seems to be on the wrong side of 1.62 anyway. It tells us that it is better to have 10 NOEC values than 30, which in its turn is better than an infinite number of test data available. This is suspect, because the Z-estimates with $k = 1.62$ already overpredict for more than 50%, so, how can smaller k-values overpredict less? This does not seem realistic.

In order to test the confidence property of the Kooijman/Van Straalen extrapolation constants, the K-95% and K-50% sampling distributions have been simulated, based on the extrapolation constants in Table 1, in the same way as the Z densities have been simulated. Figure 3 displays the K-95% sampling densities for the same set of

TABLE 1

EXTRAPOLATION CONSTANTS $k_K = (3/\pi^2) \cdot d_m \cdot C_5^1$ FOR 95% SPECIES PROTECTION (COMMUNITY SIZE: $n = 1$), CALCULATED FROM KOOIJMAN (1987) FOR VARIOUS VALUES OF m, THE NUMBER OF TEST SPECIES FOR WHICH log(NOEC)s ARE AVAILABLE

m	95%	50%
2	3.33	1.00
3	3.04	1.26
4	2.88	1.40
5	2.74	1.48
6	2.62	1.50
7	2.52	1.50
8	2.43	1.51
9	2.37	1.51
10	2.32	1.52
11	2.29	1.52
12	2.26	1.53
13	2.25	1.53
14	2.24	1.54
15	2.23	1.54
20	2.18	1.58
30	2.06	1.58
∞	1.62	1.62

Note. The resulting log Hazardous Concentration is $K = \bar{x}_m - k_K \cdot s_m$, where $\bar{x}_m$ and s_m are mean and standard deviation, respectively, for a sample of size m. The two columns refer to 95 and 50% certainty, respectively.

sample sizes as before, 2, 5, 10, and 20. The authors observe considerable overprediction of log HC_5. Figure 4 shows the simulated K-50% sampling densities. These indeed seem to overpredict even more than the corresponding Z densities.

Table 2 summarizes the overprediction percentages for these four sample sizes. If K is to be interpreted as a one-sided 95% left confidence limit, the percentage of simulated samples with a K-value *above* log $HC_5 = -2.94$ should be somewhere in the vicinity of 5%. The percentages estimated (39, 22, 20, and 14%, respectively) seem to be unacceptably high. The same holds for a one-sided 50% confidence value. Overprediction should approximate 50%. These simulated values (83, 67, 65, and 60%) seem to be too high as well.

In the next paragraph, extrapolation constants that lead to estimates of log HC_5 that do have the required confidence interpretation are calculated.

New Extrapolation Constants on Two Levels of Confidence

In order to construct an expression L that calculates the 95% species protection level with true one-sided 95 and 50% confidence levels, one need not develop an essentially new methodology. In fact, if the same type of formula is used,

$$L = \bar{x}_m - k_L \cdot s_m,$$

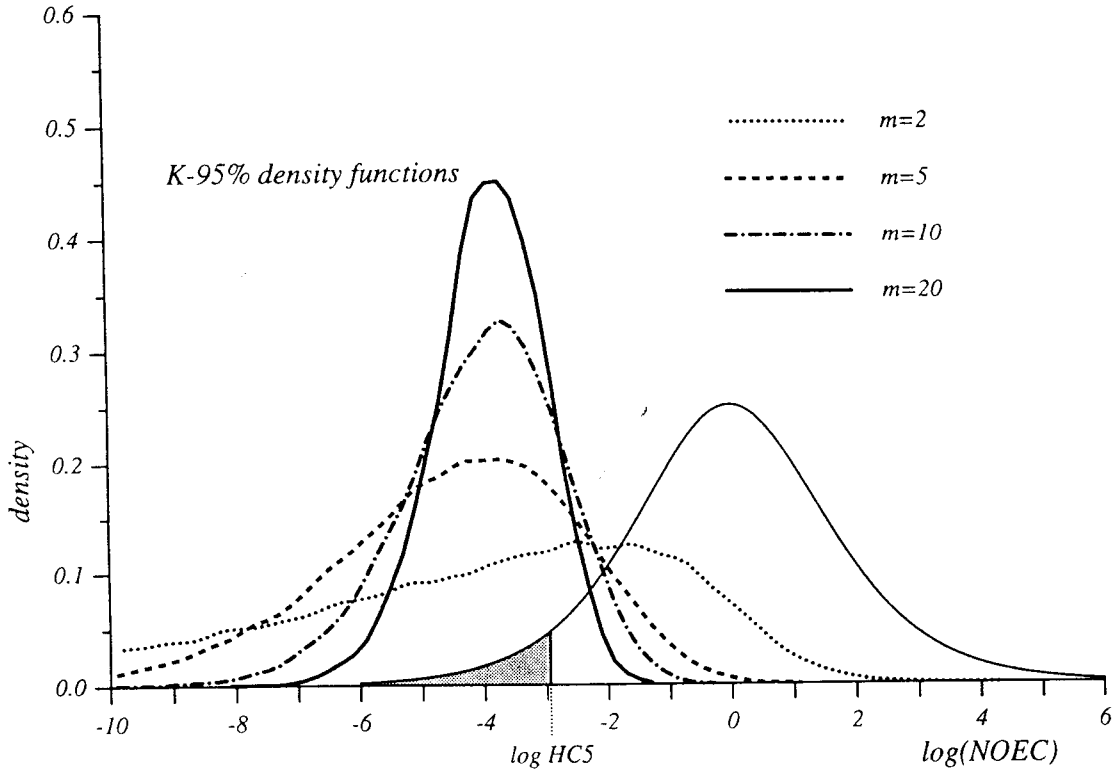

FIG. 3. Simulated variation of the answers that result from using $K - 95\%$ from Kooijman and Van Straalen (95% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $K = \bar{x}_m - k_K \cdot s_m$, with k_K from Table 1 (95%).

and if we focus on the new k_L extrapolation constants for different m, it is easy to prove that, for each m, there is just one value of k_L with the required confidence property for *any* logistic distribution (cf. Appendix). Thus, for each sample size, k_L-values have been determined through Monte Carlo simulation by generating random sample averages and standard deviations for the *standard* logistic distribution only and by adjusting k_L in such a way that a prespecified confidence level was obtained. These are tabulated in Table 3.

Figure 5 shows the sampling densities of the one-sided 95% left confidence limits (L-95%) for $m = 2$, 5, 10, and 20, as determined by the new extrapolation constants. Each one overestimates log HC_5 with 5%, as they should. Figure 6 displays the sampling densities of the one-sided 50% confidence limits (L-50%). They overpredict as well as underpredict with 50%.

Clearly, the extrapolation constants of Table 3 would pass the test of Table 2, since they are constructed that way. The percentages overprediction would be 5 and 50%, respectively. Moreover, the new constants do show the expected Student t-like blow-up for small m. Furthermore, contrary to the 50% certainty constants in Table 1, the 50% confidence extrapolation constants for finite samples are higher than the asymptotic value, i.e., 1.62 (k_Z), for "infinite" samples. This means that a one-sided 50% confidence estimate of log HC_5 must still be lower than the straightforward answer (Z), acting as if one knew the logistic parameters.

At publication, the authors came into contact with Wagner and Løkke (1991), who derived extrapolation constants for the 95% species protection level, when a *normal*

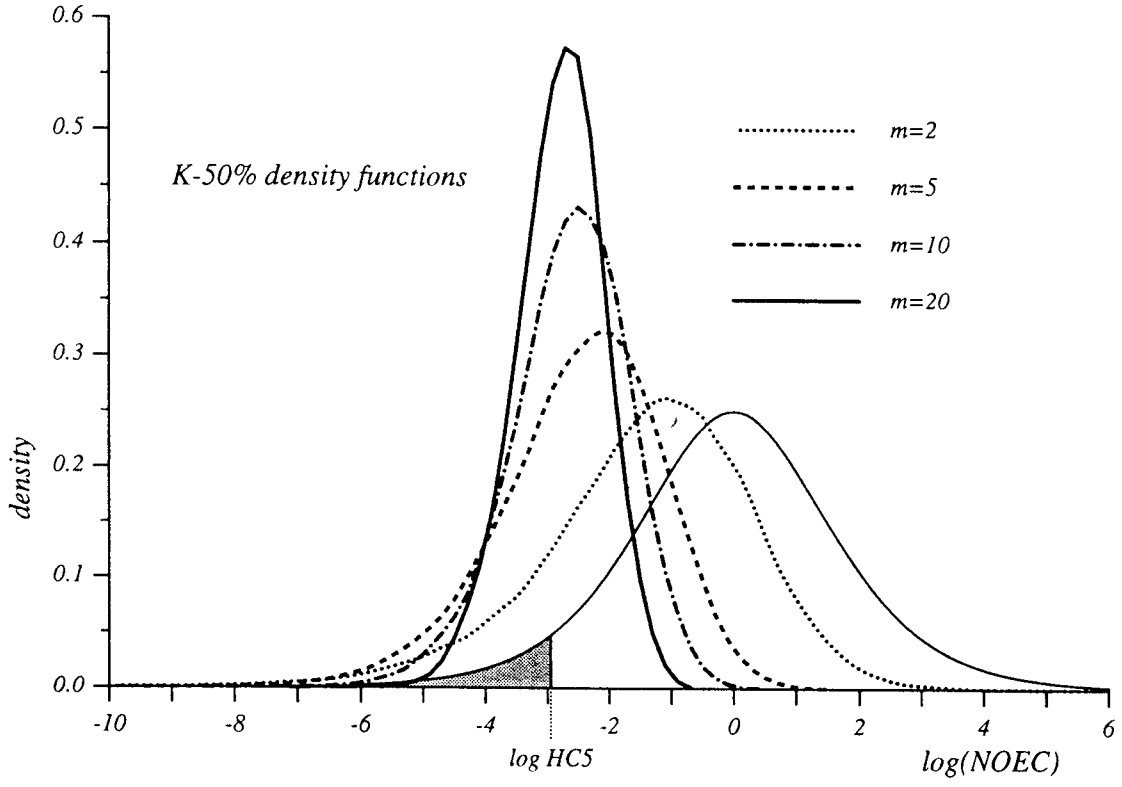

FIG. 4. Simulated variation of the answers that result from using $K - 50\%$ from Kooijman and Van Straalen (50% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $K = \bar{x}_m - k_K \cdot s_m$, with k_K from Table 1 (50%).

distribution is assumed, by using existing theory that applies to the normal distribution only. The resulting extrapolation constants are very similar to those presented here for the logistic distribution.

EXAMPLE

As an example, the HC_5 is calculated from seven NOEC values for toxicity of cadmium to reproductive parameters of various soil animals, corrected for standard

TABLE 2

MONTE CARLO SIMULATION OF THE ONE-SIDED 95 AND 50% (LEFT) CONFIDENCE LIMIT PROPERTY OF K (SEE TEXT), i.e., ln HCS IN KOOIJMAN (1987)

Sample size, m	Extrapolation constant (95%)	Percentage overprediction (95%)	Extrapolation constant (50%)	Percentage overprediction (50%)
2	3.33	39%	1.00	83%
5	2.74	22%	1.48	67%
10	2.32	20%	1.52	65%
20	2.18	14%	1.58	60%

Note. Extrapolation constants are from Table 1. The percentages overprediction should approximate 5% for 95% certainty and 50% for 50% certainty. These correspond to areas below the curves in Figs. 3 and 4, to the *right* of log HC_5.

TABLE 3

EXTRAPOLATION CONSTANTS FOR THE CALCULATION OF ONE-SIDED LEFT CONFIDENCE LIMITS FOR THE LOGARITHMIC HAZARDOUS CONCENTRATION FOR 5% OF THE SPECIES ON THE BASIS OF THE LOGISTIC DISTRIBUTION

m	95%	50%
2	27.70	2.49
3	8.14	2.05
4	5.49	1.92
5	4.47	1.85
6	3.93	1.81
7	3.59	1.78
8	3.37	1.76
9	3.19	1.75
10	3.06	1.73
11	2.96	1.72
12	2.87	1.72
13	2.80	1.71
14	2.74	1.70
15	2.68	1.70
20	2.49	1.68
30	2.28	1.66
50	2.10	1.65
100	1.95	1.64
200	1.85	1.63
500	1.76	1.63
∞	1.62	1.62

Note. Tabulated values are k_L such that a one-sided left confidence limit L for log HC_5 is given by $L = \bar{x}_m - k_L \cdot s_m$. Here $\bar{x}_m$ and s_m are mean and standard deviation, respectively, of a sample of log(NOEC) test data of size m. Constants are tabulated for two levels of confidence: 95 and 50%.

soil (Van Straalen and Denneman, 1989, Table 2) and compared to theirs. The sorted data are 0.97, 3.33, 3.63, 13.5, 13.8, 18.7, and 154 ($\mu g/g$).

After transformation with base 10 logarithms, the mean is $\bar{x}_7 = 0.9712$ and standard deviation $s_7 = 0.7028$, respectively. The Kooijman/Van Straalen estimate of the HC_5 for 95% certainty is

$$10^{(0.9712-2.52\cdot 0.7028)} = 0.16\ (\mu g/g).$$

Note that the authors directly employ the Kooijman (1987) extrapolation constant 2.52 from Table 1, entry number 7 in this paper. Second, it is easy to demonstrate that the base of the logarithm does not matter. When using the mean and standard deviation on the basis of natural logarithms, i.e., 2.236 and 1.618, respectively, one arrives at the same result,

$$e^{(2.236-2.52\cdot 1.618)} = 0.16\ (\mu g/g).$$

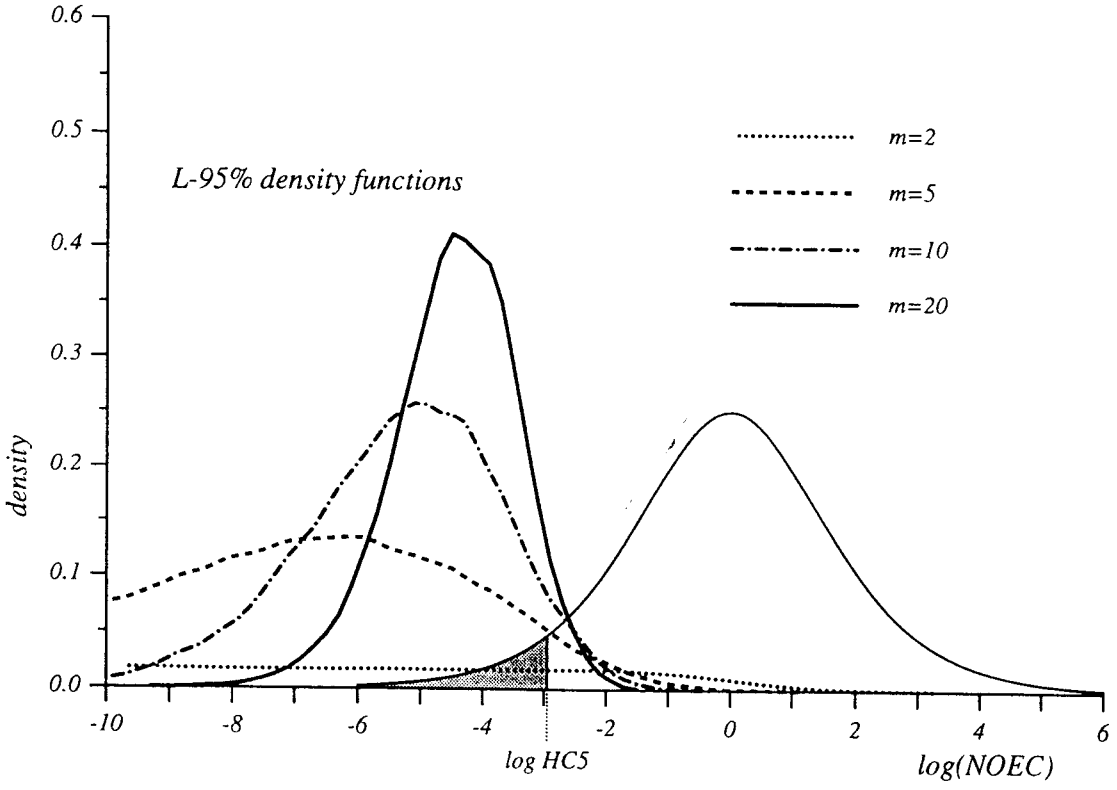

FIG. 5. Simulated variation of the answers that result from using $L - 95\%$ (this paper, 95% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $L = \bar{x}_m - k_L \cdot s_m$, with k_L from Table 3 (95%).

(Van Straalen and Denneman, 1989, Table 3). And this is true in general of course. By using the new Table 3 extrapolation constants, 3.59 and 1.78, for a sample size of 7, one arrives at the 95% left confidence limit of

$$10^{(0.9712-3.59\cdot 0.7028)} = 0.03\ (\mu g/g).$$

while the 50% confidence estimate of HC_5 is

$$10^{(0.9712-1.78\cdot 0.7028)} = 0.53\ (\mu g/g).$$

Note that the 95% lower confidence limit (0.03) and the 50% confidence, or "median," estimate (0.53), embrace the Kooijman/Van Straalen estimate (0.16). This will *always be the case,* as can easily be seen by comparing the 95% column from Table 1 with the 95 and 50% columns of Table 3. The former k constant is always between the latter two for corresponding sample sizes.

It is interesting to observe that if one really wants to limit the probability to overestimate HC_5 to only 5%, a safety factor of

$$T = 10^{(3.59\cdot 0.7028)} = 333$$

has to be applied, instead of 59, as estimated by Van Straalen and Denneman (1989, Table 3), for this example.

Hence, one may conclude that, if one wants to have 95% confidence to not overestimate the 95% species protection level, one has to calculate values that are generally lower than those calculated up to now.

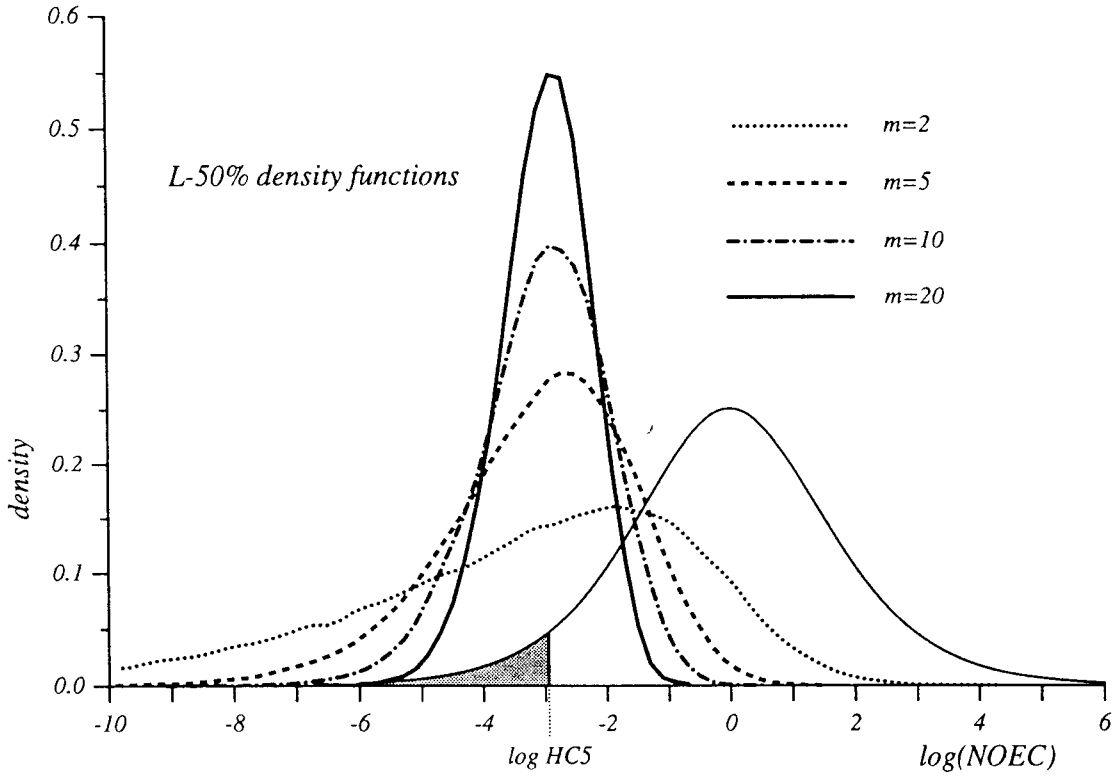

FIG. 6. Simulated variation of the answers that result from using L – 50% (this paper, 50% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $L = \bar{x}_m - k_L \cdot s_m$, with k_L from Table 3 (50%).

DISCUSSION

The authors have given extrapolation constants on two levels of confidence: 95 and 50%. The larger confidence level of the two can easily be motivated thus: one wishes to protect *at least* 95% of the species, hence one wants to limit overprediction of the true log HC_5 to a small percentage, say 5%. But why calculate a 50% confidence estimate?

First of all, there is a practical reason. The authors have found it confusing to present one left confidence limit value as *the* single answer to an extrapolation exercise. Users start asking for a confidence interval for it and forget that it is already a confidence limit. So there is a need for a middle value, which could as easily be too high as it could be too low. Thus, in analogy with a classical two-sided confidence interval for the mean of a normal distribution, e.g., a value ± a half-range, we could use the 50% confidence value as the middle value and the 95% confidence value as a one-sided left confidence limit. Which one of these values to use in an assessment of a particular situation, with all kinds of practical considerations involved, eventually is a matter of policy or decision making. However, in this decision process, the following, more theoretical, considerations should be taken into account.

The presently followed approach to estimate hazardous concentrations for ecosystems from a small set of single species data illustrates the basic principle of risk analysis in the face of uncertainty. In this situation two levels of risk must be dealt with. The primary risk is what one is interested in and what one wants to estimate (or keep low).

In the present paper, the primary risk is the percentage of species that is actually harmed.

The secondary risk is the risk that the estimate of the primary risk is wrong. In this paper, the secondary risk is set by the confidence level. If the results of the analysis are to be used as a basis for action, e.g., to determine a maximum tolerable concentration for ecosystems, the secondary risk should be taken into account. Both Kooijman and Van Straalen felt that the secondary risk should be low (5%). Yet, there have been recent discussions on the necessity of this low value; it has even been suggested that a confidence level of 50% should be accepted as the single answer to work with. However, it does not seem to make much sense to demand a low value for the primary risk and at the same time allow a high secondary risk.

Figures 7 and 8 illustrate the risks of using a 50% confidence level on the basis of 5% harm to the species for calculating maximum tolerable concentrations for ecosystems. Figure 7 shows the primary and secondary risks for a sample of five test species. The fraction unprotected species is shaded. The risk of overprediction, and thereby of harming a higher fraction of the species, is hatched horizontally. When using the 50% confidence estimate to estimate log HC_5, this risk is 50%. The risk that more than a certain percentage is harmed, on the basis of this 50% confidence estimate, can be calculated, by varying the percentage unprotected, and hence percentage overprediction, that is, by shifting the vertical line at log HC_p in Fig. 7 to the right, while keeping the sampling density, i.e., the distribution of results for five test species (dashed line), fixed. While the risk that more than 5% of the species is not protected is 50%, the risk that more than 10% of the species is not protected is 30%, whereas the risk that even 20% or more of the species is not protected is still 12%. Figure 8 shows the risks of harming larger percentages of species, for several values of m (number of species tested), constructed in this way.

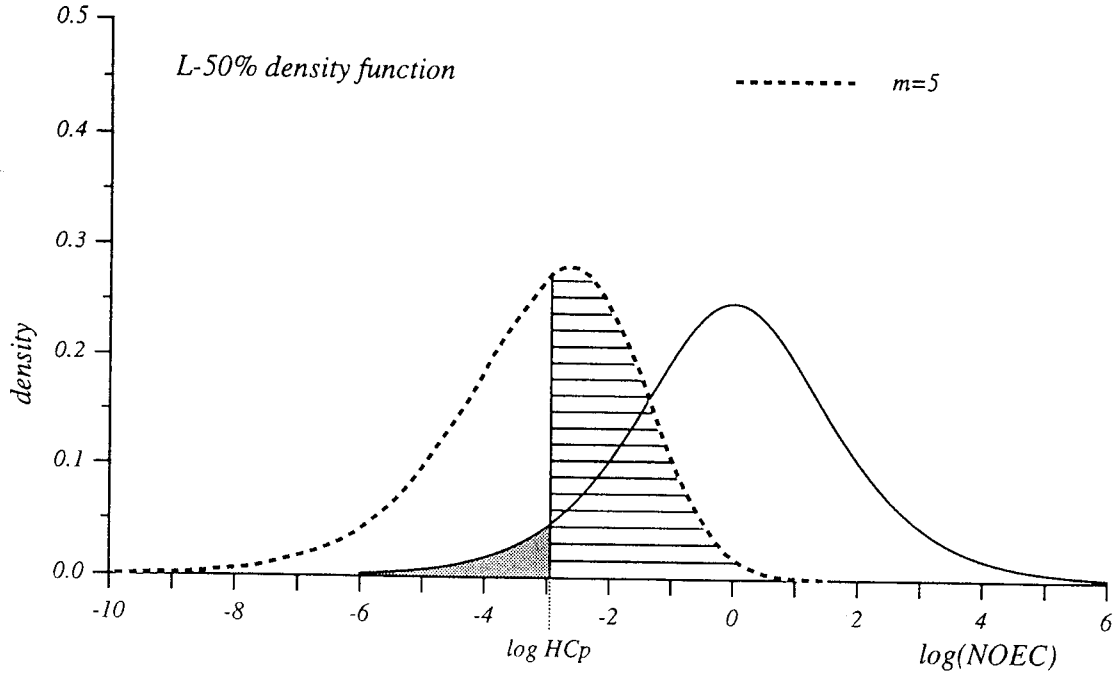

FIG. 7. Primary risk, i.e., the fraction of species not protected (shaded region), and secondary risk of overprediction and thereby harming a larger fraction of species (hatched region) in case of a 50% confidence estimate for log HC_5 for a sample of five test species.

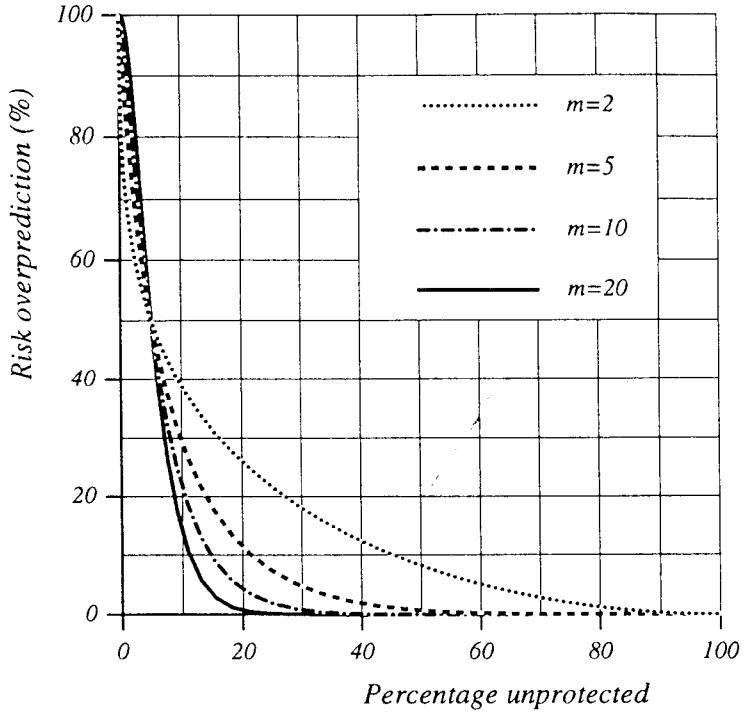

FIG. 8. Risk (ordinate) that more than a given percentage of the species (abscissa) is unprotected when using the 50% confidence estimate of the log Hazardous Concentration for 5% of the species. Due to the nature of this estimate, all curves necessarily intersect at the 5% harm, 50% overprediction point. However, the risk that higher percentages are harmed may not be negligible.

The authors, therefore, suggest the calculation of both the 50% and the 95% confidence value. The first value can be regarded as the best estimate of the hazardous concentration, whereas the latter may be taken as the "safe" value (given the assumptions underlying the calculations, of course). Comparison of these values can be used for deciding to examine more species: large differences between both values indicate considerable uncertainty.

The great virtue of regarding the 95% confidence value as the safe value is that it tends to outweigh ecological and economical interests. If this safe value, based on the available data, appears to be low enough to have important economical drawbacks, one would not hesitate to investigate more species, since the associated reduction of uncertainty might quite well result in higher values for the safe concentration. On the other hand, using the 50% confidence value as an indication of the safe value results in a strong bias toward economical interests. One would test a minimum number of species, hoping that the coin falls on the right side; if not, one could always extend the number of test organisms afterward. Obviously, this situation would be quite harmful from an ecological viewpoint.

CONCLUSIONS

The method of Van Straalen and Denneman (1989), as adapted from Kooijman (1987), to predict Hazardous Concentrations for toxic substances has a risk of overpredicting these concentrations that does not match the intended level of confidence of the estimates. As a result, larger percentages of species may get harmed than one

would accept. The refined estimates in this paper have a risk of overprediction that correspond to given levels of confidence, e.g., 50 and 95%, for which extrapolation constants are presented.

APPENDIX: COMPILATION OF SOME MATHEMATICAL ASPECTS

In this Appendix, some of the more mathematical technicalities have been compiled. The probability density function (f) of species toxicity data is supposed to be logistic:

$$f(x) = \frac{1}{\beta} \cdot \frac{\exp(-(x - \alpha)/\beta)}{(1 + \exp(-(x - \alpha)/\beta))^2}.$$

Here, x stands for logarithmic NOEC data (the base of the logarithm does not matter), α is the location parameter, and β is the scale parameter. The mean (also median), to be called μ, and the standard deviation, to be called σ, can be expressed in α and β:

$$\mu = \alpha$$

$$\sigma = \beta \cdot \frac{\pi}{\sqrt{3}} = 1.8138 \cdot \beta.$$

So, the standard deviation of the logistic distribution is roughly two times as large as the value of β. The standard logistic distribution, used in the simulations, has $\alpha = 0$ and $\beta = 1$ and therefore a standard deviation $\sigma = 1.8138$.

The cumulative distribution (F) of species NOEC toxicity data describes the probability for those log NOEC values to be smaller than x:

$$F(x) = \frac{1}{1 + \exp(-(x - \alpha)/\beta)}.$$

One of the advantages of the logistic distribution over the normal distribution is the fact that this distribution can be represented in the explicit form stated. For example, for purposes of simulation, we need to generate many random logistic data. Due to the explicitness of the cumulative distribution, these can be easily generated with

$$x^{\alpha,\beta} = \alpha - \beta \cdot \ln\left(\frac{1 - U}{U}\right),$$

where U is a uniform random number. Note that

$$x^{\alpha,\beta} = \alpha + \beta \cdot x^{0,1}$$

A second example where the explicitness of the cumulative distribution comes in handy is the calculation of the log Hazardous Concentration for p% of the species under complete knowledge of the distribution. Then, one can equate $F(x)$ to $p/100$ and solve explicitly for x,

$$x = \log \mathrm{HC}_p = \alpha - \beta \cdot C_p^1,$$

where

$$C_p^1 = \ln\left(\frac{100 - p}{p}\right).$$

For example, for $p = 5$, that is 95% species protection, we have

$$C_5^1 = 2.9444.$$

But one can also express log HC_5 in μ and σ as follows:

$$\log HC_5 = \alpha - \beta \cdot C_5^1 = \mu - \sigma \cdot \frac{\sqrt{3}}{\pi} \cdot C_5^1 = \mu - 1.6234 \cdot \sigma.$$

This expression allows the calculation of the log Hazardous Concentration, if mean and standard deviation of the distribution are known.

In the original approach of Kooijman, the calculations are similar. The probability that the log NOEC of the most sensitive of n species is smaller than x is (Kooijman, 1987)

$$F_n(x) = 1 - (1 - F(x))^n,$$

with $F(x)$ the single species cumulative distribution given before. (The notation here differs from Kooijman's.) Equating this to $q/100$ (called δ_1 in Kooijman) and solving for x gives the log Hazardous Concentration for Sensitive species,

$$x = \log HCS_q^n = \alpha - \beta \cdot C_q^n,$$

where

$$C_q^n = \ln\left(\frac{(1 - q/100)^{1/n}}{1 - (1 - q/100)^{1/n}}\right).$$

When comparing C_p^1 with C_q^n, it easily follows that for $n = 1$, $C_p^1 = C_q^1$ if and only if $p = q$. This demonstrates the mathematical relationship between the Van Straalen and Denneman's (1989) hazardous concentration for p% of the species and Kooijman's (1987) hazardous concentration for p% of the most sensitive of "communities" of *one* species:

$$HC_p = HCS_p^1.$$

In all estimates, the sample mean and sample standard deviation are used to estimate mean and standard deviation of the supposed distribution:

$$\hat{\mu} = \bar{x}_m = \sum_{i=1}^{m} \frac{x_i}{m},$$

$$\hat{\sigma} = s_m = \sqrt{\left(\sum_{i=1}^{m} \frac{(x_i - \bar{x}_m)^2}{m - 1}\right)}.$$

A simple estimate for log HC_5 neglecting uncertainty due to a limited sample size is

$$Z = \bar{x}_m - 1.6234 \cdot s_m.$$

Table 3, in the 50% column, in fact shows that this estimate overpredicts in more than 50% of the cases.

Instead of $k = 1.6234$, other constants may be derived to account for uncertainty. These necessarily depend on m. The extrapolation constant due to Kooijman (1987),

as applied by Van Straalen and Denneman (1989) with community size of 1 and 95% species protection is

$$k_K = \frac{3}{\pi^2} \cdot d_m \cdot C_5^1,$$

with d_m as tabulated in Table 1 of Kooijman (1987). These k_K constants are tabulated in Table 1 of this paper for two levels of certainty that correspond with Kooijman's $\delta_2 = 0.05$ and $\delta_2 = 0.5$. With these constants, the Kooijman algorithm for calculating a left certainty limit (our terminology) of log HC_5 becomes

$$K = \bar{x}_m - k_K \cdot s_m,$$

A new extrapolation constant k_L is tabulated in Table 3 for calculating a one-sided left confidence limit of log HC_5, called L:

$$L = \bar{x}_m - k_L \cdot s_m.$$

L satisfies the required confidence level.

However, the determination of these constants turned out to be a surprisingly hard numerical exercise. Each constant in Table 3 is an average of 20 such simulations with roundabout 250,000 sample points each, e.g., 30,000 samples for $m = 8$ (cf. 500 in Kooijman, 1987). That means that each constant is based on roughly five million drawings from the standard logistic distribution. We still cannot guarantee every second decimal in k_L, though, but the true confidence level will be closely approximated.

The simulated densities depicted in Figs. 2–6 are estimated as follows. We generated 60,000 samples of size $m = 2$ and 5, plus 30,000 of size 10, plus 10,000 of size 20. All data were drawn from the standard logistic distribution. For each sample, the mean $\bar{x}_m$ and standard deviation s_m was calculated, along with Z-, K-, and L-values. These were sorted and converted to histogram densities with bin width 0.2. The histogram midpoint values were smoothed with three-point running means with weights 1:2:1 and plotted.

Next follows the proof, referred to in the main text, that if k_L were the proper extrapolation constant for a particular sample size in the case of the *standard* logistic distribution, then $L = \bar{x} - k_L \cdot s$, for that same sample size, would have the correct confidence property for *any* logistic distribution.

Suppose $\bar{x}^{0,1}$ is a standard logistic sample average (sample size m), $s^{0,1}$ is a standard logistic sample standard deviation, and suppose that

$$L^{0,1} = \bar{x}^{0,1} - k_L \cdot s^{0,1}$$

overestimates the true log $HC_5^{0,1} = -C_5^1$ with known probability. Now, given the sample size, consider the statistic,

$$L^{\alpha,\beta} = \bar{x}^{\alpha,\beta} - k_L \cdot s^{\alpha,\beta},$$

with $\bar{x}^{\alpha,\beta}$ and $s^{\alpha,\beta}$ the sample mean and sample standard deviation, respectively, for some arbitrary logistic distribution. Then, the probability that it overestimates log HC_5 is

$$\begin{aligned}\Pr\{L^{\alpha,\beta} > \log \mathrm{HC}_5^{\alpha,\beta}\} &= \Pr\{\bar{x}^{\alpha,\beta} - k_L \cdot s^{\alpha,\beta} > \log \mathrm{HC}_5^{\alpha,\beta}\} \\ &= \Pr\{\alpha + \beta \cdot \bar{x}^{0,1} - \beta \cdot k_L \cdot s^{0,1} > \alpha + \beta \cdot \log \mathrm{HC}_5^{0,1}\} \\ &= \Pr\{\bar{x}^{0,1} - k_L \cdot s^{0,1} > \log \mathrm{HC}_5^{0,1}\} \\ &= \Pr\{L^{0,1} > \log \mathrm{HC}_5^{0,1}\},\end{aligned}$$

which was assumed to be known. Hence, L for any arbitrary logistic overestimates the corresponding log HC_5 with that same probability. Therefore, for each m, one has to calculate k_L only once, e.g., for the standard logistic distribution.

ACKNOWLEDGMENTS

We acknowledge oral and written discussion with Bas Kooijman and Nico van Straalen. Janneke Hoekstra commented on the manuscript.

REFERENCES

DGEP (1988–1989). *Premises for Risk Management: Risk Limits in the Context of Environmental Policy.* Directorate General for Environmental Protection at the Ministry of Housing, Physical Planning and Environment. Second Chamber of the States General, session 1988–1989, 21137, No. 5.

HEALTH COUNCIL (1989). *Assessing the Risk of Toxic Chemicals for Ecosystems.* Report No. 28/E, Health Council (*Gezondheidsraad*), 's-Gravenhage, The Netherlands.

KOOIJMAN, S. A. L. M. (1987). A safety factor for LC_{50} values allowing for differences in sensitivity among species. *Water Res.* **21,** 269–276.

VAN STRAALEN, N. M., AND DENNEMAN, C. A. J. (1989). Ecotoxicological evaluation of soil quality criteria. *Ecotoxicol. Environ. Saf.* **18,** 241–251.

WAGNER, C., AND LØKKE, H. (1991). Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Res.* **25,** 1237–1242.

E_TX 1.3a

A Program to Calculate Confidence Limits for Hazardous Concentrations Based on Small Samples of Toxicity Data

Tom Aldenberg

rivm
National Institute of Public Health &
Environmental Protection, P.O.Box 1,
3720 BA Bilthoven, The Netherlands

May 1993 719102015

DISCLAIMER: Neither RIVM, Bilthoven, The Netherlands, nor the programmer, Tom Aldenberg, will be liable for any direct, indirect, special, incidental, or consequential damages arising out of the use of or inability to use this software or documentation. In particular, neither RIVM, nor Tom Aldenberg is responsible for any costs including, but not limited to, those incurred as a result of lost profits or revenue, loss of use of software, loss of data, the costs of recovery such software or data, the costs of substitute software or data, claims by third parties or for other similar costs.

E_TX 1.3a

A Program to Calculate Confidence Limits for Hazardous Concentrations Based on Small Samples of Toxicity Data

Tom Aldenberg

rivm
National Institute of Public Health &
Environmental Protection, P.O.Box 1,
3720 BA Bilthoven, The Netherlands

May 1993 719102015

Contents

I. Introduction

Welcome to E_TX, the Ecotoxicological Extrapolation Program from **rivm**!
E_TX is a computer program running on MS-DOS computers. E_TX, short for EcoToX, currently version *1.3a*, handles the extrapolation of laboratory toxicity data to values, that may be of interest to policy makers in setting standards for environmental protection. The program may also be of use in other areas, e.g. in human health-oriented problems.
As a short motivating example, suppose one is confronted with the next seven NOEC (No Effect) concentrations for Cadmium for various different soil fauna species:

```
0.97   3.33   3.63   13.5   13.8   18.7   154
```

in some unit. These are real data from Van Straalen & Denneman (1989), who adapted this extrapolation technique from Kooijman (1987). Then, E_TX calculates `0.53` as the estimate of the 5th percentile of the hypothetical statistical distribution from which the data are thought to derive. This 5th percentile is the so-called *hazardous* concentration, above which 95% of the species seems relatively safe. We also speak about 95% species protection. This hazardous concentration is also indicated as the HC_5.
The estimate of the hazardous concentration just employed is a so-called *median* estimate: if everyone would calculate this estimate for similar batches of seven data in the same manner, for instance by using E_TX, then the median of the distribution of answers, of which `0.53` is one particular instance, would equal the hazardous concentration.
E_TX also calculates a second value, `0.03`, for this example, that, if everyone

would do so for their data, would result in a distribution of answers of which the 95th percentile would equal the hazardous concentration. That is, we are *confident* that we *under*estimate the hazardous concentration by 95%. One can either think of this estimate as the *left confidence limit* of a 90% double-sided confidence interval, or as the one-sided 95% confidence underestimate.
The basic idea, henceforth, is that laboratory species display different sensitivities with regard to the adverse effects of a particular toxic substance, as expressed by NOEC concentrations or LC_{50} concentrations. If nothing is assumed mechanistically, these species NOECs, or whatever, are thought to derive from some statistical distribution. 'Extrapolation', as it is called, amounts to estimating percentiles of this distribution with a certain confidence from a perhaps small set of toxicity data.
The statistical theory behind extrapolation to percentiles is treated in Erickson & Stephan (1985), Kooijman (1987), Van Straalen & Denneman (1989), Wagner & Lokke (1991), and Aldenberg & Slob (1993). E_TX is conceived as a tool for the statistical analysis of toxicity data sets. Although E_TX is relatively user-friendly, and can be run by decision makers, it is not specifically designed as a decision support system for setting environmental standards. E_TX does not care about what the nature of the data is, one feeds to it. It may be used on NOEC concentrations, EC_5 concentrations, LC_{50} concentrations, etc. In fact, the data may not be toxicity data at all, but one should be aware of the fact that the data are always log transformed. The toxicity data may pertain to any taxon level: species, genera, or even higher taxa (e.g. Crustaceans, Mammals, etc.). This means that a toxicity data set may consist of a batch of species toxicity data, or a batch of toxicity data, one for each genus, a batch of phylum data, and so on.
Nor does E_TX know about any specific environmental protection terminology, such as permissible risk levels, maximum allowable concentrations, safe or reference or background concentrations. E_TX is a program to experiment with different toxicological data sets, different confidence levels, and different species protection levels. Right now, it handles three types of statistical distribution, i.e. the log-logistic, the log-normal, and the log-triangular, one species protection level (95%), and two levels of confidence of underestimation (95% and 50%). But these may be extended in future versions. Hence, E_TX may develop over time, as more analysis tools are incorporated. We would welcome any user remarks, that could lead to improvement.
The reader who wants more information about the practical and theoretical considerations in a decision makers framework is referred to Slooff (1992) and OECD (1992).
Next to the estimation of hazardous concentrations from laboratory toxicity data,

or *extrapolation*, initial steps have been taken in E_TX to incorporate the estimation of species hazard at given environmental or experimental concentrations, here called *hazard assessment*. Whereas extrapolation goes from a pre-set species protection level and a batch of toxicity data to concentrations to be declared as environmental standard, or objective, hazard assessment goes from current or predicted environmental concentrations to estimated species protection levels. Here also, a confidence approach would be implied, but more work has to be done on that. Extrapolation is treated in the literature cited, but hazard assessment, as defined here in a statistical extrapolation-oriented framework, is treated in Appendix A of this Manual.

Acknowledgments.

I would like to thank *Kees van Leeuwen* for being the motor behind this program, and for his incredible patience; *Hans Canton, Wilbert Slooff*, and *Wout Slob* for their stimulating interest. *Esther Guinée, Janneke Hoekstra, John Janus, Robert Luttik, Erik van de Plassche, Marieke Polder*, and *Theo Traas* commented on earlier versions of the program. The correspondence with *Nico van Straalen* is acknowledged. *Maarten 't Hart* helped with the final testing and documentation.

TA

II. User's Guide

A. Installation

E_TX can be run on MS-DOS Personal Computers, or so-called Compatibles, right out of the box from floppy disk. So, strictly speaking, there is no obligation to install E_TX on a hard disk. Obviously, it is more convenient to do so.
Installation under MS-DOS is very simple. Be sure, you see the DOS prompt `C:>` after starting the PC. If wanted, make a directory to place E_TX in, by typing:

```
MD ETX,
```

or some other directory name. Go to this directory:

```
CD ETX
```

Put the E_TX floppy disk in drive `A:` or `B:`. Copy all the files from the E_TX floppy disk to the hard disk (`C:`) by typing:

```
COPY A:*.*
```

or:

```
COPY B:*.*
```

After the files have been copied, remove the floppy disk from the floppy disk unit.
Now you are ready to run E_TX from the hard disk.

B. Running E_TX

Running E_TX is even simpler. If the file **ETX.EXE** is in the current directory, just type:

```
ETX
```

Otherwise, first go to the directory, where **ETX.EXE** is located. If the MS-DOS PATH is set by you, or someone else, to scan the **ETX-directory** for commands, then typing **ETX** will work from any directory. In the current version, data files that are saved by, or are to be used by E_TX are put into the same directory, as where the program runs. This might change in a future release.
E_TX is operated through a menu system. You may be familiar with menu systems by other programs. If not, learning to operate the E_TX menu system should not cause any problem. 'Menu' options constitute alternative choices for the user of a program to control a program.
Menu items are displayed vertically in E_TX. There may be ten items, but usually less. Each list of menu options makes a menu screen. Menu options ending with a slash (/) lead, when activated, to a submenu containing new options. Those ending with a period (.), when activated, cause some action to be taken by E_TX. Options within parentheses, if any, are currently not implemented.
The main menu screen of E_TX after starting the program reads:

```
ETX 1.3a:

     1. Data/
     2. Statistics/
     3. Extrapolation/
     4. Hazard/
     5. Results/
     6. Quit/

[Edit Keys] to Select; [Enter] to Activate:
```

All options have trailing slashes, so each choice, when activated, gives rise to a new menu. These six options form a logical sequence of doing an extrapolation analysis: getting data, study them, carry out extrapolating exercises

and hazard assessment, collect output, and quit the program. But they need not be chosen in this exact order. If you try to do illogical things, however, such as extrapolation without entering toxicity data, *ETX* will complain.
One can activate a menu item in either of two ways. The first is by pressing the number that precedes the item. *ETX* should immediately respond to that. Or, secondly, one can use the arrow keys, or edit keys: **[Down]**, **[Up]**, and so on, to highlight next and previous options. In this case, no action is taken, until you press the **[Enter]** key. The last line of the menu screen, the Active Key bar, indicates what keys are active. On the screen and in this manual, brackets denote a key to be pressed, not a sequence of characters. So, **[Enter]** is the Enter or Return key.
When you are in a submenu, or sub-submenu, pressing **[Esc]**, or **[PgUp]**, brings you up one level. Pressing **[F10]** brings you up to the main menu from anywhere in the tree. Pressing initial characters does *not* activate a menu item.
Above the menu options, you find the name of the program (**ETX**) and its current version number (**1.3a**). If it says also something like **Beta**, followed by a number, then you do not have an official release, but a version distributed for review.
This title bar of a menu screen is also used to indicate where the user is located in the menu hierarchy. When you are not at the main, or top menu level, then previous menu choices have been added to the program and version indicator, separated by slashes (**/**). This results in a representation of the path followed through the menu hierarchy (tree), similar to the way operating systems display directory hierarchies. In this way, going down the menu tree (if main menu is thought to be the highest level) makes the menu path string to grow, while going back up the hierarchy does make it shrink, until you are back at the top level, and only the program and version indicator remain.
Hence, if the path string at the top of a menu screen reads:

```
ETX 1.3 a/ Statistics/ Logarithms/ Toxicity:
```

one has apparently chosen option **2. Statistics** from the *ETX* main menu, then **1. Logarithms** from the Statistics menu, then **1. Toxicity** from the Logarithms menu, and one is currently facing the two options and Active Key bar:

```
     1. As Is.
     2. Sorted.

[Edit Keys] to Select; [Enter] to Activate; [Esc] to Quit:
```

These options have trailing periods, so they will do something for you. Option

1 will show you a list of the $\log_{10}$s of the toxicity data, if there are any, in the order as they have been entered. Option 2 will do the same thing, only in sorted order from smallest to largest. Since the action follows immediately, and the menu screen will be cleared, these action options are not added to the menu path string anymore. Note the extended Active Key bar, with **[Esc] to Quit** added, indicating that **[Esc]** (the Escape key) brings you back up.
In the Reference Manual, we will represent subsequent menu choices by the path traveled through the menu tree, as given by the menu title bar, but with two slight modifications: the leading program version indicator, **ETX 1.3a**, is abbreviated to **ETX**, and last menu options with trailing periods that lead to an action are added to the path. So, referring to the previous example, the act of listing the toxicity data logarithmically in the original order is effectuated through this sequence of menu choices:

ETX/ Statistics/ Logarithms/ Toxicity/ As Is.

From now on, we refer to such strings as the menu path, or path, of a command. Uncomplete paths, ending with a slash, are also called paths.

C. Leaving E_TX

As soon as you are done, you can quit (leave) E_TX from main menu through the menu choices: **ETX/ Quit/ Leave ETX!** Note that we have used the path indication method explained in the previous paragraph. On the screen, menu option **Leave ETX** has a trailing exclamation mark, which indicates that there is no mercy in case you have not saved the data that you may have entered. There is *no extra warning* if you haven't done so!
You can find out about the status of the data under **ETX/ Data/ Show/**, that will display the current data set, as well as inform the user, where the data came from, and whether they were saved.
A very crude way to leave the program at any point in the menu hierarchy is pressing **[Ctrl]+[End]** and then **[Ctrl]+[Y]**. If you miss **[Ctrl]+[Y]**, you are back where you were. Use this only in a great hurry.
A real crash route is **[Ctrl]+[Break]**, which aborts the program at once.

III. Reference Manual

The Reference Manual consists of two parts:

A. an overview of the E_TX Menu hierarchy or tree, including a short description of what each option means or does, and

B. an alphabetical Reference list for lookup of terms relating to E_TX: menu options and general issues like Files, Version, and so on.

A. Menu tree

The E_TX menu system is a tree-like hierarchy. At the so-called root of the tree, usually considered the highest level of the menu, we have E_TX itself. This level is activated by typing **ETX** at the MS-DOS prompt. After clearing the opening screen, we arrive at the first level down the tree, with six branches. The menu screen reads:

```
ETX 1.3a:

1. Data/
2. Statistics/
3. Extrapolation/
4. Hazard/
5. Results/
6. Quit/
```

followed by the Active Key bar. The **Data** menu leads, when activated (see User's Guide), to another menu screen at depth two. This reads:

```
ETX 1.3a/Data:

1. Enter/
2. Read/
3. Show/
4. Edit/
5. Save/
```

Note the growing menu path at the first line. Now, activating menu option **Enter** leads to just another menu screen at depth three.
Below follows a complete representation of the tree with both the layout and decimal numbering indicating the hierarchical structure. Each option keyword is shortly explained between parentheses. If an option does not have a submenu, then it denotes some action to be taken by E_TX, often leading to further prompts for user action, like entering or modifying data, actions to save data in disk files, listing of results, statistical output, etc.

E_TX menu tree:

```
1 Data/                  (data management)
     1.1 Enter/          (enter new data through the keyboard)
          1.1.1 Toxicity.     (enter laboratory toxicity data)
          1.1.2 Exposure.     (enter environmental exposure data)
     1.2 Read/           (read data from files)
          1.2.1 Toxicity.     (read toxicity data)
          1.2.2 Exposure.     (read exposure data)
     1.3 Show/           (look through the data)
          1.3.1 Toxicity/     (show toxicity data)
               1.3.1.1 As Is.      (in the original order)
               1.3.1.2 Sorted.     (in sorted order)
          1.3.2 Exposure/     (show exposure data)
               1.3.2.1 As Is.      (in the original order)
               1.3.2.2 Sorted.     (in sorted order)
     1.4 Edit/           (modify entered or read data)
          1.4.1 Toxicity.     (edit toxicity data)
     1.5 Save/           (save data to files on disk)
          1.5.1 Toxicity/     (save toxicity data)
               1.5.1.1 As Is.      (in the original order)
               1.5.1.2 Sorted.     (in sorted order)
          1.5.2 Exposure/     (save exposure data)
               1.5.2.1 As Is.      (in the original order)
               1.5.2.2 Sorted.     (in sorted order)
2 Statistics/        (look at statistical data evaluations)
     2.1 Logarithms/     (show log10 transformed data)
```

```
          2.1.1 Toxicity/      (logarithmic toxicity data)
               2.1.1.1 As Is.       (in the original order)
               2.1.1.2 Sorted.      (in sorted order)
          2.1.2 Exposure/      (logarithmic exposure data)
               2.1.2.1 As Is.       (in the original order)
               2.1.2.2 Sorted.      (in sorted order)
     2.2 Basic Statistics/(look at simple data summaries)
          2.2.1 Toxicity/      (of the toxicity data)
               2.2.1.1 Logistic.    (the logistic distribution)
               2.2.1.2 Normal.      (the normal distribution)
     2.3 Goodness-of-Fit/(Goodness-of-Fit of distributions)
          2.3.1 Toxicity/      (fitted on the toxicity data)
               2.3.1.1 Logistic.    (the logistic distribution)
               2.3.1.2 Normal.      (the normal distribution)
3 Extrapolation/     (look at extrapolation results)
     3.1 Logistic.          (for the logistic distribution)
     3.2 Normal.            (for the normal distribution)
     3.3 Triangular.        (for the triangular distribution)
4 Hazard/            (look at hazard assessment results)
     4.1 Logistic/          (for the logistic distribution)
          4.1.1 Fixed Hazard.(exposures at fixed hazard levels)
          4.1.2 At Exposure/  (at the exposure data)
               4.1.2.1 As Is.       (in the original order)
               4.1.2.2 Sorted.      (in sorted order)
5 Results/           (collect the results)
     5.1 Save/              (save results on disk)
          5.1.1 All.              (save all results)
6 Quit/              (quit ETX)
     6.1 Main Menu.         (no, back to main)
     6.2 Leave ETX!         (yes, definitely leave ETX)
```

B. Reference List

The Reference List serves as an explanatory alphabetic lookup table for E_TX related keywords. Among these are all menu options, and features of more general interest to the E_TX user. The Reference List is both indexed in the Table of Contents, as well as in the General Index at the back of the Manual. Some of the entries refer to related issues elsewhere in the list. If a keyword is a menu option, then the relevant menu paths are given first. The meaning and notation of menu paths, or paths for short, is explained in the User's Guide (Running E_TX).

1. All

Path: **ETX/ Results/ Save/ All.**

All is the only option under **ETX/ Results/ Save**, indicating that one can only save all the results to a file, not a selection of them. **All** prompts for a file name (default extension **.ETX**). One can edit the line that contains **.ETX** pre-written, including the extension. If no extension is given, extension **.ETX** is added by *E_TX*. Existing files can never be overwritten by *E_TX*.

2. As Is

Paths:
ETX/ Data/ Show/ Toxicity/ As Is.
ETX/ Data/ Show/ Exposure/ As Is.
ETX/ Data/ Save/ Toxicity/ As Is.
ETX/ Data/ Save/ Exposure/ As Is.
ETX/ Statistics/ Logarithms/ Toxicity/ As Is.
ETX/ Statistics/ Logarithms/ Exposure/ As Is.
ETX/ Hazard/ Logistic/ At Exposure/ As Is.

As Is always refers to the original order in which the data have been entered or read from a file. As can be seen from the above paths, **As Is** refers to the last menu screen to fire the complete command. On this last menu screen, it is always the first option of two, the second being **Sorted**. Hence, pressing **[Enter]** activates it as the default. What **As Is** actually does, depends on the previous menu options, to which the reader is referred..

3. At Exposure

Path: **ETX/ Hazard/ Logistic/ At Exposure/**

At Exposure is the second option of **ETX/ Hazard/ Logistic/**. It displays the estimated hazard at the entered or read exposure concentrations, given the logistic density estimate fitting the current toxicity data set. See: **Hazard**. The alternative option of **At Exposure** is **Fixed Hazard** (see there, and see **Hazard**, as well as Appendix A for the full story).
The two options of **At Exposure** are: **As Is** and **Sorted**. These refer to the on-screen representation order of the data, i.e. unsorted or sorted.

4. Basic Statistics

Path: **ETX/ Statistics/ Basic Statistics/**

Basic Statistics is the second option of **Statistics**. Most of the statistics shown on screen are used in the extrapolation and hazard calculations. There is no need to study or record these statistics, unless you want to check the extrapolation and hazard estimates by hand, study the influence of outliers on means and standard deviations, see where the data are located in log space, and so on.

Basic Statistics has only one option: **Toxicity**. No statistics are calculated in E_TX for the exposure data set. The exposure data are just used to estimate the hazard at each individual exposure value. The statistics are displayed with respect to two distributions: the logistic and the normal distribution. Both screens show the toxicity sample statistics: mean and standard deviation, and the respective extrapolation constants.

These extrapolation constants are given in Tables B1 (logistic) and B2 (normal). Table B1 is derived from Aldenberg & Slob (1993). Table B2 (95% confidence) is derived from Wagner & Lokke (1991). The median extrapolation constants for the normal distribution (Table B2: median) have been kindly provided by dr. R.J. van Wijk (AKZO Research CRL, Arnhem). These constants have been determined through simulation by drawing 5000 random samples for each sample size. We have not been able to check the performance of these constants, but they seem to match the median extrapolation constants for the logistic distribution quite well. No such table seems available in the statistical literature. Extrapolation constants for sample sizes not in Table B1 or B2, are estimated in E_TX by linear interpolation.

The logistic basic statistics screen presents some parameter estimates of the logistic distribution parameters, α and β. It is important to note that all are $\log_{10}$ values. Hence, they should be compared with the values as given under **ETX/**

Statistics/ Logarithms/. The moment estimates are those of Kooijman (1987):

$$\hat{\alpha} = \bar{x}$$

and

$$\hat{\beta} = s_n \cdot \frac{\sqrt{3}}{\pi}$$

In fact, for a toxicity data set of size n, s_n is the bias-corrected sample standard deviation on the basis of $(n - 1)$. Hence, one could speak of bias-corrected moment estimates.
The maximum likelihood estimates of α and β are more difficult to estimate. We need them for the goodness-of-fit calculations for the Logistic distribution. Now, α and β have to be solved from the nonlinear equations:

$$\sum_i \frac{1}{1 + \exp((x_i - \alpha)/\beta)} - \frac{n}{2} = 0$$

and

$$\sum_i \frac{(1 - \exp((x_i - \alpha)/\beta)).(x_i - \alpha)/\beta}{1 + \exp((x_i - \alpha)/\beta)} + n = 0$$

The equations are given by D'Agostino & Stephens (1985). We did check out, that they indeed follow from a maximum likelihood argument. These equations must be solved numerically. We have used a discrete Newton-Raphson procedure to do so. A numerical subtlety turned out to be those cases where the x_i are symmetrically located around their average, for example: -1 and +1. Then the first equation with α equal to the x-average is uniformly satisfied for all β, which blows up the iterative procedure. We decided to put α equal to x-average, in these cases, and to iterate only the second equation to solve for β.
The logistic basic statistics screen further displays so-called HC_5 fitting estimates for the logistic parameters. These are described in Appendix A. See also **Hazard**. The idea is that, if an **exposure** value happens to be equal to the median (extrapolation) estimate of the HC_5, on the basis of the toxicity data,

then the estimated hazard at that exposure concentration should be 5%.Appendix A explains that this leads to the estimate:

$$\beta = k_L \cdot s_n / C_5$$

The **Basic Statistics** display screen for the Normal distribution contains analogous parameter estimates for location and scale parameter μ and σ. The moment estimates, corrected for bias, are identical to the sample mean and sample standard deviation. The maximum likelihood estimate of σ is the raw standard deviation of the sample, not corrected for bias, i.e. on the basis of n, the number of toxicity tests. These are not used by E_TX, but only given for completeness. For the normal case, goodness-of-fit calculations are based on the sample mean and sample standard deviations, not on the maximum likelihood estimates, as is the case with the logistic distribution.

5. Batch mode

The previous version of E_TX: ETX 1.2A did have a batch mode. You could type under MS-DOS a command line like: **ETX MYDATA.DAT 3**, in order to process the third line of the data in file **MYDATA.DAT**. Then, the menu system was by-passed, and an output file **MYDATA.3** was automatically generated. Hence, data files could consist of different sets of toxicity data, e.g. one line per toxic substance. A MS-DOS **.BAT** file, could consist of several such ETX command lines.
In the current version (*1.3a*), reading and saving of files has been much improved. Now, there is only one toxic substance per file, and different commented toxicity data are on separate lines. Again, a batch mode option would be feasible, but it has not been incorporated again. We would like to know whether there is a need for it.

6. Data

Path: **ETX/ Data/**

Data is the first option of the E_TX main menu screen. It is analogous to the first entry 'File' on the menu bar of many programs. **Data** gives rise to a submenu. One can enter data interactively through the keyboard (**Enter**), read data from disk files (**Read**), look through data, that were entered or read before (**Show**),

modify data (**Edit**), and save data to disk files (**Save**). The reader is referred to these options for further information.

7. Edit

Path: **ETX/ Data/ Edit/**

Edit is the fourth option of the **ETX/ Data/** menu screen. **Edit** allows the modification of a data set that has been entered through the keyboard, or read from a disk file. Currently **Edit/** has only one option: **Toxicity**. In this version, it is not possible to edit the exposure data from inside E_TX.
It is important to keep in mind, that **Edit** is not capable of editing data sets in the way an ASCII editor, or wordprocessor can do that. **Edit** is meant for correcting typing errors in data just entered, or to add comment strings to poorly annotated data; it can be used for sensitivity analysis, e.g. by changing data a little and calculating difference quotients, etc. For instance, it is not possible, to delete data, or to add extra data to the set, or extra comment lines, through **Edit**.
Data modifications, that do change the size of the data set, need an outside full-screen editor, such as the MS-DOS Editor, Turbo Pascal Editor, WordPerfect, or other word processors. In those cases, edit the disk file that contains the data. Hence, for data sets just entered: save them, leave E_TX, and start the editor of choice. Save the augmented data to a disk file, then open E_TX again and read in the modified file.
For modifications of existing data, **Edit** is fine. Initially, **Edit** works just like **Show** (see: **Show**). One can browse through the data page by page, entry by entry, until the entry to be modified is found, and highlighted, i.e. printed inversely. Then, pressing the **[Space]** bar, opens the possibility to edit the current entry. This Line edit mode is indicated by the extension of the highlight to the full length of the line (full highlight). In this state, the Edit keys (see: Edit keys) are active, and the entry can be modified, or extended. Invalid entries are not accepted. These are non-positive values, values not separated from the comment by white space (spaces and/or tabs), lines without a numeric 'head', and so on. In particular, one cannot enter a comment line starting with an exclamation mark. Pressing the **[Esc]** key, when still in Line edit mode (full highlight), recovers the previous data line.
After completing the modifications, pressing the **[Enter]** key makes the changes permanent. One stays in Browse mode, just as under **Show**, however. So, one can travel through the data again, entrywise, or pagewise, find a new

entry to be modified, press the **[Space]** bar to activate Line edit mode (full highlight), make modifications, press **[Enter]** to confirm, and so on, until you are done.
Now, there are two ways to leave **Edit**: through the **[Esc]** key, or through the **[Enter]** key. The **[Esc]** key restores the old data set, while **[Enter]** confirms all modifications. There are some extra questions asked for confirmation. Hence, there are several occasions in **Edit**, where you can change your mind and restore the previous situation.
One cannot edit exposure data right now. Apart from using an outside editor, there is a work-around in E_TX, if you know what you are doing (not recommended though). Read the **.EXP** file as toxicity data through **ETX/ Data/ Read/ Toxicity** (be sure to save the previous toxicity data!). Then, edit, save again, with extension **.EXP**, E_TX will not protest against that. The first line of the header of the file written by E_TX, however, is erroneous then, since it will read: **! ETX Data File: Toxicity Tests**. However, this information is not seen by E_TX, while reading the file as an Exposure data file. It is the responsibility of the user to correct this erroneous comment in the file afterwards with a real-world editor. Afer reading in the modified exposure set, be sure to erase the internal toxicity data set, that is still the exposure set just saved, by reading in a fresh genuine toxicity data set, or by entering new toxicity data.

8. Enter

Path: **ETX/ Data/ Enter/**

Enter is the first option of the **ETX/ Data/** menu screen. **Enter** allows the interactive entry of data through the keyboard, or of new data, overwriting existing data, previously entered, or previously read from a disk file.
Enter has two options: **Toxicity** and **Exposure**. **Toxicity** allows the entry of toxicity data, **Exposure** allows the entry of exposure data. Although the procedures of entering each category are identical, except that **Toxicity** asks for a Toxic substance name, while **Exposure** asks for an Exposure substance name, both sets are treated very differently inside E_TX. The toxicity data set is 'extrapolated' to HC_5 estimates, e.g. for setting standards, while exposure values are evaluated through their estimated hazard to species, on the basis of the toxicity data.
Enter is rather primitive, working on a line by line basis. It is impossible to go back to previous lines, in order to change them, or to delete previous entries. Typing errors, however, can be corrected through **Edit** (see **Edit**), after

completing data entry of the whole set. At present, one cannot add data to an already completed set, from within E_TX, nor can one delete one or more entries. This might change in future releases.
More flexible, full-screen editing, can be accomplished, however, with the aid of an editor or wordprocessor outside of E_TX. One has to save the data from E_TX, then, in order to apply these editors to the data.
Enter first asks for a toxic substance name, or exposure substance name. A highlighted entry line signifies that the Line edit mode is on (full highlight). Line edit mode is pretty sophisticated, though. One can enter a substance name, and edit the line with the Edit keys, until satisfied. Pressing **[Esc]** blanks the whole entry line. **[Enter]** terminates Line edit mode. If **[Enter]** is pressed immediately, without entering a name, **Unnamed substance** is assumed.
There is no check on the validity of a name entered, from the point of view of E_TX file reading conventions (see Files), so, if you do not follow these conventions, **Enter** may continue without complaint, the name may even be correctly saved to a file through **Save**, but E_TX may not be able to read in the file successfully lateron. The convention is: do not start the name with a numeric 'head', followed by white space (spaces and/or tabs). A numeric head, immediately followed by other non-white characters is fine (e.g. **2,4-dimethyl ...**). The reason is that E_TX will interpret the string with the white space as a concentration value followed by a comment, and assumes that the substance name was not given (**Unnamed substance**).
After successfully entering the substance name, one is prompted for the first numerical entry through **1:** followed by a full highlight, which means that Line edit mode is on again. This Line edit mode is exacty the same as the one for the toxic substance. Now, a numeric non-zero, non-negative, value, e.g. a concentration value, is expected. E_TX will respond with **Entry xx is invalid**, if it doesn't like it, and after pressing the **[Enter]** key, the entry line is blanked, and restored in full highlight. **[Esc]** blanks a non-empty entry line. If the entry is found OK, the next entry is prompted with **2:**. Up to a maximum of 300 data values can be entered for each data set.
One may enter bare numbers, and E_TX will not complain. But, one can make each entry self-documenting by adding additional information about the nature of the entry, such as the unit of measurement, a species name, a reference, anything informative. This information may be entered directly following the numeric value, i.e. in Edit line mode, but separated from it by at least one space or tab. The value and its comment string, as it is called, stay together as one data record. When the data is saved to a disk file, the complete data records are saved on a line by line basis.
To stop the entering process, press the **[Enter]** key while in Edit line mode,

with nothing entered on the highlight. E_TX will ask for confirmation, one may resume entering mode by pressing the **[Esc]** key. Pressing **[Esc]**, while in Edit line mode, with an *empty* highlight, will cause E_TX to ask you, whether you want to leave **Enter**ing mode, and restore the old data. If you confirm this, everything entered so far is deleted, and the previous situation (perhaps without data) is restored. To clear previous data, start **ETX/ Data/ Enter**, press **[Enter]** at the first numerical (empty) prompt, and press **[Enter]** to confirm. After completing **Enter**, E_TX immediately starts calculating all statistics and extrapolation estimates, as well as hazard estimates, if exposure data are available. Hence, there are no menu options or paths, that by themselves trigger calculations to be done. They have been done, as soon as data have been entered, edited, or read.
After having entered a complete data set, either toxicity data, or exposure data, or both, one can do several things. One can inspect the data, and browse through them, with **Show**. **Show** only lets you look through the data, while pointing (highlighting) individual entries. These are not full highlights, so you are not in Line edit mode. One can also **Edit** the data, that is make modifications, add comments and so on. It is impossible, right now, to add extra data to the sets, or to delete data. This has to be done outside of E_TX, with the aid of an ASCII editor, a wordprocessor, or a spreadsheet. A third possibility is to **Save** the data just entered to a disk file, leading to a permanent storage of the data, e.g. for distribution to colleagues, for coordinated data management within a work group, for outside editing, and so on. A fourth possibility is to go straight on to **Results/ Save**, and save the results of the calculations. A **Results** file does contain the data entered or read, so if you have a results file, the data can always be recovered. This is quite easy in fact, because the initial section of a results file is similar to a stand-alone E_TX toxicity or exposure data file. And of course, if you are just experimenting, you can walk the menu tree to study the results, and quit E_TX as soon as you are done. Then the data can not be recovered afterwards.

9. Exposure

Paths:
ETX/ Data/ Enter/ Exposure.
ETX/ Data/ Read/ Exposure.
ETX/ Data/ Show/ Exposure/
ETX/ Data/ Save/ Exposure/
ETX/ Statistics/ Logarithms/ Exposure/

Exposure always refers to an exposure data set, comprising similar concentration

values relating to one particular toxic substance under study. These values may be environmental or experimental concentrations of that substance, either measured, estimated, proposed, legally imposed, or whatever. The purpose of the exposure data is, that, given an extrapolation exercise, that is, given an impression of the distribution of species sensitivities for a specific toxic substance, one would like to assess the percentage of species that is harmed, or is safe, with respect to a set of given exposure concentrations relating to the toxic substance under study. These may be exposures in the past, at present, to be expected in the future. They may come from a survey, a set of scenario predictions, etc. What E_TX does is to try to estimate percentiles of the species sensitivity distribution at the given exposure concentrations. These need not be raw data, they may be calculated from other data, e.g to make them comparable to the nature of the toxicity data. One may also massage the toxicity data first to make them comparable to the environmental data. That is pretty much the responsibility of the user.

E_TX can have one exposure data set of maximally 300 values, including their comments. One may **Enter** them, **Read** them from a disk file, browse through them after entering or reading (**Show**), **Save** them to a disk file after entering or reading, study their log transformed values (**Logarithms**) on which all statistics are based. The reader is referred to these options.

Exposure data cannot be **Edit**ed at present from within E_TX. If the exposure data need to be edited, then save them, and use an outside editor to make modifications. There is a work-around to do it inside E_TX, by retreiving them as toxicity data (see **Edit**), but this is only feasible if you understand the structure of E_TX files.

10. Extrapolation

Path: **ETX/ Extrapolation/**

Extrapolation is the third option of the E_TX main menu screen. Extrapolation, in E_TX, is the estimation of the hazardous concentration at the 5th percentile of the distribution of species sensitivities for a toxic substance, and forms the main focus of the program.

Activating **Extrapolation** leads to a submenu with three options, since there are three statistical distributions involved: logistic (**Logistic**), normal (**Normal**), and triangular (**Triangular**). However, the statistical treatment of the triangular distribution differs in several respects from those of the logistic and normal distributions. See **Triangular** for some extra information about that.

The logistic and normal distribution version of extrapolation are very much akin. In both cases, the screen reports the species protection level (95%), that cannot be changed, the number of toxicity tests involved (minimum number is two), and the Median Estimate of the HC_5, printed bold on the screen, as well as the 95% Underestimate of the HC_5. These are based on extrapolation constants, reported under **Basic Statistics**. Tables of the extrapolation constants are reproduced in Appendix B.
In general, there is very little difference between the extrapolation answers of the respective distributions. Note that the HC_5 estimates are in the *original units* of the data. Hence, the $\log_{10}$ transformation of the data, that had been applied to do the statistics, has been removed in the HC_5 estimates. These estimates can be either directly used for setting environmental standards or objectives, or, depending on other considerations of a scientific, or policy nature, further calculations, e.g. involving extra safety factors, or partition coefficients, etc., may be applied to arrive at the final answers wanted. These considerations are outside the scope of E_TX.

11. Files

E_TX data files can come into existence in essentially two ways. One is by saving data from inside E_TX to a file. The menu path is: **ETX/ Data/ Save/ Toxicity/ As Is**, to save E_TX toxicity data, and: **ETX/ Data/ Save/ Exposure/ As Is**, in order to save E_TX exposure data.
The second way of coming into existence of an E_TX data file is by constructing it yourself with an editor, word processor or spreadsheet. Both possibilities are treated in the next two paragraphs.
In the current version of E_TX, files are saved into the same directory as where **ETX.EXE** is located. One can only read files from this same directory. Hence, do not try to change directory from within E_TX. This might change in future versions.

a. E_TX generated files

Files saved by E_TX can be read in again by E_TX. Saving a data file through E_TX, has the advantage that E_TX automatically includes all kinds of information in the header of the file, to be shown below. This identifies the file internally, e.g. date, time, a save serial number, etc.. Since they are readable ASCII files, this information can be inspected, printed, and edited.
The advantage of self-constructed files is that you can include a lot more

additional information into the data file than is currently possible through E_TX: how the data were selected, scope of the project, extra comments between data lines to identify taxa, or circumstances, and so on.

E_TX data files can be made almost self-documenting. There is no excuse anymore for unannotated toxicity data. Of course, both ways of constructing E_TX files can be combined. For example, one may enter data interactively in raw order, then save them in sorted form, and edit the file through adding additional annotation. If the very flexible rules of E_TX data formatting are maintained, the resulting file can be read again by E_TX. When more data become available, they may be added to the file, with ample space for documentation.

E_TX data files can have any extension. However, toxicity data saved by E_TX have default extension **.TOX**, while exposure data have default extension **.EXP**. Existing data files can never be overwritten by E_TX.

Both toxicity and exposure data files obey the same rules of formatting. E_TX cannot make a distinction between them, so reading an exposure file into the toxicity data structure is possible, but likely to result into non-sensical calculations.

Table B3 displays a toxicity data file as saved by E_TX. In fact, the lines starting with a ! have been added by E_TX automatically. These are comment lines and serve as comments. They tell us that it is an E_TX toxicity data set, that it was saved with E_TX *1.3a*, the name of the file (chosen by the user), date, time, a data save serial number as an extra lable in case of trouble, the name of the toxic substance and additional information about it, the number of toxicity data, and then the seven data values.

Note, that these are annotated data. The annotated data strings were entered from the keyboard in E_TX, including the lay-out to align decimal points and unit names. The comment strings following the data may have any form, and may contain any information. We have copied the information from Table 2 in Van Straalen and Denneman (1989). Unit, species, data reference, special circumstances: everything fitting in one line (80 characters) may be included in the comment string. If you enter the data through E_TX, save through E_TX, and read back into E_TX, it is impossible to separate the data values and their comment strings, although the latter are optional. Moreover, from inside E_TX it is impossible to overwrite a file. It simply refuses to save data to an existing file.

Table B4 displays an exposure data file as saved by E_TX. It has the same basic structure as an toxicity data file. The exposure substance is named differently here. The environmental or exposure concentrations may be of a different nature than the laboratory data.

b. Editor generated files

Editor files are data files meant to be read by E_TX, but not made through E_TX. The advantage, as already mentioned, is flexibility in documenting the data sets. Moreover, data files can be managed by a data manager independently of E_TX. Any ASCII editor can be used, e.g. MS-DOS 5.0 Editor, Turbo Pascal Editor, etc. Also, wordprocessors, spreadsheet, and data base programs. Do not use the internal format of wordprocessors. Use there ASCII export options.
Table B5 shows a fancy E_TX data file composed with WordPerfect. The lines are saved as 'ASCII Text (DOS)'. In this case, we have not tried to mimic the E_TX data file header, although one can take one from an E_TX data file and adapt it. The purpose of this example is to show how flexible E_TX data formatting is. Blank lines (lines with spaces, tabs, and a Carriage Return), and Comment lines with an exclamation mark in front, perhaps preceded by white space (spaces and/or tabs), can be put at any place in the file: at the beginning, at the end, and anywhere in between.
The first item to alert E_TX, while reading a data file, is the toxic substance name, or exposure substance name. If one lacks, E_TX assumes **`Unnamed Substance`**. These names are not preceded be an exclamation mark, and may contain spaces, and any additional information fitting on the line. Substance names may start with numeric information, but *not followed by spaces or tabs.* This is because a number followed by white space and additional characters is interpreted by E_TX as data: a value followed by a comment. So, as a substance name: **`2,4,5-T`** is fine, but **`2 4,5-T`** would be read as the value **`2`** and comment **`4,5-T`**. After a number has been seen, no names without exclamation marks may follow anymore. They are considered erroneous and E_TX will complain about an error in line number xx.
Hence, E_TX data, e.g. annotated concentrations of toxicity tests, consist of a number, optionally followed by a comment string, separated by white space (blanks and/or tabs). One blank suffices. Note that at least one blank, or tab, is obligatory now. So, **`15mg/l`** is *not* OK, in order to prevent typographic errors to pass by unnoticed, e.g. **`1o.7 ug/l`**. E_TX counts the data automatically. It can correctly distill a few numbers hidden within a heavily commented file.
Separate data must be on separate lines: the leading number on a line is taken as the value, while what follows, after white space, is interpreted as comment string. So, the line: **`2.0 1.0 5.0 ug/l (three reproduction values for Daphnia)`** contains *one* E_TX data value (**`2.0`**), and its comment string: **`1.0 5.0 ug/l (three reproduction values for Daphnia)`**.
There is only one substance per file. One cannot use a subset of the numbers in a file. If that is wanted, use an editor to select the numbers and save them in a

new file.
Since both the substance name, as well as the data comment strings are optional, a minimum E_TX data file consists of a bare list of anonymous numbers, separated by Carriage Returns. This is perhaps useful for numerical experimentation.

c. E_TX Results files

With the option **ETX/ Results/ Save/ All.**, one can save the results of an E_TX session to a disk file. The results saved always refer to the latest run. Hence, if one enters some toxicity data, and later reads new ones from a file, all internal results relate to the latest run. The single option **All** indicates that every result, whether it has been on the screen or not is saved. One cannot save part of them, e.g. just the extrapolation results.
In fact, there is no need to study any result on the screen. One can enter or read in the data, and go straight on to **ETX/ Results/ Save/ All.**, to save a printable ASCII file with everything in it, including the data themselves.
Table B6 shows the contents of the E_TX results file **CDVSTRAA.ETX** that corresponds to the E_TX data files **CDVSTRAA.TOX** (Table B3) and **CDVSTRAA.EXP** (Table B4). These files are also on the distribution disk.
The Results file is an ASCII printable file and gives an overview of all possible screen output appended to each other, with some additional comments. One can import this file into an editor or word processor, edit the text, and print it as a whole or in parts.
Results files have default extension **.ETX**. This can be changed, when prompted, but giving no extension is overruled by E_TX by adding **.ETX**. E_TX cannot overwrite existing files.

12. Fixed Hazard

Path: **ETX/ Hazard/ Logistic/ Fixed Hazard.**

Fixed Hazard is the first option of **ETX/ Hazard/ Logistic/**. The second option is **At Exposure** (see there, and see **Hazard**, as well as Appendix A for the full story).
Fixed Hazard refers to the display of estimated hypothetical exposure values at a range of fixed hazard percentages (1%, 2%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 98%, and 99%). This yields an impression of what the range of concentration values is that gives rise to these hazards. The median estimate of

the HC_5 is printed bold on the screen.
One may also print this list (from the **Results .ETX** file) for a given toxic substance of special interest, and use it as a lookup table for future exposure concentrations.

13. Goodness-of-Fit

Path: **ETX/ Statistics/ Goodness-of-Fit/**

When fitting distributions to data, in order to estimate percentiles, and the HC_5, it is important to assess, whether the data indeed seem to derive from the hypothesized distribution. This can be done with a goodness-of-fit test. One may be familiar with the well-known Kolmogorov-Smirnov test for goodness-of-fit (see almost any textbook on statistics). However, this test is designed for situations where the distribution generating the data is known, as well as its parameter values. In our case, the distribution parameters are estimated from the data. Hence, we arrive at a problem of circularity: the distribution is estimated from the data, can we test whether the data derive from the distribution?
D'Agostino & Stephens (1986) show that one can approach this matter with the same Kolmogorov-Smirnov test statistic:

$$D \cdot \sqrt{n}$$

only with modified critical values for this statistic, given a pre-set significance level. D signifies the maximum distance between the empirical distribution function (staircase) of the data, and the estimated distribution. Their tests, especially for the logistic distribution, distinguishes between different sample sizes: 5, 10, 20, 50, and infinity. (D'Agostino & Stephens, 1986, p.158), which seems quite relevant to the usual size of toxicity data sets. The table is reproduced as Table B7 in Appendix B. The parameters of the logistic distribution must be estimated from the data by maximum likelihood. Thus, we had to estimate the parameters this way (see **Basic Statistics**).
For the smallest sample size feasible for an extrapolation exercise, $n = 2$, we found that for any values of the two data points, the test statistic equals 0.458. It seems impossible to derive critical values in that case. We reasoned that the interpolating curve of critical values for intermediate sample sizes should all intersect at 0.458 at sample size two. Hence, we added this case to the table of critical values (Table B7, Appendix B).
Intermediate critical values are derived in E_TX through linear interpolation. E_TX

reports the goodness-of-fit for all four significance levels, and prints whether the hypothesis that the data derive from the logistic should be rejected, or not. The choice of significance level is up to the user.
For the normal distribution, D'Agostino & Stephens (1986, p. 123), present a modified Kolmogorov-Smirnov test statistic:

$$D \cdot (\sqrt{n} - 0.01 + 0.85 / \sqrt{n})$$

Critical values of this statistic for the same four significance levels are reproduced in Table B8 of Appendix B. Now, no further differentiation for low sample size is presented. We are unclear about the validity of the test for sample sizes much below 20. Hence, E_TX does present the goodness-of -fit for the normal distribution for small sample sizes, but the warning is printed: **`Below n = 20, this test may not perform well.`**'

14. Hazard

Path: **`ETX/ Hazard/`**

`Hazard` is the fourth option of the E_TX main menu. Hazard assessment, as defined in E_TX, is discussed in Appendix A. The purpose of hazard assessment is to estimate the hazard to species at given environmental exposure concentrations, that may be independent of the toxicity data. The exposure data set may refer to any predicted or measured data, perhaps adapted for comparison to the laboratory toxicity data. The only option of **`Hazard`**, currently, is **`Logistic`**, since the primary emphasis of the HC_5 has been on the logistic distribution.
The density estimate of the logistic distribution employed is calculated in such a way that, if an exposure value happens to be equal to the median estimate of the HC_5 on the basis of the current toxicity data set, then the estimated hazard is equal to 5%. Clearly, exposure values below the estimated HC_5 lead to hazards smaller than 5%, while those above lead to larger hazard percentages.
Under **`Logistic`**, two options are offered: **`Fixed Hazard`** and **`At Exposure`**. The first option displays hypothetical exposure values at a range of fixed hazard percentages (1%, 2%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 98%, and 99%). This yields an impression of what the range of concentration values is that gives rise to these hazards. One may also print this list (from the **`Results .ETX`** file) for a given toxic substance of special interest, and use it as a lookup table for future exposure concentrations.

The second option takes the exposure data as entered, or read from file, and evaluates the hazard at these values.

15. Leave E_TX

Path: **ETX/ Quit/ Leave ETX!**

After pressing **Quit**, you can either change your mind, through the first option (default): **Main Menu**. But, if you are definitely sure to leave E_TX, you can do so by activating the second option of **Quit**, i.e. **Leave ETX**. The trailing exclamation mark in a permanent warning that all internal data, or results, are lost by doing so! This warning is independent of whether you have saved the data, or the results, or not. If you haven't, no extra warning, or prompt follows, so, take care, when leaving E_TX.

16. Logarithms

Path: **ETX/ Statistics/ Logarithms/**

Logarithms is the first option of **Statistics**. E_TX data are entered or read as raw concentration values, densities, accumulated amounts, or other toxic substance measures. All statistical calculations and percentile estimates are carried out with the $\log_{10}$ transformed data. One can examine the data converted to logarithms through **Logarithms**. Further options allow one to choose between **Toxicity** and **Exposure**.
The data are always log-transformed, this cannot be circumvented once inside E_TX. If log transformation is not wanted, one may construct a raw data set consisting of ani-logs, i.e. powers to the base 10. The logs are always taken with respect to the base 10. Log transformation requires that the raw data are positive (non-zero, non-negative). Non-positive concentrations, while entered or read, lead to **Invalid Entry** errors. If the original data do contain zero values, correct them before feeding them to E_TX, by adding a small value, e.g. half the detection limit. One may add this 'starter' value to all the data if wished. Do not forget to substract these small amounts from the extrapolation estimates afterwards. These estimates are in the original units.
If **Logarithms** is activated, and the appropriate data set chosen, the screen displays log concentrations followed by an arrow (**<--**), followed by the original data, including their comments. One may browse through the log values in the

original order (**As Is**), or in sorted order (**Sorted**), by using the arrow keys **[Up]**, **[Down]**, **[Home]**, and **[End]**. If there are more than 20 values, one can page through the data set with **[PgUp]**, **[PgDn]**, and/or **[Up]** and **[Down]**. The entry highlight moves with the arrow and paging keys. It only indicates where one in located in the set, no editing is possible. Consecutive entries are numbered with a number followed by a colon (**:**). The end of the data set is indicated with: **(no more)**. End of page, with more values to follow, is indicated with: **(more)**.

17. Logistic

Paths: **ETX/ Statistics/ Basic Statistics/ Toxicity/ Logistic.**
ETX/ Statistics/ Goodness-of-Fit/ Toxicity/ Logistic.
ETX/ Extrapolation/ Logistic.
ETX/ Hazard/ Logistic/

Logistic refers to the so-called logistic distribution. This is a statistical distribution of a certain mathematical form. See Aldenberg & Slob (1993) for a review of some logistic mathematics. The logistic distribution looks very much like the normal distribution. It has two parameters α and β, closely related to the mean and standard deviation of the distribution.

Basic Statistics displays some sample statistics of the toxicity data and different estimates of α and β.

Goodness-of-Fit shows measures of departure of the toxicity data from the logistic distribution. If these measures are too high, doubt is thrown on the hypothesis that the logarithms of the toxicity data derive from the logistic distribution.

Extrapolation treats the estimation of the hazardous concentration (HC_5), being the 5th percentile of the logistic distribution, from the current toxicity data. Extrapolation can also be based on the normal and on the triangular distribution.

Hazard handles the assessment of the percentage of potentially harmed species at the current exposure data set. Hazard assessment is only done for the logistic distribution in the current version. See the options just mentioned for more details.

18. Main Menu

Path: **ETX/ Quit/ Main Menu.**

If you want to quit E_TX, but change your mind, e.g. to see if the data were saved, you can return to the E_TX main menu, by activating **Main Menu**. The alternative option is **Leave ETX**, see there.
[F10] brings you up to the main menu from anywhere in the menu tree.

19. Normal

Paths: **ETX/ Statistics/ Basic Statistics/ Toxicity/ Normal.**
ETX/ Statistics/ Goodness-of-Fit/ Toxicity/ Normal.
ETX/ Extrapolation/ Normal.

Normal refers to the well-known normal distribution from ordinary statistics. The normal distribution has two parameters μ and σ, called mean and standard deviation.
Basic Statistics displays some sample statistics of the toxicity data and different estimates of μ and σ.
Goodness-of-Fit shows measures of departure of the toxicity data from the normal distribution. If these measures are too high, doubt is thrown on the hypothesis that the logarithms of the toxicity data derive from the normal distribution.
Extrapolation treats the estimation of the hazardous concentration (HC_5), being the 5th percentile of the normal distribution, from the current toxicity data. Extrapolation can also be based on the logistic and on the triangular distribution. No hazard assessment is incorporated in E_TX as yet for the normal distribution. See the options just mentioned for more information.

20. Printer

There is currently no direct printer support in E_TX (version 1.3 a). However, all results can be saved into an ASCII printable file. This file can be imported into an ASCII Editor, or a wordprocessor for editing and printing. See **Results**, or **All**.

21. Quit

Path: **ETX/ Quit/**

Through **Quit** one can leave E_TX. E_TX won't let you do so immediately. A submenu follows with the first option: **Main Menu**, as default, meaning: return to the E_TX main menu screen. The second option **Leave ETX!** does end E_TX, without any further delay. The exclamation mark is a reminder that no warning is given, if you haven't saved your data or results. Any data entered, or results obtained, but not saved, is lost after activating **Leave ETX**.
A quick way to get out, from anywhere in the menu tree, is **[Ctrl]+[End]**, and **[Ctrl]+[Y]** to confirm. This certainly destroys all data entered, or results obtained. An emergency stop is given by **[Ctrl]+[Break]**.

22. Read

Path: **ETX/ Data/ Read/**

Read is the second option of **ETX/ Data/**. With **Read**, one can read data into E_TX from disk files. E_TX data files are ASCII printable files that are either saved through E_TX, or files made trough an ASCII Editor, a wordprocessor, or a spreadsheet. Use: save (export) as ASCII, or as a print file (often extension **.PRN**), in these programs.
The E_TX file conventions are explained under Files of the Reference List.
Read has two options: **Toxicity** and **Exposure**. A data file read under **Toxicity** fills the E_TX internal toxicity data set. A data file read through **Exposure** is put into the E_TX internal exposure data set. There is no essential difference between a toxicity and an exposure data file. Both have the same structure. While saving, default extensions **.TOX**, and **.EXP** are standard, but not obligatory, E_TX conventions. One may develop other conventions. Additional internal comment lines in data files, saved by E_TX, further indicate, whether we have a toxicity data file or an exposure data file, but this information is not interpreted by E_TX. Hence, there is good reason to stick to some extension convention, preferably that of E_TX, next to the informative and critical annotation of the toxic or exposure substance names and of the individual data entries.
Here we assume that there are valid E_TX files available in the directory where **ETX.EXE** is located. After activating **Read**, and choosing the type of data set, the menu screen clears and E_TX responds with:

Edit Search Profile for Data File to be Opened:
***.TOX**

This means that the MS-DOS search profile conventions, including wild cards, are supported. The highlighted edit line indicates that Line edit mode is on, with ***.TOX** already filled in as a default search profile. It means: list all the files with extension **.TOX**. Pressing **[Enter]**, accepts the default search profile, and indeed lists the files. With the Edit keys, one may change the search profile, or fill in a filename of specific interest. When no file conforms to the user-supplied search profile, E_TX says: **No file found on this Search Profile.**
It is currently *not* possible to specify a file including its directory path, such as **MYDATA*.TOX** (for files in the subdirectory of the one where E_TX is located), or **C:\RIVM\ALD*.TOX** (for complete paths). This feature is high on the list to be changed in a future version.
Once you have a list of files on the screen, you can browse through them with the arrow keys. The highlight indicates the position of the list you are pointing at. After having found the file to load, you can load it by pressing **[Enter]**. E_TX tries to read the file, as explained under Files, and if it runs into an offending line, if any, it will beep and say: **File Reading Terminated: Invalid Input at Line *xx*** and then, print the invalid line. The file is not processed any further. See Files for the expectations E_TX has about the structure of the information in E_TX data files.
Otherwise, if no errors are found, the file is read, and the data are stored internally. The screen says: **File Reading Successful: *n* Data Read**, pauses for a while, because E_TX will now calculate all statistics immediately. This may take some time for a large data set.

23. Results

Path: **ETX/ Results/**

Results is the fifth option of the main E_TX menu screen. In the present version of E_TX, **Results** is only used for saving the results of the calculations to a disk file. Hence, the only option is **Save**. Other options may be added in future versions, e.g. printer support, results selection, and so on. **Save** in turn has the only option **All**, to indicate that it is only possible now to save all results relating to the current data sets (toxicity data and exposure data).
Collecting results is only necessary, if one wants to save the data and the results in one file, for documentation, or later printing. There is no control over what is saved, and what not, nor about the order of the items saved. Everything that

E_TX calculates is saved, whether it *has been on-screen, or not.* Hence, a very quick way to do an extrapolation exercise, and hazard assessment, is to enter data, or load (read) a file, and save all results. The resulting **.ETX** file can be printed out, and studied. No menu tree walking, or on-screen displaying, is really necessary to obtain the results. As soon as the data are entered, or read, all calculations are done.

24. Save

Paths: **ETX/ Data/ Save/**
ETX/ Results/ Save/

ETX/ Data/ Save/ lets us save the data to disk, that previously have been entered, or read from disk files. In the current version, it is only possible to save files to the directory where the program runs.
E_TX may contain two data sets at once, differing in nature. The one set is formed by the (laboratory) toxicity data set; the other is the (environmental) exposure data set. Hence, **ETX/ Data/ Save/** has two further options: **Toxicity** and **Exposure**. The first option saves the toxicity data, the second the exposure data, but not before in each case the sorting order has been selected (**As Is**, or **Sorted**).
For saving either data set, a filename has to be chosen. A default extension, **.TOX**, or **.EXP** is given on an inverse edit line. One can use the Edit keys (see there), to enter/edit the filename, e.g. **CADMIUM.TOX**, **24DMP.TOX**, etc. The extension may be changed to something else, e.g. **.DAT**, or **.TX1**, etc., or it can be removed. But E_TX will add **.TOX**, or **.EXP**, as appropriate, if no extension is given.
E_TX is designed not to overwrite any existing file. If you try, E_TX will refuse, and say so. This is done with an eye to current trends in Good Laboratory Practice. Of course, one can fiddle in the files through MS-DOS. In fact, one must do so to add, or remove, data to an *existing* data set. Hopefully, the lack of the possibility to overwrite files from inside E_TX, as well as the in-file information that is written, should catch most common accidents.
ETX/ Results/ Save/ refers to the writing of all statistical, extrapolation, and hazard assessment results, *including* the data, to a disk file. See **Results** and **All**. Here also, an inverse edit line is offered with a pre-written extension **.ETX**. One can use the edit keys to edit the filename, e.g. **CADMIUM.ETX**, etc. The extension may be changed, or removed. If no extension is given, E_TX will nevertheless add **.ETX** to the filename. Filenames longer than eight characters are trimmed. An existing file of the same name cannot be overwritten.

25. Show

Path: **ETX/ Data/ Show/**

Show is the third option of **Data**. After having entered data (**Enter**), or having read data from file (**Read**), one may want to inspect the data, in order to see what they look like, whether there are any typing or reading errors, whether the comment strings annotating the data have to be extended or modified, and so on (**Show**).
Show allows one to browse through both the toxicity and the exposure data sets. Hence, **Show** has two options: **Toxicity** and **Exposure**. One may further choose to browse the data in the original order (**As Is**), or in sorted order (**Sorted**). **Sorted** only refers to the on-screen representation. The data are kept in the original order internally.
One may browse through the data by using the arrow keys **[Up]**, **[Down]**, **[Home]**, and **[End]**. If there are more than 20 values, one can page through the data set with **[PgUp]**, **[PgDn]**, and/or **[Up]** and **[Down]**. The entry highlight moves with the arrow and paging keys. It only indicates where one in located in the set, no editing is possible. Consecutive entries are numbered with a number followed by a colon (**:**). The end of the data set is indicated with: **(no more)**. End of page, with more values to follow, is indicated with: **(more)**.

26. Sorted

Paths:
ETX/ Data/ Show/ Toxicity/ Sorted.
ETX/ Data/ Show/ Exposure/ Sorted.
ETX/ Data/ Save/ Toxicity/ Sorted.
ETX/ Data/ Save/ Exposure/ Sorted.
ETX/ Statistics/ Logarithms/ Toxicity/ Sorted.
ETX/ Statistics/ Logarithms/ Exposure/ Sorted.
ETX/ Hazard/ Logistic/ At Exposure/ Sorted.

Sorted always refers to the order of the data after sorting from smallest to largest. As can be seen from the above paths, **Sorted** refers to the last menu screen to fire the complete command. On this last menu, it is always the second option of two, the first being **As Is**, meaning unsorted. Hence, **As Is** is the default, while **Sorted** needs an extra key press (**[Down]**) before it can be activated with the **[Enter]** key. What **Sorted** actually does, depends on the previous menu options, to which the reader is referred.
An additional subtlety of **Sorted** is that, for the on-screen actions (all except

Save), no physical sorting is done: it only refers to the on-screen representation. The data themselves are kept in the original order. In case of **ETX/ Data/ Save/**, however, the respective data are really saved in sorted order to disk. If this option is used, without an analogous action through **As Is**, the original order of the data can never be recovered from the sorted files.

27. Statistics

Path: **ETX/ Statistics/**

Statistics is the second option of the E_TX main menu screen. It allows the user to look at some statistical summaries of the data. **Statistics** gives rise to a submenu. One can look through the logarithms of the data, which essentially enter the statistical analysis (**Logarithms**), inspect some basic statistical summaries of the data, like means, standard deviations, and distribution parameter estimates (**Basic Statistics**), and one can examine the goodness-of-fit of some statistical distributions fitting the data (**Goodness-of-Fit**).

28. Toxicity

Paths:
ETX/ Data/ Enter/ Toxicity.
ETX/ Data/ Read/ Toxicity.
ETX/ Data/ Show/ Toxicity/
ETX/ Data/ Edit/ Toxicity.
ETX/ Data/ Save/ Toxicity/
ETX/ Statistics/ Logarithms/ Toxicity/
ETX/ Statistics/ Basic Statistics/ Toxicity/
ETX/ Statistics/ Goodness-of-Fit/ Toxicity/

Toxicity always refers to a toxicity data set, comprising similar concentration values relating to one particular toxic substance under study. These may be laboratory toxicity data for this toxic substance for different species or other taxa, either raw or averaged. A toxicity data set may also be a set of mesocosm, or field experiment, toxicity data for the particular toxic substance. The prime requirement is that it makes sense to assume that the toxicity data derive from a log-logistic, log-normal, or log-triangular distribution.
E_TX will try to estimate the 5th percentile for these distributions, based on the toxicity data set. This is called extrapolation. Furthermore, E_TX will try to estimate the distribution itself, in order to be able to estimate the hazard to

species, that is the percentage potentially harmed, at given environmental exposure concentrations. Hence, next to a toxicity data set, E_TX can have an exposure data set. Both sets refer to the same toxic substance, but their names (toxic substance name and exposure substance name) may differ, as well as their exact nature (total concentrations versus dissolved, or other fractions; conversion to standard conditions, e.g. sediment or soil characteristics).
Toxicity data may derive from a diversity of sources, as long as they refer to the same toxic substance. Some may be raw data, others calculated from other data, in order to obtain a consistent set, that is thought to derive from one statistical distribution. One may convert the toxicity data to values comparable to the exposure data, or adapt the exposure data to the laboratory conditions relating to the toxicity data. That is pretty much the responsibility of the user.
E_TX can have one toxicity data set of maximally 300 values, including their comments. One may enter them (**Enter**), read them from a disk file (**Read**), browse through them after entering or reading (**Show**), edit them to make modifications (**Edit**), or to extend the comment string (annotation); one may **Save** them to a disk file after entering or reading, study their log transformed values (**Logarithms**) on which all statistics are based.
One can study basic statistics of the toxicity data, as well as the distribution parameter estimates of distribution fitting them (**Basic Statistics**), and one can examine the goodness-of-fit of these distributions based on the toxicity data (**Goodness-of-Fit**) . The reader is referred to these options.

29. Triangular

Path: **ETX/ Extrapolation/ Triangular.**

Triangular is the third option of **Extrapolation**, and refers to the triangular distribution. The method implemented here is that of Erickson & Stephan (1988), called the FAV (Final Acute Value). See also OECD (1992). In the spirit of E_TX, the nature of the data is the responsibility of the user. Hence, the FAV may be applied to Chronic data, as well, and may be targeted on the taxon level preferred.
The statistics of the triangular distribution differs from the logistic and normal distributions. While the logistic and normal statistics are based on all the data points, the FAV is applied to the lower four points only. Since the four points are taken to derive from the lower tail of the distribution, they must be below the median of the data set. Hence, one needs a minimum of eight points to apply this method. The HC_5 is estimated from linear regression on these four points.

The advantage is that species insensitive to a toxic substance (the species at rank five or higher) have little influence on the estimate. On the other hand, the influence of the more sensitive species is increased.
One could device similar tail-oriented methods for the logistic and the normal distribution, and one could also develop all-points treatments for the triangular. We have not done so yet, since there is no need for a combinatorial proliferation of methods. In OECD (1992), it was found that the tail-oriented FAV on the basis of the triangular distribution yields results that correlate very well with the all-points median estimates of the HC_5 for the logistic and the normal distributions.
Perhaps, FAV falls into a caregory of its own. In a future release we might reorganize the **Extrapolation** section.

30. Version

The current program version is *E_TX* 1.3a. It is displayed on all *E_TX* menu screens as the first part of the menu title bar, e.g. **ETX 1.3a/ Data/ Save/**.

Appendix A Hazard Assessment

In this Appendix, our implementation of Hazard Assessment at given exposure concentrations is discussed, as well as some problems with a related formulation in the literature.

By Hazard Assessment in E_TX, we mean estimations of the hazard to species at one or more given concentrations, for example experimental or environmental concentrations. Such concentrations we call exposure concentrations.

For extrapolation, we work with the (laboratory) toxicity data at hand, and estimate the hazardous concentration for two levels of confidence. To estimate the hazard at concentrations that may or may not be equal to the hazardous concentration, or HC_5, we need an estimate of the statistical distribution that fits the toxicity data best.

Even if we confine ourselves to the logistic probability density function, there are several ways of fitting this distribution of species sensitivities to the toxicity data set. The first particular estimate, that comes to mind, is the density estimate based on the maximum likelihood estimate of the parameters of the logistic distribution. Another is based on the moment estimates of the parameters, and a third may be the one based also on the moment estimates, but now with the variance corrected for bias, i.e. $n/(n - 1)$ * variance.

All these estimates lead to different estimates of the logistic parameter β, and therefore different density estimates. These in turn result into different estimates of the hazard at given concentrations of exposure.

We decided that the density estimate should be the most accurate at concentrations around the HC_5, i.e. the hazardous concentration for 5% of the species.

In order to find the best β, we reasoned as follows. The 'best' estimate of the

HC_5 is taken here as the median estimate, that in the long run will overestimate the HC_5 as often as it underestimates it. Now the median estimate of the logHC5 is calculated as:

$$\log HC_5 = \bar{x} - k_L \cdot s_n$$

with k_L depending on sample size as tabulated in Appendix B (Table 1).
If we take

$$\alpha = \bar{x}$$

then we can calibrate β of the unknown density in such a way that the HC_5 of this calibrated density corresponds exactly to the median estimate of the true HC_5 on the basis of the sample. Thus we must solve beta from the identity

$$\alpha - \beta \cdot C_5 = \bar{x} - k_L \cdot s_n$$

where:

$$C_5 = \ln(\frac{95}{5}) = \ln(19) = 2.9444$$

See Aldenberg & Slob (1993) for a review of some logistic formulae.
This leads to the estimate

$$\beta = k_L \cdot s_n / C_5$$

Hence, β of the estimated density is proportional to the sample standard deviation.
For example let's take the seven Cd data (Van Straalen & Denneman 1989) again: 0.97, 3.33, 3.63, 13.5, 13.8, 18.7, 154 [ug/g]. We have (after

transformation with $\log_{10}$):

$$\bar{x} = 0.9712$$

and

$$s_7 = 0.7028$$

Hence, with

$$k_7\ (50\%) = 1.78$$

it follows that

$$\alpha = 0.9712$$

and

$$\beta = 1.78 * 0.7028 / 2.9444 = 0.4249$$

Now that we have estimated α and β, we can estimate percentages of hazard for species, p, at given $\log_{10}$-transformed concentrations:

$$x = \log_{10}(\textit{Exposure concentration})$$

This can be easily done from the explicit cumulative logistic distribution function:

$$p = \frac{100}{1 + \exp(-(x-\alpha)/\beta)}$$

So, the percentage of species *un*protected at a proposed reference value (Van

Straalen and Denneman (1989), Table 3) of 0.8 [ug/g], i.e.:

$$x = \log_{10}(0.8) = -0.0969$$

yields:

$$p = \frac{100}{1 + \exp(-(-0.0969-0.9712)/0.4249)} = 7.5\%$$

Similarly, at 3.5 [ug/g] (between 2nd and 3rd NOEC), we would calculate a species hazard of:

$$p = 26.8\%$$

At 15 [ug/g] (between 5th and 6th NOEC), the hazard is

$$p = 61.8\%$$

These hazard estimates seem in line with the scatter of the seven toxicity data. Van Straalen and Denneman, however, calculate 15% hazard for the case where we have 7.5% (Table 3: 85% protection). The explanation is that the Van Straalen and Denneman formula (9) is based on the 95% confidence estimate of the HC_5 instead of the median estimate used here. Writing their formula (9) as:

$$p = \frac{100}{1 + \exp((\alpha - x)/\beta)}$$

with:

$$\alpha = \bar{x}, \quad x = \log_{10}(c), \quad \beta = 3 \cdot d_n \cdot s_n / \pi^2$$

we observe that the formulae are identical, except that their beta is proportional to the sample standard deviation with a different constant of proportionality. The essential difference, apart from the mathematical form, is that the constant dm refers to the 95% 'confidence' column in Kooijman (1987, Table 1, delta2=0.05).

To illustrate, for the $n = 7$ case above, Van Straalen and Denneman have:

$$\beta = 3*2.82*0.7028/3.1416^2 = 0.6024$$

while we employ:

$$\beta = 1.78*0.7028/2.9444 = 0.4249$$

Indeed, by using the larger beta, we also arrive at p = 14.5%, i.e. 85.5% protection. Hence the difference between the two hazard estimates boils down to Van Straalen and Denneman having a broader density estimate on the basis of the same data. One would be inclined to think that hazard percentages sensu Van Straalen and Denneman at given concentrations tend to overestimate the true hazard percentages, and therefore, due to the 95% confidence factor d_n, could act as right confidence limits of these hazards. But, this has turned out not to be the case. If the log transformed exposure concentration happens to be equal to the sample average, both hazard percentages become 50%, while for exposure concentrations above the sample average, the Van Straalen and Denneman hazard percentages are *smaller* than those calculated on the basis of the median estimate. We think that an estimate of the species hazard at given exposure concentrations should not be based on an estimated standard deviation of the logistic distribution that results from matching the 5th percentile of that distribution with a deliberate underestimate of the true HC_5.

Appendix B Tables

Table B1. Extrapolation constants (Logistic distribution) for the 95% confidence underestimate and median estimate of the log HC_5.

n	95% confidence	median
2	27.70	2.49
3	8.14	2.05
4	5.49	1.92
5	4.47	1.85
6	3.93	1.81
7	3.59	1.78
8	3.37	1.76
9	3.19	1.75
10	3.06	1.73
11	2.96	1.72
12	2.87	1.72
13	2.80	1.71
14	2.74	1.70
15	2.68	1.70
20	2.49	1.68
30	2.28	1.66
50	2.10	1.65
100	1.95	1.64
200	1.85	1.63
500	1.76	1.63
inf	1.62	1.62

Table B2. Extrapolation constants (Normal distribution) for the 95% confidence underestimate and median estimate of the log HC_5.

n	95% confidence	median
2	26.206	2.35
3	7.656	1.94
4	5.144	1.82
5	4.210	1.78
6	3.711	1.77
7	3.401	1.76
8	3.188	1.74
9	3.032	1.72
10	2.911	1.70
11	2.815	1.69
12	2.736	1.68
13	2.670	1.68
14	2.614	1.68
15	2.566	1.68
20	2.396	1.67
30	2.220	1.67
50	2.065	1.67
100	1.927	1.65
200	1.840	1.65
500	1.763	1.645
inf	1.645	1.645

Table B3. Toxicity data set as saved by E_TX in file **CDVSTRAA.TOX**.

```
! ETX Data File: Toxicity Tests
! Saved with: ETX 1.3a
! Data File Name: cdvstraa.tox
! Date: Feb. 26, 1993
! Time: 15:46:42
! Data Save Serial Number: 1

! Toxic Substance:
Cadmium NOECs (Van Straalen & Denneman, 1989)

! Number of Data = 7

154    ug/g Dendrobaena rubida. Bengtsson et al. (1986)
 13.5  ug/g Lumbricus rubellus. Ma (1982)
 13.8  ug/g Eisenia foetida.    Malecki et al. (1982)
  3.63 ug/g Helix aspersa.      Russell et al. (1981)
  3.33 ug/g Porcellio scaber.   Van Capelleveen (1987)
  0.97 ug/g Platynothrus peltif.Van Straalen et al. (1989)
 18.7  ug/g Orchesella cincta.  Van Straalen et al. (1989)
```

Table B4. Exposure data set as saved by E_TX in file **CDVSTRAA.EXP**.

```
! ETX Data File: Exposure Values
! Saved with: ETX 1.3a
! Data File Name: cdvstraa.exp
! Date: Feb. 26, 1993
! Time: 15:47:27
! Data Save Serial Number: 2

! Exposure Substance:
Cadmium

! Number of Data = 1

  0.8 ug/g Reference Value proposed
```

Table B5. WordPerfect generated E_TX data file (**ASCII Text (DOS)**).

```
! This is an ETX data file

  ! describing the data set
  ! date, time, editor used, analyst, project
       !  any comments at any place

! The toxic substance:
  2,4-dimethylphenol

  0.1  ug/l Daphnia (geom. average of 3 reproduction tests)
  1.5  ug/l Pseudomonas (Canton, 1972b)
200    ug/l Lymnea (regression estimate)

! comments, empty lines  in between...
! (No Arthropods)

  0.67 ug/l Xenopus (Slooff, 1989)

! ...and at the end, plus some empty lines to follow
```

Table B6. Results as saved by E_TX in file **CDVSTRAA.ETX**.

```
! ETX Results File
! Saved with: ETX 1.3a
! Results File Name: CDVSTRAA.ETX
! Date: May  17, 1993
! Time: 18:45:14
! Results Save Serial Number: 1

! Toxic Substance:
Cadmium NOECs (Van Straalen & Denneman, 1989)

! Number of Data = 7
! Data Read from File CDVSTRAA.TOX

154    ug/g Dendrobaena rubida. Bengtsson et al. (1986)
 13.5  ug/g Lumbricus rubellus. Ma (1982)
 13.8  ug/g Eisenia foetida.    Malecki et al. (1982)
  3.63 ug/g Helix aspersa.      Russell et al. (1981)
  3.33 ug/g Porcellio scaber.   Van Capelleveen (1987)
  0.97 ug/g Platynothrus peltif.Van Straalen et al. (1989)
 18.7  ug/g Orchesella cincta.  Van Straalen et al. (1989)

---------Exposure Data---------------------

! Exposure Substance:
Cadmium

! Number of Data = 1
! Data Read from File CDVSTRAA.EXP

  0.8 ug/g Reference Value proposed

---------Log10 of Toxicity Data-----------------

As Entered/Read:

    Log10(Data)     Data
1:  2.1875 <-- 154    ug/g Dendrobaena rubida. Bengtsson et al. (1986)
2:  1.1303 <--  13.5  ug/g Lumbricus rubellus. Ma (1982)
3:  1.1399 <--  13.8  ug/g Eisenia foetida.    Malecki et al. (1982)
4:  0.5599 <--   3.63 ug/g Helix aspersa.      Russell et al. (1981)
5:  0.5224 <--   3.33 ug/g Porcellio scaber.   Van Capelleveen (1987)
6: -0.0132 <--   0.97 ug/g Platynothrus peltif.Van Straalen et al. (1989)
7:  1.2718 <--  18.7  ug/g Orchesella cincta.  Van Straalen et al. (1989)

Sorted:

    Log10(Data)     Data
1: -0.0132 <--   0.97 ug/g Platynothrus peltif.Van Straalen et al. (1989)
2:  0.5224 <--   3.33 ug/g Porcellio scaber.   Van Capelleveen (1987)
3:  0.5599 <--   3.63 ug/g Helix aspersa.      Russell et al. (1981)
4:  1.1303 <--  13.5  ug/g Lumbricus rubellus. Ma (1982)
5:  1.1399 <--  13.8  ug/g Eisenia foetida.    Malecki et al. (1982)
6:  1.2718 <--  18.7  ug/g Orchesella cincta.  Van Straalen et al. (1989)
```

```
7:  2.1875 <-- 154    ug/g Dendrobaena rubida. Bengtsson et al. (1986)

---------Log10 of Exposure Data-----------------

As Entered/Read:

    Log10(Data)     Data
  1:   -0.0969 <--   0.8 ug/g Reference Value proposed

Sorted:

    Log10(Data)     Data
  1:   -0.0969 <--   0.8 ug/g Reference Value proposed

---------LOGISTIC---------------------------------

Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

Sample statistics (log10s):
   XAverage =   0.9712
XStDev(n-1) =   0.7028
          n =   7
Extrapolation Constant (50%) =  1.7800
Extrapolation Constant (95%) =  3.5900

LOGISTIC Distribution Parameter Estimates (log10s):

(Moment Estimates, Bias Corrected)
   AlphaHat =   0.9712
    BetaHat =   0.3875

(Maximum Likelihood Estimates)
   Alphahat =   0.9445
    Betahat =   0.3727

(HC5 Fitting Estimates)
   AlphaHat =   0.9712
 MedBetaHat =   0.4248


Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

LOGISTIC Distribution, Goodness-of-Fit.

Number of Tests (n) = 7
Goodness-of-Fit, Kolm.-Smirn.: D*sqrt(n) =  0.512

Kolm.Smirn.   Crit.val.   Signif.   Logistic?
     0.512       0.657     10  %    Accepted
     0.512       0.699      5  %    Accepted
     0.512       0.743      2.5%    Accepted
     0.512       0.780      1  %    Accepted

---------Extrapolation, Logistic------------------

Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

LOGISTIC Distribution
Species Protection Level = 95%
```

```
Number of Tests = 7

Median Estimate Hazardous Concentration  =  5.2520E-0001
Underestimate (95% Confid.) Hazard.Conc. =  2.8076E-0002

---------Hazard Assessment, Logistic--------------

Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

Exposure Subst.: Cadmium

Exposure Values at Fixed Hazard Percentages:

Hazard      Exposure Value
   1%  <--  1.0448E-0001
   2%  <--  2.0789E-0001
   5%  <--  5.2520E-0001
  10%  <--  1.0909E+0000
  25%  <--  3.1953E+0000
  50%  <--  9.3593E+0000
  75%  <--  2.7414E+0001
  90%  <--  8.0299E+0001
  95%  <--  1.6679E+0002
  98%  <--  4.2135E+0002
  99%  <--  8.3836E+0002

---------Hazard at Exposure Data------------------

As Entered/Read:

       Hazard       Exposure Data
  1:     7.49%  <--   0.8 ug/g Reference Value proposed

Sorted:

       Hazard       Exposure Data
  1:     7.49%  <--   0.8 ug/g Reference Value proposed


---------NORMAL-------------------------------------

Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

Sample statistics (log10s):
   XAverage =   0.9712
XStDev(n-1) =   0.7028
          n =   7
Extrapolation Constant (50%) =  1.7600
Extrapolation Constant (95%) =  3.4010

NORMAL Distribution Parameter Estimates (log10s):

(Moment Estimates, Bias Corrected)
      MuHat =   0.9712
   SigmaHat =   0.7028

(Maximum Likelihood Estimates)
    MuHatML =   0.9712
 SigmaHatML =   0.6506
```

```
Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

NORMAL Distribution, Goodness-of-Fit.

Number of tests (n) = 7
Goodness-of-Fit, Kolm.-Smirn.: D*(sqrt(n)-0.01+0.85/sqrt(n)) =  0.566

Beware: Below n=20, this Test may NOT Perform Well.

Kolm.Smirn.   Crit.val.   Signif.    Normal?
     0.566       0.819     10  %    Accepted
     0.566       0.895      5  %    Accepted
     0.566       0.995      2.5%    Accepted
     0.566       1.035      1  %    Accepted

---------Extrapolation, Normal----------------------

Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

NORMAL Distribution
Species Protection Level = 95%
Number of Tests = 7

Median Estimate Hazardous Concentration  =  5.4248E-0001
Underestimate (95% Confid.) Hazard.Conc. =  3.8120E-0002

---------No Hazard Assessment, Normal---------------

---------TRIANGULAR----------------------------------

---------Extrapolation, Triangular.

Toxic Substance: Cadmium NOECs (Van Straalen & Denneman, 1989)

To calculate the Final (Acute) Value, the minimum number of tests
needed is 8. For toxic substance 'Cadmium NOECs (Van Straalen & Denneman,
1989)'
you have entered only 7 tests

---------End of ETX Results (All) file---------------
```

Table B7. Goodness-of-Fit for the Logistic distribution. Critical values of

$$D \cdot \sqrt{n}$$

at four levels of significance (adapted from D'Agostino & Stephens, 1986, p.158).

n	10%	5%	2.5%	1%
2	0.458	0.458	0.458	0.458
5	0.643	0.679	0.723	0.751
10	0.679	0.730	0.774	0.823
20	0.698	0.755	0.800	0.854
50	0.708	0.770	0.817	0.873
inf	0.715	0.780	0.827	0.886

Table B8. Goodness-of-Fit of the Normal distribution. Critical values of

$$D \cdot (\sqrt{n} - 0.01 + 0.85 / \sqrt{n})$$

at four levels of significance (D'Agostino & Stephens, 1986, p.123).

	10%	5%	2.5%	1%
n	0.819	0.895	0.995	1.035

References

Aldenberg, T., Slob, W. (1993) Confidence Limits for Hazardous Concentrations Based on Logistically Distributed NOEC Toxicity Data. Ecotox.Environm.Safety, 25, 48-63.

D'Agostino, R.B., Stephens, M.A. (1986) Goodness-of-fit techniques. Dekker.

Erickson, R.J., Stephan, C.E. (1985) Calculation of the Final Acute Value for Water Quality Criteria for Aquatic Organisms. National Technical Information Service, Springfield, Virginia, USA. NTIS No.: EPA 600/3/88/018. PB88/214994.

Kooijman, S.A.L.M. (1987) A safety factor for LC_{50} values allowing for differences in sensitivity among species. Wat. Res., 21, 269-276.

OECD (1992) Report of the OECD Workshop on the Extrapolation of Laboratory Aquatic Toxicity Data to the Real Environment. OECD Environment Monographs, No. 59. OECD, Paris.

Slooff, W. (1992) RIVM Guidance Document. Ecotoxicological Effect Assessment: Deriving Maximum Tolerable Concentrations (MTC) from single-species toxicity data. RIVM Report, 719102018. RIVM, Bilthoven.

Van Straalen, N.M., Denneman C.A.J. (1989) Ecotoxicological Evaluation of Soil Quality Criteria. Ecotox.Environm.Safety, 18, 241-251.

Wagner, C., Lokke, H. (1991) Estimation of Ecotoxicological Protection Levels from NOEC Toxicity Data. Water Res., 25, 1237-1242.

Index