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)

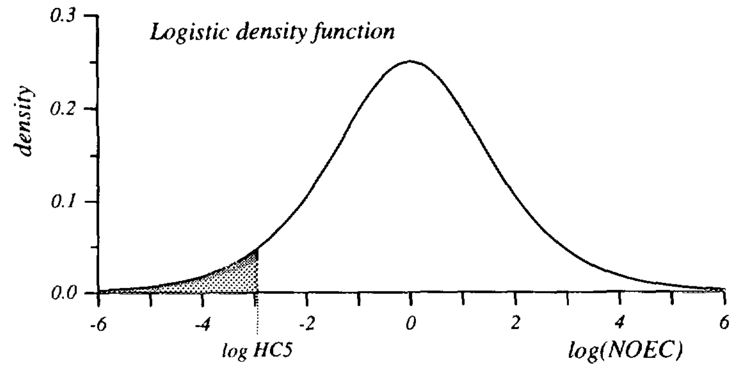

FIG. 1. The standard logistic distribution of log(NOEC) values. Log HC_5 = −2.94 is the log Hazardous Concentration to be estimated. The fraction of the species harmed is shown shaded.

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

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.

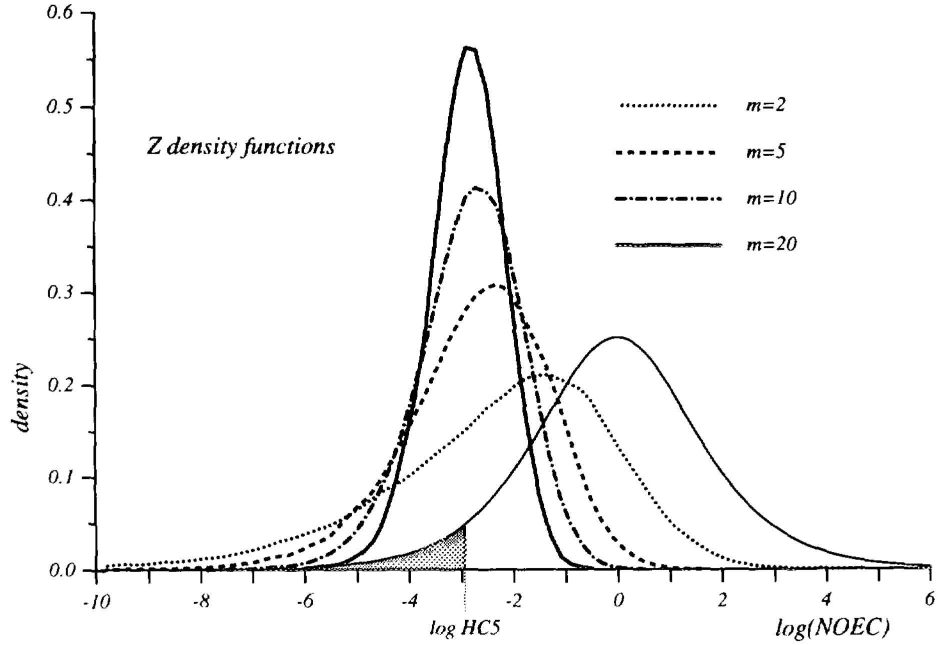

FIG. 2. Simulated variation of the answers that result from using Z to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $Z = \bar{x}_m - 1.62 \cdot s_m$. Z is the formula for estimating the hazardous concentration, as if we had complete information 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,

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

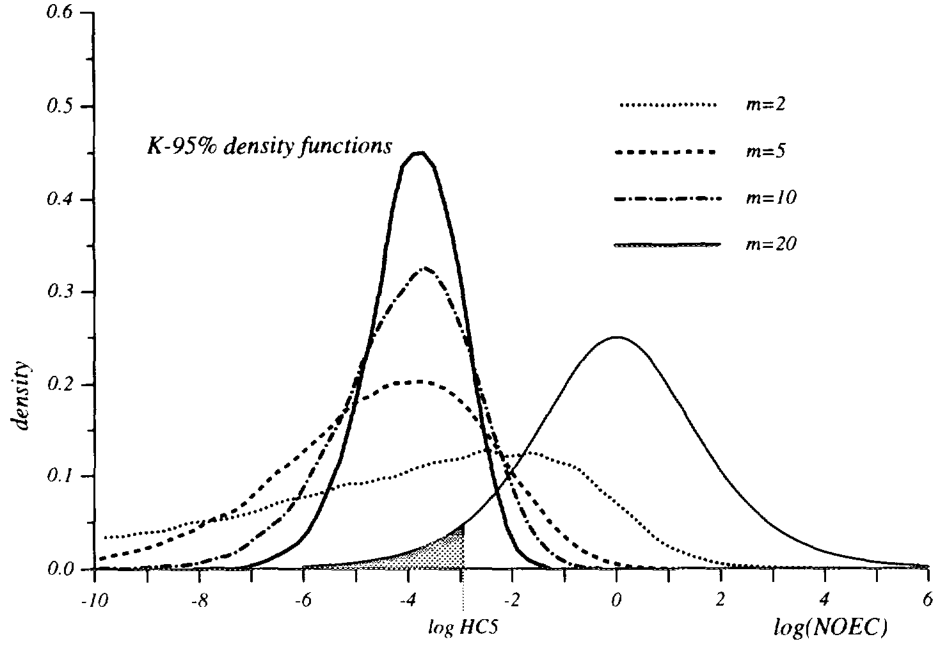

FIG. 3. Simulated variation of the answers that result from using $K - 95\%$ from Kooijman and Van Straalen (95% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $K = \bar{x}_m - k_K \cdot s_m$, with k_K from Table 1 (95%).

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*

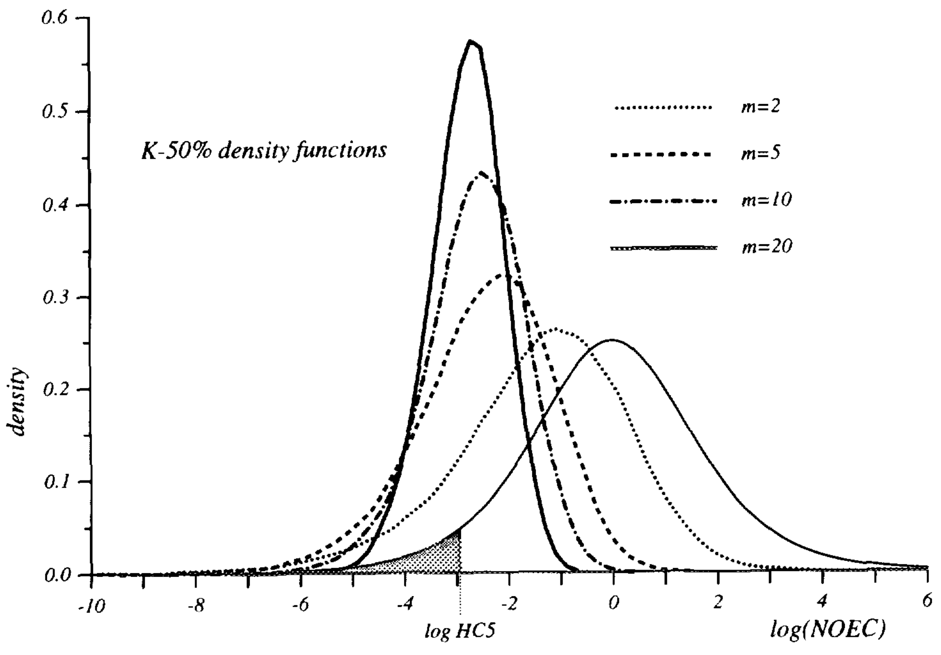

FIG. 4. Simulated variation of the answers that result from using K – 50% from Kooijman and Van Straalen (50% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $K = \bar{x}_m - k_K \cdot s_m$, with k_K from Table 1 (50%).

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 (μg/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,

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

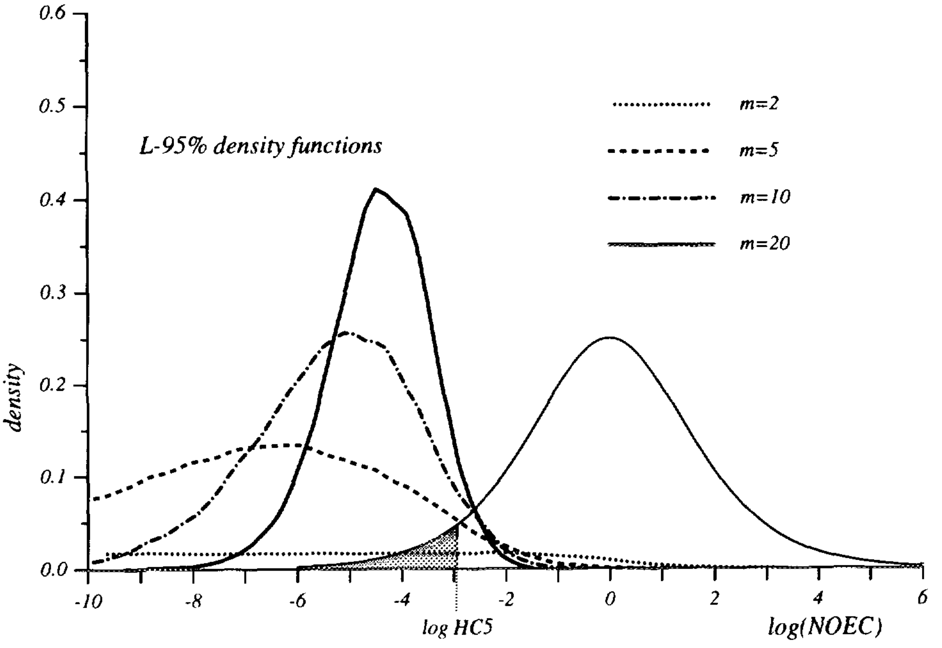

FIG. 5. Simulated variation of the answers that result from using $L - 95\%$ (this paper, 95% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $L = \bar{x}_m - k_L \cdot s_m$, with k_L from Table 3 (95%).

(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.

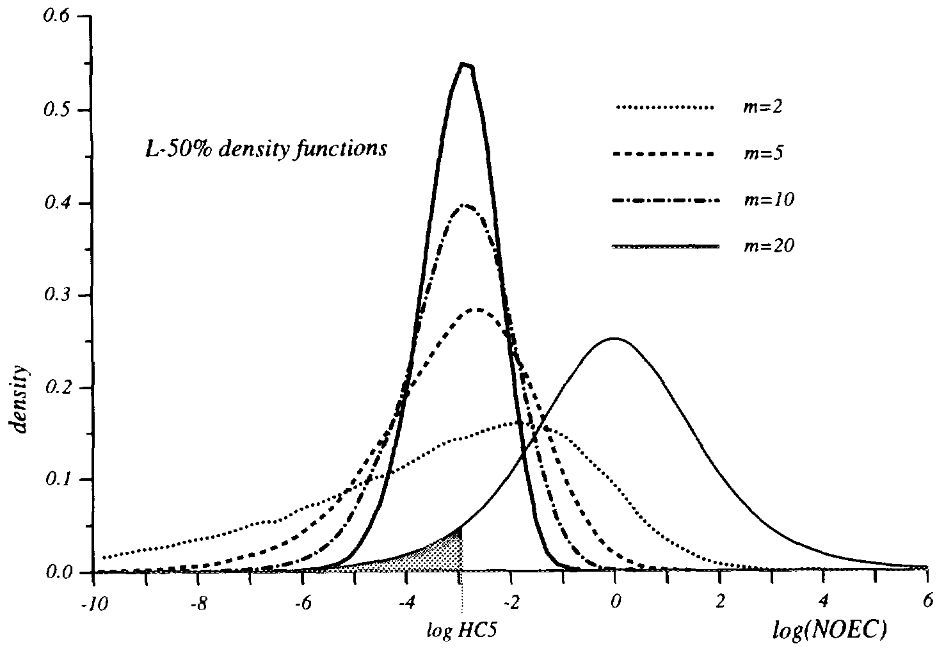

FIG. 6. Simulated variation of the answers that result from using $L - 50\%$ (this paper, 50% confidence) to estimate log HC_5 for four different sample sizes, in relation to the standard logistic density function. $L = \bar{x}_m - k_L \cdot s_m$, with k_L from Table 3 (50%).

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.

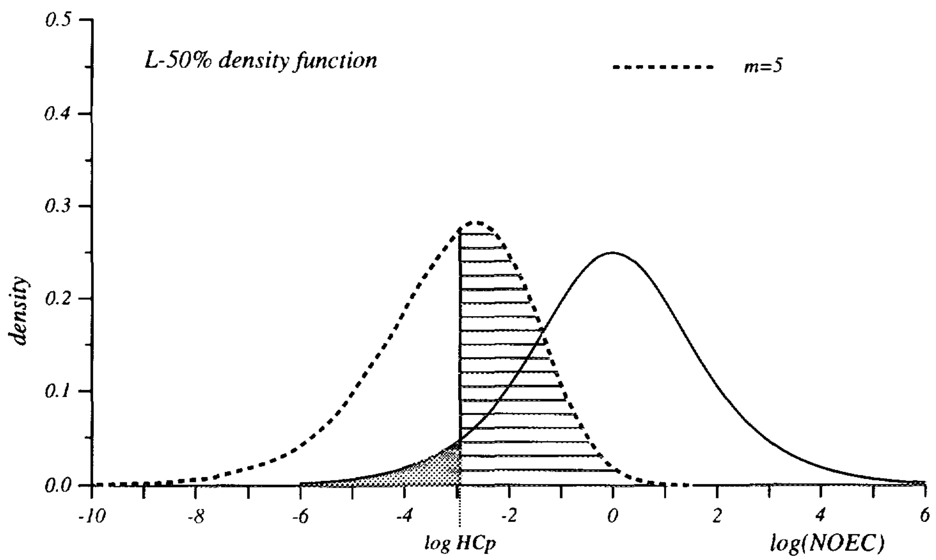

FIG. 7. Primary risk, i.e., the fraction of species not protected (shaded region), and secondary risk of overprediction and thereby harming a larger fraction of species (hatched region) in case of a 50% confidence estimate for log HC_5 for a sample of five test species.

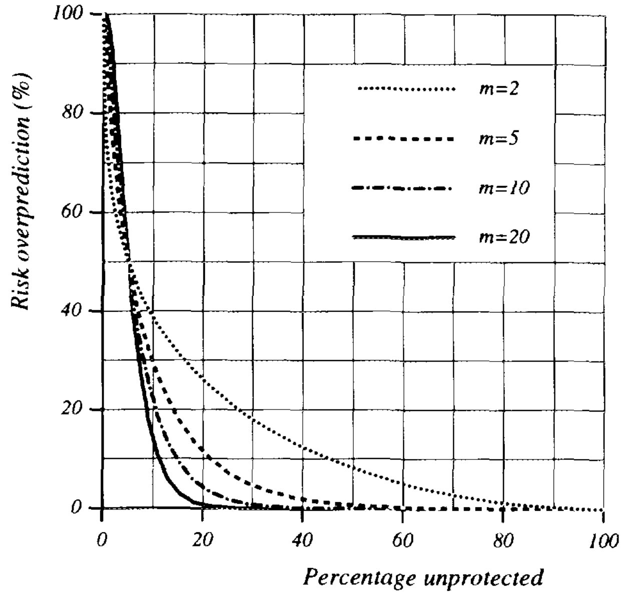

FIG. 8. Risk (ordinate) that more than a given percentage of the species (abscissa) is unprotected when using the 50% confidence estimate of the log Hazardous Concentration for 5% of the species. Due to the nature of this estimate, all curves necessarily intersect at the 5% harm, 50% overprediction point. 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,

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

where

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

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.