SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 23, No. 2, pp. 489–499, 2004

Printed in the USA
0730-7268/04 $12.00 + .00

Hazard/Risk Assessment

EFFECTS OF DATA MANIPULATION AND STATISTICAL METHODS ON SPECIES SENSITIVITY DISTRIBUTIONS

CÉDRIC DUBOUDIN,† PHILIPPE CIFFROY,*† and HÉLÈNE MAGAUD‡
†EDF, Division Recherche et Développement, Département Laboratoire National d'Hydraulique et Environnement, 6 quai Watier, 78401 Chatou, France
‡INERIS, DRC Ecot., Parc Technologique ALATA, BP 2, 60550 Verneuil en Halatte, France

(*Received* 18 *March* 2003; *Accepted* 30 *June* 2003)

Abstract—Species sensitivity distribution (SSD) methodology currently is used in environmental risk assessment to determine the predicted no-effect concentration (PNEC) of a substance in cases where a sufficient number of chronic ecotoxicological tests have been carried out on the substance, covering, for the aquatic environment with which we are concerned, three taxonomic groups: algae, invertebrates, and vertebrates. In particular, SSD methodology enables calculation of a hazardous concentration that is assumed to protect 95% of species (HC5). This approach is based on the hypothesis that the species for which results of ecotoxicological tests are known are representative, in terms of sensitivity, of the totality of the species in the environment, which raises a number of questions, both theoretical and practical. In this study, we compared various methods of constructing a species sensitivity–weighted distribution (SSWD). Each method is characterized by a different way of taking into account intraspecies variation and proportions of taxonomic groups (vertebrates, invertebrates, and algae), as well as by the statistical method of calculation of the HC5 and its confidence interval. Those methods are tested on 15 substances by using chronic no-observed-effect concentration data available in the literature. The choice of data (intraspecies variation and proportions between taxonomic groups) was found to have more effect on the value of the HC5 than the statistical method used to construct the distribution. The weight of each taxonomic group is the most important parameter for the result of the SSWD and letting literature references decide which proportions of data are used to construct it is not satisfactory.

Keywords—Risk assessment Species sensitivity distribution Weighted bootstrapping Variability

INTRODUCTION

The so-called species sentivity distribution (SSD) methodology has been used since 1989 in The Netherlands [1] for the environmental risk assessment of substances. The SSD methodology has been used since 1996 in the European Community [2] to determine the predicted no-effect concentration (PNEC) of a substance, where sufficient chronic ecotoxological tests carried out on the substance are available (a minimum of approximately 15). For the aquatic environment with which we are concerned, three taxonomic groups, algae, invertebrates, and vertebrates, need to be represented. This approach, developed by Kooijman [3], and taken up by Van Straalen and Denneman [4], Wagner and Løkke [5], Aldenberg and Slob [6], and Posthuma et al. [7], is based on the hypothesis that the species for which results of ecotoxicological tests are known are representative, in terms of sensitivity, of the totality of the species in the environment. A likely distribution of species sensitivity is then estimated from these results, which enables calculation of a concentration that is assumed to protect a given percentage of the species in the environment. The agreed European concentration is the hazardous concentration affecting 5% of species with 50% confidence (HC5_50%); equally, 95% of the species are thus protected with a confidence limit of 50%. This is the value we will use; we will simply denote it by HC5.

The SSD approach raises a number of questions (see Forbes and Forbes [8] and Forbes and Calow [9]). From a practical point of view, irrespective of validation and criteria used for selecting the results of ecotoxicological tests used in SSD (criteria that we will not discuss here), we will concentrate on the following three questions. First, how do we deal with the various test results that may exist for the same species? Indeed, should intraspecies variation be taken into account? And if so, how? Second, how do we take into account the fact that the amount of available data varies from one category of species to another (between vertebrates, invertebrates, and algae)? Third, which statistical method should be used to construct the distribution and calculate an HC5?

Various approaches are suggested here to answer to each of these three questions. They have been combined in a near-comprehensive analytical design, yielding, from the same set of data (results of ecotoxological tests), 63 distinct ways of calculating an HC5. This analytical design was tested on 15 substances, by using chronic aquatic no-observed-effect concentration (NOEC) data available for these toxicants. A comparison of these various approaches was then undertaken, to determine the respective effect of each of these questions on the calculation of the HC5: Are variations in the HC5 linked more to different ways of dealing with intraspecies variation, to differences in the amount of data between the taxonomic groups, or to the statistical method used?

METHODS AND DATA

Various ways of calculating the HC5

In answer to each of the questions raised above (regarding effects of intraspecies variation, proportions of data between the three taxonomic groups, and statistical methods), three or four approaches are proposed. They are explained in detail and justified below. Before calculating an HC5, each of these three

* To whom correspondence may be addressed (philippe.ciffroy@edf.fr).

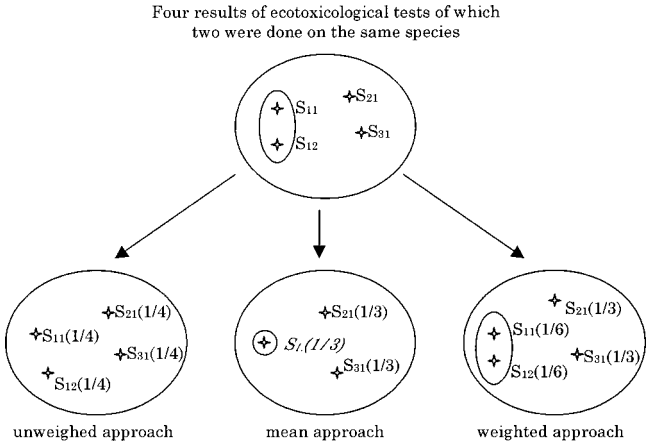

Fig. 1. Possible options for whether or not to take intraspecies variation into account. Three options are available: unweighted (the entire set of data available is treated as if all data belonged to different species), mean (a geometric mean of data for each species is first calculated; the weighting coefficients are then applied to the species mean values), and weighted (each piece of data is weighted to give each species the same weight). The S_{ij} is the result j of an ecotoxicological test carried out on species i; the associated weight it has within the species sensitivity distribution is given in parentheses. The $S_{i.}$ is the geometric mean of available data for species i.

questions must be answered, and consequently a triplet of methods must be chosen from among those suggested. Starting from the same set of data, there will be as many possible ways of calculating HC5 as there are triplets of possible methods.

Taking into account intraspecies variation. In the method commonly used by the National Institute of Public Health and the Environment in The Netherlands (RIVM) [1] and the European Commission [2], the various test results that may exist for the same species are averaged (geometric means, that is, exponential of the arithmetic mean on the log of the data); as a result, a single value per species (the geometric mean) is used to determine the SSD. The intraspecies variation between test results is frequently observed to be very high, and is sometimes equivalent to the interspecies variation. This intraspecies variation is caused by differences in experimental conditions, the criterion chosen, the strains of species, the stages of development, and other factors. These differences also are present when considering two tests carried out on separate species, and they play an implicit role in the difference observed between the results of these two tests. So on the one hand, a portion of the variation in one case is ignored (data concerning the same species), whereas on the other, part of the variation is taken into account (data concerning different species). Although it is not legitimate to ignore a priori intraspecies variation by calculating a mean of the data beforehand, neither is it satisfactory to consider the set of data as it stands without taking into account the label species, because one species is frequently much more abundant than the others with respect to the amount of data (especially for *Daphnia*). Because we wish to use all the available data without first calculating a mean, we can consider weighting the data in terms of the total number of each species present.

To study the effect of whether or not intraspecies variation is taken into account, three approaches have been adopted and compared (Fig. 1). These approaches stem directly from the above discussion, and correspond to the three answers proposed to the question about intraspecies variation (this question will be referred to as species). First, the entire set of data available is treated as if all data belonged to different species (this approach will be referred to as unweighted). Intraspecies variation is taken into account in the same way as is interspecies variation, but there is a risk of giving too much importance to one species if the amount of data available for one species is greater than the amount of data available for the others. Second, as is done in the usual method, a geometric mean of data for each species is first calculated; in this way a new set of data called species mean values is obtained, which has a single corresponding data point for each species (this approach will be referred to as mean). In this case, intraspecies variation is ignored. Third, all the data are used, but each data point is weighted to give each species the same weight within the SSD (this is the weighted approach). Intraspecies variation is taken into account, and no species is given more importance than any other.

Taking into account the three taxonomic groups: Vertebrates, invertebrates, and algae. For the determination of an aquatic PNEC, the SSD method, like the older method based on assessment factors [2], is based on breaking down species into the three taxonomic groups, vertebrates, invertebrates, and algae, which are assumed to correspond to the three trophic levels, predators, herbivores, and primary producers. In fact, this relationship is not strictly accurate, because some aquatic invertebrates are carnivores and certain fish are herbivores; however, it is accepted in practice for reasons of simplicity. From an ecological point of view, the ideal approach would be to determine a PNEC for each trophic level, that is to say in practice for each of the groups (algae, invertebrates, and vertebrates), and to retain the lowest of the values. This is the only approach that enables the three groups to be protected with the same agreed minimum threshold (e.g., 95% of species protected in each group), but the amount of available data seldom allows this approach. Thus, in most cases, it is necessary to incorporate these three taxonomic groups into the same SSD, which then raises the question of the proportions of data taken into account for each group [9].

Most research uses the data available in the literature and implicitly assumes that its distribution among the three taxonomic groups previously described is representative of the proportions existing in the environment. However, in their meta-analysis, Forbes and Calow [9] found, for the ecotoxicology data presented by Versteeg et al. [10], mean proportions of data for primary producers, invertebrates, and fish of 28, 35, and 39%, respectively. However, according to Forbes and Calow [9], the proportions of species corresponding to these three categories in the environment are 64, 26, and 10%, respectively. For cadmium and alkyl derivatives of benzenesulfonic acid (linear alkyl sulfonate), the difference between the HC5 values calculated from data found in the literature and those calculated from estimated proportions in the environment is a factor of three. As a general rule, the representativeness of laboratory species as compared to species in the environment is not guaranteed: laboratory species are chosen because they are easy to breed. They are not the result of a random sample from among all the species. Toll et al. (Parametrixs, Kirkland, WA, USA) have studied the taxonomic diversity of toxicity data for several metals as compared to that assumed for several ecosystems. Insects appear to be underrepresented in tests, whereas Cladocera (*Daphnia*) and salmonids are overrepresented. Toll et al. also showed that this has an effect on the HC5 value. In any case, letting literature references decide which proportions of data are used to construct an SSD does

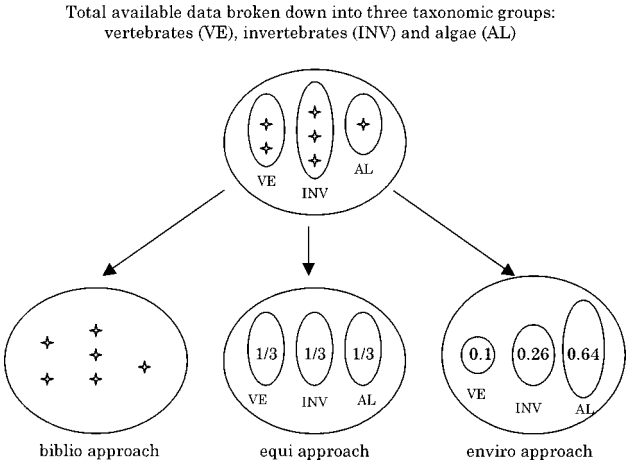

Fig. 2. Possible options for taking into account the three taxonomic groups. Three options are available: biblio (the proportions of data given in the literature are retained), equi (the data are balanced in such a way that the three taxonomic groups are equally weighted), and enviro (the data are weighted in such a way as to respect the proportions proposed by Forbes and Calow [9]). For the equi and enviro approaches, the weighting for each group in the species sensitivity–weighted distribution is given.

not seem to be very satisfactory, because these proportions vary from one toxicant to another.

Three approaches were retained in answer to the question about taxonomic groups (referred to as taxo) to study how much effect this question has (Fig. 2). First, as is done in the usual method, the proportions of data given in the literature are retained (this approach will be referred to as biblio). Ultimately, no distinction is made between the three taxonomic groups. Second, the data are balanced in such a way that the three taxonomic groups are equally weighted in the SSD (this is the equi approach); however, this does not mean that they will be equally protected. Third, the data are weighted in such a way as to respect the proportions proposed by Forbes and Calow [9]; these proportions are 10, 26, and 64% for vertebrates, invertebrates, and algae, respectively (this is the enviro approach).

Weighting coefficients. The three approaches suggested in answer to the question on intraspecies variation are combined with those suggested in answer to the question on taxonomic groups. In this way, we have nine possible ways of taking into account the available data when constructing the SSD. These nine possibilities correspond to nine different ways of weighting the data (Table 1). The weighting coefficients are applied to the whole data set for the unweighted and weighted options (intraspecies variation), and to the species mean values for the mean option (the number of data points in this latter case corresponds to the number of different species represented). The usual approach (recommended by the RIVM [1]) corresponds to the combination: species = mean, taxo = biblio. We note the following observations. The sum of the weighting coefficients corresponding to the totality of the data used is obviously equal to 1. If the number of data points available is the same for each species, the unweighted and weighted approaches are identical; if only one data point is available per species, the unweighted, mean, and weighted approaches are identical. If the number of data points in the three taxonomic groups is identical, then the equi and biblio approaches are identical. Significant differences between the approaches are to be expected when there are differences in size between

Table 1. Weighting coefficients associated with each data point, according to the approaches used for the species and taxonomic questions. Three options are available for the species parameter: unweighted, mean, and weighted (see text). Three options are available for the taxonomic parameter: biblio, equi, and enviro (see text).[a]

	Species		
Taxonomic	Unweighted	Mean	Weighted
Biblio	$\frac{1}{n}$	$\frac{1}{ns}$	$\frac{1}{n_{ij}\cdot ns}$
Equi	$\frac{1}{nc\cdot n_{i.}}$	$\frac{1}{nc\cdot ns_i}$	$\frac{1}{nc\cdot n_{ij}\cdot ns_i}$
Enviro	$\frac{p_{fi}}{n_{i.}}$	$\frac{p_{fi}}{ns_i}$	$\frac{p_{fi}}{n_{ij}\cdot ns_i}$

[a] n = total number of data; ns = total number of different species; n_{ij} = number of data associated with species j in taxonomic group i; nc = number of taxonomic groups used; $n_{i.}$ = number of data in group i; ns_i = number of different species in taxonomic group i; p_{fi} = proportion estimated by Forbes and Calow [9] for taxonomic group i (0.1 for vertebrates, 0.26 for invertebrates, and 0.64 for algae).

the datasets for the three taxonomic groups and for the species, or when the taxonomic groups occupy positions with differing sensitivities. It is the particular aim of this study to assess these differences quantitatively.

After defining the various ways in which the available data are taken into account, we will detail the ways in which the SSD proper is constructed; in fact SSD should really be called the species sensitivity–weighted distribution (SSWD).

Statistical methods for the calculation of an HC5. Ever since the SSD concept was constructed, numerous statistical methods for estimating distribution and calculating an HC5 have been suggested and compared [6,11–19]. These methods differ in the choice of the underlying statistical distribution (empirical distribution, log-normal or log-logistic distribution, and others) and in the method used to estimate the confidence interval (bootstrap, Bayesian techniques, or asymptotic theory). The results of the various methods, when applied to the same set of data, can differ by as much as a factor of three [19].

We were forced to use or suggest methods capable of being used with all the approaches suggested as answers to the questions on intraspecies variation and on taxonomic groups mentioned before and on providing a confidence interval for the HC5. We shall propose four main methods; they all enable construction of an SSWD.

In the first method, weighted data regression (WDR), a theoretical statistical distribution is fitted to weighted data, and the 5th percentile of the distribution thus obtained is used. A normal distribution on data in log base 10 is chosen. Given that the data are weighted, the formulas that give the mean (μ) and the standard deviation (σ) are

$$\mu = \sum_{i,j,k} s_{ijk} \cdot p_{ijk}$$

$$\sigma = \sqrt{\frac{n}{n-1} \sum_{i,j,k} (s_{ijk} - \mu)^2 \cdot p_{ijk}}$$

where n is the number of data points considered, s_{ijk} is the sensitivity data point k assigned to species j in taxonomic group i, and p_{ijk} is the weighting coefficient associated with data point s_{ijk}. We can make the following comments. A ratio $n/(n-1)$ is introduced into the weighted standard deviation. In this way

we obtain an unbiased estimator of the population standard deviation. If the weighting coefficient p_{ijk} equals $1/n$ (biblio-unweighted), the preceding formulas exactly correspond to the classic formulas for the mean and for standard deviation. Unfortunately, when taking weightings into account, the goodness of fit of the data to the distribution defined by μ and σ cannot always be ascertained. Other distributions are possible, in particular the log-logistic distribution, which can easily be fitted to weighted data by using a least squares regression method.

To estimate the confidence interval associated with the HC5, the approach we use is the bootstrap method [20] (resampling with replacement, *n* out of *n*: the samples selected are the same size as the initial sample). One thousand samples are generated by using the bootstrap procedure, and for each one an HC5 is calculated with the method previously described. We use the values of the HC5_5% and the HC5_95%, that is, the lower (5%) and upper limit (95%) of the 90% confidence interval of the HC5. These are obtained by using the bias-corrected and accelerated (BCa) method that is recommended by many authors, especially to estimate the confidence interval of a quantile and in the case of a nonparametric bootstrap [20].

Use of this bootstrap method has a disadvantage. If the three taxonomic groups are not present in a random sample, it can easily be seen that calculation of the HC5 is impossible. The solution to this is to reject draws that do not contain at least one data point from the three categories of species. This bootstrap method has a relatively high cost in terms of calculation time (concerning the confidence interval), because, at each draw, the existing data must be analyzed and the weightings redefined.

The second method is the direct weighted bootstrap (DWB). The bootstrap method is no longer used just to estimate a confidence interval, but also to construct samples in which the proportions of data among species and among taxonomic groups correspond to those desired. Therefore, we carry out a nonequiprobable resampling with replacement from raw data (weighted and unweighted) or from species mean values (mean option). The probability of drawing each data point corresponds to the weighting coefficient previously defined. The number of samples is 1,000 and the number of data points drawn for each one corresponds to that of the initial data set (bootstrap *n* out of *n*). To estimate the HC5 of each sample, we have two options. In the parametric approach, we make the assumption (which can be shown to be true) that the distribution of each sample follows a theoretical distribution (normal distribution in this case for log base 10 values); the parameters of the distribution used are calculated, and then the corresponding fifth percentile is calculated (DWB, normal approach). In the nonparametric approach, the HC5 of each sample is defined as the fifth percentile of the empirical distribution of each one (all values being in log base 10). A linear interpolation is used if the fifth percentile does not correspond to a value of the sample (DWB, empirical).

An HC5 distribution is obtained in all cases, from which we directly derive an HC5_50% (median value of HC5; this is the best estimate value) as well as the values corresponding respectively to the HC5_5% and to the HC5_95% (5th and 95th empirical percentiles of the distribution obtained; in this case, the BCa function has no meaning). With this approach, the desired proportions corresponding to the weighting coefficients clearly are not perfectly respected in each sample generated by the bootstrap; one only tends towards these proportions for the 1,000 samples.

The third method is the grouping weighted bootstrap (GWB). The bootstrap technique is still used to generate weighted samples, but the three taxonomic groups are separated. Draws (with replacement) are carried out for each category, vertebrates, invertebrates, and algae, respectively, in such a way that the desired proportions of each category of species are respected. At each draw the data from the three subsets are put together to determine an HC5. Two approaches are again possible: the parametric bootstrap, choosing a theoretical underlying distribution (GWB, normal), or the nonparametric bootstrap (GWB, empirical), based on the empirical distribution of each sample (see DWB approach). Within the three taxonomic groups, the various options concerning intraspecies variation are taken into account by means of weighting (in the same way as for the DWB method). From the distribution of HC5 obtained by bootstrap, we use the median (HC5_50%) as well as the 5th and 95th percentiles, as in the DWB method.

The sample size in the three taxonomic groups (vertebrates, invertebrates, and algae) can be defined in many different ways. If proportions of data different from existing ones are imposed, a bootstrap *m* out of *n* with *m* greater or lower than *n* is used. Because the bootstrap *m* out of *n* with $m > n$ does not lead to a correct estimate of the confidence interval, the following solution is used. It is desirable to draw the maximum number of data while never drawing more data than initially present in each of the groups of species. By doing this, a bootstrap *n* out of *n* is carried out in at least one group, and a bootstrap *m* out of *n* with *m* lower than *n* in the others. Mathematically, this approach is expressed as follows. If the n_i ($i = 1, 2, 3$) are the numbers present in the three taxonomic groups and the p_is are the proportions desired for each one, then $x = \min(n_i/p_i,\ i = 1, 2, 3)$ is the size of each sample drawn by using the bootstrap procedure (in fact the nearest whole round number to this value is used); $x_i = p_i \times x$ ($i = 1, 2, 3$) corresponds to the numbers drawn in the three taxonomic groups (as above).

Nonetheless, depending on the proportions desired and the number of existing data, at least one of the x_i may be zero, which comes to the same thing as not drawing data in one of the categories of species. This is not acceptable, so in this situation the rule previously used is adapted and at least one data point in each category of species is drawn, while respecting the desired proportions. In this case a bootstrap *m* out of *n* with *m* greater than *n* is carried out in at least one of the three taxonomic groups.

The fourth method is the direct weighted bootstrap *m* out of *n* with $m < n$ (DWB*m*). The DWB method is used, by drawing *x* data for each sample (with *x* having been previously defined in the GWB method), while respecting desired weightings at the species and taxo level. In this approach, as in the preceding one, the number of data points drawn is adapted to the weighting coefficients used. Both approaches, parametric bootstrap and nonparametric bootstrap, can be used to determine the HC5 (DWB*m*, normal and DWB*m*, empirical).

Data

The different approaches suggested were compared for 15 toxicants. These are substances for which we have at least three chronic NOEC-type data points in the three taxonomic groups (vertebrates, invertebrates, and algae). These data were

Table 2. Number of data for each toxicant and each taxonomic group (nb data), proportion represented by each category (prop), and number of different species represented (nb sp)

	Algae			Invertebrates			Vertebrates			Total	
Toxicant	nb data	prop	nb sp	nb data	prop	nb sp	nb data	prop	nb sp	nb data	nb sp
Boric acid	4	0.13	3	9	0.30	2	17	0.57	7	30	12
Cadmium	10	0.15	7	25	0.38	10	31	0.47	19	66	36
Copper	19	0.19	12	44	0.43	17	39	0.38	16	102	45
Nickel	6	0.29	6	6	0.29	5	9	0.43	3	21	14
Lead	12	0.28	9	8	0.19	7	23	0.53	13	43	29
Zinc	5	0.13	4	22	0.56	11	12	0.31	7	39	22
Mercury	6	0.17	6	16	0.46	9	13	0.37	4	35	19
Chromium	16	0.31	12	13	0.25	7	22	0.43	13	51	32
Dibutyl phthalate	4	0.31	1	6	0.46	3	3	0.23	2	13	6
Butylbenzyl phthalate	7	0.54	3	3	0.23	1	3	0.23	2	13	6
Dodmac	9	0.53	2	5	0.29	2	3	0.18	2	17	6
Lindane	3	0.14	2	13	0.59	9	6	0.27	5	22	16
Atrazine	17	0.45	10	10	0.26	8	11	0.29	8	38	26
Parathion	8	0.30	5	11	0.41	5	8	0.30	4	27	14
Pentachlorophenol	3	0.07	2	17	0.40	12	23	0.53	6	43	20

taken from the AQUIRE (U.S. Environmental Protection Agency: http:/www.epa.gov/ecotox/), EAT (ECETOC aquatic toxicity: http://www.ecetoc.org/), and RIVM [1] databases. The data were initially provided in micrograms per liter and converted into log base 10. With regard to the AQUIRE database, only data that were judged to be sufficiently well researched and monitored to the satisfaction of the U.S. Environmental Protection Agency have been used.

These 15 substances included seven metals (chromium, copper, zinc, cadmium, lead, nickel, and mercury), four pesticides or similar chemicals (parathion, lindane, atrazine, and pentachlorophenol), three other organic compounds (butylbenzenyl phthalate, dibutyl phthalate, and dimethyldioctadecylammonium chloride [dodmac]), and one inorganic compound (boric acid).

The number of data points for each substance is variable and lies between 13 and 102 (Table 2), with an average of 37. The number of data points in each group (vertebrates, invertebrates, and algae) is just as variable, greater than 3 and lower than 44, with an average of 8.6 for algae, 14 for invertebrates, and 15 for vertebrates. The proportion of algae is very variable and lies between 7 and 54% depending on toxicants, with an average of 27%. The proportions of invertebrates and vertebrates lie between roughly 20% and 60%, with an average of 37% for both. The proportions of data taken into account are thus close to those observed by Forbes and Calow [9] for the data of Versteeg et al. [10].

The total number of species lies between 6 and 45, depending on toxicants, with an average of 20. The number of species per taxonomic group is almost always more than 2 and less than 20, with an average of 5.5 for algae and 7 for invertebrates and vertebrates. The ratio between the number of data points and the number of species is identical for the three taxonomic groups, at approximately 5.

For each toxicant, the normality of the distribution of available data was tested (after log transformation of the data) by using chi-square, Kolmogorov–Smirnov, and Shapiro tests. Only 3 of the 15 toxicants (boric acid, copper, and lindane) were found to allow an empirical distribution of data that cannot be considered normal at the 5% threshold. For the first two of these, nonnormality is connected to the presence of outliers at low values, whereas the remaining distribution can be considered normal.

Methodologies for the analysis of results

For a given data set, calculating an HC5 comes down to choosing an approach among those suggested by way of an answer for each of the three questions raised (intraspecies variation, taxonomic groups, and statistical method). The triplet used constitutes the method for calculating HC5, and each HC5 is thus characterized by three parameters. One is the species parameter, which allows use of the three methods described previously: unweighted, mean, and weighted. The second is the taxo parameter, which allows use of the three methods biblio, equi, and enviro. The third is the statistical method parameter, which can be broken down into four main methods: WDR, DWB, GWB, and DWB*m*. The three approaches DWB, GWB, and DWB*m* each allow two possibilities for calculating HC5: parametric bootstrap (normal distribution) or nonparametric bootstrap (empirical distribution); whereas the WDR method is only associated with normal distribution. In this way, the statistical method parameter finally comprises seven methods in all; but it can equally well be considered as the combination of two subparameters. The two subparameters are the method of calculation (method = WDR, DWB, GWB, or DWB*m*) and the distribution selected (distribution = empirical or normal), with the combination of WDR and empirical not being possible (so the statistical method parameter will be called meth&dist).

Consideration of the entire set of possibilities (combining each method two at a time for each of the three parameters) leads, for the same data set, to the calculation of 63 distinct HC5s ($3 \times 3 \times 7$). Thus, 63 values of HC5, with their confidence intervals, were calculated for each of the 15 toxicants considered, from the chronic NOEC values previously presented (i.e., a total of 945 values of HC5 in all). Breaking down the HC5s according to toxicants constitutes what we shall call the fourth parameter of the analysis (toxic parameter).

Seeking to determine the effect of each of these parameters, as well as the effect of each method, on the values of HC5, we carried out several analyses of variance (ANOVAs). The principle of ANOVA is to express the result of a measurement or a calculation (in this case the various HC5s calculated) as a linear combination (model) of the parameters that were varied, and then to estimate which are the terms whose effect is statistically significant and the relative order of these terms.

Table 3. Respective effect of parameters and interactions of the analysis of variance on the hazardous concentration affecting 5% of species with 50% confidence (HC5). Coefficients of determination (r^2) of the multianalysis of variance on the HC5; ranking is given in order of importance of parameters with interactions taken into account[a]

Parameter	*df*	Sum of square	Mean of square	*F* value	Pr(*F*)	r^2
Toxic	14	911.30	65.09	10,742	0.0000	0.015 (1)
Species	2	5.61	2.80	463	0.0000	0.995 (4)
Taxo	2	1.94	0.97	160	0.0000	0.974 (2)
Meth&dist	6	6.51	1.08	179	0.0000	0.982 (3)
Toxic:taxo	28	18.34	0.66	108	0.0000	
Toxic:species	28	3.48	0.13	21	0.0000	
Toxic:meth&dist	84	5.86	0.07	12	0.0000	
Species:meth&dist	12	0.20	0.02	3	0.0000	
Taxo:meth&dist	12	0.25	0.02	3	0.0009	
Species:taxo	4	0.04	0.01	2	0.1779	
Residuals	752	4.56	0.01			

[a] *F* = Fisher's statistic; Pr(*F*) = probability that the effect of each parameter is zero. Toxic = the different toxicants; species = the question of intraspecies variation; taxo = three taxonomic groups; meth&dist = statistical method and distribution.

The model tested takes into account first-order interactions between the parameters. The model enables the main effects of the parameters to be distinguished from those of their interactions two at a time. To rank the parameters in order of their effect, while taking into account their interactions, we developed a multi-ANOVA approach. This is based on the coefficient of determination measuring the fit between the initial data and the values predicted by the linear model associated with the ANOVA. The principle is as follows. As many ANOVAs are carried out as there are parameters, with one parameter being removed each time, and these analyses are then ranked according to the coefficients of determination, in increasing order. We thus obtain a ranking of the parameters in order of effect. If the effect of one parameter on the HC5s is important, then its absence in the ANOVA will yield a low coefficient of determination.

We analyzed the HC5_50% and the HC5_5% (lower limit of the 90% confidence interval of HC5) as well as the amplitudes of the confidence intervals (HC5_95%/HC5_5%) in log base 10. In each case, we carried this out in two stages: analyses of the whole set of toxicants simultaneously and analyses of each toxicant separate. In addition to this, we also will study the amplitude of possible variations in the HC5 and the percentage of tests lying below each one (the real percentage of tests affected). All the results are presented in log base 10 (with concentration in μg/L).

RESULTS

Analyses of variance

Analysis of HC5_50% all substances. Analysis of variance of the whole set of HC5_50% (all toxicants and all methods) shows a good fit between the calculated values and the values predicted by the linear model (coefficient of determination is 0.9952 when taking into account the four parameters and their interactions). The effect of the toxic parameter appears to be much more important compared to that of the others (Table 3): the various HC5s calculated (945) particularly stand out because of the degree of toxicity of the substances; but it is the analysis of the other parameters that concerns us.

The main effects of the parameters are the most important; nonetheless, interactions containing the toxic parameter are statistically significant (in particular the toxic–taxo interaction). The effect of the different options for constructing an SSWD thus depends on the toxicants, that is, on the distribution and proportions of data available for each one. Therefore, broadly speaking, no approach leads systematically to more protective or less protective results whatever the toxicant. On the other hand, no, or practically no, observed interactions are found between the species, taxo, and meth&dist parameters, that is, between the methods of constructing an SSWD.

Apart from the toxic parameter, the species parameter produces the strongest main effects (without interactions). Therefore, whether or not intraspecies variation is taken into account seems to have a similar and significant effect on HC5_50% values. Contrasting effects are observed between the mean and unweighted approaches, the former is the least protective (effect 0.11 on HC5 in log base 10) whereas the latter turns out to be the most protective (effect −0.08). The weighted approach is intermediate, but closer to the unweighted approach (effect −0.03).

If interactions (multi-ANOVAs: r^2 in Table 3) are taken into account, the ranking in order of importance of the parameters, after the toxic parameter, is as follows: taxo, meth&dist, and species. The effects of the three categories of the taxo parameter (biblio, equi, and enviro) are highly dependent on the toxicant; nonetheless, a global tendency exists for the biblio (effect −0.05) and enviro (effect 0.06) approaches to oppose each other. Therefore, the biblio approach appears in most cases to be more protective than the enviro approach. The equi approach lies between these two (effect −0.01).

The analysis of the meth&dist parameter makes it likely that its effect is linked more to the choice of distribution (normal or empirical) than to the method of calculation proper (WDR, DWB, GWB, or DWB*m*). Looking only at HC5s calculated by using the normal distribution, the various methods WDR, DWB, DWB*m*, and GWB appear very close (effects lying between −0.09 and −0.06), or even practically identical for DWB, DWB*m*, and GWB (effects lying between −0.07 and −0.06). With respect to empirical distribution, the DWB*m* and GWB methods are very close to each other (effect of 0.10 and 0.11, respectively); they are not very distant from the DWB method (effect 0.07). On the other hand, going from HC5s calculated by using normal distribution to those calculated by using empirical distribution, we go from an effect

Table 4. Coefficients of determination (r^2) of multi–analyses of variance (ANOVAs) carried out toxicant by toxicant on the hazardous concentration (HC5_50%). For a given toxicant, the most important parameter is the one that generates the lowest r^2 in the analysis (given in italics). The species parameter corresponds to the question of the intraspecies variation; taxo, to the three taxonomic groups; and meth&dist, to the statistical method and distribution

Toxicant	r^2 complete ANOVA	r^2 ANOVA without parameter		
		Species	Taxo	Meth&dist
Boric acid	0.9689	0.6609	*0.2899*	0.8714
Cadmium	0.9790	*0.3905*	0.7036	0.7893
Copper	0.9889	0.8620	0.6767	*0.3974*
Nickel	0.9363	0.8536	*0.3970*	0.5120
Lead	0.9954	0.7544	*0.2323*	0.8329
Zinc	0.9953	0.7562	0.9627	*0.2461*
Mercury	0.9868	0.7615	*0.1953*	0.9434
Chromium	0.9520	*0.3149*	0.6887	0.7858
Dibutyl phthalate	0.9435	*0.2783*	0.5721	0.8355
Butylbenzyl phthalate	0.9679	0.6451	*0.5834*	0.6439
Dodmac	0.9870	*0.4128*	0.7734	0.6989
Lindane	0.9964	0.9569	*0.3433*	0.5682
Atrazine	0.9697	*0.2959*	0.7304	0.8326
Parathion	0.9907	0.9683	0.7690	*0.2199*
Pentachlorophenol	0.9842	0.7999	*0.3146*	0.8239

lying between −0.09 and −0.06 (depending on the method) to an effect lying between 0.07 and 0.11. The choice of distribution (normal vs empirical) thus has a much greater effect on the calculation of HC5 than the way in which the desired weighting is applied. It can also be observed that use of a normal distribution (as opposed to an empirical distribution) tends to make the HC5 more protective.

Six outliers (combinations for which a bad fit exists between the calculated HC5 and the value predicted by the linear model associated with the ANOVA) were identified. Boric acid is present in them four times, and in three cases with the biblio approach (no distinction between taxonomic groups) associated with normal distribution.

Analysis of HC5_50% substance by substance. Subsequently, we carried out a multi-ANOVA toxicant by toxicant, for the three parameters taxo, species, and meth&dist. This confirmed to us the strong interactions previously observed between toxicants and methods used to construct the SSWD, because the ranking in order of importance of these three parameters is not the same for all the toxicants. A ranking of the parameters can nonetheless be established for the 15 toxicants (Table 4). This yields the following order: taxo, species, and meth&dist (distribution). The taxo parameter is the most important parameter for variations in HC5 for 7 out of 15 toxicants, and is ranked second 6 times out of 15. The species parameter is the most important parameter for 5 out of 15 toxicants; it is also ranked second 5 times out of 15. These two parameters thus have a similar degree of influence on the 15 toxicants. The meth&dist parameter is observed to be the most important 3 times out of 15 and is ranked second 4 times out of 15. Similar results were observed for HC5_5% (lower limit of HC5); details are not given here.

Analysis of the amplitude of the confidence interval of HC5 for all substances. As we did for the HC5_50% and the HC5_5%, we carried out an ANOVA on the amplitudes of HC5 confidence intervals calculated for all toxicants and all methods (Table 5). The fit between the calculated amplitudes and those predicted by the model associated with the ANOVA is not as good as with HC5_50% (the coefficient of determination is 0.848 with all parameters and their interactions taken into

Table 5. Respective effect of parameters and interactions of the analysis of variance on the hazardous concentration affecting 5% of species with 50% confidence (HC5). Coefficients of determination (r^2) of the multianalysis of variance on the HC5; ranking is given in order of importance of parameters with interactions taken into account[a]

Parameter	*df*	Sum of square	Mean of square	*F* value	Pr(*F*)	r^2
Toxic	14	64.66	4.62	106.94	0.0000	0.111 (1)
Species	2	0.56	0.28	9.83	0.0001	0.770 (4)
Taxo	2	3.64	1.82	63.35	0.0000	0.717 (3)
Meth&dist	6	5.78	0.96	33.60	0.0000	0.624 (2)
Toxic:species	28	8.76	0.31	10.90	0.0000	
Toxic:taxo	28	10.49	0.37	13.06	0.0000	
Toxic:meth&dist	84	20.39	0.24	8.46	0.0000	
Taxo:meth&dist	12	4.14	0.35	12.03	0.0000	
Species:meth&dist	12	1.35	0.11	3.92	0.0001	
Species:taxo	4	0.20	0.05	1.74	0.1388	
Residuals	752	21.58	0.03			

[a] *F* = Fisher's statistic; Pr(*F*) = probability that the effect of each parameter is zero. Toxic = the different toxicants; species = the question of intraspecies variation; taxo = three taxonomic groups; meth&dist = statistical method distribution.

Table 6. Coefficients of determination (r^2) of multi–analyses of variance (ANOVAs) carried out toxicant by toxicant on the amplitude of the hazardous concentration (HC5) confidence interval. For a given toxicant, the most important parameter is the one that generates the lowest r^2 in the analysis (given in italics). The species parameter corresponds to the question of the intraspecies variation; taxo, to the three taxonomic groups; and meth&dist, to the statistical method and distribution

Toxicant	r^2 complete ANOVA	r^2 ANOVA without parameter		
		Species	Taxo	Meth&dist
Boric acid	0.9404	0.8633	0.4260	*0.0941*
Cadmium	0.9548	0.8395	*0.4025*	0.4345
Copper	0.8889	0.7625	0.4962	*0.2208*
Nickel	0.9435	0.7569	0.2840	0.3013
Lead	0.9262	0.8191	*0.2476*	0.4895
Zinc	0.9701	0.6049	0.6684	*0.3335*
Mercury	0.9459	0.8747	*0.2912*	0.4558
Chromium	0.9641	0.6742	0.8972	*0.2205*
Dibutyl phthalate	0.9417	*0.4467*	0.7166	0.5046
Butylbenzyl phthalate	0.9201	0.4965	0.4097	*0.3770*
Dodmac	0.9895	*0.0639*	0.9401	0.7951
Lindane	0.9837	0.9640	*0.3683*	0.3984
Atrazine	0.8948	0.4305	0.4951	*0.3383*
Parathion	0.9804	0.8861	0.6149	*0.1969*
Pentachlorophenol	0.9088	0.7197	0.5693	*0.2259*

account). Thus, more uncertainty exists in the amplitude of the HC5 confidence interval than in the HC5_50% proper. The toxic parameter, although still the most important, has less effect on the confidence intervals than on the HC5s.

In the same way as for HC5_50%, interactions between the toxic parameter and the other parameters are significant; they even sometimes predominate over the main effects. If only main effects (without interactions) are taken into account, and with the exception of the toxic parameter, the taxo parameter is the most important. The biblio and enviro approaches still pull in opposite directions. The former tends, in most cases, to reduce the amplitude of the confidence interval (effect −0.08 in log base 10) whereas the latter increases it (effect 0.07). The equi approach lies between the two (effect 0.01).

On the other hand, if interactions are taken into account, the meth&dist parameter is the most important (multi-ANOVA: r^2 in Table 5); the species parameter appears to have very little effect. The analysis of the meth&dist parameter shows this time that the effect of the method of calculation (WDR, DWB, GWB, and DWB*m*) is greater than that of the distribution (normal or empirical). The DWB*m* method tends to increase the amplitude of the confidence interval (effect between 0.1 and 0.12 depending on the distribution) compared to the GWB method (effect between −0.02 and 0.02), with the latter coming before both the DWB and WDR methods (effect between −0.05 and −0.1). Out of six outliers identified, three correspond to copper with the equi approach and the empirical distribution.

Analysis of the amplitude of the confidence interval substance by substance. The multi-ANOVA toxicant by toxicant of the amplitude of the HC5 confidence interval yields the following results (Table 6). The most important parameter still differs from one toxicant to another but, for the 15 toxicants, a ranking in order of importance can be drawn up: meth&dist (method), taxo, and species. The meth&dist parameter is the most important for 8 out of 15 toxicants; it is ranked second 7 times out of 15. Taxonomy (taxo) is ranked first for 5 out of 15 toxicants, and, again, it is ranked second 5 times out of 15. Finally, the species parameter is ranked first 2 times out of 15 and second 3 times out of 15. The differences between the amplitudes of the HC5 confidence intervals are chiefly connected to the various methods of calculation of the HC5 (WDR, DWB, GWB, and DWB*m*) and to the options of the taxo parameter.

Amplitude of variation in HC5s and percentage of tests affected

The range (i.e., maximum–minimum in log base 10) of the HC5_50%s calculated by the 63 methods varies from one toxicant to another (Table 7 and Fig. 3). It lies between 0.3 (in log base 10) for butylbenzyl phthalate and 1.19 (i.e., more than one order of magnitude for concentrations) for boric acid; it exceeds 0.5 (factor of three for concentrations), in 12 cases out of 15. The deviation from the usual method (meth = WDR, dist = normal, species = mean, taxo = biblio) can reach 0.2 (in log base 10) for butylbenzyl phthalate and 1 for boric acid. The deviation between the usual method and the method consisting of a quadruplet (DWB, normal, weighted, equi) does not exceed 0.32, that is, a factor of two on the concentration. The deviation between the usual method and the (DWB, normal, weighted, enviro) method is greater. The deviation can be as high as 0.84 in log base 10, that is, a factor of seven on the concentrations, and it is greater than 0.25 for 50% of the toxicants.

The amplitudes of the confidence intervals lie between 0.1 and 2.66 in log base 10 depending on methods and toxicants (Table 8). The amplitude is lowest on average for butylbenzyl phthalate (between 0.11 and 0.63 depending on methods) and greatest for boric acid (between 0.24 and 2.66).

The percentage of data lying below HC5_50% also is variable, depending on methods and toxicants and varies between 0 and 40%. This percentage is lowest on average for boric acid, not exceeding 10% (even if deviations among HC5s for this toxicant are the greatest), and greatest for dodmac. This percentage of tests affected can exceed 20% for one half of the toxicants; it is always under 18% with the usual method, and under 15% with the (DWB, normal, weighted, equi) method. However, this percentage may be greater for the (DWB, normal, weighted, enviro) method. The percentage of data lying under HC5_5% can still reach 30% for dodmac, but it is systematically under 15% for one half of the toxicants. This

Table 7. Statistical characteristics of results associated with 63 methods of calculation of the hazardous concentration (HC5 in log base 10) toxicant by toxicant[a]

Toxicant	Mean (63)	Deviation to the mean				Amplitude of the results		Proportions of data below HC5		
		Min	q25	q75	Max	q75–q25	Max–min	Min	Mean	Max
Atrazine	0.65	−0.41	−0.16	0.13	0.48	0.29	0.89	0.00	0.10	0.18
Boric acid	2.62	−0.58	−0.28	0.27	0.61	0.56	1.19	0.00	0.06	0.10
Butylbenzyl phthalate	1.96	−0.11	−0.05	0.04	0.19	0.09	0.30	0.08	0.11	0.23
Cadmium	−0.23	−0.21	−0.07	0.09	0.19	0.15	0.41	0.05	0.09	0.17
Chromium	1.08	−0.27	−0.09	0.11	0.37	0.21	0.65	0.04	0.07	0.12
Copper	0.43	−0.31	−0.18	0.16	0.30	0.33	0.61	0.05	0.08	0.15
Dibutyl phthalate	2.13	−0.21	−0.13	0.03	0.42	0.17	0.63	0.00	0.14	0.23
Dodmac	1.61	−0.35	−0.15	0.13	0.30	0.28	0.64	0.06	0.12	0.41
Lead	1.21	−0.24	−0.10	0.09	0.29	0.20	0.53	0.05	0.12	0.21
Lindane	0.29	−0.53	−0.37	0.31	0.54	0.68	1.07	0.00	0.08	0.23
Mercury	−0.32	−0.54	−0.21	0.31	0.59	0.52	1.13	0.06	0.17	0.31
Nickel	0.85	−0.45	−0.17	0.15	0.50	0.32	0.95	0.05	0.06	0.14
Parathion	−1.24	−0.56	−0.30	0.26	0.60	0.57	1.16	0.00	0.07	0.22
Pentachlorophenol	1.11	−0.38	−0.17	0.13	0.40	0.31	0.78	0.05	0.14	0.34
Zinc	1.33	−0.20	−0.12	0.06	0.25	0.18	0.46	0.00	0.07	0.18

[a] min = minimum; q25 = quantile 25%; q75 = quantile 75%; max = maximum.

percentage does not exceed 5% for the (DWB, normal, weighted, equi) method and 10% for the usual method.

DISCUSSION

Differences between HC5_50% are naturally, and fortunately, connected first of all to differences in toxicity between the substances. The least toxic substance is separated from the most toxic substance by four orders of magnitude. However, different methods can be separated by one order of magnitude for the same toxicant (Fig. 3). Regardless of differences in toxicity between substances, the HC5s differ by the options used, first in terms of taxonomic groups and intraspecies variation, and to a lesser degree in terms of the choice of distribution. On the other hand, whatever method of calculation is used (WDR, DWB, GWB, or DWB*m*), the HC5s obtained are very similar. Depending on the toxicant, the order in which these parameters have an effect can vary, and generally the effect of the methods is toxicant-dependent, and depends on the breakdown and the proportions of available data for each method. This is particularly true for taxonomy. The effect of the proportions determined for each taxonomic group depends on the proportions that are found in the data. The proportions proposed by Forbes and Calow [9] tend to be less conservative; this is connected to the fact that a very high weight is allocated to algae, which, for the 15 toxicants examined, are rarely the most sensitive species. Conversely, taking into account intraspecies variation systematically increases the range of the distribution, but without modifying its average level; the result of this is to lower the HC5. Taking into account the whole set of data and weighting the data with respect to species is thus more conservative (legitimately) than calculating means for each species, regardless of the toxicant.

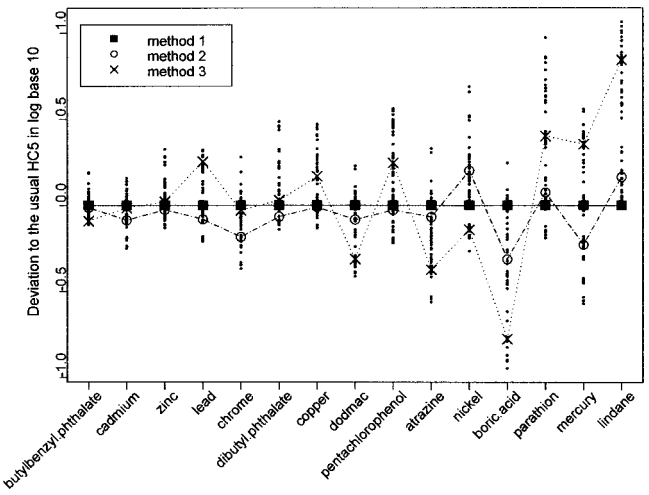

Fig. 3. Deviation from the usual approach (method 1) for each method of calculation of a hazardous concentration (HC5 in log base 10), for the 15 toxicants in order of increasing intermethods divergence (standard deviation). The species option taken into account is, respectively, for method 1: mean (a geometric mean of data for each species is first calculated), and for methods 2 and 3: weighted (each piece of data is weighted to give each species the same weight). The taxonomic option taken into account is, respectively, for method 1: biblio (the proportions of data given in the literature are retained), for method 2: equi (the data are balanced in such a way that the three taxonomic groups are equally weighted), and for method 3: enviro (the data are weighted in such a way as to respect the proportions proposed by Forbes and Calow [9]). The statistical option taken into account is, respectively, for method 1: weighted direct regression with the normal distribution, and for methods 2 and 3: direct weighted bootstrap also with the normal distribution.

Even if the method of calculation used (WDR, DWB, GWB, and DWB*m*) has little effect on the value of HC5, its effect on the amplitude of the confidence interval associated with each HC5 is nonetheless the most important. With the DWB and DWB*m* methods, the three taxonomic groups are not necessarily present as regards each draw of the bootstrap; the proportions desired and defined by the choice of approaches concerning intraspecies variation and taxonomy are not necessarily scrupulously respected for each draw either (these proportions are approached over 1,000 draws of the bootstrap). Conversely, with the GWB method, the number of data points drawn in each taxonomic group is the same in each sample, and with the WDR method the desired weightings are always totally respected. These choices do not have a major effect on the calculation of the HC5_50%; they have a greater effect on its confidence interval. Nonetheless, the most important factor seems to be the number of data points drawn in each sample

Table 8. Statistical characteristics regarding amplitudes of the hazardous concentration (HC5) confidence intervals and proportions of data below the hazardous concentration affecting 5% of species with 5% confidence (HC5_5% in log base 10)[a]

	Amplitude of confidence interval			Proportions of data below HC5_5% (lower limit)		
Toxicant	Min	Mean	Max	Min	Mean	Max
Atrazine	0.19	0.61	0.86	0.00	0.04	0.16
Boric acid	0.24	1.21	2.66	0.00	0.03	0.07
Butylbenzyl phthalate	0.11	0.35	0.63	0.00	0.06	0.15
Cadmium	0.25	0.49	0.89	0.02	0.04	0.11
Chromium	0.47	0.89	1.80	0.02	0.03	0.08
Copper	0.22	0.83	1.93	0.02	0.04	0.10
Dibutyl phthalate	0.42	0.71	1.03	0.00	0.07	0.15
Dodmac	0.03	0.68	1.11	0.00	0.09	0.29
Lead	0.22	0.56	1.03	0.02	0.05	0.14
Lindane	0.39	0.96	2.31	0.00	0.02	0.14
Mercury	0.52	0.93	1.84	0.00	0.08	0.20
Nickel	0.82	1.17	1.59	0.00	0.03	0.05
Parathion	0.59	1.03	2.07	0.00	0.02	0.04
Pentachlorophenol	0.43	0.78	1.29	0.02	0.06	0.16
Zinc	0.10	0.33	0.57	0.00	0.04	0.15

[a] min = minimum; max = maximum.

of the bootstrap, because the most pronounced contrast is between the two methods DWB and DWB*m* (the only difference between the two is the size of the sample). When we go from the DWB method to the DWB*m* method, the reduction in the size of the samples drawn in the bootstrap generates a greater intersample variation, and consequently an increase in the amplitudes of the confidence intervals. The GWB method lies between the two preceding ones. The reduction in the size of the samples drawn increases the amplitudes of the confidence intervals, but this increase is compensated for by the fact that the data points are drawn separately in the three taxonomic groups. Because of this, the number of data points belonging to each of the three categories is the same in each sample generated by the bootstrap, which naturally tends to reduce variation among the samples. Paradoxically, the WDR method yields results that are very similar to the DWB method. However, it should be observed that, whereas the amplitude of the HC5 confidence interval depends mainly on the method of calculation, the HC5_5% (lower limit of the HC5 confidence interval), like the HC5_50%, mainly depends on taxonomy. Variations in the mean level of HC5 are greater than those of its confidence interval.

The amplitudes of HC5 confidence intervals calculated for some toxicants such as butylbenzyl phthalate are abnormally low. The number of data points available is too low for these toxicants. A minimum of approximately 20 data points appears to be necessary to use the bootstrap technique. Two toxicants (copper and boric acid) were found to behave in a different way to the others in ANOVAs (outliers). We will simply note that, for these two toxicants, the distribution of available data could not be considered normal because of an excess of outliers at low values.

CONCLUSION

Thus, the result of an SSD largely depends on the way the data are processed and differences of about one order of magnitude may be observed between the methods for the same substance. In concrete terms, constructing an SSD means making choices and constructing hypotheses that we have attempted to clarify (hence the change in name from SSD to SSWD). Some of these choices are now suggested, and some possibilities are therefore discarded.

Regarding redundant data for each species, rather than using the standard method that consists in calculating a geometric mean beforehand, we favor the weighted method, the aim of which is to retain intraspecies variation while giving each species the same weight within the SSWD. This uses the entire set of data, but does not favor one species over the others.

With regard to the statistical method, we favor approaches that enable us to construct various hypotheses about the probability distribution of the data (empirical distribution, log-normal or log-logistic distribution, and others), and to test these hypotheses. This is the case for methods based on the weighted bootstrap, called DWB, DWB*m*, and GWB. The WDR regression approach is more costly in terms of computing time than the others, while giving results similar to those obtained with the DWB method, is therefore discarded. The GWB approach also is discarded, because it is better suited to the regrettably rare cases where there are a considerable number of data points (e.g., more than 15) in each taxonomic group; its results are nonetheless intermediate between those of the DWB and DWB*m* methods, which we retain. What is the difference between these two methods? When a weighting is chosen that is somewhat in conflict with proportions of available data, we find few data points with a very high weight within the SSWD. These data will tend in the DWB method to be present with very high redundancy in samples generated by the bootstrap, which does not yield a satisfactory distribution. The DWB*m* method compensates for this drawback. However, when the number of data points is very low, the value of the HC5 rests to a large extent on the few large weighted data, and uncertainty is no longer spread over the whole data set. Estimating HC5 and its confidence interval properly becomes difficult whatever method is chosen.

Making the right choices regarding taxonomy (weighting of data corresponding to the three taxonomic groups) is the trickiest question, particularly because this is the parameter with the greatest effect on the result of the SSWD. Expertise about ecological communities should nonetheless be included

in the construction of the SSWD, and testing various weightings for the proportions found in the literature for the three taxonomic groups appears to be essential.

By doing this, depending on the choice of statistical method (DWB or DWB*m*), of probability distribution, and of the weightings regarding the three taxonomic groups, several values of HC5 are obtained for one substance. Nothing is absurd about this, and indeed it may be an asset for risk assessment. The various values should be discussed by experts with the aim of defining a PNEC. In any case, the various values make it possible to put into perspective the confidence interval obtained for each value, or to define a safety factor that can be applied to one of them.

REFERENCES

1. National Institute of Public Health and the Environment (RIVM). 1999. Environmental risk limits. RIVM Report 601640001. Bilthoven, The Netherlands.
2. European Commission. 1996. Technical guidance document in support of commission directive 93/67/EEC on risk assessment for new notified substance and commission regulation 5EC no 1488/94 on risk assessment for existing substance. Part II. Office for Official Publications of the European Communities, Luxembourg.
3. Kooijman SALM. 1987. A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Res* 21:269–276.
4. Van Straalen NM, Denneman CAJ. 1989. Ecotoxicological evaluation of soil quality criteria. *Environ Manag* 14:779–792.
5. Wagner C, Løkke H. 1991. Estimation of ecotoxicological protection levels from NEOC toxicity data. *Water Res* 25:1237–1242.
6. Aldenberg T, Slob W. 1993. Confidence limits for hazardous concentrations based on chronic NOEC toxicity data. *Ecotoxicol Environ Saf* 25:48–63.
7. Posthuma L II, Suter GW II, Traas T, eds. 2002. *Species Sensitivity Distributions in Ecotoxicology*. Lewis, Boca Raton, FL, USA.
8. Forbes TL, Forbes VE. 1993. A critique of the use of distribution-based extrapolation models in ecotoxicology. *Funct Ecol* 7:249–254.
9. Forbes VE, Calow P. 2002. Species sensitivity distributions revisited: A critical appraisal. *Hum Ecol Risk Assess* 8:3:473–492.
10. Versteeg DJ, Belanger SE, Carr GJ. 1999. Understanding single-species and model ecosystem sensitivity: Data-based comparison. *Environ Toxicol Chem* 18:1329–1346.
11. Jagoe RH, Newman MC. 1997. Bootstrap estimation of community NOEC. *Ecotoxicology* 6:293–306.
12. Newman MC, Ownby DR, Mézin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson B-A. 2000. Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species. *Environ Toxicol Chem* 19:508–515.
13. van der Hoeven N. 2000. Estimating the 5-percentile of the species sensitivity distribution without any assumptions about the distribution. *Ecotoxicology* 10:25–34.
14. Aldenberg T, Jaworska JS. 2000. Uncertainty of the hazardous concentration and proportion affected for normal species sensitivity distributions. *Ecotoxicol Environ Saf* 46:1:1–18.
15. Shao Quanxi. 2000. Estimation for hazardous concentrations based on NOEC toxicity data: An alternative approach. *Environmetrics* 11:583–595.
16. Verdonck FAM, Jaworska J, Thas O, Vanrolleghem PA. 2001. Determining environmental standards using bootstrapping, Bayesian and maximum likelihood techniques: A comparative study. *Anal Chim Acta* 446:429–438.
17. Wheeler JR, Grist EPM, Leung KMY, Morritt D, Crane M. 2002. Species sensitivity distributions: Data and model choice. *Mar Pollut Bull* 45:192–202.
18. Van Straalen NM. 2002. Threshold models for species sensitivity distributions applied to aquatic risk assessment for zinc. *Environ Toxicol Pharmacol* 11:167–172.
19. Grist EPM, Leung KMY, Wheelers JR, Crane M. 2002. Better bootstrap estimation of hazardous concentration threshold for aquatic assemblages. *Environ Toxicol Chem* 21:1515–1524.
20. Efron B, Tibshirani RJ. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY, USA.