



METALS ENVIRONMENTAL RISK ASSESSMENT GUIDANCE

FACT SHEET

MERAG

03

EFFECTS ASSESSMENT: DATA COMPILATION, SELECTION AND DERIVATION OF PNEC VALUES FOR THE RISK ASSESSMENT OF DIFFERENT ENVIRONMENTAL COMPARTMENTS (WATER, STP, SOIL, SEDIMENT)



January 2007

Check you have the most recent fact sheet by visiting www.metalsriskassessment.org

CONTENTS

01. Introduction	01
02. Data compilation and selection	02
2.1 Criteria for data reliability and data relevance	02
2.1.1 Data reliability	02
2.1.1.1 Type of test	03
2.1.1.2 Description of test material and methods	04
2.1.1.3 Description of physico-chemical test conditions	04
2.1.1.4 Chemical analysis	05
2.1.1.5 Concentration-effect relationships	06
2.1.2 Data relevancy	07
2.1.2.1 Biological relevancy of the endpoint used	07
2.1.2.2 Relevancy of the test substance	08
2.1.2.3 Relevancy of the test species	08
2.1.2.4 Relevance of the exposure duration	08
2.1.2.5 Acclimation / Adaptation	09
2.1.3 Conclusions	10
03. Derivation of the Predicted No Effect Concentration (PNEC)	11
3.1 Aggregation/selection of L(E)C ₅₀ /NOEC data	11
3.2 Approach for the derivation of the PNEC	13
3.2.1 Introduction	13
3.2.2 Calculation of the PNEC using assessment factors (data-poor substances)	13
3.2.3 Calculation of the PNEC using statistical extrapolation methods (data rich substances)	13
3.2.3.1 Choice of appropriate distribution model	15
3.2.3.2 Specific requirements to be fulfilled before using the statistical extrapolation method	17
3.2.3.2.1 Number of data	17
3.2.3.2.2 Data treatment	20
3.2.3.3 Estimating the chronic effects distribution from the acute effects distribution	20
3.2.3.4 Uncertainty management	21
3.2.4 Effects of weight-of-evidence	21
Annex A – Use of assessment factors according to the Technical Guidance Document (TGD, 2004)	23
Annex B – Use of assessment factors according to the OECD Manual for Investigation of HPV Chemicals (OECD, 2004)	27
References	29

1. INTRODUCTION

In the framework of environmental risk assessment the main goal is long term protection of the environmental compartment under consideration. Typical compartments that are considered for the inland environment are the aquatic (including the sediments) sewage treatment plants (STPs) and terrestrial ecosystems. For each of these compartments a Predicted No-Effect Concentration (PNEC) needs to be derived. This PNEC is considered the concentration below which an unacceptable effect will most likely not occur.

Data selected for PNEC derivation should comply with the requirements (criteria) for data quality and data relevance taking into account metal specific considerations. Therefore, it is deemed necessary to develop first a set of metal specific reliability and relevance criteria against which to evaluate the ecotoxicity data to be used. In the subsequent sections guidance is provided on data quality, aggregation, interpretation, derivation of the PNEC value.

The content of the MERAG fact sheets reflect the experiences and recent progress made with environmental risk assessment methods, concepts and methodologies used in Chemicals Management programs and Environmental Quality Standard setting (soil, water, sediments...) for metals. Since science keeps evolving these fact sheets will be update on a regular basis to take into account new developments. To be sure you have the most recent fact sheet on the current subject check our website: www.metalsriskassessment.org

2. DATA COMPILATION AND SELECTION

Ecotoxicity data can be drawn from data required for regulatory purposes (e.g. IUCLID) as well as from relevant literature and/or internationally recognized databases. Because the data quality of the extracted information may vary considerably between individual source documents it is very important to evaluate all ecotoxicity data with regard to their adequacy for PNEC derivation and risk assessment purposes. In general, this evaluation involves a review of how well each study was conducted (see Table 1 below) and how the results are interpreted in order to accept (or reject) a study in accordance with the purpose of the assessment. The term adequacy covers here both the *reliability* of the available data and the *relevance* of the data for environmental risk assessment/environmental quality setting purposes in general and for metals/metal compounds in particular. These two basic elements are defined as follows:

- *Reliability*: covering the inherent quality of a test relating to test methodology and the way that the performance and results of a test are described.
- *Relevance*: covering the extent to which a test is appropriate to be used for the derivation of a PNEC for a given region/place and ecosystem type.

Only those data that can be considered of sufficiently high quality should be retained for the assessment and used for the PNEC derivation. Guidance on how to screen and select the most appropriate data in the framework of PNEC setting is outlined further. Metal specific issues and new concepts are highlighted.

2.1 Criteria for data reliability and data relevance

The term *reliable* or *relevant* can be assigned to a study if the study complies with a number of criteria.

2.1.1 Data reliability

A checklist for evaluating the general quality of ecotoxicity studies is provided in Table 1. These criteria are mostly not metal-specific: they simply adhere to the principles of good study conduct.

The criteria mentioned below should be met, for a study's results to be considered reliable. An experiment can be classified as *reliable (Q1)* if it has been carried out according to all criteria, or is missing one or two less important criteria. If one important criterion, or several less important criteria are missing the experiment should be classified as *less reliable (Q2)*, while an experiment should be classified as *unreliable (R)* if several important criteria are missing. They are outlined in more detail further, together with some more metal specific focus points.

Table 1: Checklist of criteria for the evaluation of the reliability of ecotoxicity studies used for risk assessment and/or EQS setting

<p>Type of test</p> <ul style="list-style-type: none"> • standard test or non-standard test • endpoint used reported • test duration reported • static or flow through
<p>Description of test material and methods</p> <ul style="list-style-type: none"> • test set-up, measuring chamber/device • test material (including purity), solutions, dilution water if applicable • test organism, including size (age), origin, number of organisms per replicate • test design (# replicates should be used) • type of food given (chronic tests)
<p>Description of physico-chemical test conditions</p> <ul style="list-style-type: none"> • proper description and control of physico-chemical conditions (e.g. pH, major cations and anions) that may influence the outcome of a test (validity criteria should be met at the end of the test)
<p>Chemical analysis</p> <ul style="list-style-type: none"> • test concentrations during the test are measured • test concentrations are not measured, but indication is given that the nominal concentrations are close to actual concentrations • evidence is given that concentrations were maintained during the test (< 30% variation)
<p>Concentration-effect relationship</p> <ul style="list-style-type: none"> • acceptable control mortality, reproduction, growth. • sound statistics used, 95 % confidence limits reported or data on the relationship given amenable to further analysis to derive a suitable L(E)C_x value • concentration range is given • at least 2 different concentrations must have been tested besides the control • a concentration related response should be clear (a progressive effect should be observed as a function of the dose) • hormesis effect observed or not

2.1.1.1 Type of test

- Both standard test organisms and non-standard species can be used in the framework of a risk assessment. In general, toxicity data generated from standardized tests, as prescribed by organizations such as OECD and USEPA will need less scrutiny than non-standardized test data,

which will require a more thorough check on their compliance with reliability criteria before being used. GLP¹ and non-GLP tests can be used provided that the latter fulfill the stipulated requirements.

- In the aquatic environment both static and semi-static tests and flow through tests can be used. The results of the latter should be handled with care for metals/metal compounds and it should be evaluated if enough equilibration time was provided to allow for equilibrium partitioning of the bioavailable metal fraction (see MERAG background document). For sediment testing, semi static test designs are preferred but static test designs may also be used. Note that under static conditions the sediment may act as a source of dissolved metals to the overlying water column, (Wang et al, 2004). If the latter is observed the tests should not be considered reliable.
- For terrestrial soil testing and sediment testing, adequate time should elapse between mixing metal or metal compounds into the test medium and introducing biota (plants or soil/sediment invertebrates). Both short equilibration times and high spiked metal concentrations in sediments/soils will accentuate partitioning of metals disproportionately to the dissolved phase and increase the probability of exposure and/or toxicity via dissolved metals (Lee et al, 2004, Simpson et al., 2004) making it more uncertain to use results of laboratory spiked studies in determining the partitioning of metals to different binding phases of sediment/soils in the field.. Equilibration time is also important for spiking procedures in water only tests. If for example dissolved organic carbon (DOC) is present allowing enough equilibrium time is important. However, equilibration in a water only system will be reached within days while for sediments and soils chemical equilibrium and the aging process on the partitioning metals will lead to much more equilibration times in sediments/soils than in water. As a consequence, it is recommended that the overlying water of sediment toxicity tests is measured for the test substance, and that testing is initiated only when overlying water concentration reaches acceptable levels, e.g. ambient concentrations. Simulated aging and weathering processes may also be desirable but currently are not required in standard test protocols,

2.1.1.2 Description of test material and methods

- A detailed description of methods employed in the study should be provided. This description should include at least the method of test medium preparation, time of spiking, recorded observations. To calculate free ion concentrations with speciation codes the concentrations of dissolved major anion and cations, Fe, Mn, Al, dissolved organic carbon, pH are required. Furthermore the organisms used should be uniform in age and represent a sensitive life stage. The test results should allow a proper statistical analysis and the experimental design should provide sufficient replicates per test concentration to derive a high quality L(E)C_x/NOEC value².

2.1.1.3 Description of physico-chemical test conditions

- In Table 2 an overview is given of physico-chemical characteristics for each compartment that should preferably be reported and fall within the tolerance limits of the test organisms. If these limits are exceeded the test has to be considered not reliable.

¹ GLP = Good Laboratory Practice

² L(E)C_x = the concentration that causes x % change in response (e.g. mortality, immobility) during a specified time interval. NOEC = No Observed Effect Concentration is defined as the test concentration below the lowest concentration that did result in a significant effect (LOEC = Lowest Observed Effect Concentration) in the specific experiment

Table 2: Physico-chemical parameters that should preferably be reported

Water	Sediment	Soil
<ul style="list-style-type: none"> • temperature • oxygen • hardness • salinity • pH 	<ul style="list-style-type: none"> • temperature • particle size • ammonia/ammonium • organic carbon • oxygen • salinity • pH 	<ul style="list-style-type: none"> • temperature • particle size • organic carbon • pH • clay content

- In addition to the above mentioned parameters, abiotic parameters, e.g. dissolved organic carbon concentration (DOC) in the test water, Acid Volatile Sulfides (AVS) in sediments, Cation Exchange Capacity (CEC) and soil pore water chemistry (DOC, cations and anions) that govern the speciation and hence the bioavailability of some metals are to be considered in case correction for bioavailability is required. Furthermore, in the case of testing essential metals a proper description of the culture conditions, specifically related to the level of essential metals added or already present in the culture media could give valuable insight on issues such as acclimation.

2.1.1.4 Chemical analysis

- There is a strong preference for using measured data. Analytical measurements of the metal concentrations in the test solution allow to (1) exclude human error related to the preparation/addition of test substance solutions; (2) since metals are a natural elements it is therefore important to know the total metal concentrations organisms are exposed to, including the metal background levels in the control/dilution test medium. In the case river waters are used, the metal levels in control waters can already be relatively elevated in comparison to the metal added as test solution. In this respect it is important to also consider that organisms adapt to the culture media not test media
- If it is not mentioned whether the reported toxicity values are based on measured or nominal concentrations, they should be considered as nominal concentrations. In cases where no measured data are available the use of nominal concentrations could be considered as long as soluble metal salts have been used and the reported effect levels are well above the background in the test medium. However, if the effect levels are close to reported metal background concentrations in the specific test medium (i.e. $\mu_{\text{nominal}} - 1.95 \sigma_{\text{nominal}} \leq \mu_{\text{bg}} + 1.95 \sigma_{\text{bg}}$) or the test concentrations are close to the essentiality levels only measured values should be used. For sparingly soluble metals (e.g. Sb_2O_3) measured data on the dissolved fraction³ are always required. If the solubility is exceeded the test has to be considered as unreliable. Results from tests where a visual precipitation is observed should be discarded. The absence of a visual precipitation does not exclude that sometimes colloids may still be present that could still affect the test results.

³ Different definitions for the dissolved fraction exist. Most often the dissolved fraction in ecotoxicity tests refers to the fraction that passes through a filter of 0.45 μm . It should be noted, however, that this definition may not necessarily refer to the metals in solution. In the range of 0.01-0.45 μm colloid inert particles that remain suspended may exist and these could account for 50 % or more of the "dissolved" 0.45 μm fraction

2.1.1.5 Concentration-effect relationships

With regard to the acceptability of the test results the following recommendations can be formulated (these recommendations are not metal specific):

- Minimal requirements for endpoints such as mortality, growth, seed germination, reproduction (e.g. control mortality < 10 %) are often given in standard procedures. When these requirements are not met studies should be considered as not reliable.
- When adverse effects are observed in the different treatment groups a clear and consistent (increasing effect with increasing dose) concentration-effect relationship should be present. If no concentration-effect relationship can be established the test should be considered not reliable.
- Sometimes a hormesis effect is observed (i.e. increased performance in for example growth, reproduction) at low metal doses. Such effects can be important especially for trace nutrients such as Fe, Zn, Cu. In such cases, as positive effects should not be considered in the derivation of EC_x other models than the conventional log-logistic dose-response model should be used to fit the toxicity data. For example the linear-logistic model of Brain and Cousens (1989) has been extended to allow EC₅₀ and EC₁₀ calculations (Van Ewijk and Hoekstra, 1993; Schabenberger et al., 1999, Cedergreen et al, 2005) in the case of hormesis.
- Because effect concentrations are statistically derived values, information concerning the statistics should also be used as a criterion for data selection. If no methodology is reported and no raw data are reported or if values are 'visually' derived, the data have to be considered unreliable. In absence of sound statistics or no L(E)C_x or NOEC has been calculated or reported in the study itself, the study could still be used if data are available amenable to further analysis that allow to derive a suitable L(E)C_x or NOEC/LOEC value.
- Test concentration intervals should bracket the NOEC with concentrations that are as closely spaced as practical. Increasing the size of the test concentration intervals leads to reduced statistical power for the test. Following new OECD guidelines (e.g. OECD, 2001) test concentrations should preferably differ by no more than a factor of 2.

With regard to the proper use of NOEC/LOEC values and L(E)C_x values the following recommendations can be made (these recommendations are not metal specific):

- For acute studies L(E)C_x values should be estimated using appropriate statistical analysis (e.g. probit analysis or linear regression).
- For chronic studies concentration-response modeling such as regression models to calculate L(E)C_x⁴ are generally preferred over the classical hypothesis testing (p < 0.05) used to derive NOEC values. The latter method has indeed a number of limitations (Moore and Caux, 1997). Since the NOEC is by definition an applied dose, its value is to some degree dependent upon the choice of the experimenter. Secondly the NOEC depends upon the variability of the organism to a single dose. Organisms which are particularly sensitive to small variations in their environment, and hence display a greater variability of response to a given dose, are likely to have higher NOECs than if they were less sensitive, independent of their sensitivity to the toxicant. The use of a regression based approach offers the advantage that all of the information in the concentration-response curve is used and furthermore precludes the use of poor quality information because in those cases an inadequate model fit will be obtained.
- In case a benchmark dose (L(E)C_x) is calculated using a regression based approach and this value is to be used as an equivalent for a NOEC value, then typically a cut-off level should be identified representing a low effect percentile. This cut off value to be used should be derived based on the plausibility to detect a statistical significant difference and is depending on the inherent variability observed in the control test. The choice of the appropriate effect level is still an area under discussion and more research is needed (ISO, 2004; Environment Canada, 2005). A concentration that causes a low level of reduction, such as an EC₅ or EC₁₀, is rarely statistically

⁴ Usually L(E)C₁₀ values are selected, but the use of other L(E)C_x values (e.g. L(E)C₂₀) could also be warranted

significantly different from the control treatment. Therefore in some guidance documents the EC₂₀ is sometimes proposed as a compromise representing a low level of effect that is generally significantly different from the control treatment (US-EPA, 1999a). Whatever effect level is chosen it is recommended that the L(E)C_x value should not be extrapolated below the lowest applied (non-zero) concentration. According to Reiley et al (2003) and the draft ISO document (ISO, 2004) estimation of L(E)C_x values outside the concentration range tested introduces a great deal of uncertainty. If the resulting L(E)C_x value should be below the lowest applied control level (background level) or essentiality level, its reliability/relevance has to be questioned (another confounding factor in this respect is the hormesis phenomenon which for essential metals can be very important). Before estimating the L(E)C_x value it should also be checked, case-by-case if the experimental design is appropriate to be used for regression methods. The statistical design needed for a proper L(E)C_x derivation are more doses with fewer replicates at each dose. For estimating an L(E)C_x value three concentration groups, as well as the control group, is an absolute (theoretical) minimum. However, if there are only three treatment groups and one fails to show any (partial) effect the test would be considered inadequate. Therefore more concentration groups are recommended in practice (ISO, 2004). Many of the older toxicity data do not fulfill the statistical requirements in order to derive an L(E)C_x value. In those cases the conventional NOEC and LOEC values should be used. NOEC values could be in the natural range but LOEC values should not.

- If only a LOEC \geq 20% effect is reported (i.e. no NOEC could be derived as the lowest test group produced a response significantly different from the control group) and a distinct concentration/effect relationship is apparent, the L(E)C_x is calculated and should be evaluated if it can be regarded as the NOEC. If the effect percentage of the LOEC is unknown, no NOEC can be derived. Such an approach is only recommended if insufficient bounded NOECs are available.
- In general, the use of unbounded NOEC values is not recommended. Unbounded NOEC values should only be considered in specific cases. For example, if other toxicity values are not available for a particular species. In that case an unbounded NOEC could be used as a conservative estimate for the 'real' NOEC.

2.1.2 Data relevancy

After qualifying effect data as reliable, it also should be checked for relevancy for the risk assessment purpose. This is a step that is particularly important for metals/metal compounds. For example, the exposure duration could be insufficient for use in PNEC derivation or the pH or hardness of the test medium may be outside the boundaries of the physico-chemical conditions encountered in a specific environment under investigation. These relevancy issues should be considered carefully. A summary of the main attention points is given hereunder.

2.1.2.1 Biological relevancy of the endpoint used

- For risk assessment purposes the use of non-standardized endpoints (enzyme activity, morphological changes, etc.) of which the influence on the survival or reproduction of the species is not clear, should be done with caution. Preference should be given to toxicological criteria that may affect the species at the population level (e.g. survival, growth and reproduction). For microbial testing in soils, the use of NOECs based on functional parameters (e.g. C- and N-mineralization) are preferred over NOECs based on enzymatic processes (e. g. phosphatase, urease,...).

2.1.2.2 Relevancy of the test substance

- Since impurities can have an effect on the toxic properties of the substance under investigation or have toxic effects themselves, studies involving test substances in which impurity levels are >1% should not be used. High purity soluble metal salts should be used for the purpose of deriving a PNEC for risk assessment.

2.1.2.3 Relevancy of the species

- For preliminary, screening-level risk assessment purposes in general both endemic and non-endemic species are usually considered relevant. However, the use of endemic species that are not relevant for the area under consideration should not be evaluated with care. If a non-endemic species is particularly sensitive⁵ and therefore causes the PNEC to be significantly lower, the origin and ecological relevance of the test species should also be carefully evaluated before deciding to keep the species in or out of the assessment. For example, it could be that the species does not occur in the region under evaluation and is not representative of any taxa of concern. The species assemblage should focus on representative species for the different compartments: i.e. primary producers (higher plants & algae), primary consumers (invertebrates) and secondary consumers (fish, amphibians) for the aquatic ecosystem; bacteria and ciliates for STPs; sediment-dwelling organisms with different exposure routes, feeding habits and ecological niche for the sediment compartment; primary producers (plants), consumers (invertebrates) and decomposers (microbial mediated processes) for the terrestrial environment.
- If a risk assessment is conducted for a certain region (e.g. eco-region) than test species that are not able to thrive in the region should not be used. Species should also not be used if they have been tested under conditions that are not representative for the region and no bioavailability model is present to account for the differences in bioavailability.

2.1.2.4 Relevancy of exposure duration

- Both acute and chronic data can be used for the derivation of PNEC values. Preference should be given to the use of chronic data if available. Acute/chronic exposure depends upon the exposure duration and is also a function of the life cycle of the test organisms. *A priori* fixed exposure durations are therefore not relevant and should instead be related to the species, their typical life cycle and to the recommended exposure duration as described in standard ecotoxicity protocols (e.g. acute: 24/48 h for Daphnids (OECD n° 202), 96 h for fish (OECD n° 203); chronic exposure (e.g. 7 days for Ceriodaphnids (ASTM, 2004), 21 days for daphnids (OECD, 1998), 30 days for fish (OECD, 1992), 28-42 days for *Hyalella azteca*, 42 days for *Enchytraeus albidus* OECD, 2000, 56 days for *Eisenia foetida*, ISO, 1995). The 72h algal growth inhibition test is a chronic test but the EC₅₀ is treated as an acute value for classification purposes. Following the latest OECD requirements, relatively short-term studies, focusing on sensitive life stages rather than focusing on the full life stage are also deemed chronic studies.
- When there is a lack of chronic data it may be possible to use acute data in combination with appropriate acute to chronic ratios. Quantitative ion character-activity relationships (QICARs) or quantitative cationic-activity relationships (QCARs) could be used in the complete absence of experimental data (Owiny and Newman 2003, Walker et al. 2003.) as is the case for some data poor inorganic substances. However, more research efforts are needed in this field to develop and validate appropriate models. If no appropriate models are available the PNEC has to be derived from acute data.

⁵ if the opposite would be true, i.e. the species is particularly insensitive than the effect on the PNEC setting will be less since the best fit is focused on the lower tail of the SSD

2.1.2.5 Acclimation/adaptation

- The fact that metal/metal compounds are naturally occurring substances should be taken into account when selecting toxicity data if phenomena such as acclimatization and adaptation are of importance. These concepts are described extensively in the MERAG background document. In short, due to the ubiquitous presence of metals in the natural environment, organisms have become conditioned to these backgrounds since they have evolved in the presence of the natural metal background concentrations. For this reason, exposure of organisms to the natural background level reflects in fact the theoretical lower limit of the predicted no effect concentration (PNEC) i.e. a concentration, which from an evolutionary perspective, does not present a potential disruption of the genetic pool composition of a species. This theory is applicable for all metals and is even more crucial for essential metals (EE⁶). As a result, the sensitivity of organisms to metals is determined to a large extent by the bioavailable concentration that the organism experienced before testing and their developed capability to cope with this concentration. Moreover, organisms cultured in media with a low essential metal concentration⁷ may also exhibit an overall decreased fitness (deficiency issues) and become more sensitive to stress, including exposure to metals, even essential ones. Conversely, organisms cultured in media with elevated metal concentrations (both essential and non-essential metals, e.g. natural waters or contaminated waters) may become less sensitive. This phenomenon is related to the recently introduced “biogeochemical region” concept (Fairbrother and McLaughlin 2002).
- Ideally only those data sets where background concentrations in the culture medium (ideally both essential as non-essential metals) are similar to the clearly defined, relevant conditions of the biogeochemical -region⁸ under investigation and are also representative for natural conditions suitable for the organism under testing should be used for PNEC derivation. However, it is acknowledged that this type of information is rarely reported and hence difficult to use as a selection criteria. If the information is available (occurring especially for the major metals) the information can be used to consider not using test results where the organisms were cultured under natural background conditions that deviate from the conditions encountered in the environmental compartment under consideration. It is recommended that the essential metal concentration in the culture medium should be at least equal to the minimal concentration not causing deficiency for the test species used (lower boundary of OCEE, see background document). In case of multi-species tests (microcosm, mesocosm), the lower boundary of the No Risk Area (NRA) of the species tested could be used as the minimal concentration. Concentrations of non-essential metals should fall within the natural variation of these metals. Defining minimal levels of metal background for selection of relevant culture media should only be performed in case there is scientific evidence that acclimation/adaptation phenomena are relevant for the metal under investigation. If no direct information is available on the background concentrations of the metals in the culture medium, second line evidence (e.g. metal concentrations in river water used for maintaining the cultures could have been measured in other studies) can be used to support any decision taken on this issue.

⁶ An element is considered essential when (1) it is present in living matter; (2) it is able to interact with living systems; (3) a deficiency results in a reduction of a biological function, preventable or reversible by physiological amounts of the element (Mertz, 1974)

⁷ This is especially the case in artificial media, since these media contain no or very little (essential) micronutrients.

⁸ The biogeochemical-region approach arises from the fact that different eco-regions can be identified based upon climatic factors, latitude and elevation. Within eco-regions, sub-eco-region (also called biogeochemical-regions) can be differentiated based upon the natural background concentration (see also Reimann and Garret, 2005 on the means to determine background concentrations) of the metal under consideration and the presence of well-defined abiotic factors that influence metal bioavailability.

2.1.3 Conclusion

Only ecotoxicity data that comply with the above-mentioned criteria can be considered valid and may be used for risk assessment purposes for metals/metal compounds. However, the proposed quality criteria could be used in a flexible manner using expert judgment⁹. Main acceptability and relevancy criteria are a clear concentration-relationship, measured test concentrations, proper statistics, acceptable test performance and representativeness for the environmental compartment under investigation. For transparency reasons, it should be clearly documented which studies are being rejected and on what ground.

⁹ For example information on metal concentrations in the culture medium will most often not be available. In those cases the toxicity data of studies lacking this information could still be used when no other information is available.

3. DERIVATION OF THE PREDICTED NO EFFECT CONCENTRATION

Typical compartments that are considered for the inland environment are the aquatic (including the sediments), sewage treatment plants (STPs) and terrestrial ecosystems. For each of these compartments a Predicted No-Effect Concentration (PNEC) needs to be derived. This PNEC is considered the concentration below which unacceptable effects are unlikely to occur.

3.1 Aggregation/selection of L(E)C₅₀ /NOEC data

For data rich substances such as metals/metal compounds multiple data points can be available from reliable studies for a given species. These results will be subject to variability from several sources such as differences in geochemical characteristics of the test media, which can affect metal speciation and bioavailability, inter- and intra- laboratory variability, as well as inherent intra-specific heterogeneity in test organism sensitivity.

The most straightforward way to handle situations in which multiple data points exist for a given test species / endpoint, is to use the lowest value, e.g., the lowest NOEC/E(L)C_x. The use of the lowest value provides a conservative approach, especially when a wide range occurs between the lowest and highest data points for a given species. However, it should be realized that some of the lower toxicity values reported in literature may be the results of poor organism health, operational conditions or may just reflect differences in abiotic test conditions (bioavailability), and may therefore not reflect the intrinsic sensitivity of the organisms to a given toxicant.

When it is apparent from the data that the observed differences in test results for one species are due to differences in bioavailability in the test media then the use of the lowest toxicity value should be avoided whenever possible, and data aggregation approaches (grouping) should be used instead. In these approaches data are aggregated into geometric mean NOECs/EC₁₀ values when multiple data are available from the same species, test duration and endpoints¹⁰.

Prior to appropriate aggregation (grouping), it is recommended that intrinsic information relative to the effects of specific metals, e.g. mechanisms of toxicity and factors affecting bioavailability, be taken into account when aggregating data from multiple tests. Bioavailability differences should be taken into account by normalizing the data, prior to further processing, according to the best level of scientific knowledge available (e.g. organic carbon normalization, hardness correction, bioavailability models).

In case bioavailability models are available (e.g. Biotic Ligand Models) the scope of the data gathering can be broadened (provided that the models are validated over a broader range of conditions). Guidance on the principles and the way how bioavailability can be incorporated are given in separate fact sheets (fact sheet 5 and fact sheet 6) dealing with this issue.

When it is apparent from the data that the observed intra- species variability in toxicity test results can be assigned to differences in bioavailability and no bioavailability model is available to normalize the data, the effect data should be grouped by similar ranges of abiotic factors that control the bioavailability of metals. This grouping should preferentially be conducted such that it reflects the range of abiotic factors encountered in the region under evaluation (e.g. soft water scenario). It is recommended to define region-specific boundaries of these physico-chemical parameters for the selection of relevant test media. In this regard both natural and artificial test media are acceptable, provided that major physico-chemical characteristics that alter the bioavailability of the metal (i.e. pH, major cations, anions, DOC for the water compartment, pH/CEC, pore water dissolved anions, cations and DOC for the soil compartment or AVS and information on organic content for the sediment compartment) are similar to the range of the physico-chemical conditions encountered in the waters, soils or sediments under investigation.

If acclimation/adaptation is important test results should be grouped on the basis of the similarity of the background in the culture medium with the background of the environment under evaluation. This type of

¹⁰ The Technical Guidance Document (2003) states that: "For equivalent data on the same endpoint and species, the geometric mean should be used as the input for the calculation." In technical discussions, the meaning of "equivalent data" has been clarified to mean data collected from tests conducted under similar physical and geochemical conditions.

grouping should be done in cases where the biogeochemical region concept is relevant and can be applied. In this concept it is recognized that background concentrations of a metal in a given region can differ between ecosystems, resulting in different sensitivities to the toxic effects of metals due to acclimation or adaptation. In this way effect datasets can be divided into different biogeochemical region groups. Typical biogeochemical regions for the metal under consideration should be based upon clearly distinguished ranges of natural background concentrations and PNECs should be derived for each of these biogeochemical regions.

Summary grouping rules of selected data

In general, the following grouping rules can be applied:

- If for one species more than one L(E)C₅₀/chronic NOEC values based on the same toxicological endpoint are available, these values are averaged by calculating the geometric mean, resulting in the “species mean” NOEC/L(E)C₅₀. In case of a flawed dataset: e.g. only two data points are available and one represents a very low value and another a high value it is recommended to repeat testing and take the geometric mean of all data.
- If for one species several acute L(E)C₅₀/chronic NOEC values based on different toxicological endpoints are available, the lowest value is selected. The lowest value is again determined on the basis of the geometric mean if more than one value for the same endpoint is available.
- In some cases, L(E)C₅₀/NOEC values for different life stages of a specific organism are reported in the same study. If from these data it becomes evident that a distinct life stage is more sensitive, the result for the most sensitive life stage is selected. The life stage of the organisms is to be indicated in the tables as the life stage at the start of the test (e.g. fish: yearlings) or as the life stage(s) during the test (e.g. eggs → larvae, which is a test including both the egg and larval stages).
- In case different endpoints are given it is recommended to use the most sensitive endpoint.
- If acclimation/adaptation is important test results should be grouped on the basis of the similarity of the background in the culture medium with the background of the environment under evaluation.
- When it is apparent from the data that the observed intra- species variability in toxicity test results can be assigned to differences in bioavailability and no bioavailability model is available to normalize the data, the effect data should be grouped by similar ranges of abiotic factors that control the bioavailability of metals. The grouping should preferentially be conducted such that they reflect the range of abiotic factors encountered in the region under evaluation (e.g. soft water scenario).

3.2 Approach for the derivation of the Predicted No Effect Concentration (PNEC)

3.2.1 Introduction

PNEC values have to be derived for different environmental compartments: aquatic, sediment, terrestrial, microbial activity in STPs. Depending on the size of the database these PNEC values are calculated from the available data:

- When limited number of data are available (data poor substances), PNEC setting is based on the use of assessment factors reflecting the degree of uncertainty in extrapolating from laboratory toxicity test data for a limited number of species,
- If sufficient ecotoxicological data are available (data rich substances), the use of a statistical extrapolation method is recommended.

In any case the relevance of the PNEC derived using the assessment factor approach should be evaluated. For example a reality check should be conducted to evaluate if the PNEC is below or above the natural background of the metal under consideration or in case of essential metals if the PNEC is not situated in the deficiency levels.. If the PNEC is below or close to background/essentiality levels due to use of assessment factors there is a need for additional data testing.

A pragmatic way to deal with different backgrounds is to use a PNEC add. This approach has been further developed in annex 1 of fact sheet 3.

3.2.2 Calculation of PNEC using assessment factors (data poor substances)

For some metals/metal compounds the amount of data available for predicting ecosystem effects will be limited. In these circumstances either additional tests are performed in order to fulfill the requirements to use the statistical extrapolation method (section 3.2.3) or empirically derived assessment factors must be used. It should be recognized that these factors do not have a strong scientific validity and have been rather used as rule of thumb.

Typically, PNEC values are calculated from the lowest acute LC_{50} or EC_{50} or, preferably, from the lowest chronic $NOEC/L(E)C_x$, plus the application of an assessment factors that depend on the amount of toxicity data available. It has never been recommended to use fixed "al-purpose" assessment factors. Rather they should be tempered with whatever information is available in a given situation (Environment Canada, 1999). In general the size of the applied assessment factor will decrease as confidence in the data set increases. The requirements to be fulfilled for the different environmental compartments may differ. Several sets of assessment factors have been proposed to date. Some examples of currently used assessment factors are given in Annex 1 and 2.

Thus lower assessment factors will be used with larger and more relevant data sets (e.g. data available for a number of trophic levels, different feeding strategies etc.). In most frameworks this is, however, not infinite. If the data set already fulfill the requirements in order to use the lowest assessment factor extending the data set will lead only to the derivation of a lower PNEC/EQS because the likelihood that one may identify more sensitive species will increase while the same assessment factor (i.e. same uncertainty) is applied typically to the lowest value/geometric mean in the dataset. Logically, one would expect a further decrease in uncertainty as more information is collected.

3.2.3 Calculation of PNEC using statistical extrapolation methods (Data rich substances)

When a large data set for different taxonomic groups is available, the PNEC can be calculated using the statistical extrapolation method in which the susceptibility of a set of species for a given toxicant can be described by some statistical distribution (i.e. Species Sensitivity distribution or SSD). A SSD can be visualized as a cumulative distribution function (Figure 1). The cumulative distribution function curve follows the distribution of the sensitivity data obtained from ecotoxicological testing, plotting effect concentrations derived from acute or chronic toxicity tests, for example LC_{50} values and No Observed Effect Concentrations (NOECs), respectively.

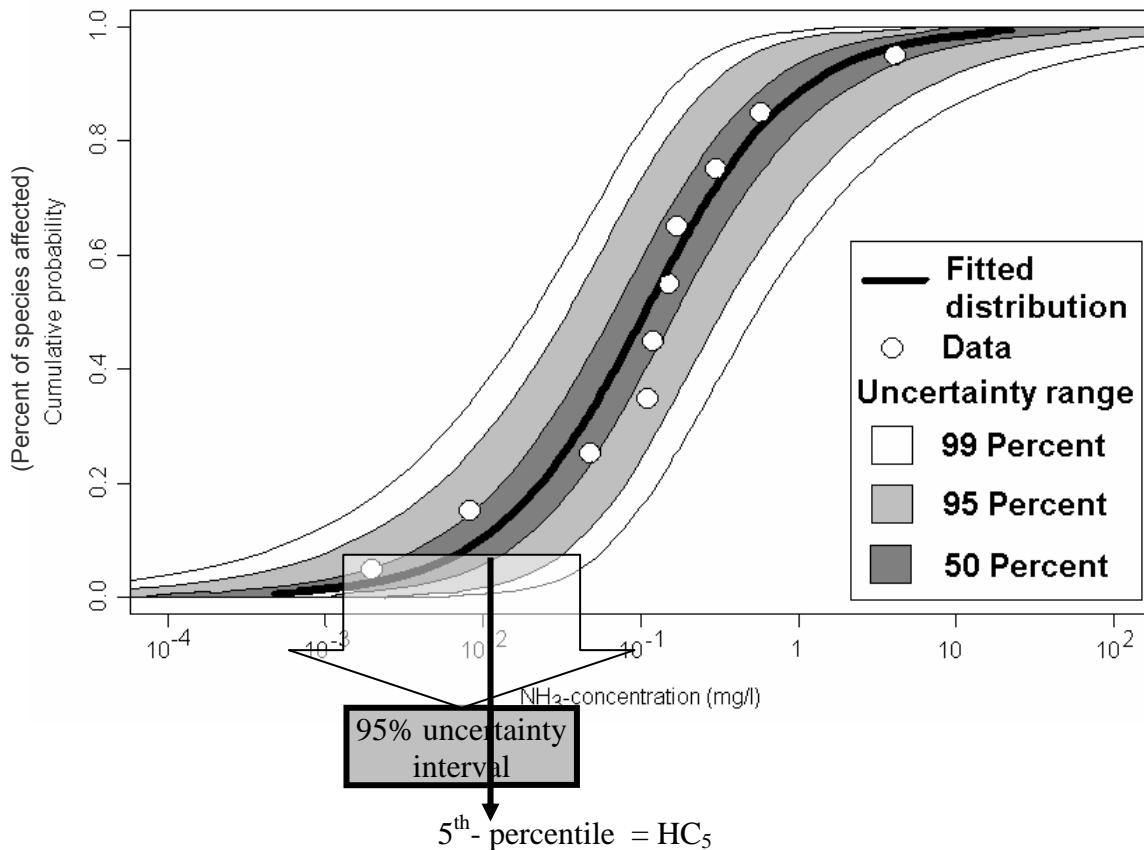


Figure 1: Example of a SSD (Species Sensitivity Distribution - loglogistic distribution) with uncertainty band and its HC₅ (Hazardous Concentration at 5 %)

SSDs were originally proposed to derive environmental quality standards in the late 1970s and mid-1980s in the United States and Europe, respectively, and their importance in ecotoxicity evaluations has steadily grown (Van Straalen and van Leeuwen, 2002 and Suter II, 2002). A more recent application is the use of SSDs in environmental risk assessments (ERA). The most common current approach is to derive the Predicted No Effect Concentration (PNEC) from the 5th percentile of SSD (EU-TGD, 1995) as shown in Figure 1. Historically that value is known as Hazardous Concentration at p-protection level or HC_p. A cut-off percentage p is chosen (to protect 1-p percent of species), and the desired “safe” concentration HC_p is calculated. The 5th percentile of a chronic toxicity distribution has been chosen in the earliest methods as a concentration that is protective for most species in a community (namely 1-p %), but the value of p is a policy decision, not science. In popular use of the method, the complementary value of p has become known as the 95% (100-p) protection criterion. Researchers also started to determine a confidence or uncertainty interval on the HC₅ (Figure 1). This was mainly done because the median HC₅ is a conservative estimate of the HC₅ calculated without uncertainty (Aldenberg & Slob, 1993). Note that Aldenberg & Jaworska (2000) extended the calculation of uncertainty to both HC_p and p at a given concentration. A confidence or uncertainty interval can quantify the sampling error in the HC₅ estimate.

Intensive discussions have taken place on principles, statistics, assumptions, data limitations, and the use of SSDs in the field of risk assessments (Posthuma et al, 2002, Forbes and Calow, 2002). Since an SSD is populated by a variety of chronic endpoints it has been questioned if using solely an SSD for establishing an acceptable risk level is appropriate in a risk assessment context where the main aim is protecting ecosystem structure and function. The ambiguity in the relationship of SSDs to chronic assessment endpoints is due in part to the lack of guidance from regulatory agencies since the conventional chronic endpoints represent rather thresholds for statistical significance and generally lack

biological interpretation. These concerns refer, however, to the use of chronic endpoints in general and pertain also to other approaches such as the assessment factor approach.

Most critiques have tended to focus on the technical aspects of constructing an SSD avoiding the inferential difficulties of defining and estimating assessment endpoints. For example the choice of distribution functions, dependence of SSDs on the amount and quality of available data and the choice of protection level has been particularly debated.

Despite the criticisms, the SSD concept is currently used in an array of decision-making processes. Improvements in SSD design and uses can be traced to three important topics (Posthuma et al., 2002). First, the specificity of SSDs can be improved by tailoring the fundamental design of the analysis to the problem using ecological, toxicological and environmental chemical information. This includes ecological aspects regarding test endpoint, identifying and correcting for nonrandom species selection, correcting for lack of independence between toxicity data, ecological aspects regarding auto-ecological features of species and biogeography and statistical aspects. Second, SSD techniques and associated techniques that are applied in the assessment can yield improved assessment accuracy, by using concepts from environmental chemistry, ecology, toxicology, biogeography, and taxonomy. Third, the fundamental statistical techniques can be improved, and regularities in toxicity databases can be used to address the problems of small sample size. The evolution of SSDs along these lines may limit the relevance of criticisms for certain SSD approaches.

3.2.3.1 Choice of the appropriate distribution model

Numerous methods have been proposed for developing species sensitivity distributions (SSDs) and there is no consensus on the most appropriate method. One of the key aspects is the selection of an appropriate distribution model. Many users of SSDs simply employ a standard distribution such as the log-logistic and the lognormal distribution because these have been historically used. In selecting these functions statistical arguments have been used more frequently than ecological arguments. Aldenberg and Slob (1993) chose the logistic function based on its inherent properties rather than their fit to data. The model is more conservative than the normal distribution (generates lower HC5 values) and is computationally tractable (Posthuma et al, 2002).

Newman et al. (2000) evaluated 30 published toxicity data sets and found that the null hypothesis of a log-normal distribution was rejected ($\alpha = 0.05$) in one-half of the data sets according to the Shapiro-Wilk's test. The authors cautioned that the defensibility of the SSD approach may be compromised if a fundamental assumption behind the approach is frequently violated (i.e., assumption that the data are log-normally distributed). In cases where the data are not log-normally distributed, the use of an SSD based on a log-normal approach would not be defensible and can lead to SSDs that badly fit the data and hence cast doubt on the appropriateness of the method.

Considering that different chemicals have different mechanisms of toxic action and that different organisms react differently to the same chemical there is a need for flexibility in choosing the appropriate distribution type for developing an SSD. Experts in the field of SSDs recommends against the *a priori* selection of a specific statistical model for an SSD and, rather, suggests that there is considerable latitude for developing appropriate SSDs for various applications (Posthuma et al. 2001).

It is preferable to select functions based in goodness-of-fit or other statistical comparisons of alternative functions. Goodness-of-fit tests (e.g. Anderson-Darling and Kolmogorov-Smirnov tests) are formal statistical tests of the hypothesis that the data represent an independent sample from an assumed distribution. These tests involve a comparison between the actual data and the theoretical distribution under consideration. The calculated goodness-of-fit statistic measures how good the fit is: critical values are calculated and used in order to determine whether a fitted distribution should be accepted or rejected at a specific level of confidence. Typically, these values depend on the type of distribution fit, the number of data points and the confidence interval. The level at which one distinguishes between likely and unlikely values of the test statistic is a matter of judgement. A significance level of 0.05 is most often used, implying that a value of the test statistic below the 95th percentile of the distribution for the statistic is acceptable and leads to the inability to reject the hypothesis. A value of the calculated A-D/K-S statistic above the 95th percentile of the distribution leads to the rejection of the null hypothesis, i.e. the distribution

is not a good fit (Cullen & Frey, 1999). In case of lack of fit at the 95% confidence level, the statistical extrapolation method should not be used.

The (A-D) test places most emphasis on tail values whereas the (K-S) test investigates the data fit for the whole distribution curve to the same extent. Care must be taken when evaluating results of best-fit analyses, since one goodness of fit test statistic (e.g. A-D) may indicate that one distribution offers the best fit, while another goodness of fit test statistic (e.g. K-S) may indicate that a different distribution has the best fit. This can influence the choice of the distribution, and also the derivation of the HC₅. Anyway it is recommended that SSD functions should not be too complex (2-3 parameters functions are preferred)¹¹.

For the purpose of deriving the HC₅ estimate (i.e. hazardous concentration at which above 95 % of the species is protected) to be used in a risk assessment context, preference could be given to the outcome of the A-D test because it places more emphasis on tail values and as such reduces uncertainty in this estimate. If the whole SSD is used as in the case of estimating the "Potentially Affected Fraction" it could be argued from an ecological viewpoint that the K-S test statistic is equally important. Anyway, the left tail of the distribution should always be analyzed carefully. If a subgroup of species can be identified as particularly sensitive the role should be assessed of this species in terms of their function in the ecosystem.

Next to statistical and ecological arguments knowledge of the chemical may also guide the choice of model. For example from a conceptual viewpoint the use of threshold models can be considered in the case of natural elements such as metals. Indeed assessment of metal SSDs requires consideration of several unique aspects, such as background concentrations, which organisms have evolved with, and essentiality for normal metabolic functions. Metals taken up by active transport have a threshold metal concentration below which the organism cannot uptake the metal from the environment. Accordingly, the *a priori* use of a model such as the normal (or log-normal) distribution, with tails extending to infinity, may result in unrealistically low HC₅ estimates that are within the range of typical background concentrations or, in the case of essential metals, potentially HC₅ estimates that may lie within the range of metal deficiency for some organisms.

Distribution types largely driven by central tendency and variability in toxicity values, such as the mean and standard deviation for the normal distribution, can also result in very long SSD tails if the toxicity data are highly variable. This is also an important issue for metals because many metals are regulated by aquatic organisms differently. Some organisms are highly tolerant because they can store the metals in non-toxic forms and other organisms are very sensitive because they do not have the same detoxifying mechanisms, thereby resulting in a range in toxicity values for aquatic organisms that can be very large. Practical assessment of the resulting elongated lower tail is further confounded by the fact that many metals tend to exhibit a threshold response in the lower tail, as discussed above. For this reason, it has been suggested that a threshold model for SSD development may be more appropriate for metals in general, and essential elements in particular (Brix et al. 2001; Van Straalen 2002; Van Sprang et al. 2005). Van Straalen (2002), for example, found that the triangular distribution provided the best fit of four finite distributions fit to zinc toxicity data, while Brix et al. (2001) and Van Sprang et al. (2005) used a Pareto model to characterize the threshold response observed in chronic copper and zinc toxicity data, respectively.

Other factors could also constrain the choice of a distribution. For example, as unrealistic values (e.g. NOEC values above the solubility product of the considered metal) may bias the estimation of the threshold value, truncating the tails of a distribution should be considered. In all cases, it is essential to explain clearly and fully the reasoning underlying the choice of a specific distribution.

¹¹ In statistics, overfitting is fitting a statistical model that has too many parameters. An absurd and false model may fit perfectly if the model has enough complexity by comparison to the amount of data available. A perfect fit can therefore always be obtained by using for example a high degree polynomial distribution. However, one should not forget that the NOECs in a SSD represent only a small sample of all sensitivities encountered in an ecosystem and as such the true distribution of sensitivities will always be unknown.

3.2.3.2 Specific requirements to be fulfilled before using the statistical extrapolation method.

3.2.3.2.1 Number of data

The number of data to construct SSDs on may vary widely, between a few data ($n > 3$) to more than 50 or 100 sensitivity values (for data rich metals). An appropriate question to consider while evaluating the data set as a candidate for the statistical extrapolation approach is 'how many data are needed?' to fit a Species Sensitivity Distribution (SSD) model with sufficient confidence using all available acute/chronic NOEC values as input.

Generally, the larger the sample size, the greater one's confidence in the choice of a probability distribution and the corresponding estimates. Conversely, for small sample sizes, goodness-of-fit statistics will often fail to reject any of the hypothesized probability distribution function. In general, there is no rule of thumb for the minimum sample size needed to specify a SSD. Increasing sample size may however be an important consideration when making decisions about uncertainty (US EPA, 1999). Nonparametric approaches for estimating 5th percentiles can be unreliable at small sample sizes ($N < 10$) because assumptions on extrapolating below the smallest data point are difficult to make.

A sufficient number of data are needed as a prerequisite for selecting a distribution function with adequate confidence but this is not the only requirement that has to be fulfilled in order to use the SSD approach properly. First, under the intrinsic assumption that the available ecotoxicological data are an independent and identically distributed sample of the real SSD, the available data should be representative in terms of ecological relevance (e.g. include key species), and include the appropriate number of taxonomic groups and trophic levels. Secondly, the more entry points (number of NOECs or L(E)C₅₀ values), the more precise the estimation of the HC₅ will be. Furthermore data quality is equally important as data quantity. Adding more species to a SSD will have no or little impact on the spread of the SSD if there is a representative sample of species. However, by adding more species, the uncertainty on the SSD mean 5th percentile will decrease.

With regard to the minimum species requirements when using the SSD approach for the aquatic compartment the London workshop (2001) formulated some recommendations. The SSD should cover at least 8 taxonomic groups containing at least 10 NOECs (preferably more than 15) for different species (Table 3). In reality for some metal/metal compounds and in particular for the sediment and soil compartment it will be difficult to obtain 10 NOEC data. In those cases an SSD could still be constructed as long as the associated sampling uncertainty with the HC₅ estimate is properly quantified¹². Aldenberg and Jaworska (2000) provide a simple lookup-table from which a confidence interval of the HC5 can be easily derived for a lognormal distribution. In essence, the HC5 (with alfa % confidence) can be derived using following formulae (note that the mean and standard deviation refer to the log-transformed data):

$$HC5 = 10^{(\text{mean} - K_{\text{alfa}} * \text{standard deviation})}$$

The K-values are tabulated according to the sample size and the level of confidence (alfa %) in Aldenberg and Jaworska (2000).

The 90% uncertainty bound decreases with increasing number of data points as depicted in Figure 2 for a hypothetical but realistic situation

¹² See Aldenberg & Jaworska (2000) for an example to quantify uncertainty.

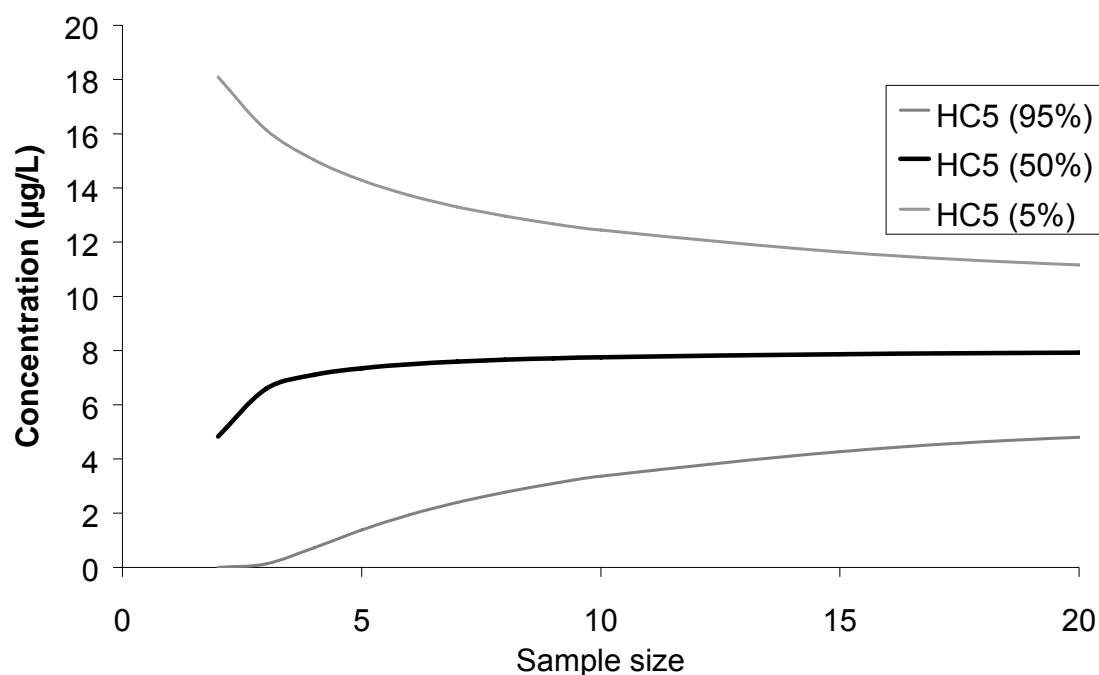


Figure 2: Relationship between the 90 % uncertainty band of the HC₅ estimate and the sample size. Hypothetical case (Lognormal distribution, mean = 1.4, SD = 0.3). (Ref to be completed)

The minimum group requirements to be fulfilled according to the London workshop for the aquatic compartment are summarized in Table 3. Deviations from these recommendations can be made on a case-by-case basis through consideration of sensitive endpoints, sensitive species, mode of toxic action and/or knowledge from structure-activity considerations.

Table 3: Minimum taxonomic group requirements for the derivation of the PNEC water (freshwater) using the statistical extrapolation technique

	Taxonomic groups
1	Fish (usually tested species like salmon, bluegill, channel catfish, etc.)
2	A second family in the phylum Chordata (fish, amphibian, etc.)
3	A crustacean (e.g. cladoceran, copepod, ostracod, isopod, amphipod, crayfish etc.)
4	<i>An insect (e.g. mayfly, dragonfly, damselfly, stonefly, caddisfly, mosquito, midge, etc.)</i>
5	A family in a phylum other than Arthropoda or Chordata (e.g. Rotifera, Annelida, Mollusca, etc.)
6	A family in any order of insect or any phylum not already represented
7	Algae
8	Higher plants
	Recommendation
	Ideally the SSD should cover at least 8 taxonomic groups containing at least 10 NOECs (preferably more than 15) for different species (London workshop, 2001).

Guidance on the minimum number of taxonomic groups needed to apply the statistical extrapolation technique for the terrestrial and sediment compartment was not part of the London workshop. For these compartments and sediment in particular, a more pragmatic approach should be developed based upon

expert judgement since the number of taxa for which internationally accepted test protocols are available is limited.

Some recommendations on taxonomic groups that could be covered for the sediment (freshwater) and soil compartment is given in Table 4. While for the aquatic compartment the minimum taxonomic group requirements have already been discussed and a common consensus have been reached on the applicability/usability of the criteria the validation and discussion of the minimum number of species needed for an SSD in sediments and soil is still needed. In this discussion biodiversity and representativeness of species for both compartments needs to be taken into account.

Table 4: Recommendations for taxonomic group to be included for the derivation of the PNEC soil and sediment using the statistical extrapolation technique

	Sediment (freshwater)	Soil
	Taxonomic groups	Taxonomic groups
1	A crustacean (e.g. amphipods, copepods etc.)	Microbe mediated processes (e.g. respiration, denitrification, N- mineralization etc.)
2	An insect (e.g. diptera.)	An insect (e.g. collembola)
3	An oligochaete (e.g. tubificidae, Lumbriculidae)	An oligochaete (e.g. eisenia, enchytreus)
4	A family in any order of insect or any phylum not already presented	A family in any order of oligochaete or any phylum not already presented
5		Higher plants (monocotyle)
6		Higher plants (dicotyle)
	Recommendations	Recommendations
	The SSD should cover three different taxonomic groups (minimum) listed above and contain at least 4 NOEC values for different species. It is also recommended that different living and feeding strategies: (1) surface deposit and/or filter feeders; (2) sub-surface feeders and (3) burrowing species with a combined surface and subsurface feeding behavior should be represented.	The SSD should cover at least three different taxonomic groups listed above, one of which is a plant and contain at least 4 NOEC values for different species.

For soils both the tests on terrestrial species (plants and invertebrates) and the tests on microbial functions can be used to derive the SSD for the terrestrial compartment. According to the TGD (2003) data on microbe mediated processes and single species tests should be considered separately due to fundamental differences between these tests (functional vs. structural test, multi-species vs. single species, adapted indigenous microbe community vs. laboratory test species, variability of test design and different endpoints, etc.). The basis of the SSD approach is, on the other hand, the inter-species difference, which makes splitting the terrestrial data set redundant. A split of the dataset may, however, be required depending on the mode of action making certain species very vulnerable (e.g. herbicides/pesticides). If it can be shown that the different types of tests have a similar sensitivity for the metal of concern no split is required and the data from plants/invertebrates/microbial functions should be pooled.

With regard to microbial multiple-species (function) test results it is unclear if they can be considered to yield a result as if it were a single-species test, namely they yield a single NOEC for each test. In favor of this approach is that each tested community is unique, like each species in the structure-based approach. So, a range of such tests yields a range of sensitivities of communities, especially regarding functions that can be taken up as individual points in the SSD.

For sediments the SSD could encompass a range of taxa covering different potential exposure routes (e.g. pore water vs ingesting sediment particles) and living strategies. With regard to feeding strategy a distinction can be made between: (1) surface deposit and filter feeders (e.g. *chironomids*) ; (2) burrowing sub-surface feeders (e.g. *oligochaetes*) and (3) burrowing species with combined surface and subsurface feeding behavior (e.g. *Hexagenia/Hyalella azteca*).

An additional recommendation for all compartments is that all individual NOEC data from one trophic level should not be below the HC₅ estimate. If it appears that all such NOEC values are lower than the HC₅, then this could be an indication that a particularly sensitive group exists, implying that some of the underlying assumptions for applying the statistical extrapolation method may not be met. In this respect, bioavailability considerations can also be important.

3.2.3.2 Data treatment

Where multiple reliable data are available for one species/endpoint and it is apparent from the data that the observed difference in test results for one species is due to differences in bioavailability it is recommended to process (e.g. geometric mean, normalization) prior to the SSD fitting (see section 3.1). If no bioavailability tools have yet been developed for the metal/metal compound under consideration, a pre-selection should be performed in relation to realistic environmental parameters for the compartment under investigation.

As such the potential bias introduced through the over-representation of ecotoxicological data from one particular species, included in the SSD without further processing, can be avoided¹³. For a risk assessment the focus should be on interspecies variability and not on intraspecies variability.

Another approach, that can be applied in order to limit the impact of only a few species on the outcome of the entire SSD, is a weighted analysis that takes into account redundant data for each species, so that all data are used, and intra-species variation is taken into account, but no species is given more importance than another, and all data points contribute evenly to the SSD. It is a more complex approach and so far there is, however, no or little experience with this approach under a regulatory framework.

3.2.3.3 Estimating the chronic effects distribution from the acute effects distribution

In the case of paucity of chronic data for a metal/metal compound it could be worthwhile to explore the use of acute data to estimate chronic toxicity using an appropriate acute to chronic ratio (ACR). An acute to chronic ratio is calculated by dividing a species' acute mean LC₅₀ by the geometric mean of the chronic no-observed effect concentration and lowest-observed-effect concentrations of the same species (Stephan et al, 1985). Preferably the acute and chronic values used to derive an ACR should be calculated based on data from side-by-side experiments, or at a minimum from studies using the test media with similar characteristics. The data used for calculating the ACR should be evaluated carefully with regard to reliability and relevancy.

Typically, a constant ACR is used but it has been observed for some metals (e.g. copper) that the ACR tends to increase or decrease in relationship to the species' mean acute value (SMAV) (Brix et al, 2001). If this is the case it is recommended that this relationship be quantified and a variable ACR used. This will improve the translation from acute to chronic values which than can be used to expand the SSD. Recently, a new methodology have been proposed based on the relationship between the weighted mean and standard deviation statistics of both the acute and chronic SSD (Duboudin et al, 2004). However, this concept still needs more testing and validation before its use can be recommended.

¹³ Generally, internationally acknowledged test species such as *Daphnia magna*, *Pimephales promelas*, or *Raphidocelis subcapitata* generate a large amount of toxicity data, and can thus have much greater importance in the SSDs if all data points are included without further processing

3.2.3.4 Uncertainty management

Once a HC5 is chosen it is recommended to look at the remaining uncertainty. If deemed appropriate an additional assessment factors on the HC5 value could be applied. Typically in defining assessment factors the size of an assessment factor depends on the confidence with which a $PNEC_{water}$ can be derived from the available data. This confidence increases if data are available on the toxicity to organisms at a number of trophic levels, taxonomic groups and with lifestyles representing various feeding strategies. Thus lower assessment factors can be used with larger and more relevant datasets than a base-set data." Thus, among other factors, the size of the assessment factor is mainly driven by the number of species (covering sufficient taxonomic groups). A way forward to objectively evaluate and if needed define an assessment factor to be applied on an HC5 is looking at the decrease in confidence interval surrounding the HC5 in function of the number of entries in an SSD. In this regard it has been suggested by Verdonck et al. (2006) to use the ratio between the HC5 (50 %) and HC5 (5 %) as a surrogate for the assessment factor to cover uncertainty due to limited number of species (assuming that several taxonomic groups are covered).

During this process care should be taken that the level of conservatism embedded in the recommended assessment factors is consistent with the relation to the number of species used which can be done by recalibrating the factor according to the widely-accepted level of conservatism already in use (e.g.. in the EU a factor of 10 for 3 NOEC data is used in the classical AF approach).

3.2.4 Effects weight-of-evidence

Once a PNEC has been derived for the different compartments, either through use of assessment factors or by using a statistical extrapolation method, the weight-of-evidence from all other available data including mesocosm/field data, use of alternative sediment effect levels (such as AET, PEL/TEL values¹⁴) etc should be evaluated in a final tier. This is especially important for the validation of assessment factors used on the lowest ecotoxicity value.

If the results of laboratory and field (model) ecosystem studies show that effects on ecosystems are unlikely to occur at the derived PNEC level, the assessment factor may be decreased. On the other hand, if it is clear that the PNEC is under-protective for the ecosystem the mesocosm value could be used.

At the moment clear quality criteria to select high quality mesocosm and/or field data are lacking. The interpretation of these studies is also difficult. According to Van Leeuwen et al. (1994) and a draft guidance document on Freshwater Lentic Field test the following quality criteria should be fulfilled:

1. A distinct concentration-effect relationship should be obtained;
2. A reliable multi-species (MS) NOEC should be derived (appropriate statistics & test concentration series);
3. Several taxonomic groups, in more or less natural ecosystems, should be exposed to test concentrations for a period longer than standard laboratory toxicity test durations (i.e. > 28d) and duration of the study should be appropriate to the life cycle of the organisms);
4. Several concentrations should be tested in each experiment, consisting of one control and at least two test concentrations;
5. Each test concentration should have by preference more than one replicate;
6. The concentration of the test compound should be measured analytically several times during the experiment; including the beginning and end;
7. Physico-chemical parameters like pH, temperature and other abiotic factors, relevant for bioavailability considerations, should be measured;
8. Apart from effect parameters like population density and biomass, effect parameters on higher integration indexes such as species diversity and species richness should also be measured.

¹⁴ AET: Apparent effect level, PEL: Probable effect level, TEL: threshold effect level

9. In addition preference should be made to effects on those species that are natural occurring in the area of interest

According to Van Leeuwen et al. (1994) a tiered approach should be followed in assigning quality criteria to these tests because otherwise most mesocosm and field studies would be classified as unreliable. Criteria 1, 2, 3, 6 and also criterion 7 can be considered as the most important ones.

Next to the use of mesocosm/field data the results of direct toxicity testing could also be evaluated. Guidance on the use of weight-of-evidence approaches will be dealt with in a separate facts sheet. Weight-of-evidence should not only be used to validate the PNEC but has also its merits to validate if identified risks are indeed occurring. If specific field or mesocosm NOECs are below the PNEC derived from single species studies, it is important to understand the observed differences through evaluation of

- Exposure system : static or flow through
- the physico-chemistry of the test media: is (are) the physico-chemistry of the mesocosm/field study (ies) inside/outside the range of physico-chemistry used for the single –species tests?
- background levels : are the field/mesocosm populations acclimated to background conditions similar to the ones used for the single species SSD?,
- dietary exposure : can dietary exposure routes explain the observed difference between the single species sensitivity and mesocosm/field NOECs?
- species considered : is the field/mesocosm NOEC driven by particular sensitive species, not used for the derivation of the singly species PNEC? Is the species naturally occurring in the area of interest?

This analysis, will allow to attach an appropriately interpretation/weight to the NOECs derived in the mesocosm/field studies (a new fact sheet on this issue is under development).

ANNEX A: USE OF ASSESSMENT FACTORS ACCORDING TO THE TECHNICAL GUIDANCE DOCUMENT (TGD, 2003)

Derivation PNEC aquatic compartment with AF method

The assessment factors used in the EU (TGD, 2003) deriving the PNEC aquatic are summarized in Table A.1. These assessment factors are generic values that can be altered under certain conditions. The use of these arbitrary assessment factors to translate laboratory test results to field conditions needs to be carefully considered as they often generate very conservative values. For example in most toxicity tests the bioavailability is increased compared to natural water because there is no or low Dissolved Organic Carbon (DOC) present, they are performed with soluble metal salts, no pre-acclimation to metal background level) and are conducted with generally recognized sensitive species. When other assessment factors are being used, clear justification should be given (i.e. weight of evidence or to avoid calculating a PNEC that would be below background).

Table A.1: Assessments factor to derive the PNEC aquatic (TGD, 2003)

Available data	Assessment factor
At least one short-term L(E)C ₅₀ from each of three trophic levels of the base set (fish, Daphnia and algae)	1,000 ^a
One long-term NOEC (either fish or Daphnia)	100 ^b
Two long-term NOECs from species representing two trophic levels (fish and/or Daphnia and/or algae)	50 ^c
Long-term NOECs from at least three species (normally fish, Daphnia and algae) representing three trophic levels	10 ^d

Notes to Table A1 (TGD, 2003):

a) The use of a factor of 1,000 on short-term toxicity data is a conservative and protective factor and is designed to ensure that substances with the potential to cause adverse effects are identified in the effects assessment. It assumes that each of the uncertainties identified above makes a significant contribution to the overall uncertainty. For any given substance there may be evidence that this is not so, or that one particular component of the uncertainty is more important than any other. In these circumstances it may be necessary to vary this factor. This variation may lead to a raised or lowered assessment factor depending on the available evidence. A factor lower than 100 should not be used in deriving a PNEC_{water} from short-term toxicity data except for substances with intermittent release. Variation from a factor of 1,000 should not be regarded as normal and should be fully supported by accompanying evidence.

b) An assessment factor of 100 applies to a single long-term NOEC (fish or Daphnia) if this NOEC was generated for the trophic level showing the lowest L(E)C₅₀ in the short-term tests. If the only available long-term NOEC is from a species (standard or non-standard organism) which does not have the lowest L(E)C₅₀ from the short-term tests, it cannot be regarded as protective of other more sensitive species using the assessment factors available. Thus the effects assessment is based on the short-term data with an assessment factor of 1000. However, the resulting PNEC based on short-term data may not be higher than the PNEC based on the long-term NOEC available. An assessment factor of 100 applies also to the lowest of two long-term NOECs covering two trophic levels when such NOECs have not been generated from that showing the lowest L(E)C₅₀ of the short-term tests. This should, however, not apply in cases where the acutely most sensitive species has an L(E)C₅₀ value lower than the lowest NOEC value. In such cases the PNEC might be derived by using an assessment factor of 100 to the lowest L(E)C₅₀ of the short-term tests.

c) An assessment factor of 50 applies to the lowest of two NOECs covering two trophic levels when such NOECs have been generated covering that level showing the lowest L(E)C₅₀ in the short-term tests. It also applies to the lowest of three NOECs covering three trophic levels when such NOECs have not been generated from that trophic level showing the lowest L(E)C₅₀ in the short-term tests. This should however not apply in cases where the acutely most sensitive species has an L(E)C₅₀ value lower than the lowest NOEC value. In such cases the PNEC might be derived by using an assessment factor of 100 to the lowest L(E)C₅₀ of the short-term tests.

d) An assessment factor of 10 will normally only be applied when long-term toxicity NOECs are available from at least three species across three trophic levels (e.g. fish, Daphnia, and algae or a non-standard organism instead of a standard organism). When examining the results of long-term toxicity studies, the PNEC_{water} should be calculated from the lowest available NOEC. Extrapolation to the ecosystem effects can be made with much greater confidence, and thus a reduction of the assessment factor to 10 is possible. This is only sufficient, however, if the species tested can be considered to represent one of the more sensitive groups. This would normally only be possible to determine if data were available on at least three species across three trophic levels. It may sometimes be possible to determine with high probability that the most sensitive species has been examined, i.e. that a further long-term NOEC from a different taxonomic group would not be lower than the data already available. In those circumstances, a factor of 10 applied to the lowest NOEC from only two species would also be appropriate. This is particularly important if the substance does not have a potential to bioaccumulate. If it is not possible to make this judgment, then an assessment factor of 50 should be applied to take into account any interspecies variation in sensitivity. A factor of 10 cannot be decreased on the basis of laboratory studies..

Derivation PNEC sediment with AF method

Only a limited number of internationally agreed sediment test guidelines are available. Most of them are short-term tests with a duration of 10 days and mortality as the endpoint. For the derivation of a sediment PNEC, results from long-term tests with sub-lethal endpoints such as reproduction, growth and emergence are regarded as most relevant due to the generally long-term exposure of benthic organisms to sediment-bound substances. Long-term sediment toxicity tests typically have a duration of 28 days or more but only a few test protocols are available (e.g. OECD 218). Since sediment dwelling organisms may be exposed through different routes (e.g. via pore water or ingesting sediment particles) the feeding regime and behavior of the species should be taken into account. The assessment factors used in deriving the PNEC sediment are summarized in Table A2.

Table A2: Assessments factor to derive the PNEC sediment

Available data	Assessment factor
At least one short-term test (LC ₅₀)	1,000
One long-term test (NOEC or EC ₁₀)	100
Two long-term tests (NOEC or EC ₁₀) with species representing different living and feeding conditions	50
Three long-term tests(NOEC or EC ₁₀) with species representing different living and feeding conditions	10

It should be noted that these assessment factor are tentative only, since a deeper understanding of the difference between short-term and long-term tests for several taxonomic groups is needed. The final PNEC should be above background.

Derivation PNEC soil with AF method

Standardized test methods exist for the soil compartment, most commonly for plants and earthworms. In addition bioassays with springtails and enchytraeids are now standardized and commonly used in Europe (OECD, 2000). Typically, a data set comprising of primary producers (plants), consumers (invertebrates) and decomposers (microbes) is preferred. For the latter group, the use of toxicity data with enzymatic endpoints should be interpreted with caution¹⁵. The assessment factors used in deriving the PNEC soil are summarized in Table A3.

Table A3: Assessments factor to derive the PNEC soil

Available data	Assessment factor
L(E)C ₅₀ short-term toxicity test(s) (e.g. plants, earthworms or micro-organisms)	1,000
NOEC for one long-term toxicity test (e.g. plants)	100
NOEC for two long-term toxicity tests representative of two trophic levels	50
NOEC for three long-term toxicity tests representative of three trophic levels	10

The final PNEC should be higher than the background.

Derivation PNEC STP with AF method

Potential effects on microbial activity in STPs are assessed using test systems such as the respiration inhibition test and the nitrification inhibition test. In general, short-term exposure using a mixed inoculum is preferred. Standard tests such as the respiration inhibition tests (e.g. OECD 209), the nitrification inhibition tests (e.g. ISO 9509) but also laboratory/pilot scale activated sludge simulation tests (e.g. OECD 303) could be used for the PNEC STP derivation. Additionally, test results with ciliated protozoa representative for the functioning of STPs should also be considered. To that purpose, growth inhibition tests with *Tetrahymena* or other representative ciliated protozoa should serve as the basis for the PNEC STP derivation. The final PNEC STP is derived from the test results obtained in the most sensitive test system available using the assessment factors as mentioned in Table A.6

¹⁵ Unless more definitive linkages are made between specific microbial enzymatic activities and an adverse condition for typical assessment endpoint species, enzymatic endpoints will continue to have limited use in risk assessments; i.e. they will not drive the process as primary assessment endpoints".

Table A.6: Assessments factor to derive the PNEC STP

Available data	Assessment factor
L(E)C ₅₀ from respiration inhibition tests inhibition tests	100
L(E)C ₅₀ from respiration inhibition tests AND/OR nitrification inhibition tests or ciliate growth inhibition tests	10
NOEC from respiration inhibition tests and/or nitrification inhibition tests or ciliate growth inhibition tests or pilot scale activated sludge simulation tests	1

ANEX B: USE OF ASSESSMENT FACTORS ACCORDING TO THE OECD MAUNAL FOR INVESTIGATION OF HPV CHEMICALS (OECD, 2004)

Within the OECD High Production Volume Chemicals Programme. A PNEC is calculated using toxicity test data such as LC50, EC50, other L(E)Cx values, NOEC (no observed effect concentration) and LOEC (low observed effect concentration). MATC (maximum allowable toxicant concentration, calculated as $MATC = (NOEC \times LOEC)/2$) is also used in effects assessment.

Assessment factors are used to adjust the effect concentration from a limited data set and to estimate a PNEC. Assessment factors should be applied with care to acute data for substances which are suspected of having a specific mode of action, or which have a high log KOW or which significantly bioaccumulate. Assessment factors should reflect the following uncertainties and extrapolations:

- intra-species and inter-species variations;
- the extrapolation of short term toxicity towards long term toxicity; and
- the extrapolation of laboratory results towards the field.

The assessment factors to be used in estimating a PNEC value from the Screening Information Data Set (SIDS) – i.e. the minimum amount of data that is required for making an initial hazard assessment of HPV chemicals which has been agreed upon by OECD are summarized in Table B1.

Table B1: Summary of proposed assessment factors for estimating an PNEC (OECD, 2004).

Case	Data available	Range of assessment factor
(a)	EC ₅₀ algae (72h) EC ₅₀ <i>Daphnia</i> (24-48h acute test) LC ₅₀ fish (96h)	100-1,000
(b)	NOEC <i>Daphnia</i> (14-21d chronic test) NOEC algae (72h) NOEC fish (chronic toxicity test)	10-100

a) In case (a), all three data should be included in the SIDS

b) In case (b), NOEC algae is a SIDS element and NOEC *Daphnia* or NOEC fish may also be included in the SIDS for certain chemicals

When only acute toxicity data in the SIDS are available, an assessment factor of between 100 and 1,000 is applied to the lowest L(E)C50 [i.e. case (a)]. A factor of 1,000 is a conservative and protective factor and applied when only limited data are available, i.e. this value may be reduced to 100 if evidence is available to suggest that this may be a more appropriate factor. Such evidence would include:

- availability of data from a wide variety of species including those which are considered to represent the most sensitive species;
- information, from structurally similar compounds or QSAR, to suggest that the acute to chronic ratio is likely to be low;
- information to suggest that the chemical acts in a non-specific or narcotic manner, with little inter-species variation in toxicity; and
- information to suggest that the release of the chemical is short-term or intermittent, and that the chemical would not be persistent in the environment.

When chronic toxicity data are available in addition to acute data, an assessment factor of between 10 and 100 is applied to the lowest NOEC [i.e. case (b)], taking the following situation into account:

- if chronic NOEC is available from one or two species representing one or two trophic levels (i.e. fish, *Daphnia* or algae), a factor of 100 or 50 is applied to the lowest NOEC. In this case, a PNEC value derived from chronic data should be compared to that derived from the lowest acute data. It is then the lowest value that is used in the assessment.
- if chronic NOECs are available from three species representing three trophic levels (i.e. fish, *Daphnia* and algae), a factor of 10 is applied to the lowest NOEC. If there is convincing evidence that the most sensitive species has been tested, a factor of 10 may also be applied to the lowest NOEC from two species representing two trophic levels (i.e. fish and/or *Daphnia* and/or algae).

Use of different assessment factors should be clearly justified in the assessment report.

REFERENCES

Aldenbergh T. & Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, 25, 48-63.

Aldenbergh T. & Jaworska J. S., 2000. Estimation of the hazardous concentration and fraction affected for normally distributed species sensitivity distributions. *Ecotoxicology and Environmental Safety*, 46, 1-18.

ASTM, 2004. Guide for conducting three brood renewal toxicity tests with *Ceriodaphnia dubia*. E1295 (11.05).

Brain, P. and Cousens, R. 1989. An equation to describe dose responses where there is a stimulation of growth at low doses. *Weed Research* 29: 93-96.

Brix K., D.K DeForest and W.J. Adams, 2001. Assessing acute and chronic copper risks to freshwater aquatic life using species sensitivity distributions for different taxonomic groups. *Environmental Toxicology and Chemistry*, 20 (8), 1846-1856.

Cedergreen, N., C. Ritz and J.C. Streibig, 2005. Improved empirical models describing hormesis. *Environmental Toxicology and Chemistry*, 24 (12), 3166-3172.

Draft Guidance Document on Simulated Freshwater Lentic Field Tests, 2003

Duboudin C., P. Ciffroy and H. Magaud, 2004. Acute-to-chronic species sensitivity distribution extrapolation. *Environmental Toxicology and Chemistry*, 23 (7), 1774-1785.

Environment Canada, 1999. Guidance document on application and interpretation of single species tests in environmental toxicology, EPS 1/RM/34.

Environment Canada, 2005. Guidance document on statistical methods for environmental toxicity tests. EPS 1/RM/46 March 2005. Method development and applications section, Environmental Technology Center, Environment Canada.

Fairbrother A and McLaughlin M. 2002. Fact Sheet on Environmental risk assessment 12: Metalloregions. International Council on Mining and Metals (ICMM), Publ. London, UK. January 2002.

GHS, 2003. Globally harmonized system of classification and labeling of chemicals (GHS). United Nations, New York and Geneva, 2003.

ISO, 1995. Soil Quality - Effects of Pollutants on Earthworms (*Eisenia fetida*) – Part 2: Determination of Effects on Reproduction, Draft International Standard, International Organization for Standardization, ISO/DIS 11268-2.2..

ISO, 2004. Working draft, Water Quality- Guidance document on the statistical analysis of ecotoxicity data, March 2004. ISO/CD20281.

Lee J.S., B.G. Lee, S.N. Luoma and H. Yoo (2004). Importance of equilibration time in the partitioning and toxicity of zinc in spiked sediment bioassays. *Environmental Toxicology and Chemistry*, 23 (1), 65-71.

London, workshop, 2001. Report of the expert consultation workshop on statistical extrapolation techniques for environmental effects assessments. London, UK, European Chemicals Bureau, 17-18 January 2001.

Newman, M.C., D.R Ownby, L.C.A. Mézin, D.C. Powell, T.R.L. Christensen, S.B. Lerberg, and B.-A. Anderson. 2000. Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient number of species. *Environ. Toxicol. Chem.* 19(2):508-515.

OECD, 1984. Guidelines for Testing of Chemicals. No. 207, Earthworm, Acute Toxicity Tests.

OECD, 1992. OECD Guidelines for Testing of Chemicals 210. Fish, Early-life stage toxicity test.

- OECD, 1998. OECD Guidelines for Testing of Chemicals. *Daphnia magna* reproduction test.
- OECD, 2000. OECD Guidelines for Testing of Chemicals. *Enchytraeid reproduction test*. Guideline No. 220. OECD, Paris, draft
- OECD (2001). OECD guidelines for the testing of chemicals proposal for a new guideline 218 sediment-water Chironomid toxicity test using spiked sediment.
- OECD (2004). Manual for investigating of HPV chemicals, September, 2004.
- Owny and Newman, 2003. *Combi. Sci.* 22: 241-246.
- Posthuma, L., T.P. Traas, D. De Zwart, and G.W. Suter II. 2001. Conceptual and technical outlook on species sensitivity distributions. Pages 475-508 in L. Posthuma, G.W. Suter II, and T.P. Traas, eds. *Species sensitivity distributions in ecotoxicology*. Lewis Publishers, Boca Raton, FL. 587 pp.
- Posthuma L., Suter II G. W. & Traas T. P., editors. 2002. *Species sensitivity distributions in ecotoxicology*. Lewis publishers.
- Reily et al, 2003. *Water Quality Criteria Revisited*. Book SETAC Pellston workshop-Chapter 3.
- Reimann C. and Garret R., 2005. *Geochemical background: concept and reality*. *Science of the Total Environment* 350: 12-27.
- Schabenberger, O., Tharp, B.E., Kells, J.J. and Penner, D., 1999. Statistical tests for hormesis and effective dosages in herbicide dose response. *Agronomy Journal* 91: 713-721.
- Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman, and W.A. Brungs. 1985. *Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses*. NTIS No. PB85-227049. EPA 440-5-85-501. EPA Environmental Research Laboratory, Duluth, MN.
- Suter II G. W. 2002. North American history of species sensitivity distributions. Pages 11-17 in L. Posthuma, G. W. Suter II & T. P. Traas, editors. *Species sensitivity distributions in ecotoxicology*. Lewis publishers, London, UK.
- TGD, 1996/2003. *Technical Guidance Document on Risk Assessment in support of Commission Directive 93/67/EEC on Risk Assessment for new notified substances, Commission Regulation (EC) No 1488/94 on Risk Assessment for existing substances and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. Part II: Environment*.
- US-EPA, 1999a. Environmental Protection Agency, 1999 update of ambient water quality criteria for ammonia, EPA-822-R99-014.
- US-EPA (1999b). *Draft Risk assessment guidance for Super Fund. Volume 3-Part A, process for conducting probabilistic risk assessment*.
- Van Leeuwen K., Emans H.-J., Van de Plassche E. & Canton H., 1994. Chapter 27: The role of field tests in hazard assessment of chemicals. P. 425-437. In: *Freshwater field tests for hazard assessment of - chemicals*. Edited by: Hill I., Heimbach F., Leeuwangh P., Matthiessen P., Lewis publishers.
- Van Ewijk, P.H. and Hoekstra, J.A. ,1993. Calculation of the EC50 and its confidence interval. *Ecotoxicology and Environmental Safety* 25: 25-32.
- Van Sprang, P.A., F.A.M. Verdonck, P.A. Vanrolleghem, M.L. Vangheluwe, and C.R. Janssen. 2004. Probabilistic environmental risk assessment of zinc in Dutch surface waters. *Environ. Toxicol. Chem.* 23(12):2993-3002.
- Van Straalen, N.M. 2002. Threshold models for species sensitivity distributions applied to aquatic risk assessment for zinc. *Environ. Toxicol. Pharmacol.* 11:167-172.

Van Straalen N. M. & van Leeuwen C. J. 2002. European history of species sensitivity distributions. Pages 19-34 *in* L. Posthuma, G. W. Suter II & T. P. Traas, editors. Species sensitivity distributions in ecotoxicology. Lewis publishers, London, UK.

Verdonck F.A.M, M.L. Vangheluwe and P.A. Van Sprang, 2006. Uncertainty management for data rich natural substances: how to balance statistical and proportionality concerns (under development).

Walker et al, 2003. Quantitative Cationic-Activity relationships for predicting toxicity of metals. *Environmental Toxicology and Chemistry* 22 (8): 1916-1935.

Wang F., Goulet R.R., and Chapman P.M. 2004. Testing sediment biological effects with the freshwater amphipod *Hyaella azteca*: The gap between laboratory and nature. [Chemosphere](#).57(11):1713-24.