For Official Use **ENV/JM/HA/RD(2009)2**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **27-Feb-2009**

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/HA/RD(2009)2
For Official Use

Task Force on Hazard Assessment

REPORT FROM WHO

26-27 March 2007, starting at 14h00 on 26 March, OECD Conference Centre, 2 rue André Pascal, 75775 Paris Cedex 16

JT03260329

English - Or. English

The present document was prepared by WHO. It provides information about on-going work relevant to the terms of reference of the Task Force, *inter alia* on the status of development of a Risk Assessment Toolkit as well as the findings of a study comparing environmental risk assessment methods.

ACTION REQUIRED: ***The Task Force is invited to note the information and provide comments as appropriate.***

INTRODUCTION

1. At past meetings of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Biocides and Biotechnology, and the Working Group on Pesticides, the World Health Organization (WHO/IPCS) has reported on its activities including those likely to be of most interest to those bodies. In the same vein, the present paper aims to provide a summary of on-going work relevant to the Terms of Reference of the Task Force on Hazard Assessment, and in particular to its task to collaborate with relevant IOMC organizations as appropriate in its work on hazard assessment methodologies.

2. A comprehensive summary of activities undertaken by WHO on Chemical Safety can be found in its Activity Report 2008 see http://www.who.int/entity/ipcs/about_ipcs/activity_report_2008%20.pdf. This includes Applied Risk Assessment (chemical assessment reports); Risk Assessment Methodology; Poisons Prevention, Information and Management; Chemical Emergencies; Children's Environmental Health; International Conventions and Agreements, and Capacity Building.

RISK ASSESSMENT OF SPECIFIC CHEMICALS

3. WHO risk assessment work on non-food related chemicals centres around the production of Concise International Chemical Assessment Documents (CICADs) and International Chemical Safety Cards (ICSC). This work is undertaken: to establish the evidence-base on priority environmental agents; to empower individuals, health care professionals and public health officials by providing authoritative information on the risks to health from environmental exposures; to strengthen and inform risk management choices to protect human health; to focus public health policies on disease prevention and on priority hazards; and to identify research needs to fill gaps in our protection of human health and the environment.

4. Current activities include:

i. A project to select and establish the current evidence base for 10 chemicals of major public health concern, update assessments where needed, identify gaps in the evidence base and estimate costs to the health system of the unaddressed problem (where feasible). Short information documents aimed at decision-makers will be developed for the chemicals, enabling the development and implementation of public health policies to address the concerns.

ii. Preparation of a number of CICADs. See:
http://www.who.int/ipcs/publications/cicad/en/index.html

5. The above-mentioned web page also includes the agreed "Rules of Thumb" for helping to avoid duplicative activity between IPCS and OECD assessment programmes.
http://www.who.int/entity/ipcs/publications/cicad/en/rules_of_thumb.pdf.

HARMONIZATION OF RISK ASSESSMENT METHODOLOGIES

6. The IPCS "*Project on the Harmonization of Approaches to the Assessment of Risk from Exposure to Chemicals*" (commonly referred to as the "Harmonization Project") aims to harmonize global approaches to risk assessment through both increased understanding and agreement on basic principles, and developing international guidance documents on specific issues. The Project has a global Steering Committee, which includes experts drawn from national risk assessment agencies, representatives of supra-

national bodies (such as the EU (ECB/JRC), EFSA, and the OECD), and representatives of non-governmental organizations in official relations with WHO and working in the field of chemical risk assessment (ECETOC and ILSI/RSI). The Steering Committee meets every 2-3 years to recommend the Project workplan.

7. Harmonization Project publications continue to be taken up across the range of assessment sectors, i.e. industrial chemicals, biocides, pesticides, veterinary products, pharmaceuticals, occupational and public health, and are known to be used by many national and supra-national risk assessment bodies, e.g. Australia, EFSA, the European Union, Canada, Japan, OECD, UNECE (GHS guidance), United Kingdom, and the United States.

8. The Project web site (http://www.who.int/ipcs/methods/harmonization/en/index.html) contains, inter alia, both the current Project workplan and Brochure http://www.who.int/entity/ipcs/methods/harmonization/brochure2007.pdf.

9. Activities undertaken in 2008 and planned for 2009 include:

Completed Activities

i. Publication of the **Framework for Analysing the Relevance of a Non-Cancer Human Mode of Action for Humans.** http://www.who.int/ipcs/methods/harmonization/areas/non_cancer/en/index.html

ii. Publication of guidance on **Data Quality in Chemical Exposure Assessment** (by the time of the OECD Task Force meeting this document will be available at: http://www.who.int/ipcs/methods/harmonization/areas/exposure/en/index.html

iii. **Mutagenicity Testing** for Chemical Risk Assessment: Update of the WHO/IPCS Harmonized Scheme was submitted for publication in a journal.

iv. The report of the 2006 Berlin **Skin Sensitization** workshop was published. http://www.who.int/ipcs/methods/harmonization/areas/sensitization/en/index.html

Activities Underway

v. Development of a **Framework to Consider Combined Exposures to Multiple Chemicals**. The draft framework will be released for public and peer review in the coming months.

vi. Development of Harmonized Guidance for **Immunotoxicity Risk Assessment** is underway. The draft guidance will be released for public and peer review in 2009.

vii. Guidance on "**Principles of Characterizing and Applying PBPK Models**" was released for public and peer review. It will be finalized in 2009. http://www.who.int/ipcs/methods/harmonization/areas/pbpk/en/index.html

viii. **Identifying important life stages for monitoring and assessing risks from exposures to environmental contaminants.** An international symposium is planned to address this topic. A workshop planning group meeting is scheduled for March 2009**.**

ix. **Characterizing and communicating uncertainty and variability in hazard assessment.** A thought-starter on this topic is being developed under the joint authorship of experts from the US EPA and RIVM, Netherlands.

x. **Guidance on Interpreting Effects that may be modest or adaptive.** The development of a thought-starter on the interpretation of effects that may be modest or adaptive commenced in 2008. The Fraunhofer Institute, Hanover, is preparing the first draft.

xi. **Dermal absorption**. IPCS contributed to an OECD led process on this topic, with a view to development of harmonized international guidance. The work, which complements the IPCS Environmental Health Criteria Document on this topic, will continue in 2009, with the assistance of the Fraunhofer Institute, Hanover.

10. Also, two **Environmental Health Criteria** in the (yellow cover) methodology series are in development or final preparation for publication, i.e.: **Dermal Exposure** (first draft being prepared by the Fraunhofer Institute, Hanover); **Dose-Response Characterization** (in press). All EHC publications are available on the IPCS web site http://www.who.int/entity/ipcs/publications/ehc/en/index.html

WHO/IPCS RISK ASSESSMENT TOOLKIT

11. Chemical risk assessment involves the application of risk assessment methodologies and tools. There are a range of international chemical risk assessment methodologies available, for example, those of WHO/IPCS and OECD. Many of these tools are available on the internet and in print. However risk assessment bodies in developing countries and countries with economies in transition do no necessarily have the resources to locate these tools and become informed about their possible applications in their own countries and in support of their obligations under international agreements on chemicals. In parallel, information on the hazards of chemicals is becoming more readily accessible, for example, the WHO INCHEM database and a number of other databases through the OECD eChemPortal.

12. The aim of this WHO/IPCS project is to: make the international tools available on chemical risk assessment more readily-accessible in the form of a "toolkit", develop priority new tools, and develop related training materials including case study examples of use of the tools. The toolkit is being prepared with developing countries in mind as an important user group. However it aims to be useful to risk assessors worldwide.

13. The draft risk assessment toolkit is being prepared following an expert meeting in March 2008 in Montreux, Switzerland. The next phase of the project will involve pilot testing in countries during 2009, to inform further development of the toolkit. The report of the Montreux Meeting can be found at: http://www.who.int/ipcs/methods/harmonization/areas/ra_toolkit/en/index.html

INTERNATIONAL COMPARISON OF METHODOLOGIES FOR ENVIRONMENTAL RISK ASSESSMENT

14. WHO/IPCS CICADs provide summaries of the relevant scientific information on the potential effects of chemicals on human health and/or the environment. A project on *'Development of Methodology on Environmental Risk Assessment for the CICADs Programme'* was undertaken as a practical step to developing further transparency and possibly harmonization of methodology in this area. This work was coordinated by Dr Torsten Hahn, Fraunhofer Institute Toxicology and Experimental Medicine (ITEM),

Hanover, Germany), in collaboration with international partners. WHO presented the initial findings of this study to the 41st Joint Meeting, and offered to present the final conclusions to the OECD at a future date.

15. The study authors have submitted their findings to a journal for publication. A summary of the study is provided at Annex 1. The study compared ecological hazard assessment methodologies for aquatic systems. Experts applied the methodologies of the OECD, Australia, EC, Canada and the US to a number of well-studied chemicals. Large variability in the derived PNECs was observed, sometimes exceeding three orders of magnitude. An important source of this variability was different choice of critical study, due to different interpretations as to whether a study was acute or chronic. These findings bear on the issue of sharing of national hazard assessments and avoiding duplication of effort, in particular developing country interpretation of existing hazard assessments.

COLLABORATION BETWEEN WHO/IPCS AND OECD

16. One of the Tasks of the Task Force on Hazard Assessment, is to collaborate with relevant IOMC organizations as appropriate in its work on hazard assessment methodologies. To date OECD and WHO/IPCS have engaged in work on chemical risk assessment in the following ways:

i. Experts from OECD countries participate in WHO/IPCS meetings and comment on documents.

ii. OECD and IPCS Secretariats participate in each other's meetings and comment on documents, as appropriate.

iii. OECD may adopt already completed WHO/IPCS documents for its own use (e.g. as occurred with the Mode of Action publication included in OECD Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies, OECD Series on Testing and Assessment No. 35, 2002).

iv. Joint activities, involving a joint development process, with both WHO and OECD adopting/agreeing the documents (e.g. IPCS/OECD Key Generic Terms used in Chemical Hazard/Risk Assessment).

v. Project workplans are shared, both in relation to risk assessment methodology and preparation of risk assessments for specific chemicals.

vi. Ad hoc activities, such as the planned WHO/OECD Training Workshop on Risk Assessment Methodology: Evaluating the human relevance of modes of action in animals and replacing default uncertainty factors with data derived values. This Training Workshop, supported by the Joint Meeting is scheduled for 5-6 October 2009, at OECD.

17. WHO contact Ms *Carolyn Vickers*, email: vickersc@who.int .

Annex 1

ARTICLE SUBMITTED FOR PUBLICATION: DO NOT CITE, QUOTE OR COPY

Development of Methodology on Environmental Risk Assessment for the CICADs Programme

Sources of variability in environmental hazard assessment of chemicals in aquatic systems: An international analysis

Torsten Hahn[1], Jenny Stauber[2], Stuart Dobson[3], Paul Howe[3], Janet Kielhorn[1], Gustav Koennecker[1], Jerry Diamond[4], Chris Lee-Steere[5], Uwe Schneider[6], Yoshio Sugaya[7], Ken Taylor[8], Rick Van Dam[9] and Inge Mangelsdorf[1]

[1]Fraunhofer ITEM, Hannover, Germany, [2]Centre for Environmental Contaminants Research ,CSIRO Land and Water, Sydney, Australia, [3]Centre for Ecology and Hydrology, Monks Wood, UK, [4]Tetra Tech Inc., USA , National Standards and Guidelines Office, Environment Canada, [5]Canberra, Australia,, [6]National Standards and Guidelines Office, Environment Canada, Canada, [7]National Institute for Environmental Studies, Ibaraki, Japan, [8]Existing Substances Division, Environment Canada, Canada, [9]Environmental Research Institute of the Supervising Scientist, Darwin, Australia.

1 Introduction

The WHO International Program on Chemical Safety (IPCS), together with the International Labour Organisation (ILO) and the United Nations Environmental Program (UNEP) regularly publishes Concise International Chemicals Assessment Documents (CICADs) on the environmental and human health impacts of individual chemicals. The aim of the CICADs is to characterise the hazard and dose-response of exposure to chemicals and to provide examples of exposure estimation and risk characterisations for application at the national or local level. The CICADS include a sample environmental risk assessment which compares the predicted exposure concentration (PEC) for that chemical in an environmental compartment (air, water, sediment or soil) with the PNEC to give a Hazard Quotient (HQ). For chemicals with HQs > 1, there may be potential toxicity and the chemical may warrant further investigation on a site-specific basis. Because these sample risk assessments are conducted and reviewed by many participants from different countries, there are considerable inconsistencies in both the way the PNECs are derived and the actual PNEC values.

Few studies have examined the factors such as different assessment frameworks and methodologies, assessor subjectivity in data interpretation and key study selection, and choice of assessment factor magnitude, which can all contribute to variability in chemicals risk assessments. Even the use of one method e.g. a species sensitivity distribution (SSD) on an individual chemical, can give rise to a 10-fold difference in the derived PNEC (Duboudin et al., 2004). These authors found that the choice of data (intraspecies variation and proportions between the taxonomic groups) had more of an effect on the SSD-derived PNEC than the statistical method used to construct the distribution. Various sources of uncertainty in the use of these probabilistic methods, including lack of knowledge, natural variability in time and space, and choice of model with differing underlying assumptions, has been discussed in Jager et al. (2001).

In an attempt to understand the sources of variability in these hazard assessment we undertook an international comparison of ecological hazard assessment methodologies for aquatic systems. The aim was

not to derive new PNECs nor to re-evaluate existing assessments, nor to critique or harmonise the different approaches, but rather to understand the different hazard assessment methodologies used internationally and their major sources of variability.

Participants from Europe, North America, Asia and Australasia independently performed assessments on five chemicals with different chemical properties for which sufficient high quality ecotoxicological information was available –, ethylene glycol, trichloroethylene, hexachlorobenzene, nonylphenol and copper. All copper freshwater data were corrected to a hardness of 30 mg $CaCO_3$/L according to ANZECC/ARMCANZ (2000). These chemicals were chosen to represent well studied chemicals with a range of physico-chemical properties and for which there were large databases of quality-checked, non-contentious ecotoxicological data. The hazard assessment methods included that of EC (2003) and OECD (2002), the Australasian Water Quality Guidelines approach (ANZECC/ARMCANZ, 2000), two US approaches - USEPA OPPT (National Science and Technology Council, 1999) and US Office of Water (Stephan et al., 1985), and two Canadian approaches - Water Quality Guidelines method, 2006 and the Canadian Existing Substances method (Environment Canada, 1997).

2 Methods

Environmental effects data for the selected chemicals were taken from one or two peer reviewed data sources, thus data quality was not further scrutinised. Nevertheless, some remarks on test design from the source documents were included into the data set tables to support the decision process in the hazard assessment. Data sources were IPCS (2000) and Greene and Kocan (1997) for EG, EU (2004) for TCE, NHW/DOE (1993) for HCB, EU (2002) and Environment Canada (2000) for NP, and Stauber and Davies (2000) and Brix et al. (2000) for copper. We are aware that many more high quality data might exist for these five substances. However, no further comprehensive literature search was done and the contributing scientists agreed not to use any other data than those provided with the data sets. It was neither the aim of this project to re-evaluate existing risk assessments, nor to use any of the resulting PNECs in a regulatory context. Rather, our aim was to describe and analyse hazard assessment variability within and between different methodologies, and the chosen substances and data sets served merely as examples.

In real assessments of chemicals, there is often only limited ecotoxicological data available. In order to examine whether data richness would lead to lowered or increased variability in hazard assessments compared to limited data availability, assessments were performed for both scenarios. For this purpose assessors applied their methodologies not only on the Full Data Set, but also to a subset of data that was compiled for each chemical from the Full Data Set. This subset was termed the 'Base Data Set' because it contained typical base data used in ecotoxicological hazard assessment, i.e. results from algal-, daphnid- and fish-tests, (i.e. three trophic levels) and included both acute and chronic data.

Both data sets were provided to all participants, who documented their PNEC derivations using both their own national methods and common methods across jurisdictions (e.g. EC, 2003). All participating scientists performed their assessments independently from each other, and the assessments using the Base Data Sets were performed before those using the Full Data Sets, in order to exclude knowledge of the complete data set which may otherwise have influenced the decision process on the Base Data Set.

3 Results and Discussion

Theoretically, 34 hazard assessments were to be conducted on each chemical, 17 using the full data set and 17 using the base data set. Not all participants performed all assessments on each chemical, e.g. because the base data set was considered insufficient for a particular approach, so that between 29 and 31 PNECs were derived for each organic chemical. For copper, 28 assessments using hardness-corrected freshwater

data and 25 assessments using marine data were carried out. Many participants used the same methodology (e.g. OECD, 2002), enabling comparisons of variability between individual assessments using the same methodology and the same data sets.

Full data sets – base data sets

Using the full data sets, there was large variability in the derived PNECs for each chemical, sometimes exceeding 3 orders of magnitude (Figure 1). It was expected that the PNECs derived from these large data sets would be higher (i.e. less conservative) than those derived from limited data (base data sets) because uncertainty related to inter-species differences is lower. However, PNECs derived from full data sets were not always consistently higher or lower than those derived from the base data sets and in most cases variability was still large.

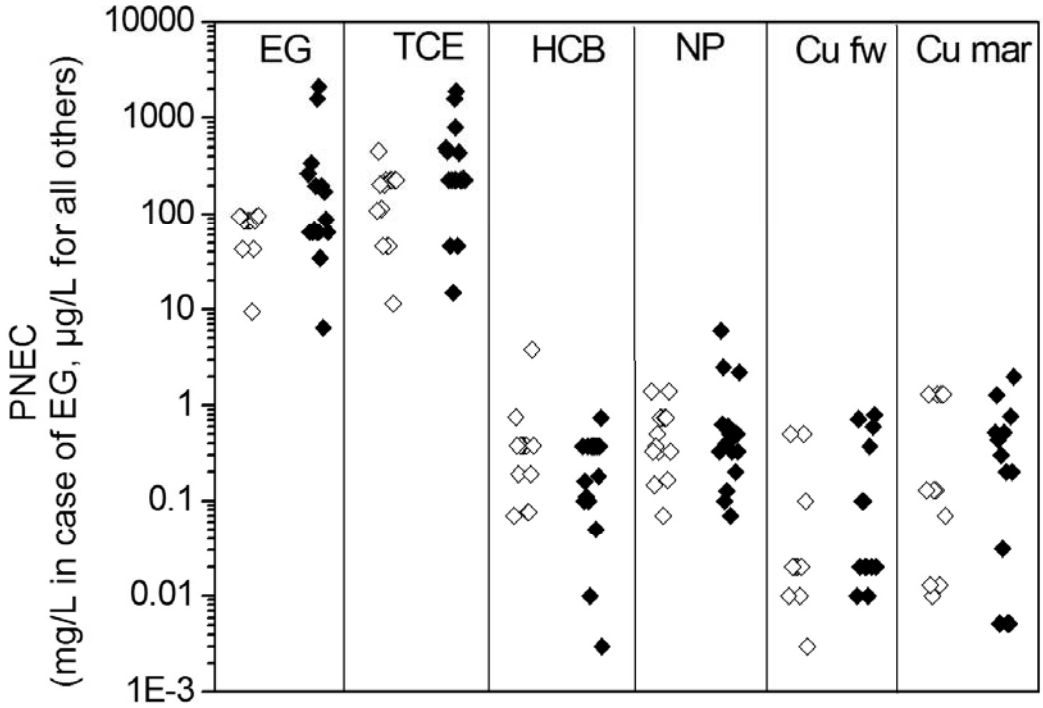


Figure 1. Comparison of PNECs derived from full data sets (closed symbols) versus base data sets (open symbols) for each of the five chemicals: EG – ethylene glycol, TCE – trichloroethylene, HCB – hexachlorobenzene, NP – nonylphenol, Cu fw – copper freshwater, Cu mar – copper marine

The range between the lowest estimate (Minimum) and highest estimate (Maximum) is at least one, and up to three, orders of magnitude (Figure 1). Despite the large variation in values of PNEC derived, medians for the base-set/full-set pairs for each chemical are very similar (Table 1), indicating clustering of estimates

around similar values. The medians also indicate little effect of using the 'full' over the 'base' set in the absolute value of PNEC derived by any of the methods.

Probabilistic approaches – deterministic approaches

Standard deviation and coefficient of variation has been calculated on log10 transforms of the data in Table 1; log transforms stabilise the variability in situations where standard deviation of the response is proportional to the mean. This approach is reasonable for summarising the random variability in a homogeneous population. However, we know that the data in Table 1 are not homogeneous since some of the methods are deterministic and some probabilistic in approach. Coefficient of variation has, therefore, been calculated both for the whole dataset and for deterministic approaches alone.

Table 1. Summary statistics of the PNEC values derived for each method and chemical.

	Chemical and dataset											
	EG base set	EG full set	TCE base set	TCE full set	HCB base set	HCB full set	NP base set	NP full set	Cu fresh base set	Cu fresh full set	Cu marine base set	Cu marine full set
Median	85.9	78.8	208	230	0.38	0.37	0.44	0.48	0.02	0.1	0.13	0.3
Min	9.5	6.5	11.5	11.5	0.07	0.003	0.07	0.07	0.003	0.01	0.0052	0.0052
Max	97.5	2096	460	460	0.76	0.74	1.4	6	0.5	0.9	2	2
Max:Min	10.3	322.5	40	40	10.9	246.7	20	85.7	166.7	90	383.6	384.6
SD (all)	0.283	0.610	0.431	0.529	0.324	0.627	0.373	0.494	0.651	0.741	0.814	0.932
CV	0.65	1.40	0.99	1.22	0.74	1.44	0.86	1.14	1.50	1.71	1.87	2.15
SD (det)	0.295	0.444	0.338	0.435	0.347	0.694	0.383	0.496	0.375	0.630	0.754	0.960
CV	0.69	1.02	0.78	1.00	0.80	1.60	0.88	1.14	0.86	1.45	1.74	2.21

SD denotes standard deviation of log_{10} transformed responses; the coefficient of variation (CV), the ratio of the standard deviation to the mean, is SD multiplied by 2.303. Standard deviations are calculated for all data (SD(all)) and for deterministic approaches only (SD(det)). EG = ethylene glycol, TCE = trichloroethylene, HCB = hexachlorobenzene, NP = nonylphenol, Cu = copper.

In addition, for each chemical ('full' data sets only), the means of the log-transformed PNECs derived from deterministic and probabilistic methods were compared; 'base' data sets were never sufficiently large to meet the criteria for applying probabilistic approaches. Despite the small sample sizes for probabilistic methods, statistically significant differences were observed for the full-sets for ethylene glycol (p 0.004), trichloroethylene (p 0.012) copper in freshwater ($p < 0.001$) and copper in seawater (p 0.047). Probabilistically-derived PNECs were higher (less conservative) in each of these instances. In the other two cases (hexachlorobenzene and nonylphenol), no statistically significant difference between the means was seen (Figure 2). Overall, the effect of including probabilistic sources on variability in the data set is small; the frequently held view that probabilistic methods always lead to less conservative PNEC values is not borne out by the datasets in the current study.

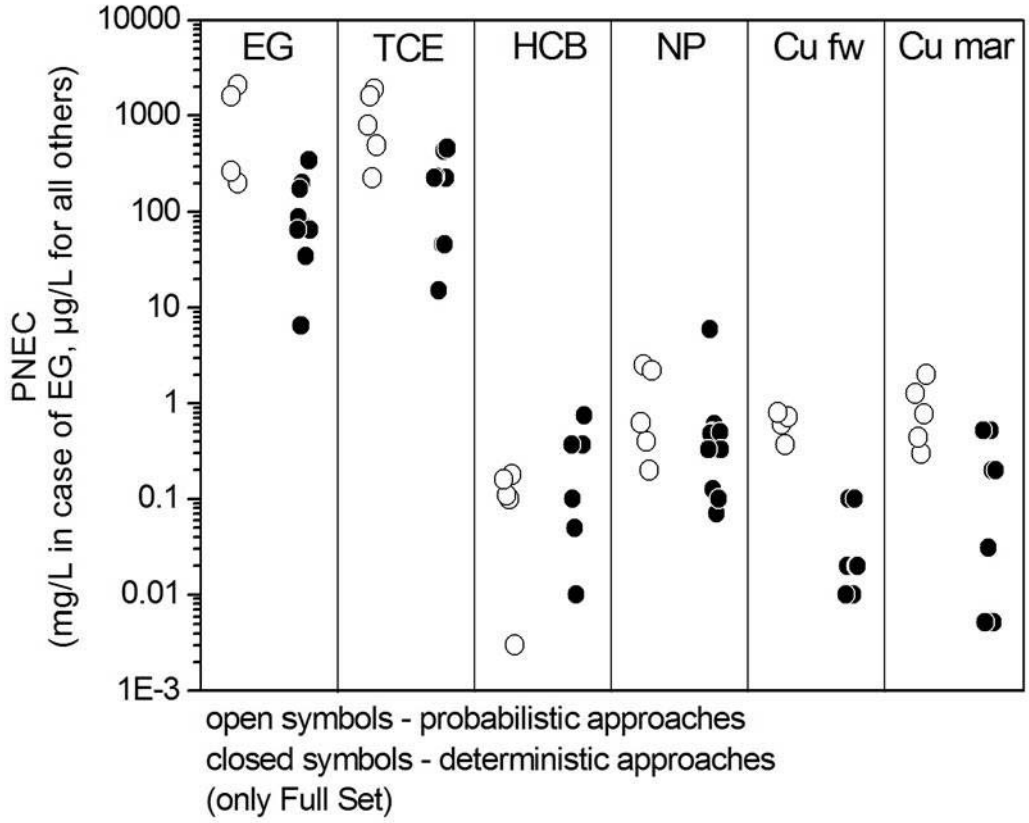


Figure 2. Comparison of PNECs derived from probabilistic and deterministic approaches (only 'full' data sets)

Comparing the coefficient of variation for the base-set against that for the full-set shows greater variability in the full-sets for all chemicals, irrespective of whether the calculated coefficient is based on all data or just deterministic approaches (Table 1). Given that all regulatory systems are predicated on the view that more data provides more certainty of the result, this might appear surprising.

Older chemicals are likely to have more tests conducted on them simply because they have been available for longer. Research groups looking at new test systems or organisms will selectively choose chemicals already tested with as wide a range as possible of species, taxonomic groups and test types as possible, so they have good comparisons with their new data. Such undirected increase in data richness does not necessarily concentrate further testing on key species or endpoints. Large numbers of tests on related species increase the likelihood that regulators will choose different test results on which to base PNECs. We might, therefore, be less surprised that more data consistently increases rather than decreases variability.

'Protection goals'

The aim behind the regulatory system might influence the variability of the outcome. Figure 3 compares general chemical assessment guidelines (existing and new substances) with water quality guidelines.

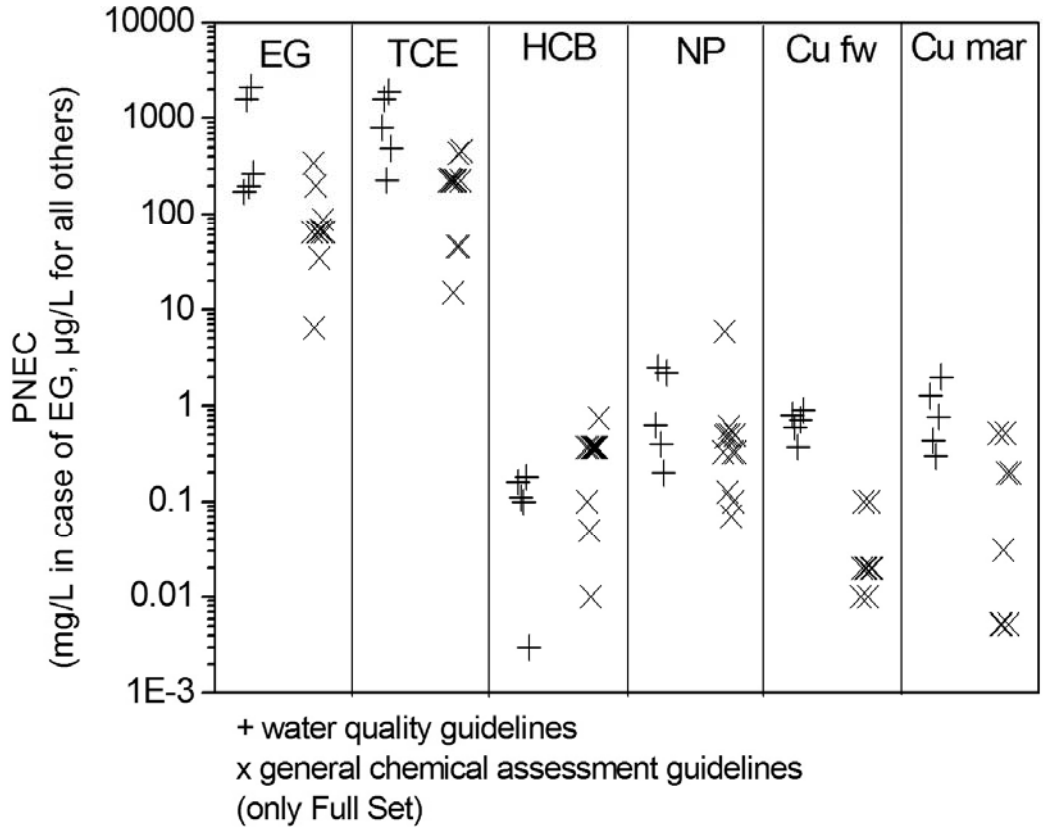


Figure 3. Comparison of PNECs derived from water quality guidelines and general chemical assessment guidelines (only 'full' data sets)

If we compare figures 2 and 3 it reveals that the data points are virtually the same. The reason behind this is the fact that, with only few exceptions, assessments according to water quality guidelines applied probabilistic methods, while assessments following chemical assessment guidelines used the 'classical' deterministic approaches.

Decision trees

None of the factors considered so far can account for a substantial part of the variability seen in the derived PNECs. Details of the decision process for each PNEC derived were plotted out as 'decision trees'. Space precludes the presentation of all of these but as an example, Figure 4 shows the range of decisions taken by different participants using the full dataset for nonylphenol. A total of 17 PNECs were derived from this dataset using 7 distinct regulatory systems. Three of the derived PNECs came from a probabilistic approach; key studies are not identified with this methodology. One value for PNEC was derived from acute data only according to the guidelines followed. From the 12 deterministically derived PNECs, 5 different key studies were chosen by participants in the study. The reasoning behind every derived PNEC was examined using the decision trees and the following sources of variability were identified:

Differing decisions on the key study often related to different perceptions on the validity of the 'chronic' studies available. For nonylphenol, for example, long-term studies were available for an alga S.

subspicatus, two invertebrates D. magna and a copepod, and one fish O. mykiss. Six participants disregarded the algal study because the endpoint was algal biomass rather than growth; guidelines identified growth as the only acceptable endpoint. A further participant also rejected the algal study because, at 72 hours, the study was too short to be considered 'chronic' in the guideline followed. Four participants chose the algal study as key regarding it as the lowest of 3 valid chronic studies. Nine participants rejected the fish study as a valid 'chronic' test because, although the duration of the test was 91 days, the endpoint was a NOEC for mortality.

'Chronicity' influences the derived PNEC in more than one way:

Many regulatory systems preferentially take chronic test data over acute to derive guidance values. Whether or not a study is regarded as 'chronic' influences whether or not it is chosen as a key study in deterministic approaches. The converse is also true; some regulatory systems only use acute test results and calculate an acute to chronic factor at a later stage.

In deterministic approaches, availability of chronic tests for different taxonomic groups defines the assessment factors applied to derive the PNEC from the NOEC in the key study. A different key study could, therefore, lead to a different NOEC plus a different assessment factor which would change the PNEC easily by 1 or even 2 orders of magnitude.

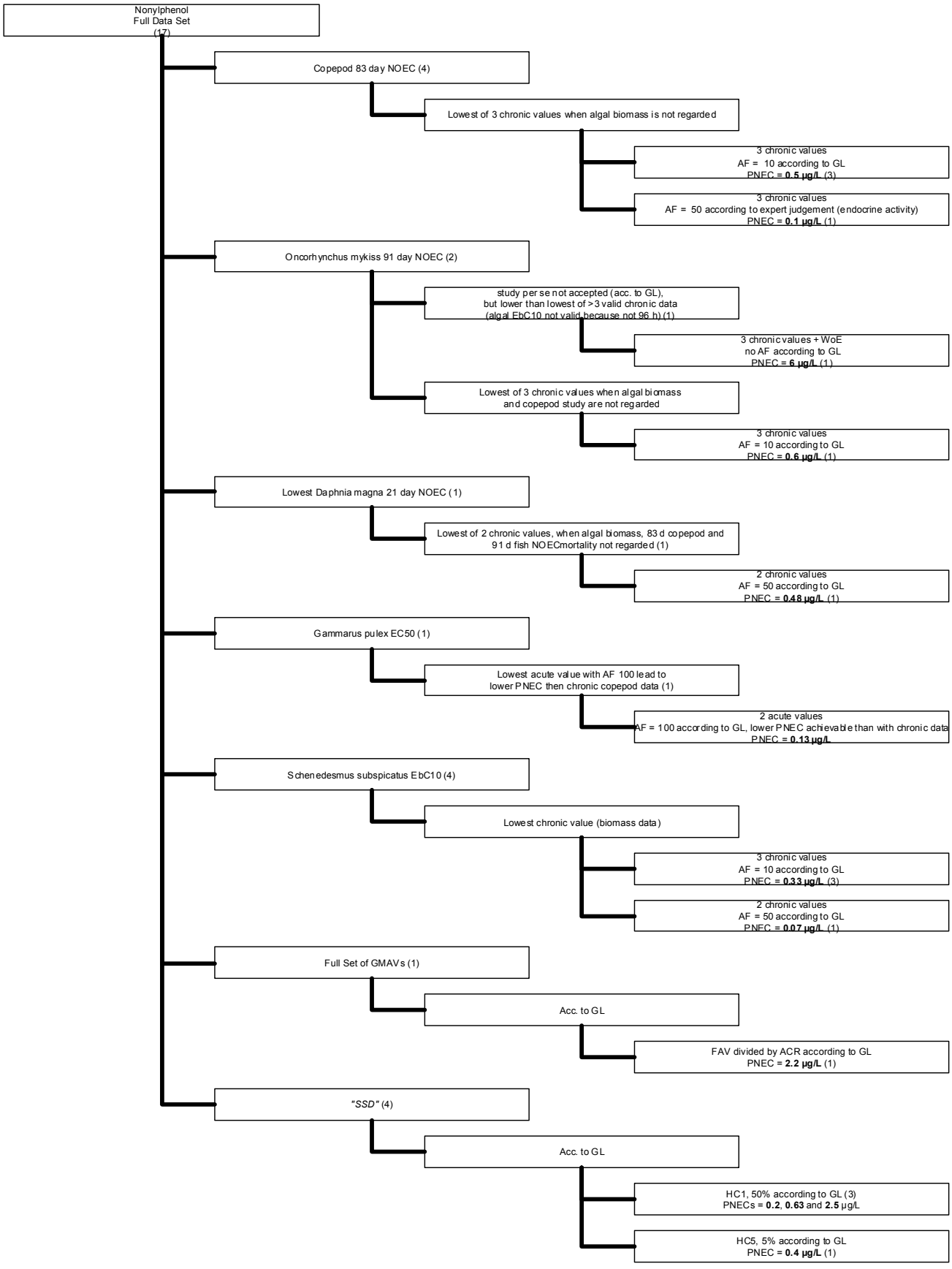
The number of available 'chronic' test results influences whether a probabilistic method can be applied or not. Criteria differ between regulatory systems but all depend on number of chronic tests and number of different taxonomic groups represented in the dataset. Some probabilistic approaches allow or encourage 'chronic' values to be derived from acute ones either at the beginning or the end of the process. Definition of particular studies as acute or chronic, therefore, influences both the fit of the derived curve and the error on the estimate. If acute to chronic factors are applied after fitting acute test results to a probabilistic distribution, definition of chronicity influences the final correction of the PNEC value.

Acute study– chronic study

'Chronic' is almost universally considered 'long-term' and usually defined as representing a significant proportion of the organism's life-span. What a 'significant proportion' is, however, is seldom defined exactly and will clearly differ between types of test organism.

A short test period using algae or microorganisms will expose the cells to the chemical over many generations because of very short generation times. However, even here, definition of acute and chronic differs between administrations with some, for example Australasia, defining 72 hour tests on algae as 'chronic' and others, such as North America, requiring 96 hour exposure to classify the result as 'chronic'. In the European Union, 72 hour algal tests are regarded as acute when the EC50 for growth rate is used but chronic when the NOEC from the same test is used in deriving PNEC. As already mentioned for nonylphenol, the endpoint is also important in determining if algal tests are valid. Growth rate is the only acceptable endpoint in Europe and Australasia. Biomass is an acceptable endpoint in the US and is normally the measure taken in US test guidelines.

Figure 4. Decision tree for nonylphenol('full' data set) to illustrate the range of decisions and the reasoning behind them for PNEC derivation. Numbers in parentheses – number of approaches, AF – assessment factor, Acc GL – according to guideline, HC1/HC5 – hazard concentration derived from a probabilistic dstribution ('SSD') that impairs 1% (5%) of the species with 5(50)% confidence.


Nonylphenol
Full Data Set
(17)
Copepod 83 day NOEC (4)
Lowest of 3 chronic values when algal biomass is not regarded
3 chronic values
AF = 10 according to GL
PNEC = 0.5 µg/L (3)
3 chronic values
AF = 50 according to expert judgement (endocrine activity)
PNEC = 0.1 µg/L (1)
Oncorhynchus mykiss 91 day NOEC (2)
study per se not accepted (acc. to GL),
but lower than lowest of >3 valid chronic data
(algal EbC10 not valid because not 96 h) (1)
3 chronic values + WoE
no AF according to GL
PNEC = 6 µg/L (1)
Lowest of 3 chronic values when algal biomass
and copepod study are not regarded
3 chronic values
AF = 10 according to GL
PNEC = 0.6 µg/L (1)
Lowest Daphnia magna 21 day NOEC (1)
Lowest of 2 chronic values, when algal biomass, 83 d copepod and
91 d fish NOECmortality not regarded (1)
2 chronic values
AF = 50 according to GL
PNEC = 0.48 µg/L (1)
Gammarus pulex EC50 (1)
Lowest acute value with AF 100 lead to
lower PNEC then chronic copepod data (1)
2 acute values
AF = 100 according to GL, lower PNEC achievable than with chronic data
PNEC = 0.13 µg/L
Schenedesmus subspicatus EbC10 (4)
Lowest chronic value (biomass data)
3 chronic values
AF = 10 according to GL
PNEC = 0.33 µg/L (3)
2 chronic values
AF = 50 according to GL
PNEC = 0.07 µg/L (1)
Full Set of GMAVs (1)
Acc. to GL
FAV divided by ACR according to GL
PNEC = 2.2 µg/L (1)
"SSD" (4)
Acc. to GL
HC1, 50% according to GL (3)
PNECs = 0.2, 0.63 and 2.5 µg/L
HC5, 5% according to GL
PNEC = 0.4 µg/L (1)

For cladocerans, the exact species influences definition of chronicity. Daphnia species are frequently used in tests and require 21 days to produce multiple generations; Ceriodaphnia species achieve the same number of generations within 7 days. Most regulatory systems recognise the difference and define chronicity accordingly. The criteria for daphnids also reflect the common view that 'chronic' implies that reproduction can occur within the test duration. Many invertebrate species reproduce within practical test periods. Ostensibly, tests using invertebrates should be straightforward to classify as acute or chronic using the reproductive generations criterion. However, even this is not universally applied with some 'reproductive' parameters, such as fertilisation in oysters, accepted by some administrations as 'chronic; here the 'reproduction' tag is being judged more indicative of longer-term effects than the strict length of the test (1 hour). Growth and survival of invertebrates are also considered valid 'chronic' endpoints in tests which are longer-term by Australasia and, sometimes North America.

The problem becomes much more significant when fish are being tested. Full, multi-generation tests are essentially impossible in fish because the time frame would be much too long. Many test designs replace the 'significant proportion of life-span' criterion with 'significant portion of the life-span'; the early life stages from fertilisation to juveniles over 30 days in warm water and 60 days in cold water species. Growth and survival are the normal endpoints along with reproductive effects (largely fertilisation and hatch rates). However, some administrations allow survival and growth endpoints over much shorter time frames (>96 hours) as chronic whilst others disallow survival as a 'chronic' endpoint even over extended test periods.

This combination of allowable time periods and allowable endpoints explains the choice of 5 different studies as 'key' from the full dataset for nonylphenol. The number of valid chronic studies identified by different participating groups was 3, 2 or 0 and these perceptions led to application of assessment factors of 0, 10, 50 or 100. Importance of these factors is valid across all of the chemicals included in the present study. As a consequence, the most obvious way to reduce variation between PNECs derived using different national/regional regulatory systems would be to harmonise definitions of acute and chronic toxicity test data across jurisdictions. This could substantially reduce the differences in choice of key study and the magnitude of the assessment factor applied in deterministic methodologies.

Small print

Regulatory systems have evolved over time. The headline approach of all deterministic systems is similar – take a critical study and apply assessment factors to derive PNECs. It is perhaps surprising that more of the variation can be explained by aspects of this headline than by subtle differences between systems or practitioners. However, some of the differences seen in the PNECs derived by the participants in the present study can be ascribed to subsequent revisions of basic regulatory systems - the small print. Small print in regulatory guidelines often derives from application of expert judgment. Over time, what was early expert judgment gradually becomes written into revisions of the guidelines but the process continues. We did see one example of the process still in operation. Participants from the region of one regulatory system applied a new rule (sub-acute oyster larval tests acceptable as chronic); other participants using the same guidelines did not because the new practice had not yet been written into guidelines. The process would, presumably, be transparent within the jurisdiction doing the regulation; different stakeholders within the region would have been aware of the development which was, again presumably, agreed between them or imposed by the regulator. Global transparency is almost impossible in these circumstances and makes international attempts at risk assessment particularly problematic. Widespread peer-review of international documents is essential if these local factors are to be identified.

Expert judgement

Participants in the present study were asked to explain, and justify, decisions taken in deriving PNECs; 10% of all PNECs derived included "expert judgement" as part of the justification. In the present context,

'expert judgment' was very narrowly defined and related strictly to the small print sections of the guidelines followed. What participants were identifying as expert judgement was the application, or not, of extra factors which were optional in the guidelines – they were what the guidelines themselves identified as needing 'expert judgment'. There were no examples of participants overriding guidelines completely because of specific expert knowledge or experience. Clearly, as guidelines evolved, occasions must have arisen where experts considered individually, and eventually by local consensus, that the guidance had to be modified. Equally, the chance of such an occurrence within the present study was small, given that the chemicals chosen for the study were old, widely tested and had had locally, and globally derived PNECs for some time.

Examples of the options which participants identified as requiring expert judgment included dealing with bioaccumulating chemicals, endocrine disruptors, essentiality and background concentrations for naturally occurring elements. Global consensus on how to handle these factors is advanced even if detail differs slightly between regulatory systems. This level of expert judgement does not seem to impact on the process of international risk assessment.

Other factors identified as requiring expert judgement were more problematic. Some jurisdictions recognised specific quirks in deterministic approaches, for example chronic testing carried out on taxonomic groups which were less sensitive than others should not allow reduction in assessment factors applied to the key study. In such cases, the process should revert to the acute test results with the application of larger assessment factors. Commonly, but not universally, it is acceptable or suggested that marine risk assessments based on few data should use more sensitive freshwater species to derive PNECs. Some jurisdictions gave greater weight either to locally relevant species or to species with local economic value. The latter factors are difficult, if not impossible, to apply globally. International risk assessment must either remove these factors during the internationalisation process or clearly state what these factors are so that users worldwide can apply them or not in their local situations.

4 Way forward

International documents such as CICADs that present examples of hazard assessments for chemicals based on national assessments need to be consistent and transparent in the way PNECs are derived. International programmes, such as OECD, which have their own methodology for deriving guidance values need to be aware of the differences between their own and national systems to avoid reader confusion from different headline guidance values in the literature.

Having identified a number of factors that contributed to the overall variability in the present study, the next step should be to develop clear guidelines on how these factors, and the decisions that are based on them, should be presented in international documents. Exact descriptions of each of the stages and factors used to derive a value should be included in such documents and the new guidelines would indicate all of the issues to be addressed. Guidelines should enable authors to work more consistently in an international context. Guidelines would also allow readers to identify if different guidance values derived from different data or simply from different interpretations of the same data.

It should be recognised that an internationally derived guidance value will regularly differ from a national one on which it is based; this has proved the case in the CICAD programme. This is not problematical if the decision-making process is clear and transparent. Complete harmonisation would not be possible without changes in a wide range of national and regional approaches; the latter is not realistic.

All decisions made on the basis of such international guidelines should be unequivocally science-based. The primary goal of any hazard assessment methodology is to derive an environmentally safe (but not

overprotective) concentration. Achieving an internationally harmonised approach for hazard assessment is worthless if it does not reflect the actual ecological situation.

5 References

1. ANZECC/ARMCANZ (2000) Australian and New Zealand guidelines for fresh and marine water quality. Australian and New Zealand Environment and Conservation Council and Agricultural and Resource Management Council of Australia and New Zealand.
2. Chapman PM, Fairbrother A, Brown D (1998) A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environmental Toxicology and Chemistry*, 17(1):99-108.
3. Duboudin C, Ciffroy P, Magaud H (2004) Effects of data manipulation and statistical methods on species sensitivity distributions. *Environmental Toxicology and Chemistry*, 23(2):489-499.
4. EC (2003) Technical Guidance Document on Risk Assessment: Part II. European Commission, EUR 20418 EN/2. http://ecb.jrc.it/Documents/TECHNICAL_GUIDANCE_DOCUMENT/EDITION_2/tgdpart2_2ed.pdf
5. Environment Canada (1997) Environmental assessments of priority substances under the Canadian Environmental Protection Act. Guidance manual version 1.0 – March 1997. Chemicals Evaluation Division, Commercial Chemicals Evaluation Branch, Hull, Quebec (EPS 2/CC/3E).
6. Environment Canada (2006) Draft protocoll for the derivation of Water Quality Guidelines for the protection of aquatic life.
7. Hahn, T., Stauber, J., Dobson, S., Howe,P., Kielhorn. J., Koennecker, G., Diamond, J., Lee-Steere, C., Schneider, U., Sugaya, Y., Taylor, K., Van Dam, R., Mangelsdorf, I. (2009). Reducing uncertainty in ERA: Clearly defining acute and chronic toxicity tests. "Learned Discourses" in Integrated Environmental Assessment and Management, 3(1) 175-177.
8. Jager T, Vermeire TG, Rikken MGJ, van der Poel P (2001) Opportunities for a probabilistic risk assessment of chemicals in the European Union. *Chemosphere*, 43(2):257-264.
9. Keating Jr., (Eds.). Re-evaluation of the State of the Science for Water-Quality Criteria Development. SETAC Press, Florida, USA, pages 162-170.
10. Keating Jr, FJ, (2003) USA Ambient Water-Quality Criteria. In M.C. Reiley, W.A. Stubblefield, W.J. Adams, D.M. Di Toro, P.V. Hodson, R.J. Erickson, and F.J. Maltby, L. (2006) Environmental risk assessment. In: Issues in Environmental Science and Technology, No. 22 Chemicals in the Environment: Assessing and Managing Risk, The Royal Society of Chemistry, UK, pp 84-101.
11. Monforts, MHMM (2006) Assessment of persistency and bioaccumulation in pesticide registration frameworks within the Organisation for Economic Cooperation and Development. Integrated Environmental Assessment and Management 2(1):13-21.
12. National Science and Technology Council (1999) Ecological risk assessment in the federal government. Committee on Environment and natural resources of the National Science and technology Council Report CENR/5-99/001, Washington DC.
13. OECD (2002) Manual for the Investigation of HPV Chemicals. Chapter 4.2 Guidance for the Initial Assessment of Aquatic Effects. http://www.oecd.org/dataoecd/6/14/2483645.pdf
14. Roman G, Isnard P, Jouany JM (1999) Critical analysis of methods for assessment of predicted no-effect concentration. *Ecotoxicology and Environmental Safety*, 43(2):117-125.
15. Stephan, C.E., Mount, D.L., Hansen, D.J., Gentile, J.H., Chapman, G.A., and Brungs, W.A. (1985). Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. United States Environmental Protection Agency Report No PB-85-227049, Washington DC, USA, 98 pages