DERIVATION AND EXPRESSION OF WATER QUALITY STANDARDS

Opportunities and constraints in adopting risk-based approaches in EQS setting

R&D Technical Report P2-157/TR

P Whitehouse, M Crane, E Grist, A O'Hagan and N Sorokin

Research Contractors

WRc-NSF Ltd
Royal Holloway, University of London
University of Sheffield

Publishing Organisation
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol BS32 4UD
Tel: + 44 (0) 1454 624400 Fax: + 44 (0) 1454 624409
Website: www.environment-agency.gov.uk

ISBN 1 85705 8577

Dissemination Status
Internal: Released to Regions
External: Publicly Available

Statement of Use
This report considers how recent developments in the conduct and interpretation of toxicity data and the use of species sensitivity distributions could be incorporated into the derivation of EQSs for the protection of aquatic life. It considers how more ecologically relevant endpoints can be utilised and whether or not risk based standards could be derived. It will be used by Agency staff to help inform future decisions on the derivation and implementation of EQSs for the protection of aquatic life.

Key Words
EQS, risk, population growth rate, aquatic toxicity, species sensitivity distribution, SSD, safety factor, extrapolation, EQS

Research Contractor
This document was produced under R&D Project P2-157 by:
WRc-NSF Ltd, Henley Road, Medmenham, Marlow, Buckinghamshire. SL7 2HD
Tel: + 44 (0) 1491 636500 Fax: +44 (0) 1491 636501

Environment Agency's Project Manager
The Environment Agency's Project Manager for R&D Project P2-157 was: Helen Wilkinson, NCEHS, Evenlode House, Howbery Park, Wallingford, Oxon OX10 8BD

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre, c/o
WRc, Frankland Road, Swindon, Wilts SN5 8YF

tel: 01793-865000 fax: 01793-514562 e-mail: publications@wrcplc.co.uk

CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Environmental Quality Standards (EQSs) for the protection of aquatic life have been derived in the UK using the same basic approach for approximately twenty years. It is based on application of safety factors to a critical ecotoxicological dataset, representing the most sensitive species/endpoints. However, there have been significant advances in the conduct and interpretation of aquatic testing in recent years and also a questioning of the relevance of much of the test data that is routinely generated. Two themes are noteworthy:

- The ecological relevance of the endpoints that are measured.
- A trend toward risk-based approaches in regulatory decision-making.

In this report we have explored whether these developing themes may be usefully incorporated into the way EQSs are derived and expressed.

The main limitations of the current approach to deriving EQSs are a lack of clarity about protection objectives, sub-optimal use of information, a tendency toward conservative standards and their expression only as pass/fail thresholds. However, some of these limitations are ameliorated by emphasis on interpretation of critical data and flexibility in the size of safety factors applied.

Ecotoxicity test data usually describes the responses of individuals to chemicals but we argue that protection should ideally be aimed at population sustainability. More useful estimates of chemical effects can be obtained by predicting the intrinsic rate of population increase, r, or its log equivalent, λ. Values of r of zero imply no increase in population over time and a negative value, implies a projected decline, ultimately leading to extinction. Approaches based on Life Tables for estimating r from long-term studies with aquatic organisms have been refined and tested with data from aquatic toxicity studies.

The approach is workable in practice although requires certain data to be reported and a significant input of specialist modelling expertise. It has proved difficult to obtain the data required for analysis for more than a few species and so the approach cannot be used to replace conventional endpoints. However, based on data for the steroid oestrogen, 17α-ethinyloestradiol, we have shown how this approach can be used to inform the standard-setting process.

Defining concentrations of chemicals that will have no adverse effects on survival is also important. However, conventional fixed-time effects concentrations (e.g. a 96h LC50) require substantial extrapolation to derive EQSs. More useful summaries of survival data can be generated through time to event analyses of survivorship data. We have estimated time-independent LC0 values using a refinement of a two-step linear regression method described by Mayer *et al* (2002). Suitable datasets for a range of species are more readily obtained than are suitable datasets for estimation of r. The resulting analyses may be used to inform thresholds derived for particularly sensitive species but sufficient datasets may be collected to construct formal species sensitivity distributions, SSDs.

SSDs explicitly relate chemical concentration to a meaningful measure of biological impact (species diversity). However, they are subject to some important criticisms, particularly about how representative they are of assemblages of organisms in nature. We have been able to construct SSDs using time-independent LC0 values for ammonia and chlorpyriphos. Species 'representativeness' has also been investigated through Bayesian methodology to incorporate both expert judgement and empirical data into SSDs. Our experience with chlorpyriphos shows that very similar results resulted when SSDs were based on empirical data alone. The use of SSDs in addressing both the costs and benefits of different risk reduction options has been illustrated through a series of worked examples. However, sufficient data of adequate taxonomic diversity will not always be available. Furthermore, it is important to be clear about the level of impact (e.g. the proportion of species affected) that is acceptable in a given situation.

Ideally, EQSs would be based on ecologically relevant summaries in a risk-based framework with minimal reliance on default safety factors. There are clearly limitations to achieving this goal but useful progress can be made through a tiered approach. This entails the development of:

(1) A generic EQS derived using the critical data/safety factor paradigm and expressed as a simple pass/fail threshold. It is against this generic EQS that receiving water quality would be assessed and discharge consents derived, much as they are now.

(2) In the event of marginal or non-compliance with this generic EQS, a risk-based approach is invoked in which the relationship between chemical concentration and biological impact is explicitly described using an SSD.

For the generic EQS, recommendations for generating endpoints that are more relevant to the protection objectives are offered. Default safety factors should be used as a last resort with maximum use of the data to define these factors where possible i.e. data should supersede default values. The tiered approach treats the generic EQS as a trigger for more detailed investigation but, where data permit, the incorporation of a risk-based step such as an SSD should allow more informed decisions about:

- costs and benefits
- the practicalities of achieving particular levels of environmental protection
- the uncertainty that is inherent in the risk assessment process
- the environmental objectives for the catchment in question

1. INTRODUCTION

1.1 Background

It is self evident that the environment receives inputs of many chemicals through a variety of natural and man-made sources. Events of the past fifty years also show the adverse effects to the environment that some of these substances can produce if their release is not properly controlled. It is against this background that standards for individual substances of concern have been developed. Essentially, these are derived for hazardous substances with the intention of defining an acceptable concentration which will ensure that different 'uses' of the environment such as the abstraction of water for potable supply, or protection of aquatic life or terrestrial ecosystems are not compromised. Those developed for the protection of surface water quality are referred to as Environmental Quality Standards (EQSs) in the UK.

This report is concerned with the methods used to derive EQSs and the way they are expressed and used in a regulatory context. Although standards are available for other environmental media (soil and air) in this report we are concerned entirely with standards for the aquatic compartment. Furthermore, EQSs often define criteria for potable water supplies, but in this project we have been concerned only with the protection of aquatic life.

The derivation of EQSs is based on an appraisal of biological effects (ecotoxicity) data for the substance of concern. Ecotoxicology is a relatively young science and in recent years the theory and practice of ecotoxicology have advanced significantly. Whilst ecotoxicological principles already feature in the derivation of EQSs, some of these recent advances do not, yet they are highly pertinent to the way in which EQSs are derived and expressed. It is therefore timely to consider what lessons might be learned from these technical, and in some cases, philosophical, developments and how they might be applied to EQS-setting in practice.

The following Section describes the recent history of EQSs in the UK. The strengths and limitations of the approaches currently used to derive EQS are then considered and notable technical developments of relevance are highlighted. Arising from this, we define the characteristics of an 'ideal' EQS and in the subsequent Chapters address the practicalities of achieving these characteristics.

1.2 Legislative Background

The use of Environmental Quality Standards (EQSs) in the regulation of water quality stems from a suite of national (and now, international) legislation concerned with the protection of water resources. During the 18th and 19th centuries, increasing urbanisation and population increase in the UK led to increasing pressures on water resources and, in particular, serious issues about their sanitary quality. This led to the first UK water pollution control legislation (The 1875 Public Health Act) which made it an offence to 'pollute waters'. Further pressures led to legislation in the 20th century but only in 1974, through the Control of pollution Act, did attention begin to focus on measures to limit emissions of noxious substances in the aquatic environment.

In the 1980s and 1990s, a series of EU Directives sought to control the impacts of emissions to the environment. These included product-related (e.g. Detergent and Marketing and Use Directives) and use-related Directives (e.g. Surface Water Directive, Bathing Water Directive, Freshwater Fish Directive), Directives concerned with the activities of particular industrial sectors (e.g. Titanium Dioxide, IPPC and Urban Wastewater Directives) and, most significantly (at least as far as EQSs are concerned) the Dangerous Substances Directive and Daughter Directives which sought to control individual substances through the use of Uniform Emission Limits or EQSs.

The Dangerous Substances Directive included two lists of substances: List I dealt with substances regarded as being particularly dangerous because of their toxicity, persistence and bioaccumulation and for which statutory EQSs were effectively imposed. A much larger list of substances (List II) was also identified and covered substances that are considered less dangerous but which may have a deleterious impact on the environment. Competent Authorities within Member States were charged with eliminating pollution by List I substances and reducing pollution by List II substances. Together with substances that were regarded as of potential concern by UK regulators (the Environment Agency and the Scottish Environmental Protection Agency, SEPA), substances falling within the List II categorisation were included in a major programme of EQS development throughout the 1990s. These were developed by scientists within The National Centre for Environmental Toxicology at WRc-NSF (formerly WRc) and subjected to review by an external Steering Group.

In 2000, a significant piece of European legislation was introduced – the Water Framework Directive. This is important because it supersedes a number of the Directives mentioned above or links them strategically in a way that has not been evident hitherto. As far as EQSs are concerned it is also significant because the Directive contains an Annex of Priority Substances (PSs) and Priority Hazardous Substances (PHSs) for which pan-European EQSs are required. A procedure to be used to derive these standards has been proposed (Fraunhofer Institute, 2002) but the details of the methodology have yet to be agreed. The only latitude in methodology for deriving standards for PSs and PHSs will be that described in the agreed text. Any constraints would have to be applied to other substances as well to avoid different methods being used in the UK.

1.3 What Are EQSs and How Are They Used?

An EQS represents an acceptable concentration for a chemical in the environment. If it is not exceeded, it is assumed that the intended protection objective (e.g. aquatic life, potable water supply) will not be compromised. In practice, EQSs are thresholds based on a review of ecotoxicological data describing the effects of a chemical on aquatic organisms. It follows that EQSs:

- are biologically based;
- are expressed in terms of a chemical concentration;
- refer to a single substance, or a group of closely-related substances that act by the same mode of toxic action;

- apply to receiving waters (rather than discharges to watercourses).

Essentially, EQSs are intended to provide use-related 'yardsticks' for specific substances that can be used for the following purposes:

- environmental benchmarks against which environmental monitoring data can be assessed;
- setting goals for pollution control activities;
- acting as triggers for remedial action;
- derivation of site-specific discharge limits.

A convenient way of considering the possible types of chemical regulation is illustrated in Figure 1.1 (from Barnett and O'Hagan, 1997). This shows that controls may be applied at a variety of points between the point at which chemicals are produced through the point at which they are discharged into the environment, in the receiving environment or (rarely) at the biological receptor itself. Because they apply within the receiving environment (the 'point of contact' in Fig. 1.1), EQSs are independent of the means by which the contaminant has entered the receiving environment. In other words, an EQS is as applicable to point source inputs as it is to diffuse sources of a chemical. In practice, these standards may be translated into a critical end-of-pipe concentration ('point of emission' in Fig. 1.1) when deriving consent conditions for discharges, where the aim is to ensure that the standard is met outside a defined 'mixing zone'.

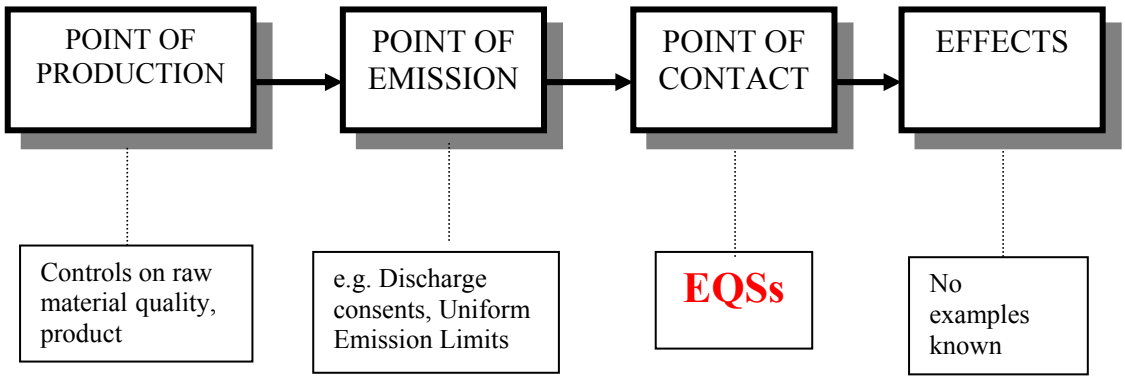


Figure 1.1: The role of standards in environmental regulation of chemicals

A number of regulatory regimes in the EU and UK require the estimation of an acceptable threshold concentration for individual chemicals e.g.

Threshold	Regulatory scheme
Predicted No-effect Concentration (PNEC)	Risk assessment of new and existing chemicals
Toxicity:Exposure Ratio (TER)	Plant Protection Products and Biocides approval
Environmental Assessment Levels (EALs)	Impact assessment under PPC
Environmental Quality Standard (EQS)	Dangerous Substances/Operational control/Water Framework Directive

In some of these cases, the thresholds may be based on only a modest dataset, representing just a few aquatic species. In contrast, derivation of EQSs tends to be based on a more substantial dataset. This is because the substances of concern tend to be ones that are known to occur in surface waters and have often been the subject of significant monitoring or research effort, particularly with respect to possible ecotoxicological effects.

EQSs are amongst the most widely applied regulatory controls on chemicals in the environment. Arguably, they represent the highest level of assessment within the range of regulatory schemes that call for an appraisal of ecotoxicological effects. It follows that the methods used to derive EQSs and the resources applied should seek to extract as much value as possible from the available data and put it to effective use in guiding regulatory decisions about the acceptability - or otherwise - of a particular level of chemical contamination. At the same time, we must remember that there may be practical constraints on what can be achieved: EQSs sometimes have to be estimated when only a limited quantity of data are available, and they are applied routinely by personnel who may have only a limited appreciation of the subtleties involved in deriving these thresholds.

1.4 Current Approach to Deriving EQSs

A generalised view of the steps involved in deriving environmental thresholds, including EQSs, is shown in Figure 1.2 (from Whitehouse *et al*, 2000).

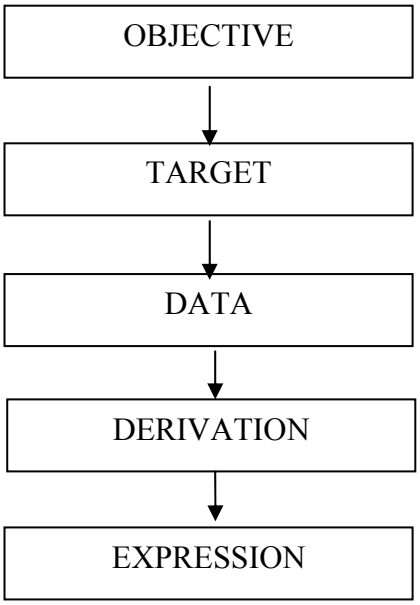


Figure 1.2: Common steps in standard-setting procedures

Normally, we would expect to see an **objective** that can be expressed in lay terms e.g. 'to protect aquatic life'. In some cases the objective is translated into a **target**, which is more precise and in some cases makes clear exactly what is to be protected and may even be quantified e.g. 'no more than x% of species affected'. At the heart of the process is an adequate ecotoxicological **dataset** but, in order to **derive a standard**, it is necessary to translate summaries of toxicity obtained from ecotoxicological studies into a concentration which will achieve the protection target. It normally entails a process of "extrapolation", effectively bridging the gap between the data available and what we are seeking to protect. The standard should then be **expressed** in a way that permits comparison with the presence of the chemical in the environment (almost invariably as a concentration[1]).

The current methodology for EQS derivation in the UK has been in place for a number of years. Current practice is described in Zabel and Cole (1999) and Table 1.1 summarises how the procedure corresponds to the steps outlined in Figure 1.2.

[1] Generally, a dissolved concentration is the most technically valid expression although for metals and some organics (e.g. weak acids) there are questions about speciation and whether the threshold – and hence compliance assessment – should be expressed in terms of particular species or forms of the substance

Table 1.1: Current UK practice for deriving EQSs (from Zabel and Cole, 1999)

Element	Stated aim/approach taken
Objective	Protection of aquatic life
Target	None stated although there is an implicit suggestion that the aim is to protect populations of aquatic organisms
Data	Comprehensive searches for existing data from open and 'grey' literature. Acute and chronic ecotoxicity data for algae and/or macrophytes, arthropods, other invertebrates (e.g. molluscs) and fish. Other taxa when there is reason to believe they will be particularly sensitive e.g. insect larvae for insecticides. Freshwater species used for freshwater EQSs and saltwater species for saltwater EQSs critical data are selected (i.e. most sensitive species identified) critical data quality-assessed (validity of test procedure, relevance of endpoint, plausible dose-response) and regarded as 'primary' data (good quality) and 'secondary' data (uncertainties remain). QSARs may be used to supplement experimental data but would constitute only 'secondary' data
Extrapolation to a standard	Flexible 'safety factors' are applied to the critical data to take account of uncertainties arising from: • Interspecies differences in sensitivity • Acute exposure to prolonged exposure • Effects at different levels of biological organisation (e.g. laboratory to ecosystem and accounting for possible sensitivities of different communities) Ranges of 'safety factors' recommended (see Fig. 1.3) but size of safety factors depends on quantity of data and information that would help reduce the level of uncertainty (bullet points above). Emphasis is placed on use of expert judgement and peer review of proposals through Steering Group, especially in selection of suitable 'safety factors' Additional 'safety factors' may be applied for highly bioaccumulative substances

Element	Stated aim/approach taken
	'Safety factors' are applied to 'primary' data and the 'secondary' data are used in a validation role.
Expression	EQSs expressed as ambient concentration, usually expressed in terms of dissolved concentrations although speciation-based standards have recently been developed e.g. for aluminium Maximum Allowable Concentration (MAC) for short-term, episodic events, based on acute effects data only Annual Average (AA) for long-term or continuous exposure, based on chronic NOECs (where available) but on acute effects data where none are available 'Tentative' status for EQSs where minimum toxicity dataset unavailable or if data for sensitive species are absent. 'Tentative' status given to saltwater EQSs when these are based on freshwater data. Effect of water quality parameters e.g. pH, hardness can have a marked effect on bioavailability and toxicity. This is reflected in pH or hardness 'bands' for some substances, notably metals and weak acids

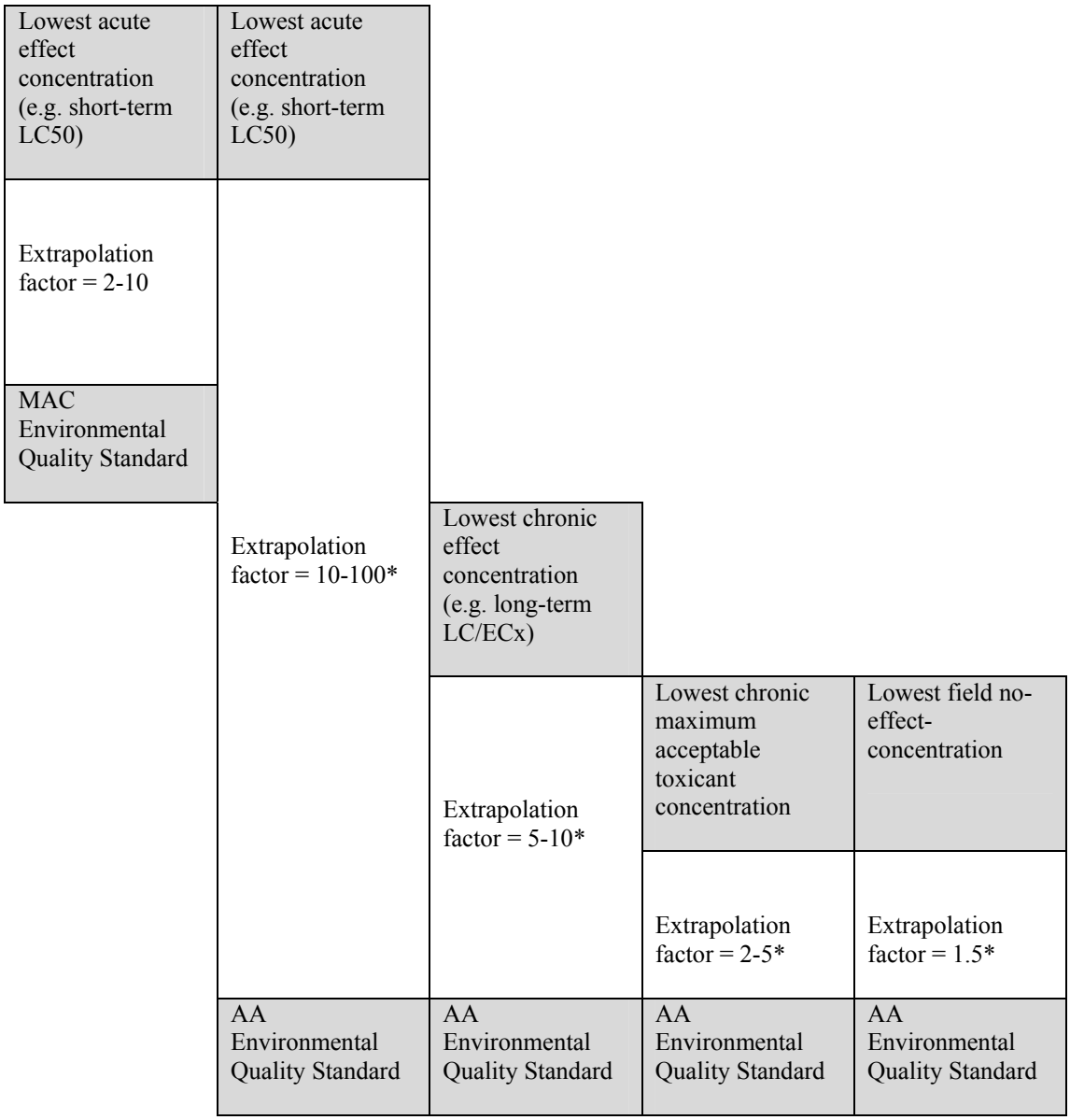
Some aspects of the current UK approach to deriving EQSs invite particular comment:

1. Like many other standard-setting schemes (Whitehouse *et al*, 2000), the target for EQSs is actually rather unclear. It is not documented for example whether the aim is to protect all species or enough species to ensure that certain key ecological processes continue (i.e. a functional target). However, it has become evident through participation in the EQS Steering Group that the aim of EQSs is one of structural protection (i.e. species diversity and abundance) as opposed to functional protection. The desired level of protection in the UK EQS scheme (i.e. what proportion of species should be protected?) is more difficult to judge and remains uncertain. However, there is also an implicit understanding that it is not necessary to protect every individual of every species but that an EQS should permit maintenance of viable populations, in which losses can be compensated for by recruitment (without having to rely on immigration from refugia).

2. There is no concession to achievability of EQSs except where analytical methods are not sensitive enough to quantify with accuracy at the proposed threshold(s). Even then, the EQS may be set at a lower level than the limit of analytical detection in the expectation that techniques will be refined in time to enable accurate quantification. The key point here is that there is no economic or social overlay to the standard-setting process at the moment – the level defined for an EQS is based purely on scientific considerations.

3. The use of ‘safety factors’ to account for uncertainties when translating the available data into a standard is a conventional approach, used by many other standard-setting jurisdictions. Whilst the use of safety factors in this way tends to lead to conservative outcomes, there is a consequent possibility of overestimating risk (Chapman *et al*, 1998). Where default factors are the norm there can also be a false reliance placed on a ‘one size fits all’ use of safety factors because no one set of factors has universal applicability (Chapman *et al*, 1998). However, in UK EQSs, a high level of emphasis is placed on the quality assessment of ecotoxicity data and the use of expert judgement in the selection of critical data and safety factors used in extrapolation. Based on a survey of other standard-setting schemes (Whitehouse *et al*, 2000), UK EQSs stand out in this regard. This goes some way to addressing legitimate concerns about the tendency to use default factors to address uncertainty even when this might be achieved through proper investigation of experimental data. This is a strength of the current approach although, as we suggest later, it may be exploited more extensively.

   The way in which ‘safety factors’ are selected in deriving UK EQSs is illustrated in Figure 1.3.



* the size of factor selected is based on expert judgement
AA = Annual Average
MAC= Maximum Allowable Concentration

Figure 1.3: 'Safety factors' used in the derivation of UK EQSs (Zabel and Cole, 1999)

1.5 Limitations of the Current Approach to Deriving EQSs

1.5.1 Technical considerations

What are we trying to protect?

In Section 1.4, we highlighted a lack of clarity in the protection objectives for EQSs and in particular whether there was a level of impact on species abundance that might be considered acceptable. This is an area of uncertainty that is common to much environmental risk assessment and is certainly not confined to EQSs.

At the heart of any attempt to define a threshold level for a chemical in the environment is the assumption that an acceptable threshold does indeed exist. Cairns (1992) explores this theme, describing different possible expressions between the level of a toxicant and the response by biota. Whether a true threshold actually exists is an interesting but largely academic debate because society expects us to define one anyway. Variability in natural systems with time and space is normal and so there is a practical problem of discerning a signal from this background 'noise'. The threshold concept is a necessary man-made concept and might reasonably be taken to mean a level of contamination that does not cause unfavourable trends in natural systems or irreversible deviations from a 'nominative state' (Cairns, 1977; Odum *et al*, 1979).

But the definition of an acceptable threshold is not entirely a scientific one. There is also a sociological component to be considered. Pollard *et al* (2000) argue that the value placed on a component at risk (here it is aquatic life) needs to be considered when assessing the significance of harm to the environment, taking account of the economic and societal value that results from impact. They go on to emphasise that the acceptability of a particular level of harm (operationalised through, for example, the % of species or individuals of a species affected) is ultimately a value judgement. To our knowledge, such considerations have not been addressed within the arena of environmental protection although they do feature in human health risk assessment for human medicines and with occupational exposure to chemicals (Illing, 1999).

Certainly, public consultation on this point would place standard setting on a sounder philosophical footing than it is currently and could have significant implications for the way we extrapolate from data to a standard. The Royal Commission on Environmental Pollution's 21st report 'Setting Environmental Standards' (RCEP, 1998) places considerable emphasis on stakeholder involvement and consultation of people's values at an early stage in developing policies and standards, and this is re-iterated in the UK Government response to that report (HMSO, 2000). Greater public consultation to help define more clearly the protection objectives would seem to be one that is sufficiently generic to many chemical regulation schemes to warrant further consideration. Perhaps the requirement of Article 14 under the Water Framework Directive to consult the public on River Basin Management Plans could provide a forum for addressing this point.

Consulting on the protection objectives for environmental standards should not be confused with incorporating social or economic aspects of particular EQSs. In other words, in setting standards, a clear dividing line should be drawn between analysis of

scientific evidence and any social or ethical issues (RCEP, 1998). Therefore it is proper to confine our review of EQSs to only technical issues. As far as defining protection objectives are concerned, we suggest that:

- protection should be primarily at the level of population sustainability

  and

- an acceptable level of protection is one that gives rise to no detectable difference in population viability from a control or reference situation

- the aim should be to protect all multicellular species (by implication, some unicellular species may not be protected)[2].

Whilst the implication is that the welfare of individuals is not a technical concern, this may be of societal or even legal significance. Certainly, for some rare species at higher trophic levels, protection of individuals would be a legitimate protection aim.

Species composition of the available dataset

Often, ecotoxicological data are available for just a few species yet a lowland river may support many hundreds of species. Even within these modest datasets, the taxonomic composition is usually highly distorted. For example, in aquatic toxicity datasets, important taxa such as annelids, amphibians and molluscs are rarely present at all but certain crustaceans and fish (e.g. water fleas of the genus *Daphnia* and fish species such as rainbow trout, fathead minnow and bluegills) are often over-represented, with a consequent loss of 'representativeness' of assemblages in nature (Wheeler *et al*, 2001). Whilst there is little we can do practically to address this (apart from commissioning new studies on under-represented species), it is important to recognise the risk of biasing outcomes by basing EQSs on such unrepresentative taxonomic arrays. Again, this is not a feature that is peculiar to the UK approach but it may be argued that extrapolation based on a small subset of the total available species diversity (as the UK approach is) may exacerbate the risk of bias in the outcomes.

Inefficient use of information

In basing standards on a small subset of the available toxicity data, much of the available information is effectively ignored. Although it may be argued that the entire dataset has been used to identify the critical data, this represents an inefficient use of valuable toxicity data.

We have already pointed out that toxicity summaries expressed as a NOEC use only a small part of the available data from a study. The use of LOECs and NOECs as the basis for EQS setting is also subject to more fundamental flaws because they are subject

[2] This would include all higher organisms but would exclude bacteria, protozoa and unicellular algae. It does not follow that these taxa would necessarily be unprotected but that they are not an explicit protection aim.

to considerable uncertainty. Some commentators (e.g. Chapman *et al*, 1998) even go so far as to suggest that thresholds based on extrapolation from a NOEC or LOEC are meaningless.

The NOEC is the highest concentration in a toxicity test producing a response that does not differ from the control when compared in a statistical significance test. It is usually calculated from concentration-response data by using Analysis of Variance (ANOVA), followed by a multiple comparison test (Newman, 1995). The perceived advantage of the NOEC is that it is easy to understand (OECD, 1996). However, there are many disadvantages to its use (Chapman *et al*., 1996):

- The NOEC must be one of the concentrations used in a bioassay, because hypothesis testing does not allow interpolation between test concentrations. It is therefore dependent on the choice of test concentrations in the study.

- The NOEC tends to increase as the precision of the bioassay decreases, thus rewarding careless experimentation that increases response variability.

- Confidence intervals cannot be calculated for a NOEC, so the precision of this value is unknown and different NOECs cannot be compared.

- A NOEC cannot be obtained if the lowest concentration tested produces a significant effect when compared with the control.

- The NOEC is not a 'safe' concentration, since large effects may still occur at this level.

- The NOEC breaks a basic rule of scientific method by attempting to *prove* the null hypothesis of 'no effect'.

- The NOEC wastes data because it does not provide information on the range of sensitivity of bioassay organisms.

- The NOEC depends upon the type I error rate chosen in significance tests and on the type of multiple range test that is selected for comparing the means: different choices produce different NOECs.

- It can be difficult to determine a NOEC if the response does not follow a monotonic trend, for example when hormesis occurs.

These shortcomings are particularly pronounced when critical data are expressed in terms of a NOEC or LOEC and are subsequently used as the basis for standard setting. Despite wide acceptance of the fundamental flaws inherent in estimating no effects concentrations from toxicity and ecotoxicity studies, they are likely to remain a key part of the available datasets in the foreseeable future.

Conventional summaries of toxicity

Statistical analyses are essential tools for interpreting the outcomes of a toxicity test (Cox and Oakes, 1984; Crane and Chapman. 1996). This is because toxicity tests use individual organisms, each with a different tolerance to toxic chemicals, which produce

a distribution of responses from the most sensitive to the least sensitive. This 'statistical tolerance distribution' applies both when data are continuous (such as measurements of growth) and when data are quantal (all-or-nothing responses such as survival). The standard procedure is to select a concentration series of test chemical and expose a separate group of test organisms to each of these. Recordings are made of the response (usually mortality, or reductions in growth or reproductions) at each concentration (Forbes, 1993). This response is then used to estimate the concentration-response curve and point estimates such as the LC/EC50 or the concentration causing no adverse effects (e.g. the 'no observed effect concentration', NOEC).

These summaries of toxicity have to be expressed in terms of effects occurring after a fixed period of exposure. They are usually estimated after a specific exposure duration. For example, estimates from acute ecotoxicity tests with the crustacean *Daphnia magna* are usually reported as a 48-h EC50, while chronic fish tests normally last for several days or weeks and report an EC value and a NOEC. Although there is general agreement that the use of a dose-response curve to estimate an ECx has many advantages over the derivation of a NOEC, the calculation of an ECx at specific time intervals still uses data sub-optimally. This is because most investigators will take some measurements during the course of a bioassay, especially if survival is the endpoint. These data from intermediate observation periods are usually not reported or used in the final estimation.

In standard setting, the effects of long-term exposure to low levels are of primary importance. Where only acute data are available, safety factors may be applied to account for the ratio between acute effects concentrations and chronic no effects concentrations. Whilst default safety factors are the norm, experimental data will sometimes be used to estimate the ratio between acute and chronic toxicity, especially when it relates to a particularly sensitive species.

A more serious criticism of the ECx is that it has no ecological relevance. This approach was originally developed as a method of ranking chemicals in order of toxic hazard, a useful exercise if one wishes to prioritise chemicals for further investigation. However, the LC/ECx is an inappropriate summary of toxicity data if the aim is to assess the risk of death, reduced population size or some other ecologically meaningful parameter. Instead, Dixon and Newman (1991), Newman and Aplin (1992) and Sun *et al*. (1995) have recommended the use of survival time modelling and accelerated life testing, and Bedaux and Kooijman (1994) have proposed theoretically-derived functions to take explicit account of the time-dependence of toxicity. These approaches have particular advantages with respect to the derivation of EQSs and are discussed further below.

Disincentives to generate data

The approach used to derive EQSs is based on the most sensitive species. Therefore there is little incentive for manufacturers or dischargers of a chemical to generate additional toxicity data because the EQS can only become more stringent. If toxicity testing shows lower sensitivity the new data will simply be ignored; if they show higher sensitivity, they will be used to drive the standard to a lower level i.e. the lowest value can only get lower.

Inherent conservatism

The 'safety factors used in extrapolation are based on precedent and on surveys of chemicals in which the ratios between acute and chronic toxicity or effects concentrations and no-effects concentrations have been examined (e.g. USEPA, 1984; ECETOC, 1993; Lange *et al*, 1998). When data for individual substances are compiled, it is clear that a distribution of ratios emerges, often with a substantial range. For example, Lange *et al* (1998) showed that acute EC50: chronic NOEC ratios for metals lay between 0.3 and 1290, giving rise to an upper 90%-ile of 192 and a median of 28.0. Safety factors to inform these extrapolation steps are generally based on the more conservative statistic (e.g. the upper 90%-ile) which will inevitably be over-conservative for most substances, perhaps by a substantial margin (Chapman *et al*, 1998).

The Precautionary Principle encourages action when there is uncertainty about a causal link between emissions and effect. At worst, this can have the effect of marginalising science and instead encourage the use of large, default safety factors that will tend to emphasise Type II errors (the risk of false positives). Whilst a conservative (i.e. precautionary approach) is entirely appropriate in the early stages of a tiered scheme (e.g. Risk Assessment of New and Existing Substances or impact assessment under PPC) or where the effect of an adverse outcome may be to trigger additional testing (e.g. approval of Plant Protection Products), over-conservatism in EQSs has more serious consequences. This is because the resulting threshold (i.e. the EQS) is adopted for regulatory purposes, with little recourse to the provision of new data. Over-conservative outcomes may place unnecessary burdens on dischargers and manufacturers whilst providing no environmental benefits.

In the UK approach, safeguards against over- or under-conservatism are addressed through review of proposed EQSs by a Steering Group who exercise judgement largely through a flexible attitude to the size of safety factors employed. Nevertheless, heavy reliance continues to be placed on the use of default safety factors, especially where data are sparse.

Expression of the standard

Although several thresholds may be proposed for a single substance (e.g. a Maximum Allowable Concentration to accommodate short-term or episodic exposure scenarios and an Annual Average designed to protect against long-term or continuous exposure) the use of safety factors applied to critical data means the values can only be expressed as simple pass/fail thresholds. This does not present the range of possible interpretations of the available evidence or the assumptions that have gone into the derivation process. Consequently, it can give a spurious impression of accuracy. This very point was highlighted in the Royal Commission on Environmental Pollution's 21st report 'Setting Environmental Standards' (RCEP, 1998) and in the UK Government Response to that report (HMSO, 2000).

1.5.2 Regulatory considerations

Current approaches to EQS setting are sub-optimal with respect to the utility of the resulting standards, particularly in site-specific risk assessments and when considering the costs and benefits of different control options. This arises because EQSs are currently expressed as simple pass/fail thresholds. If the EQS is exceeded, the regulator and discharger cannot say what impact might then result. Consequently, any debate about the significance of the exceedance in terms of possible environmental impact must remain uninformed. Indeed, even detailed scrutiny of the original dataset is unlikely to be fruitful unless predicted exposure levels coincide with the dose-response for one or more of the species for which test data are available.

Furthermore, the costs of evaluating different emission reduction options cannot take account of the environmental benefits that would result from one option or another because there is no way of predicting the biological consequences e.g. the number of species that could be affected or the abundance of a given species. A more explicit risk-based approach that relates the level of impact and chemical concentration would, we argue, provide both the regulator and discharger with the means to reach informed decisions about the balance between costs and benefits for different risk reduction options (e.g. treatment options, waste management, chemical substitution).

1.5.3 Summary

In summary, the approach currently used to derive EQSs in the UK is one based on identification of critical ecotoxicity data (usually the lowest credible effects or no-effects concentration) to which a 'safety factor' is applied to extrapolate to a concentration that should protect aquatic life in the field.

The reliance placed on expert judgement to identify critical data and the flexibility in the selection of appropriate safety factors allows greater use to be made of the available data than would be possible in a more mechanistic approach based on the same fundamentals. There is thus greater opportunity to use experimental data to inform the extrapolation process rather than rely entirely on default 'safety factors'. Nevertheless, there are some inherent weaknesses of the approach, particularly the effectiveness with which data are used and a risk of over-conservatism, especially where a default safety factor is applied. Continued reliance on toxicity data that represents effects (or no-effects) at fixed times and is expressed at the level of the individual rather than populations also widens the gulf between the data and the protection aim of EQSs. This effectively adds to the uncertainty in extrapolation.

It is also possible that current approaches to EQS setting are sub-optimal with respect to the utility of the resulting standards, particularly in site-specific risk assessments and when considering the costs and benefits of different control options. This arises because EQSs are currently expressed as simple pass/fail thresholds. If the EQS is exceeded, the regulator and discharger cannot say what impact might result and so any debate about its significance must remain uninformed. Furthermore, the costs of evaluating different emission reduction options cannot take account of the environmental benefits that would result from one option or another because there is no way of predicting the biological consequences e.g. the species that could be affected. A more explicit risk-

based approach that relates the level of impact and chemical concentration would be more useful.

1.6 Significant Technical Developments

A number of important technical developments in the field of ecotoxicology, standard-setting and risk assessment have emerged in recent years. These are briefly described here and a number of them explored in more detail in the following Sections in the context of standard setting.

1.6.1 Demise of the NOEC?

Problems with the NOEC have led many statisticians and biologists to propose that these methods of hypothesis testing are not well suited to the type of data obtained from most toxicity tests (Chapman *et al*., 1996). The estimation of an ECx value (the EC at a specified value of x – usually EC50) overcomes most of the problems associated with hypothesis testing (Chapman *et al.,* 1996), and is the usual form of analysis for acute ecotoxicity experiments. The advantages of this approach are,

- The ECx is not restricted to be one of the test concentrations;
- The precision of the ECx can be estimated; the experimental precision and the choice of the type I error rate affect only the confidence limits, not the estimation of the EC value itself;
- The regression model used to estimate an ECx allows the investigator to characterise the entire toxic response of the test organism and uses all of the data for that time period;
- Non-monotonic relationships can be modelled.

Data from fixed times of observation (usually 24, 48, 72 or 96-h in acute bioassays) are transformed so that least-squares fits can be made to linear models (Forbes, 1993). Linearity is usually achieved by logging the exposure concentration and converting the response to its probit (Bliss, 1935) or logit (Berkson, 1944). Whatever the derivation of the dose-response curve, ECx values are then estimated for the magnitude of effect that interests the investigator. This is normally an EC50 or LC50, because more precise estimates are possible at this median point. However, the x in ECx can be as large or as small as an investigator wishes, although estimates at the extremes of the probability function are likely to have very wide confidence intervals (Hartley and Sielken, 1977). Bruce and Versteeg (1992) discussed the choice of x in ECx and concluded that a value of 20% is normally protective when the natural variability of populations is taken into account. However, many authors would consider a value of 20% effect as too high (OECD, 1996). Furthermore, the choice of different ECx values often leads to differences in the toxicity ranking of samples if the response slopes are not parallel (Oris and Bailer, 1997).

1.6.2 Survival time modelling

Survival is the commonest endpoint in aquatic ecotoxicity testing. The data are usually analysed by time-specific dose-response models to produce an estimate of the LC50 at a particular time (e.g. 48h). Whilst this tells us something about effects after this exposure period we must extrapolate from these data if we are to predict effects at other times.

Approaches that reduce the time-dependence of toxicity summaries are therefore of interest. Dixon and Newman (1991), Newman and Aplin (1992), Newman and McCloskey (1996) and Sun *et al.* (1995) have recommended the use of survival time modelling and accelerated life testing, and Bedaux and Kooijman (1994) have proposed theoretically-derived functions to take explicit account of the time-dependence of toxicity (the 'DebTox' model). In essence these methods assume that the probability of dying within a given time interval depends on the toxicant concentration and the amount by which the baseline hazard (in the absence of a toxicant) remains constant with time:

$$H(t,x_i) = e^{f(xi)}h_0(t)$$

where:

$h(t,x_i)$ = the hazard at time t and toxicant concentration x_I

$h_0(t)$ = baseline (i.e. control) hazard

$e^{f(xi)}$ = the function of the toxicant acting on the baseline hazard

Toxicity is expressed in terms of the risk of dying (e^{ti}) upon exposure to different concentrations of a toxicant. In marked contrast to conventional approaches, this approach explicitly acknowledges both the influence of toxicant concentration and exposure duration on the response of organisms. In theory, much of the information needed to adopt these survival time modelling approaches will have been collected during standard tests, even if it has not been reported. This represents an inefficient use of available data and one we pursue later in Chapter 2.

1.6.3 Estimating effects at different levels of biological organisation

Recent years have seen an increasing focus on population viability of biota in the environment. In part this is because we often recognise environmental problems in these terms e.g. declines in amphibian populations in many regions of the world, declines in UK songbird populations, population consequences of oestrogenic chemicals, but also because we recognise that environmental decisions are made at the population level e.g. '…. the effects of concern to ecologists performing assessments are those of long-term exposures on the persistence, abundance and/or production of populations' (Barnthouse *et al*, 1987) and 'Environmental policy decision-makers have shifted emphasis from physiological individual-level to population-level impacts of human activities (Emlen, 1989). In some cases, regulatory agencies have made the protection of populations an explicitly stated goal of mandates and regulations (e.g. USEPA, 1991).

This interest is in marked contrast to the ecotoxicity data that are typically generated from standard ecotoxicity tests, which, with few exceptions[3], express effects at the level of the individual. This probably originates from the fact that ecotoxicity testing has its roots in mammalian toxicology (Newman, 2001). We are therefore forced to extrapolate from the data at our disposal (generally describing effects at the individual level) to predict effects at the population level (Figure 1.4).

Level of organisation	Biochemical and Physiological effects	Organs and Tissues	Individuals	Populations	Assemblages and Communities
Availability of standard test methods (e.g. OECD, USEPA, ASTM)			+++	+	
Availability of non-standard test methods	++	+	+++	+	+
Relevance to protection objectives of EQSs	Low →				High

\+ rarely encountered

++ encountered occasionally

+++ likely to be encountered frequently

Figure 1.4: Levels of biological organisation in ecotoxicity testing

Inferring effects on populations from individual-based effects data is subject to four main problems (Newman, 2001):

1. Individual-based toxicity tests often use a life stage that is expected to be particularly sensitive to toxicants. However, the most sensitive life stage may not be the most important as far as maintaining a viable population is concerned, e.g. those species with a strategy of over-production of individuals at early life stages.

2. Although there are some methods available for predicting demographic parameters from survival data (discussed in Chapter 2), the fixed time effects concentrations

[3] Provisions within standard OECD methods for toxicity testing with algae and aquatic plants of the genus *Lemna* allow for the estimation of endpoints describing changes in the growth rate of cell density (algae) or fronds (*Lemna*) which are effectively population-level endpoints.

that are normally reported e.g. '96h LC50' cannot be used in these analyses. Alternative 'time-to-event' analyses that would be suitable in these analyses are not covered in standard test guidelines and do not feature in existing regulatory chemical assessment schemes.

3. Observations of test organisms in standard tests do not usually continue beyond the period of chemical exposure and so possible ecotoxicological effects that could be of significance to population viability are missed.

4. The basic assumption underlying the probit model for analysing ecotoxicity data is that there is 'pre-defined' level of tolerance for all individuals and the distribution of individual effective doses in a population is a log-normal one. An alternative possibility is that the same random processes are occurring in all individuals and the probability of dying is actually the same for all individuals. In the context of the response of individuals it probably makes little difference but it can if we then wish to infer effects on populations.

We argue that inferring population consequences from effects on individual organisms (and to a lesser extent at the sub-organism level) is a key step in the development of environmental thresholds, including EQSs. Much of the need for extrapolation stems from the discrepancy between the data at our disposal and what we are trying to protect (populations). It follows that any measures that can help bridge this gap will inevitably lead to more robust EQSs. In doing so, a major source of uncertainty would be eroded and less reliance need then be placed on default safety factors. Chapters 2 and 3 address possible approaches in more detail.

1.6.4 Accounting for interspecies effects - species sensitivity distributions

Species sensitivity distributions (SSDs) are increasingly being used to account for interspecies sensitivity in higher level risk assessments for water and soil (e.g. Solomon *et al*, 1996; Giesey *et al*, 1999). Essentially, construction of an SSD involves ranking no-effect concentrations[4] for different species, plotting these ranks against log concentration and applying an appropriate model e.g. lognormal, log-logistic. The approach was first recommended by Kooijman (1987) and subsequently developed through application of parametric (i.e. model-based) regressions (Wagner and Lokke, 1991; Aldenberg and Slob, 1993). More recently some authors have argued that there is no reason to assume an underlying distribution for species sensitivities (e.g. Forbes and Forbes, 1993) and this led to development of a non-parametric (i.e. making no assumptions about the underlying error distribution) approach (Jagoe and Newman, 1996).

From the line of best fit applied to the data, certain useful parameters can be estimated e.g. the HC5, the concentration affecting no more than 5% of species. Because this is estimated by regression and its precision can also be estimated, the lower 50% or 95% confidence limit associated with this concentration can also be taken, the latter leading to a more conservative estimate of the acceptable concentration that takes explicit

[4] Effect concentrations e.g. EC50 or LC50 values may be used but the resulting regression parameters refer to concentrations that are predicted to affects half of the population of unprotected species.

account of statistical uncertainty in the fitted line. It follows that the lower confidence interval will be lower (i.e. the outcome is more conservative) when data are sparse.

SSDs were first used for deriving water quality criteria in the US but now also have a prominent role in the derivation of environmental risk limits in the Netherlands, Canadian soil and sediment quality guidelines and soil quality criteria in Denmark. The use of SSDs in a regulatory context is still limited although they are essential to probabilistic approaches to risk assessment. In addition, SSDs have recently been cited as options for estimating a PNEC in the latest revisions of the EU Technical Guidance Document and in proposals for deriving Quality Standards for Priority Substances under the Water Framework Directive (Fraunhofer Institute, 2001).

The relative merits of safety factor and 'modelling' approaches in extrapolating from ecotoxicity data have been the subject of some debate (e.g. Forbes and Forbes, 1993; Smith and Cairns, 1993; USEPA, 2000; Forbes and Calow, 2002). A significant difference between these approaches is the way in which the standard is subsequently expressed. When based on the application of safety factors, the resulting standard is invariably a simple threshold, inviting a simple 'pass/fail' assessment of compliance. In SSD approaches, the relationship between concentration and the proportion of species affected can be described and the regulator can subsequently superimpose the level of protection that is considered appropriate (e.g. 'protection of 95% of species'). While a threshold can still be derived e.g. the HC5, it is supported by additional information describing the effects of higher or lower concentrations, in contrast with safety factor approaches. The approach is not without its limitations however, and these are highlighted alongside a summary of the strengths of the SSD approach in Table 1.2.

Table 1.2: Characteristics of species sensitivity distribution models in defining ecological standards (compared to 'safety factor' approaches to extrapolation)

Strengths	Limitations
Unlike safety factor approaches, all the available data are used Uncertainty is explicitly quantified through the confidence limit associated with the concentration giving the selected level of species protection (e.g. 95% of species) Because large datasets result in lower confidence limits, data generation is not penalised by the risk of a more stringent standard The relationship between concentration and impact can be described in a way that the consequences of different exposure levels can be predicted	Datasets must be large to avoid confidence intervals that are large (and may extend beyond zero) Some of the assumptions inherent in the approach cannot be verified or are probably incorrect: • that test data can be regarded as independent random trials • that the distributions generated accurately reflect the variability between species that are the target of the risk assessment • that the selected model describes the pattern of species sensitivity distribution in nature • that the chosen level of protection (e.g. 95% of species) is adequate Only interspecies differences in sensitivity are dealt with

1.6.5 Probabilistic approaches to risk assessment

We have already noted that SSDs are increasingly playing a prominent role in higher level risk assessments where a probabilistic approach is being adopted i.e. where worst-case, single point estimates of concentrations predicted to occur in the environment (PEC) and predicted no-effect concentrations (PNEC) are replaced by distributions of exposure and effects data (e.g. Barnthouse, 1996; Solomon and Takacs, 2002; Warren-Hicks *et al*, 2002). Some practical applications of a risk-based approach to describing chemical effects within regulatory decision-making are illustrated in Chapter 4.

1.7 A 'Specification' for the Ideal Water Quality Standard

From the previous Sections, it is clear that advances in the scientific credibility of EQSs and also their regulatory utility could be made if certain aspects could be addressed.

From a technical viewpoint these would include:

- Expressions of toxicity data that are more ecologically relevant. Ideally, toxicity should be described in a way that describes as closely as possible the effects of a chemical on population viability rather than the responses (e.g. survival, growth, reproduction) of a group of individual organisms. In addition, a move away from summaries of toxicity that are described in terms of a fixed time e.g. '48h EC50' toward ones that are independent of time would again render the data more relevant to the protection aims of EQSs. Such measures would help close the gap between the assessment endpoint and the protection endpoint and, in so doing, should help reduce reliance on default safety factors because uncertainty (e.g. individual to population, or effects over a fixed time to effects over an unspecified time) has been reduced.

- Reduced reliance on toxicity summaries expressed in terms of a LOEC and NOEC. Instead greater reliance should be placed on the use of point estimates (ECx) describing a low level of effect e.g. EC0, EC10, EC20.

- The development of SSDs that addresses the known weaknesses of the approach e.g. ensuring that SSDs are constructed using a more diverse range of species than is possible from standard test data alone.

- Incorporation of more relevant (particularly population-level) summaries of toxicity into SSDs in which the relationship between concentration and threshold concentrations for populations of different species is described. In other words, SSDs based on population-level endpoints are developed.

A number of the recommendations made by the Royal Commission on Environmental Pollution on the derivation of environmental standards (RCEP, 1998) are also relevant to the derivation of EQSs. Specifically, the RCEP suggest that:

- the views of non-specialists should be addressed in the standard-setting process,

- there should be a clear separation between technical and policy deliberations (e.g. levels of protection, achievability),

- considerations should be documented and auditable,
- there should be a formal review process for standards,
- assumptions made in the standard-setting process should be explicit,
- Standard setting should be done in a transparent way with opportunities for peer review.

1.8 Structure of this Report

In this Chapter, we have sought to describe current practice in the derivation of EQSs in the UK and to analyse both the strengths and the weaknesses of the current methodology. We have then gone on to highlight recent technical developments and to identify a number of issues that have a bearing on the scientific robustness and utility of EQSs.

Much of the remainder of the report addresses two parallel themes:

1. More ecologically relevant expressions of toxicity, that help reduce uncertainty in extrapolation (Chapter 2).
2. A risk-based approach to describing these data based on SSDs (Chapter 3). In particular, we examine novel approaches to extending the understanding of species sensitivity to toxicants by using Bayesian techniques to incorporate expert judgement into the construction of SSDs.

Chapter 4 illustrates how costs and benefits may be addressed using a risk-based approach to standard setting, based on two worked examples. Finally, in Chapter 5, we summarise what can reasonably be achieved based on the experience gained in this research, and make recommendations for the derivation of EQSs for a range of scenarios varying in the type and quantity of data available.

2. DEVELOPMENT OF MORE RELEVANT ECOTOXICOLOGICAL ENDPOINTS

2.1 Introduction

One of the key points in Chapter 1 was the current emphasis placed on determining chemical effects on individuals and the need to predict effect on populations of organisms. In this Chapter we explore the suggestion that ecotoxicity endpoints directly measuring effects at the population level, or that allow us to infer such effects, are often to be favoured over ones that describe effects at the individual level. Methods for determining such endpoints from existing datasets have been developed and are illustrated with worked examples. The practicalities of employing such approaches to a wider range of species have been explored through an extensive data-gathering exercise. Where useful population-level effects cannot be estimated, the possibility of generating more useful summaries based on survival data represents a step forward and we go on to describe how this can be achieved, again with some worked examples and an assessment of the practicality of the approach.

Some of the data generated in these studies has then been employed in constructing SSDs, and that is covered in Chapter 3.

2.1.1 Population-level endpoints

Measurements of toxicity that either measure directly, or allow us to infer, effects on changes in abundance (population size) of a species should be highly relevant to the derivation of EQSs. This is because the gap between the measurement endpoint and protection endpoint is narrowed and much of the uncertainty involved in extrapolation eliminated[5]. The significance of such demographic endpoints in risk assessment is increasingly being recognised (Calow *et al*, 1997) and there is growing interest in making decisions on the basis of endpoints such as risk of extinction (Snell and Serra, 2000; Tanake and Nakanishi, 2000).

Direct measurements of changes in species abundance are rarely performed except in a few laboratory studies (Walthall and Stark, 1997; Sibly, 1999) and in mesocosm and semi-field studies which often monitor chemical impacts on abundance of zooplankton or benthos (e.g. Girling *et al*, 2000). Such data – where they are available – are helpful in informing the derivation of EQSs largely in a validation capacity but they are not routinely available. Their interpretation is also complicated by predator-prey interactions and in separating natural variations in population size of particular species from toxicant-induced effects, with the result that only large effects may be discerned.

A key feature of population growth (or decline) is the rate at which it occurs and this may be positive (population size increases), zero (steady state) or negative (population size decreases). A widely used summary of population growth rate in ecology is r, the

[5] There is still, of course, a question about accounting for interspecies differences in sensitivity. That is addressed in Chapter 3.

population increase per unit time, divided by the number of individuals in the population (also known as the intrinsic rate of increase or per capita growth rate) (Birch, 1948):

$$r = 1/N \, dN/dt$$

where N= the size of a population and t = time

So, if r>0, population size will continue to increase but if it falls below 0, population decline is inevitable and could ultimately lead to extinction. Integration of this equation results in the following equation. This shows that the size of a population will increase exponentially if r is constant (although in nature population growth is usually asymptotic because of density dependence). It also allows us to estimate population size at any time from r and the initial population size, N_o:

$$N_t = N_0 e^{rt}$$

Another commonly used statistic is λ, which represents the factor by which the population is increased in a given time. It is closely related to r and is defined as:

$$\lambda = e^r \text{ (alternatively, } r = \log_e \lambda)$$

If a population doubles in a year, λ = 2. With an initial population size of 100, this predicts a population of 200 in the subsequent year and 400 in the following year, and so on. In this case $r = \log_e 2 = 0.693$. Throughout this report, we tend to refer to r as the endpoint of choice but estimation of λ is equally useful (Forbes and Calow, 2002).

Interpreting changes in population size are complicated by density-dependence (typically the population growth rate declines as population density increases), ultimately stabilising at the 'carrying capacity of the environment' (Walker *et al,* 1996). The equation shown above can be modified to take this into account (Verhulst, 1838 in Newman, 2001) but useful progress can still be made by confining our attention to toxicant-induced effects on population growth rate even without addressing density-dependent effects.

For more details of the theory behind the use of demographic endpoints in ecology and ecotoxicology, and methods for estimating changes in population growth rate, the reader is referred to Newman (2001) and Kammenga and Laskowski (2001).

How can effects on population growth rate be estimated in practice?

Population growth rate depends on three factors:

1. Individuals' birth rates

2. Individuals' death rates

3. Timing of breeding events (e.g. the age at which they first reproduce)

These individual factors or traits can change with age (within a given period, old individuals are more likely to die for example) and so it is necessary to collect data on

age-specific birth and death rates. There are two methods available for analysing these data and estimating r (or λ). The first involves iterative substitution of the Euler-Lotka equation (Newman, 2001) which integrates data for fecundity for individual females, the interval between breeding events and survival time of those females. The following equation can then be rewritten to solve for r:

$$1 = \sum i_t m_t e^{-rt}$$

where:

t is the age of the cohort

i_t is the proportion of individuals surviving to age t

m_t is the number of offspring produced per adult of age t

r is the intrinsic rate of population increase

A more powerful approach involves the recording of these traits at regular intervals and tabulating them in a matrix known as a 'life table' which are then subjected to analysis to estimate individual demographic traits and also r (Caswell, 1989). The importance of individual traits to population growth rate depends strongly on the life history strategy of different species (Calow *et al*, 1997). The life table approach allows the investigator to identify the trait(s) that are most responsible for effects on r. However, it cannot be assumed that if a strong effect of reproduction is seen in one species, it follows that this is the most critical trait in other species because their life history strategies may be very different.

In an early investigation of this type, Daniels and Allan (1981) studied the effect of dieldrin on individual life history traits in the copepod, *Eurytemora affinis*. Each trait was affected by increasing concentrations of dieldrin and these resulted in a decline in r, the intrinsic rate of population increase. In this case, each trait contributed to the effect on r although theoretical examples based on species with different life history strategies (Calow *et al*, 1997) and practical studies (Kuhn *et al*, 2000) clearly show that this is not always the case. It follows that toxicity studies that assess chemical effects on single traits – survivorship or fecundity – will not always be an adequate surrogate for estimations of r. We return to this theme in Section 2.3.1.

It is clear that r is a highly relevant expression of the toxicity of a chemical and one to which more research effort is now being devoted, including the field of chemical risk assessment (e.g. Lin *et al*, 2002). In the following Sections, we describe our contribution to the development of models for estimating r that also quantifies the uncertainty associated with these estimates. We then go on to apply these models to data from two aquatic toxicity studies to assess whether meaningful endpoints can be derived in practice and, in one case, to help inform proposed water quality thresholds.

2.2 Development of Models for Predicting Population-Level Effects

2.2.1 Introduction

Ideally, the input data used in life table response experiments (LTREs) are based on full life cycle studies but these are expensive and time-consuming. However, partial life cycle studies may also generate useful data (Calow *et al*, 1997) and these are routinely employed in regulatory and non-regulatory testing e.g. with algae and invertebrates such as *Daphnia* and *Neomysis*. To be practical in helping derive EQSs, it will be necessary to employ existing data from such standard tests rather than initiate new studies.

Our aim is to generate estimates of r from existing data and to allow the uncertainty associated with this endpoint to be quantified. Uncertainty associated with a given value of r may be estimated by a resampling method to generate a bootstrap confidence interval (Caswell, 2001). Estimates of r for different chemical concentrations can be plotted against chemical concentration from which the concentration corresponding to a critical level of r (e.g. where r =0) can be estimated. However, in fitting a regression line to such data, further uncertainty is introduced, which leaves a question about the accuracy of any threshold concentrations estimated from this regression.

In Appendix A, we describe a novel 'double bootstrap' approach which can be used to generate a confidence interval for quantifying the uncertainty associated with interpolated estimates of concentrations giving rise to a particular value for r. This approach has been applied to two datasets, with the aim of testing the practicalities of the approach and in the second case, in helping inform the derivation of thresholds. The first dataset is from a chronic study carried out at the Environment Agency laboratories at Waterlooville with the water flea *Daphnia magna* exposed to zinc sulphate. The second is from a fish life cycle study with fathead minnows (*Pimephales promelas*) exposed to the steroid oestrogen, 17α-ethinyloestradiol. Details of the datasets for these two substances and the analyses are shown in Appendices B and C, respectively, but key points are summarised below (Section 2.2.2).

2.2.2 Worked examples

Daphnia and zinc sulphate

A 21-day study was undertaken at The Environment Agency's Waterlooville laboratory to investigate the effects of zinc sulphate on demographic traits (survival and reproduction) in the water flea, *Daphnia magna*. Details are given in Appendix B but the experiments were conducted according to OECD test guidelines (OECD, 1984) in which test animals were held individually so that brood dates and number of offspring could be assigned to individual animals. Zinc sulphate was introduced at a range of concentrations between 0.022 and 0.46 mg/L.

Data were used to estimate EC10, EC20 and EC50 values for reproduction inhibition (as specified in OECD test guideline 202) and also concentrations describing different levels of effect on the intrinsic population growth rate, r (ErC10, ErC20 and ErC50 values) from the LTRE analysis described in Appendix A. In the latter case, ErCx values were estimated either using a linear regression or quadratic regression. Key

statistics from this study (ECx values and associated confidence intervals) are shown in Table 2.1.

Table 2.1: Key statistics from chronic study with *Daphnia magna* and zinc sulphate – comparison of conventional endpoints and estimates of r (all as mg/l; values in brackets represent 95% confidence intervals)

	Reproduction inhibition	**Intrinsic population growth rate, r**	
x%	**ECx**	**ErCx (linear regression)**	**ErCx (quadratic regression)**
10	0.07 (0.03-0.57)	0.11 (0.04-0.16)	0.21 (0.10-0.31)
20	0.60 (0.06-0.88)	0.29 (0.19-0.44)	0.36 (0.27-0.42)
50	1.10 (0.79-1.22)	0.90 (0.55-1.48)	0.58 (0.47-0.81)

In this study, lethal effects were not apparent and even at the highest concentration tested (0.46 mg/L), one way ANOVA followed by Dunnett's test did not reveal a significant effect on reproduction and so it was not possible to estimate a NOEC. Comparing estimates of r with the conventional expression of reproduction inhibition showed that ErC20 and ErC50 values from LTRE analysis were both lower than the corresponding EC20 and EC50 values i.e. r was the more sensitive endpoint. In addition, the estimates of uncertainty (as indicated by the confidence intervals) were lower when based on ErCx than ECx determinations. This tendency is supported by Grist *et al* (2003). Although the EC10 estimate (0.07 mg/L) was lower than the corresponding ErC10 values (0.11 mg/l for the linear regression and 0.21 mg/L for the quadratic regression), this was offset by the uncertainty in their derivations: the confidence intervals for both the ErC10 estimates were wholly contained within the much wider confidence interval for the conventional EC10.

It is interesting to note that the confidence intervals using the quadratic regression were narrower than for the linear regression approach, indicating a better fit of the quadratic function to the data. Although Snell and Serra (2000) found that several models predicted similar extinction probabilities, Tanaka and Nakanishi (2001) found that the response of r to chemical exposure for a range of chemicals and species was approximately quadratic in most cases.

These results indicate that, at least in this particular dataset, r can be estimated with greater precision than conventional estimates of reproduction inhibition as specified in OECD 202. Furthermore, a decision based on the conventional reproduction endpoint may underestimate the effects of zinc sulphate by a factor of up to 2-3 (depending on the fitted model) when compared to its effects on the intrinsic rate of population increase, r.

Fathead minnow and 17α-ethinyloestradiol

In 2002, environmental thresholds for 17α-ethinyloestradiol were proposed using the EQS methodology outlined in Chapter 1(Young *et al*, 2002). Most of the critical studies on which these thresholds were based are from long-term studies with fish in which effects on individuals were monitored. Whilst there is a clear link between the endpoints measured (e.g. changes in sexual characteristics of males, altered sex ratios, fertilisation success and egg production) and the recruitment of offspring to populations, effects of 17α-ethinyloestradiol at the population level were not estimated directly.

One of the critical studies informing the derivation of thresholds for 17α-ethinyloestradiol involved a full life cycle exposure of fathead minnows (*Pimephales promelas*) (Laenge *et al*, 2001). In this study, a NOEC of 1 ng/l was estimated based on feminisation of the exposed fish. Using a set of LTREs, we estimated the effect of different concentrations of 17α-ethinyloestradiol on the intrinsic rate of population growth, λ. From this, we predicted concentrations that would bring about eventual extinction of populations of fathead minnows, along with concentrations which may be regarded as causing no significant effects. As in the previous example, these have been compared with corresponding concentrations for conventional endpoints to assess whether thresholds based on such endpoints are likely to be adequate for protecting fish populations. Details of the study and the subsequent LTRE analysis are shown in Appendix C.

Records of age-specific survival and reproduction of fish exposed to different concentrations of 17α-ethinyloestradiol (Laenge *et al*, 2001) were re-analysed according to the methods described in Appendix A. Life tables of survivorship and egg production were constructed for each test concentration (except for the 16 and 64 ng/l treatments where high toxicity meant that λ could not be usefully estimated). Data were combined into weekly age classes and the probability of surviving up to age class x determined, along with the mean number of offspring produced by an individual in age class x. Account was taken of the periodicity in egg laying (photoperiod-driven) and also the removal of individuals during the experiment. Data were used to estimate ErC20, ErC50 and ErC100 values from the LTRE analysis. ErCx values were estimated either using a linear regression or quadratic regression. Key statistics from this study (ErCx values and associated confidence intervals) are shown in Table 2.2.

Table 2.2: Estimates of r from life cycle study with fathead minnows exposed to 17α-ethinyloestradiol (values in brackets represent 95% confidence intervals)

Reduction in population growth rate *r*	E*r*Cx (ng/L)	
x %	Linear (R^2 =0.97) (95% pointwise CI)	Quadratic (R^2 =0.99) (95% pointwise CI)
20	0.78 (0.57-0.96)	1.35 (0.73-1.92)
50	1.66 (1.51-1.86)	2.33 (1.64-2.72)
100	3.11 (3.04-3.51)	3.42 (3.12-3.66)

Sensitivity analysis of the contributions made by individual traits confirms that reductions in λ are mainly due to reduced fertility rather than survival. The results project population growth at 0, 0.2, and 1.0 ng/l 17α-ethinyloestradiol. However, at 4 ng/l, there was total inhibition of reproduction leading to a negative value of λ. Under these circumstances, population extinction is assured. A 'threshold' concentration, where a zero population growth rate (i.e. λ =0) is predicted, lies at a concentration between 1 and 4 ng/l 17α-ethinyloestradiol.

The relationship between 17α-ethinyloestradiol concentration and λ has been described using both a linear function (f(x)=ax+b) and a quadratic function ($f(x)=ax^2+bx+c$). In these functions, the critical concentration i.e. zero population growth rate (λ =0) is 3.1 or 3.4 ng/l, respectively. Furthermore, ErC20 values of 0.78 ng/l (linear) or 1.35 ng/l (quadratic) are estimated. These may be regarded as an estimate of the 'no-effect' concentration for continuous toxicity data (Bruce and Versteeg, 1992) and so provide a means of comparing directly with NOECs from the original Laenge *et al*, (2001) study in which reproductive endpoints were monitored. These ErC_{20} values correspond closely to the NOEC of 1 ng/l estimated by Laenge *et al* (2001) for sex reversal in the same study, suggesting that the thresholds for phenotypic sex reversal and population growth - at least in this experiment - are very similar.

The immediate conclusion is that environmental thresholds for 17α-ethinyloestradiol based on conventional endpoints are likely to afford adequate protection at the population level, at least for fathead minnows. It is reasonable to suppose that the same principle would apply to other fish species with similar life history strategies. More generally, this analysis also shows how existing data may be used to project the effects of chemicals to population level without reverting to standard endpoints that describe effects on survival, growth or reproduction in isolation.

2.2.3 Summary

The example illustrated for 17α-ethinyloestradiol in particular shows how demographic analysis of existing ecotoxicological data can assist the decision-making process for deciding 'acceptable' concentrations of chemicals in environmental risk assessment or in deriving environmental benchmarks. Furthermore, describing the results of demographic analyses such as those described in Appendices A-C in terms of the 'risk of extinction' is appealing because it is both meaningful and easy to communicate to a lay audience. Indeed, expressing effect in this way has been used recently by other researchers (e.g. Snell and Serra, 2000; Iwasa *et al*, 2000; Tanake and Nakanishi, 2000).

The method described in Appendix A represents an advance on existing approaches to the analysis of life table data because uncertainty in threshold concentrations can be quantified. Both linear and quadratic functions have been attempted. Although the quadratic function gives rise to narrower confidence intervals and is thought to be more appropriate by other workers, the option to use either a linear or quadratic function should be retained and the one giving the best fit to the experimental data employed.

We must recognise the limitations of population-level studies in the laboratory. Density-dependent effects are rarely accounted for, nor is competition and predation from other species (although this applies equally to conventional measures of toxicity). Life history data for natural populations can be developed (e.g. Brown *et al*, in press) but these are currently the exception. The duration of laboratory studies is also usually less than the nominative state. For example, chronic studies with *Daphnia* take 21 days but the average life cycle is between 60 and 90 days and natural populations of fathead minnow have been shown to contain two- and even three-year old fish (Carlson, 1967). By underestimating the life cycle, the sustainability of the populations may have been over-estimated unless these unaccounted, older life stages contribute to the reproductive capacity of the population, in which case sustainability may have been under-estimated. Ideally, estimation of r would be calculated over at least an entire life cycle.

There remains the question 'what level of effect is acceptable?' An ErC100 would not be acceptable because it is the maximum concentration that a population can endure before a projected decline towards extinction ensues. In reality, the ErC100 is unlikely to be protective because of fluctuations in the population that will result from demographic and environmental variability and, in the longer term, genetic drift. Balanced against this is the suggestion that population density dependent effects may tend to compensate for the effects of chemical stressors because, as the population declines, more food is available for survivors (Calow *et al*, 1997). Based on studies with rotifers, Snell and Serra (2000) reported that reductions in r greater than ca. 30% (the greatest reduction induced in their studies) would give rise to near certain extinction within 100 years and that a '5% reduction in r is probably the maximum tolerable for the long-term persistence of rotifer populations'. Timescales of 100 years seem excessive and a reasonable approach might be to specify the critical concentration as one where r = 0 because, by definition, above this concentration, population decline is projected. In practice, one may then take the lower confidence limit for the concentration corresponding to r = 0 as a conservative estimate of the threshold concentration.

Where demographic data are available for a species (or several species) and there is good reason to believe that this species is amongst the most sensitive to a chemical, then

it may not be necessary to devote further effort to estimating population-level effects for other species. This could well be the case for one of the substances described above (17α-ethinyloestradiol) where an understanding of fish physiology would indicate that the most sensitive taxon to this chemical is likely to be fish. Within fish, interspecies differences are likely to be modest (Barnthouse *et al*, 1990) and so only a small safety factor need be applied to account for interspecies differences in sensitivity. In this particular case, the LTRE analysis was helpful in supporting a decision about a proposed threshold for 17α-ethinyloestradiol. However, we must be aware of recent research that suggests endocrine disruptors competing for a receptor site normally occupied by endogenous substances may not have thresholds (Sheehan, 2000).

Of course, this will not always be the case and relevant data for a range of species will be desirable. The practicality of achieving that goal is considered next.

2.3 Feasibility Study

Most of the reported examples of the use of demographic analysis of ecotoxicity data (including the worked examples described above) have been detailed analyses within species. Having established the theoretical and practical validity of the LTRE approach (Section 2.2), we now investigate whether it is possible to extend the approach across different species for the same chemical. Ideally, the taxonomic 'spread' should be sufficient to construct meaningful SSDs for these chemicals (Forbes and Calow, 2002). If so, many of the features of an 'ideal' EQS (Section 1.7) could be fulfilled because the resulting threshold benefits from both the greater ecological relevance of the endpoints discussed in this Chapter and the risk-based utility offered by the SSD.

In this Section we address the practicality of collecting, or generating, demographic data for a number of species exposed to the same chemical in laboratory studies.

2.3.1 Potential availability of demographic data

Existing estimates of r or λ

In recent years, an increasing number of aquatic studies have been designed to estimate demographic endpoints such as the intrinsic rate of population increase, r. Calow and Forbes (1999) review studies for which estimates of r or λ have been reported alongside toxicity summaries based on individual-level traits such as survival or reproduction. They identified 41 examples representing 28 different species and 44 chemicals, although a significant proportion related to soil-dwelling species. Moreover, only six invertebrate phyla were represented, belonging to a total of eight classes. Of the 41 studies examined, 17 employed one or more *Daphnia* species.

Although estimates of r or λ in the absence of data for individual traits were excluded from Forbes and Calow's (1999) analysis, it is evident that the reporting of r or λ is by no means routine. Our own analysis supports this view. Generally, there are insufficient literature descriptions of r or λ to inform the derivation of EQSs except in cases where there is strong evidence that such endpoints for particularly sensitive species are available.

Suitability of existing data for re-analysis to estimate r or λ

In practice, it will be necessary to reanalyse data to generate more ecologically relevant endpoints like r. The worked examples described above show that data extracted from aquatic tests designed to generate conventional endpoints such as NOECs for fecundity or other reproductive parameters can sometimes be used to generate estimates of r. However, these analyses require access to raw data and, specifically, observations of age-specific survival and reproduction over an extended period.

A number of standard test methods are capable of generating data that is suitable for LTRE analysis, notably invertebrate reproduction studies and, as illustrated above, fish life cycle studies. It is reasonable to suppose that, in many cases, data that are suitable for demographic analysis are probably collected but not routinely reported. In test guidelines for determining growth inhibition in unicellular algae and the aquatic plant, *Lemna*, population-level endpoints are specifically covered as an option within the test guidelines (growth rate of the test population of cells (algae) or fronds (*Lemna*)). However, this useful feature of these test methods does not seem to be widely understood. There is a tendency to express EC50 values from these tests simply in terms of the 'standing crop' (final biomass) at the end of the test, possibly because this often leads to a lower EC50 and is therefore favoured by regulators (Smrchek, USEPA, *pers. comm*.). Nevertheless, the technical merits of the growth rate endpoint in both algal and *Lemna* tests are well-documented (Nyholm, 1990; Huebert and Shay, 1993).

2.3.2 Acquisition of test data

Clearly, there is significant potential to generate more ecologically relevant endpoints using existing study data. A shortlist of substances for further investigation was developed based on the following criteria:

- A substantial quantity of acute and chronic aquatic toxicity data was known to be available.
- EQSs were available (for comparative purposes).
- The substances were amenable to chemical analysis in the event that practical validation studies were required.
- The substances were ones that field biologists were likely to be familiar with and could express a view about relative species sensitivity in the field[6].

[6] This was important for the following Chapter, dealing with construction of SSDs

The following substances were highlighted:

Substance	Key properties
Cypermethrin	Pyrethroid insecticide with predominantly acute effects. Implicated in sheep dip pollution incidents[7]
Chlorpyriphos	OP insecticide with predominantly acute effects. Extensive ecotoxicity dataset and implicated in some pollution incidents[7]
Ammonia	Common contaminant in surface waters and high familiarity by field biologists. Extensive ecotoxicity dataset, at least for fish
Zinc	Metal with extensive ecotoxicity dataset
Cadmium	Metal with extensive ecotoxicity dataset
Phenol	Organic contaminant with extensive ecotoxicity dataset for wide range of taxa
Nonylphenol	Organic contaminant generated from degradation of nonylphenol ethoxylates. Subject to current assessment under Existing Substances Regulations. Extensive body of ecotoxicity data has been generated in recent years

2.3.3 Data collection

An exhaustive search was undertaken to identify published (and unpublished) studies in which aquatic toxicity to freshwater organisms had been determined following exposure to the substances listed above. This yielded approximately 1100 studies, covering the years of 1940 to 2001. Emphasis was then placed on studies of chronic toxicity where reproduction was one of the test endpoints determined. This reduced the number of relevant studies to 385.

For each of the selected studies, authors were contacted by letter, explaining the background to our interest in the study and requesting raw data for survivorship and fecundity that we may re-analyse to generate population-level endpoints. This request was endorsed by an accompanying letter from the Environment Agency; an example of the letter sent is given in Appendix D. In total 356 letters were sent out to the authors of studies carried out between the years 1985 to 2001.

Data returned as a result of these requests was subjected to review, especially on whether the technical criteria for age-specific survivorship and fecundity could be met.

[7] Helen Wilkinson, Environment Agency, *pers. comm*.

In addition to this strategy, published (and unpublished) studies in which r or λ had been estimated for freshwater organisms exposed to the seven substances listed above were also identified, to supplement those obtained from the contact by letter.

2.3.4 Results

Appendix E summarises the raw datasets returned and their suitability for LTRE analysis. Of the 385 letters that were sent out 85 replies were returned. However, only 34 of the replies actually contained any raw data and only 6 of those contained both fecundity and mortality data suitable for LTRE analysis. The poor response was due to the fact that in many cases, data were being sought from studies that had been carried out many years previously and experimenters were no longer contactable. Furthermore, there were contractual difficulties in releasing study data from some investigations, especially where these had been performed to GLP for regulatory submission purposes.

Also included in Appendix E are the details of literature studies where r or λ had been calculated for the chemicals of interest. A total of 18 studies were identified, for five of the seven chemicals.

Examination of the data in Appendix E shows that, of the datasets made available, only a small proportion contained raw data in a form that would make them amenable to LTRE. This condition could be fulfilled to the greatest extent for cadmium (12 datasets were considered suitable for LTRE, including studies described in the scientific literature containing reported values of r or λ). However, closer examination of these studies shows they represent an extremely narrow range of species (Table 2.3). All but one of the five species for which suitable data could be located were daphnids. Although the criteria for the minimum number of datasets for constructing a SSD could be fulfilled, it was clear that to attempt to construct an SSD based on such narrow taxonomic diversity would be meaningless. Indeed, this highlights one of the main limitations of the SSD approach, that of ensuring there is a random draw of species in populating the SSD (Forbes and Calow, 2002). The preponderance of *Daphnia* data is almost certainly a reflection of the existence of test guidelines and the regulatory requirement for data for this genus in chemical approval schemes.

Turning to the other chemicals (Table 2.3), even by combining suitable datasets with literature r or λ values, only a very small range of species could be represented. Only 2 species were available for nonylphenol and phenol, and only 1 or 2 datasets were obtained for most of the other substances. In the cases of ammonia and cypermethrin, no suitable datasets were obtained.

Table 2.3: Number of species and chemicals for which 'r' or λ could be generated

Chemical	Number of species for which 'r' or λ could be calculated	Number of species for which literature values of 'r' or λ are given	Total number of species for which 'r' or λ could be estimated
Ammonia	0	0	0
Cypermethrin	0	0	0
Nonylphenol	1 - *Ceriodaphnia dubia*	1 - *Brachionus calyciforus*	2
Chlorpyriphos	1 - *Daphnia pulex*	1 - *Brachionus calyciforus*	2
Cadmium	0	5 - *Moina macrocopa, Daphnia pulex*, Brachionus calyciforus, Daphnia magna**, Daphnia carinata*	5 (from 12 studies)
Zinc	1 - *Hyalella azteca*	2 - *Moina macrocopa, Dinophilus gyrociliatus*	3
Phenol	1 - *Daphnia magna*	1 - *Brachionus calyciforus*	2

*from three separate studies **from six separate studies

This lack of suitable data is frustrating when there is every reason to believe that there are useful datasets in existence. The problem is more likely one of access than one of the data not being available. Nevertheless, this is a realistic situation and if reliance was to be placed on soliciting raw data in this way for the purposes of EQS setting, this experience suggests it may not be a fruitful exercise.

At this point, two other possibilities were considered:

1. Are conventional endpoints likely to be protective of demographic endpoints?

2. Can useful progress be made in estimating more relevant summaries of acute toxicity data?

2.4 Are Conventional Endpoints Protective of Demographic Endpoints?

Forbes and Calow (1999) undertook a comparison of the sensitivity of conventional 'individual-level' endpoints and r or λ, calculated from the same study data. Their main findings, simplified into whether r/λ was less than, equal to or more sensitive than effects based on individual traits are summarised in Figure 2.1.

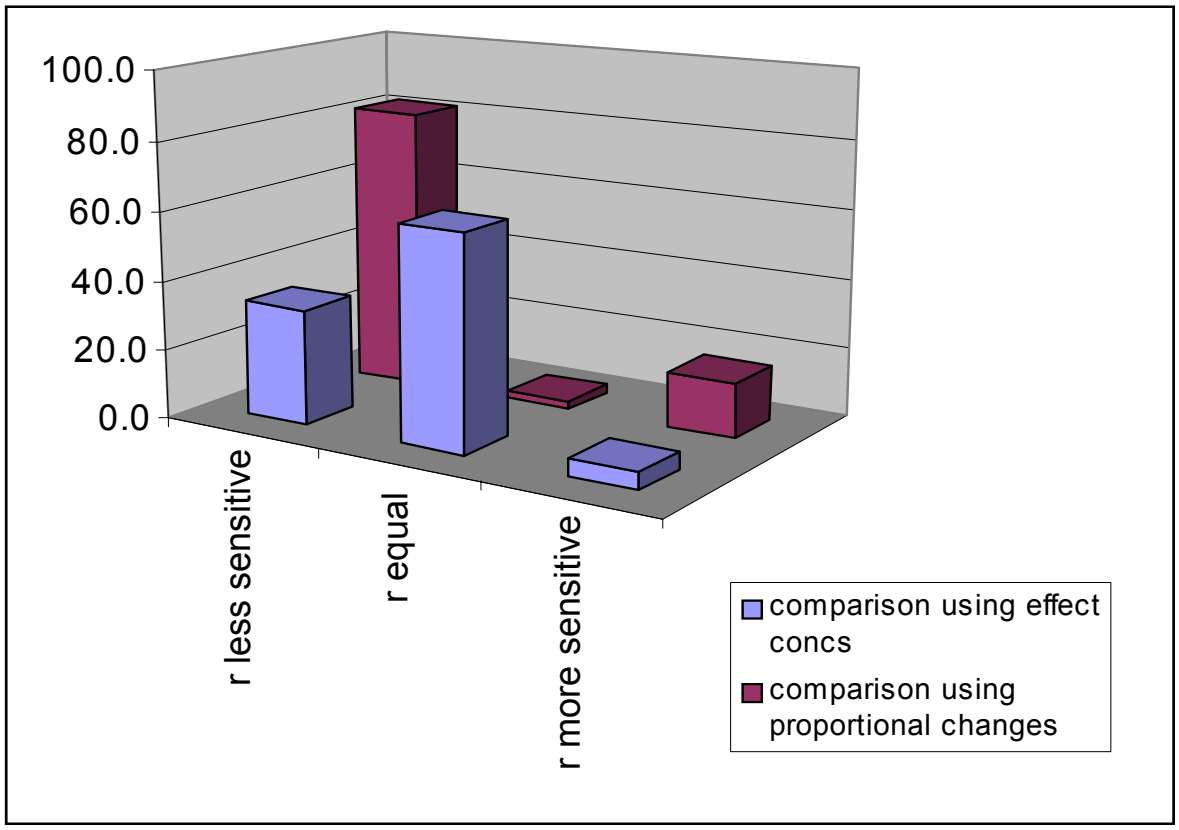


Figure 2.1: Comparison of the sensitivity of demographic endpoints and individual-level traits (Forbes and Calow, 1999)

Their data suggests that, in most cases, r or λ was equally sensitive to or less sensitive to toxicant exposure than individual life history traits contributing to it. The logical conclusion from this is that in a high proportion of cases, decisions based on conventional endpoints (survival and reproduction) would be adequately protective of effects described in terms of r, although data for both traits would need to be available. Our own experience based on the two worked examples is that one study gave rise to a similar outcome (fathead minnows and 17α-ethinyloestradiol) but that slightly greater sensitivity was seen in the study with *Daphnia* and zinc sulphate.

2.5 More Relevant Summaries of Acute Toxicity

2.5.1 Introduction

Survivorship is obviously an important endpoint because it bears directly on population size and may indeed be a protection goal in its own right, especially for species at higher trophic levels. Conventionally, survival data is obtained from short-term exposures in which the median lethal concentration (LC50) is estimated. However, these summaries are always time-dependent and, for many substances, substantial

changes in the LC50 over time are found. This gives rise to difficulties in extrapolation, especially when data from studies of different duration are available and where the protection aim does not specify any particular exposure period (as is usually the case). In short, conventional expressions of survival data are rather remote from the protection objectives of EQSs. Steps to narrow that gap, e.g. by expressing survival data in terms of 'no effects' or in a way that does not relate only to effects after a particular exposure period are therefore to be encouraged.

In the absence of full demographic data, time-to-event approaches that take account of the time course of toxicity can reduce uncertainty in an important area for risk assessors. Sprague (1969) wrote of its importance, and thresholds for median lethal *times* (LT50) were frequently reported before the 1970s. Since then, more powerful approaches than the LT50 have been developed and used in medicine, agriculture, engineering and ecology (Crane *et al.*, 2002a). There is little doubt that risk assessment in the field of environmental toxicology and chemistry can also be improved by use of time to event models (Crane *et al.*, 2002b). The availability of literature to describe different time to event approaches has increased (e.g., Crane *et al.*, 2002a), and the techniques are widely used in other fields. Cox proportional models, amongst the most commonly used semi-parametric approaches in time to event analysis, are based upon an original paper that is one of the top ten cited articles in science today (Cox, 1972; Cox and Oakes, 1984). Fully parametric and nonparametric approaches are also available (Newman and Crane 2002). Although a reasonable level of numeracy is required to use some of the more sophisticated time to event approaches, Crane and Grosso (2002) show that even the modestly numerate can gain some rewards from use of time to event in standard software packages.

There are some clear advantages to the use of time to event approaches. Time to event methods explicitly address the twin causes of toxic effects - the intensity and duration of exposure to hazardous chemicals, and they make better use of the data gathered, often at great expense, from toxicity experiments. Diverse end points in time can be addressed with time to event approaches, and individual organism characteristics can be incorporated as covariables, as shown by Newman and McCloskey (2002). Because more data are used it is possible to select more appropriate statistical models, and to relate these models to population level effects (Caswell and John, 1992), and epidemiological and exposure models (Newman and McCloskey, 2002; Karman 2002). Data from field studies can be analysed as readily as data from laboratory studies (Manly, 2002). Because of this flexibility, both prospective and retrospective risk assessments can benefit from use of time to event analyses. Time to event approaches are also likely to be most effective for analysis of toxicity from pulsed exposures, and of latent toxic effects emerging after exposure has ceased, because both of these phenomena are time-related.

An added attraction is that projection of likely toxic effects in short-term experiments to effects over the long-term is possible (Mayer *et al.* 2002). Expression of results from toxicity experiments as *risks* or as reduced life expectancy, rather than as LC50 or NOEC summaries, is likely to aid in communication of the absolute or relative dangers of particular chemical exposure scenarios (Crane and Grosso, 2002; Newman and McCloskey, 2002). Finally, from an animal welfare point of view, fewer animals may be necessary in some experimental designs if the increased statistical power from time to event approaches is traded off against reduced numbers of test organisms (Dixon, 2002).

Time to event approaches do have some negative points which could count against them. Currently they are not the approach taken by most environmental toxicologists, and recording of data at several different time points is more time-consuming than recording at one final time point. However, interval censored data can be analysed by several time to event approaches, including the life tables used for many years by actuaries and ecologists (Pyke and Thompson, 1986; Leslie *et al.*, 1955; Daniels and Allan, 1981; Day and Kaushik, 1987; Pesch *et al.*, 1991; Martínez-Jerónimo *et al.*, 1993). In addition, video technology is now inexpensive and accessible and could make the recording of exact results from toxicity tests much easier (Baatrup *et al.*, 2001).

2.5.2 Availability of survivorship data

As indicated previously a large quantity of the raw data received from our request for raw data was not suitable for estimating r or λ by LTRE analysis. However, a substantial quantity of raw data describing survivorship of test organisms was available (34 responses), much of it (17 studies) suitable for estimating time-independent summaries of survival (i.e. survival at a range of exposure times had been recorded). Table 2.4 summarises these datasets.

Although raw mortality data were available for more species than for fecundity data, the number and taxonomic spread of the species is still not ideal. Datasets for chlorpyriphos and ammonia contained the most representatives from different species. Representatives of insects, crustaceans and fish were available for chlorpyriphos whilst for ammonia, raw data were only located for fish species. However, fish are known to be particularly sensitive to exposure to ammonia and so this was regarded as a potentially valuable set of data. Of the other chemicals, only nonylphenol contained potentially useful data, with raw mortality data for four species representing crustaceans, insects and fish. Datasets for cadmium and zinc contained data for three species. In Section 2.5.3, we explain how time-independent summaries of toxicity for a range of species were estimated using datasets for chlorpyriphos and ammonia.

Table 2.4: Chemicals and species for which data describing survivorship at a range of times was collected

Chemical	Species	Reference	Total number of species
Ammonia			5
	Cyprinus carpio	Mallet *et al* (1988)	
	Onchorhynchus mykiss	Mallet *et al* (1992)	
	Onchorhyncus mykiss	WRc (no date)	
	Perca fluvatilis	WRc (1979)	
	Rutilus rutilus	WRc (1979)	
	Rutilus rutilus	Mallet *et al* (1992)	
	Salmo trutta	Seager *et al* (1990)	
Cadmium			3
	Daphnia magna	Berrata *et al* (1998)	
	Daphnia magna	Van Leuwen (1985)	
	Daphnia magna	Weltens (2000)	
	Gammarus pulex	Borgmann *et al* (1989)	
	Hyallela azteca	Borgmann *et al* (1989)	
Chlorpyriphos			9
	Daphnia pulex	Van der Hoeven (1997)	
	Chaoborus sp.	van Wijngaarden (1993)	
	Cloeon dipterum	van Wijngaarden (1993)	
	Corixa punctata	van Wijngaarden (1993)	
	Simocephalus vetulus	van Wijngaarden (1993)	
	Daphnia longiseta	van Wijngaarden (1993)	
	Gammarus pulex	van Wijngaarden (1993)	
	Gasterosteus aculeatus	van Wijngaarden (1993)	
	Pungitius pungitius	van Wijngaarden (1993)	
Nonylphenol			4
	Ceriodaphnia dubia	CMA (1995)	
	Ceriodaphnia dubia	England (1995)	
	Gammarus pulex	Sims *et al* (1998)	
	Ischnura elegans	Sims *et al* (1998)	
	Pimephales promelas	CMA (1991)	
Phenol			1
	Daphnia magna	Tisler (1999)	
Zinc			3
	Daphnia magna	Berrata (1998)	
	Daphnia magna	Scott (1995)	
	Daphnia magna	Weltens(2000)	
	Hyalella azteca	Borgmann (1993)	
	Noemacheilus barbatulus	Solbe and Flook (1975)	
Cypermethrin			0

2.5.3 Time to event analyses

Raw survival data after 48-h and 96-h exposure to chlorpyriphos were made available by Rene van Wijngaarden of Alterra for the following species: *Chaoborus* sp. (phantom midge larva), *Cloeon dipterum* (mayfly nymph), *Corixa punctata* (water boatman), *Simocephalus vetulus* (waterflea), *Daphnia longiseta* (waterflea) *Gammarus pulex* (shrimp) *Gasterosteus aculeatus* (3-spined stickleback), and *Pungitius pungitius* (10-spined stickleback).

Mayer *et al*. (2002) describe a two-step linear regression method for estimating time-independent LC0 values from raw survival data. We used this approach to analyse the van Wijngaarden data and produce estimated LC0 values and 95% confidence limits for the eight species. We also used a double bootstrap approach (Grist *et al.*, 2002) with more robust statistical properties to the original two-step regression methodology to estimate LC0 values. Essentially this endpoint describes chemical concentrations that are predicted to result in no effects on survival after an infinite exposure period. These data were used to construct SSDs and to compare them with SSDs for other survival endpoints (e.g. fixed time LC10 and LC50 values).

The same approach was used to estimate time-independent LC0 estimates for ammonia data and, again, used to construct SSDs.

2.5.4 Results

Figure 2.2 summarises the time to event analyses of van Wijngaarden's chlorpyriphos data, presenting them alongside the 96-h LC50 and LC10 data that he reported (van Wijngaarden *et al.*, 1993). The Mayer *et al.* (2002) two-step regression approach produced estimates of LC0 values very similar to those produced by the double bootstrap, except in the lower tail of the distribution, near the HC5[8], where the Mayer *et al*. approach produced a more conservative estimate. However, even this was contained within the 95% confidence interval for the bootstrapped SSD. LC50 and LC10 values reported in the original paper by Wijngaarden *et al.* (1993) generally fell outside the bootstrap SSD confidence interval except, interestingly, in the more important lower tail of the distribution.

The present study shows that for chlorpyriphos, use of a two-step linear regression or a double bootstrap time-to-event approach to estimate LC0 values leads to generally similar estimates of toxicity, with an HC5 of around 0.01 µg/L. This estimate is also close to the 96-h LC10 value of 0.02 µg/L estimated for *Gammarus pulex*, the most sensitive species tested by Wijngaarden *et al.* (1993). However, the 95% confidence interval for the double bootstrap suggests that the LC0 could be as low as 0.0001µg/L, which is considerably lower than the lower 95% confidence limit of 0.01 µg/L on the LC10 estimated by Wijngaarden *et al.* (1993). In general, it seems from other studies that LC/EC values for chlorpyriphos can be broadly predictive of longer-term toxic effects, and do not appear to over- or under-estimate them greatly. The mean acute to chronic ratio (ACR) seems to be relatively small for chlorpyriphos, as found by Giesy *et*

[8] The concentration limit required to protect 95% of species at the given level of effect e.g. LC50

al. (1999), although they find an ACR range of 1.4 to 181. Crane *et al.* (submitted) concluded that chlorpyriphos is highly toxic to arthropods, with the water flea *Ceriodaphnia dubia* the most sensitive species on the AQUIRE database with a 96-h LC50 of 0.057 μg/L. These data compare well with the estimates of HC5 concentrations found in the present study and the LC10 estimate for *G. pulex* in the study by van Wijngaarden *et al.* (1993), and suggest that concentrations should be less than 0.1 μg/L to protect against individual mortality in aquatic systems.

Figure 2.3 summarises data for ammonia based on fish survival data[9]. Only 48h LC50 data are available for comparative purposes but again, the LC0 SSD is displaced to the left, indicating a more sensitive endpoint (as expected) by approximately an order of magnitude. This time, the 95% confidence intervals exclude all the LC50 data from the same experiments.

Mesocosm results may help in 'ground-truthing' laboratory estimates, although these systems cannot fully represent the range of natural water bodies and taxa that could potentially be impacted in the natural environment (Crane, 1997). Giesy *et al.* (1999) reviewed available mesocosm data and concluded that effects on invertebrates could be measured at concentrations of chlorpyriphos >0.2 μg/L, with recovery of most populations within 2-8 weeks, and that effects on fishes occurred at concentrations >0.5 μg/L (Giddings *et al.*, 1997) These values are similar to those discussed above, which suggests that LC/EC50 estimates of toxicity from single-species studies might be reasonably protective of field populations.

[9] Individual data points cannot be shown in Fig. 2.3 because the output of the analysis is a function as shown in Fig. 2.3 rather than individual data points

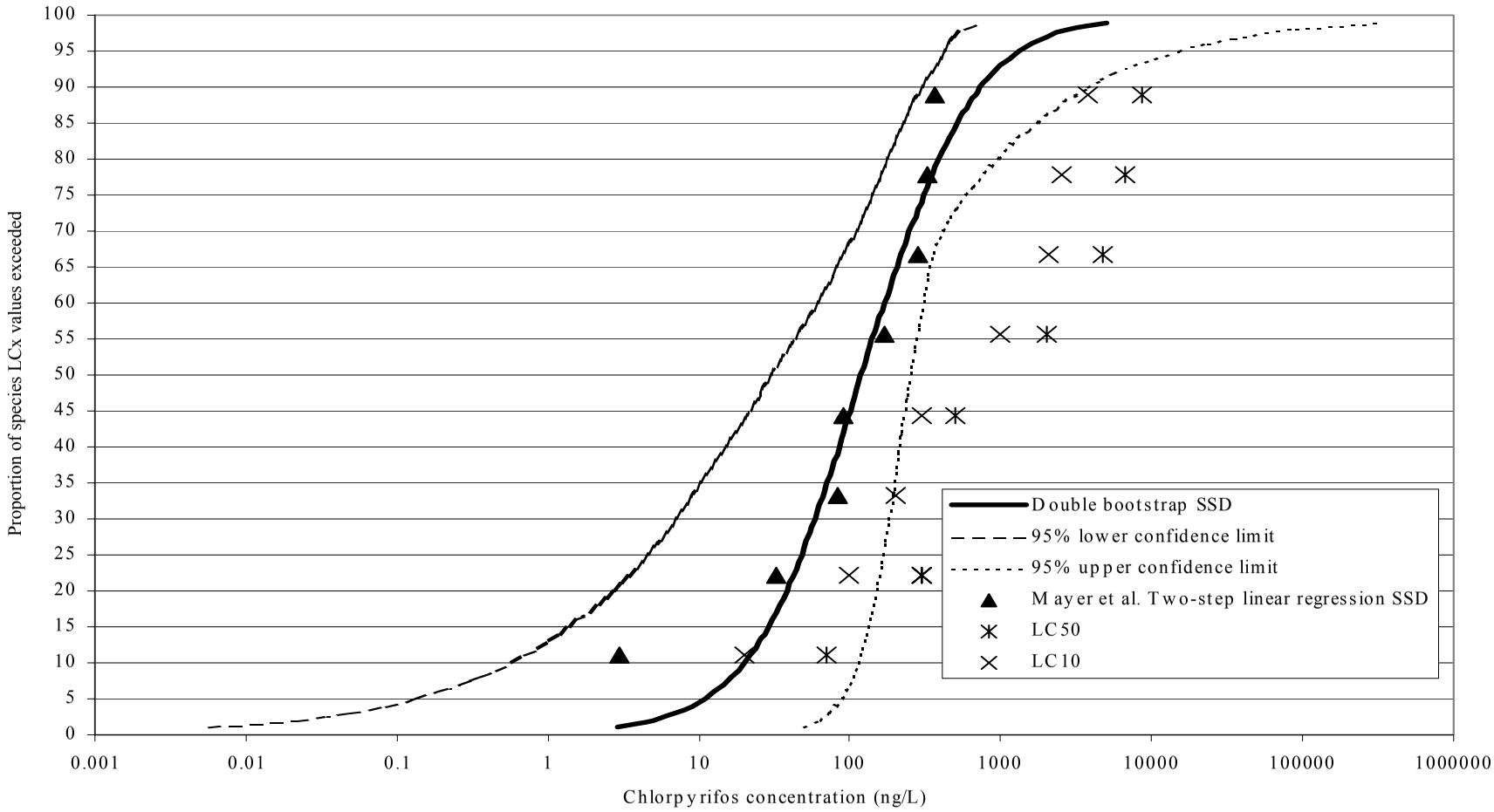


Figure 2.2: SSDs based on 96h LC50, LC10 and time-independent LC0 values for chlorpyriphos from Wijngaarden *et al* (1993) data

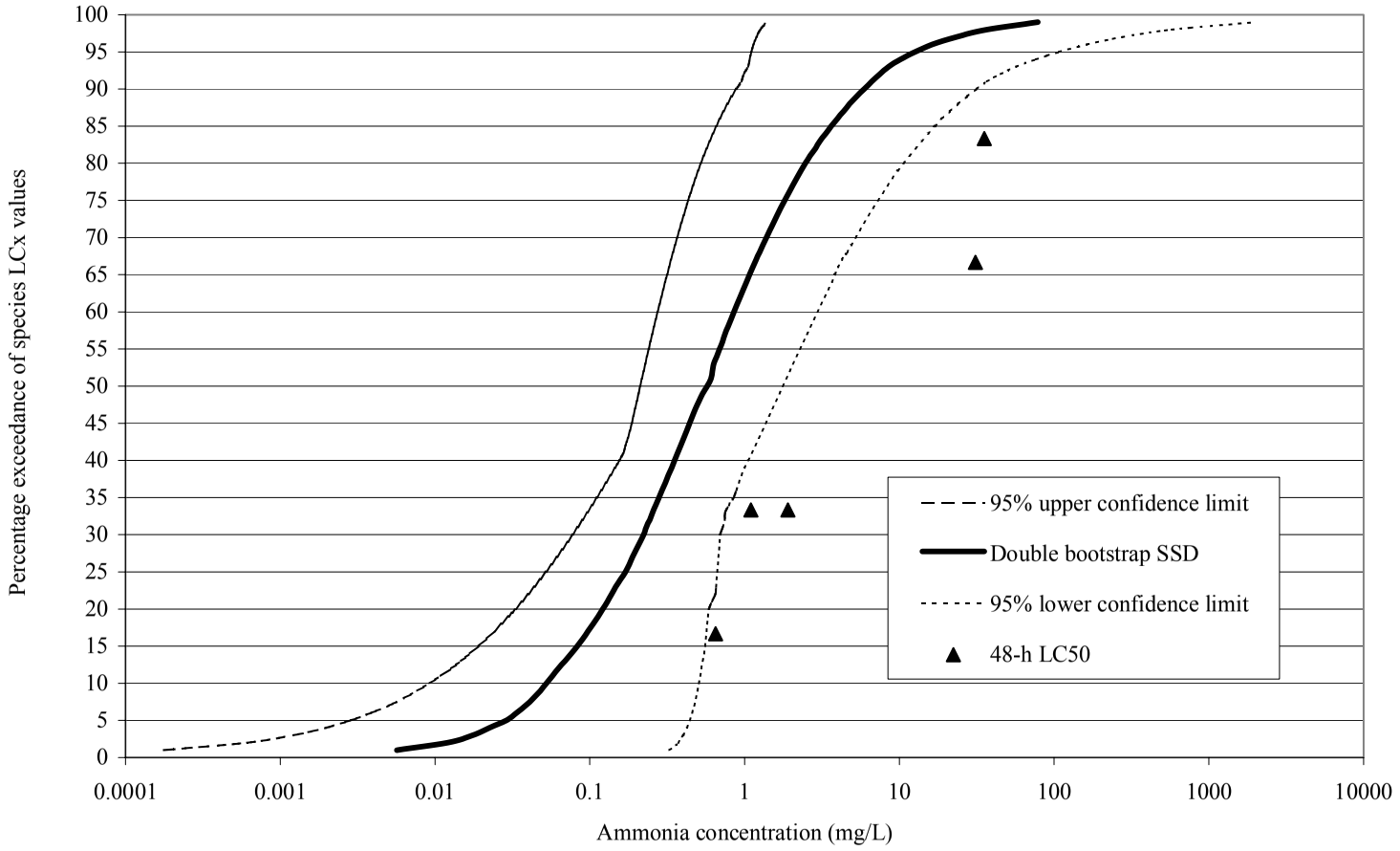


Figure 2.3: SSDs based on 48h LC50 and time-independent LC0 values for ammonia

2.6 Conclusions

There is little doubt that r (or λ) is, by definition, a superior ecotoxicological endpoint for population responses than any individual-level endpoint. It is also closer to the protection objectives for EQSs and for that reason, the use of this endpoint in decision-making should be more robust than relying on extrapolation from individual-level responses to population-level responses. However, there are indications that the most sensitive of individual-level endpoints such as survival and fecundity is often at least as sensitive as r or λ. It should be borne in mind that no general pattern exists with respect to which individual-level traits are most sensitive and so adequate protection based on these in isolation cannot be assured.

From a practical viewpoint, summaries describing r or λ are available for only a few isolated species/chemical combinations. Raw data suitable for estimation of these endpoints can be located, thereby extending the range of species/chemical combinations. Furthermore, robust methods for carrying out the analyses have been developed. However, experience shows it is unlikely to extend the range of species for which these endpoints are available substantially and it is unlikely that the data requirements for SSDs can be met in the foreseeable future with r or λ as inputs. Nevertheless, the use of these endpoints to inform EQS setting within a critical data/safety factor paradigm is a realistic proposition if suitable data for critical species can be obtained (as was the case in the fathead minnow/17α-ethinyloestradiol study, for example).

We must also recognise the limitations of population-level studies in the laboratory because of factors that are difficult to account for, such as density-dependent effects, and competition and predation from other species (although this applies equally to conventional measures of toxicity). As explained in Section 2.2, the duration of laboratory studies is also usually less than the nominative state. By underestimating the life cycle, the sustainability of the populations may have been over-estimated unless these unaccounted, older life stages contribute to the reproductive capacity of the population, in which case sustainability may have been under-estimated. Ideally, estimation of r would be calculated over at least an entire life cycle.

Generally speaking, a more fruitful area is the estimation of time-independent LC0 values based on survival data. Although not relating directly to impacts at the population level, such summaries do overcome some major limitations of conventional summaries of survival (i.e. fixed time exposures and expressed as 'effects' rather than 'no effects' concentrations). Data suitable for these analyses are more likely to be available for a range of species and in two cases, data from sufficient species were found to be able to construct SSDs. However, data for such analyses may not be widely available. Nevertheless, where the main impact of a chemical is expected to be through short-term toxicity – either because of the substance's mode of toxic action or mode of release – application of the two-step linear regression method for estimating time-independent LC0 values (Mayer *et al*, 2002) provides an effective way of generating more meaningful expressions of survival data with relatively modest levels of modelling input.

3. DESCRIBING CHEMICAL RISKS USING SPECIES TOLERANCE DISTRIBUTIONS

3.1 Introduction

In Section 1.6.4 we described the background to the use of SSDs in risk assessment and the derivation of environmental thresholds, including standards. Whilst a number of important strengths are evident, there are also some important flaws (Table 1.2). Some particular criticisms of the use of SSDs have been investigated to determine to what extent they may be overcome and what role they might play in the derivation of EQSs (or in supporting EQSs).

Forbes and Calow (2002) summarise some of the main criticisms of the approach as follows:

1. SSDs are usually constructed from what they describe as a 'haphazard' collection of literature or database values that have no direct relevance to site-specific assemblages of organisms. This may be a particular problem if site-specific exposure concentrations are then compared with these generic SSDs.

2. The endpoints used to construct SSDs are not demographically relevant, i.e. they are collections, and sometimes mixtures, of lethal or sublethal threshold concentrations, such as median lethal effects or sublethal no observed effect concentrations.

3. The technical aspects of constructing an SSD, such as model choice, selection of appropriate confidence intervals, definition of the minimum number, quality and representativeness of data points required, and selection of summary statistics such as the HC5 are often not transparent and may not clearly relate to environmental protection goals.

We begin to address these criticisms, using the organophosphorus insecticide chlorpyriphos as a model. Chlorpyriphos toxicity is rapid and intense for susceptible species, but relatively rapid degradation in the environment means that cumulative toxicity is unlikely. This means that short-term lethal toxicity may be a good predictor of long-term sublethal toxicity. A comprehensive ecological risk assessment for chlorpyriphos was recently performed for North American aquatic environments (Giesy *et al.*, 1999), providing a useful comparison for results in the present study. Mesocosm data are also available for chlorpyriphos, allowing comparisons of predictions from the present study with results from semi-field tests. Finally, there is currently a review of the use of chlorpyriphos in Europe and a separate national review in the UK, so it is timely for further analysis of possible risks from this insecticide.

In particular we have addressed points 1 and 2 above. The approach taken has been to:

1. Review available data on chlorpyriphos and selecting only those species that are indigenous to (or occur widely in) UK freshwaters for inclusion in an SSD.

2. Elicit expert information on organism tolerances to chlorpyriphos from UK field biologists, to counteract bias in the toxicity database.

3. Develop scenarios for three different aquatic habitat types containing different generic organism assemblages, in a preliminary effort to understand differences between site-specific assemblages.

4. Use non-parametric bootstrapping time-to-event and Bayesian statistical approaches to construct SSDs and associated confidence intervals based on 96-h LC50 toxicity values.

3.2 Data Collection

3.2.1 Reported and estimated data

Acute lethal data for chlorpyriphos on the USEPA AQUIRE database were collated by Crane *et al.* (submitted) and used in a simple probabilistic risk assessment without regard to data quality. From this species list, those not found in UK waters were excluded and the original remaining papers reviewed according to the usual criteria for acceptability of data when setting UK Environmental Quality Standards (Zabel and Cole, 1999).

3.2.2 Expert elicitation

Through a series of meetings with Environment Agency and Freshwater Biological Association (FBA) biologists, the views of freshwater field biologists were sought on the sensitivity to chlorpyriphos of 100 taxa representing a range of phyla found in UK freshwaters (Appendix F). Seventeen biologists employed by the Environment Agency or belonging to the FBA were asked to score the sensitivity to chlorpyriphos of each taxonomic group on a scale from 1 (insensitive) to 8 (highly sensitive). They were also asked to score their own knowledge of each taxonomic group from 0 (no knowledge) to 5 (high level of knowledge).

3.2.3 Generic aquatic assemblage scenarios

The taxonomic groups listed in Appendix F were separated into three generic assemblages on the basis of information in Fitter and Manuel (1994). These assemblages were:

a) a fast-flowing stream

b) a slow-flowing lowland river, and

c) a static pond or ditch

3.3 Construction of SSDs Based on Bayesian Analysis

The Bayesian statistical software WINBUGS was used to construct SSDs based on selected 96-h LC50 values and the results of the expert elicitation exercise. The observed log LC50 values were assumed to be normally distributed around the taxon mean with a precision defined by both inter- and intra-specific variance. Taxa mean log LC50 values were assumed to be linearly related to the experts' weighted mean sensitivity values. SSDs were constructed by running the model 7000 times, both with and without use of expert opinion from the elicitation exercise. The SSDs produced by this novel method were used to identify sensitive taxa for which data were not available.

Four species were then exposed to chlorpyriphos in the laboratory at WRc-NSF's Medmenham Laboratory to examine whether the experts had accurately predicted their sensitivity. These species, with the experts' assessment of their sensitivity in parentheses were *Ephemerella* sp. (6th), *Brachycentrus subnubilis* (8th), *Leuctra* sp. (17th) and *Hirudo medicinalis* (47th). All were exposed for 96-h in static test systems, using techniques based on OECD Test Guideline 203 (Fish, Acute Toxicity Test) (OECD, 1993). In these studies, test medium was renewed every 24-h to minimise the effects of degradation of chlorpyriphos due to its hydrolytic instability.

3.4 Results

The expert opinions on species sensitivities to chlorpyriphos, weighted according to their assessment of individual knowledge, are shown in the final column of Appendix F.

Figure 3.1 summarises the SSDs estimated through the Bayesian model, both with and without the expert opinions included. Figures 3.2 – 3.4 compare the SSDs for the three generic assemblages, both with and without expert opinions included. The results show that use of expert opinion produced lower estimates of the SSD and its 95% confidence interval, plus summary statistics like the HC5, for all three habitats. This is largely because the taxa that experts regarded as being particularly sensitive to chlorpyriphos were not well represented in the empirical toxicity dataset. For example, the HC5 was approximately 9 μg/L with expert opinion and 11 μg/L without those opinions for fast-flowing streams and slow-flowing rivers. However, 95% confidence intervals overlapped substantially for all SSDs, so the differences between assemblages and between SSDs derived with or without expert opinion did not differ significantly.

The results from toxicity tests with four additional species, designed to test the predictions made by experts, suggested that the experts were broadly correct in their assignment of sensitivities (Table 3.1). *Ephemerella* sp. was the most sensitive species, followed by *Brachycentrus subnubilus*, *Leuctra* sp. and *Hirudo medicinalis*, in the order predicted by the weighted mean expert ranking.

Table 3.1: Sensitivities of freshwater invertebrates from practical testing

Species	Taxon	48h LC50 and 95% CL (ug/L)	96h LC50 and 95% CL (ug/L)
Hirudo medicinalis	Hirudinea	No response	No response
Ephemerella sp.	Baetidae	0.07 (0.004-0.214)	0.035 (no CL)
Leuctra sp.	Leuctridae	3.69 (1.35-8.89)	0.87 (0.49-1.57)
Brachycentrus subnubilis	Brachycentridae	0.50 (0.41-0.63)	0.45 (0.49-1.57)

3.5 Discussion

The use of Bayesian methodology to incorporate expert judgement of species tolerance distributions and empirical data into SSDs produced very similar results to SSDs based on empirical data alone. This may have been because the empirical data and expert judgement happened to match each other well. If so, this should provide reassurance that the empirical toxicity data were reasonably representative of assemblage tolerance distributions. The experts were correct in their relative ranking of the four species tested as part of this study, but did make some mistakes. For example, Chironomidae were judged to have low sensitivity to chlorpyriphos, which clearly would not be the case for this insecticide. Such beliefs probably originate from the 'sanitary water quality' bias of many field biologists in the UK, and particularly those who work for the regulatory agencies. Biological monitoring in the UK has in the past often concentrated on examining the impact of sewage effluent discharges, and chironomid larvae are known to be insensitive to the relatively low dissolved oxygen and high ammonia concentrations that are present in these.

The choice of organisms used to produce the three generic assemblages in this study is open to debate. There was substantial species overlap between the three assemblages, with most Phyla represented in all three scenarios, and this will have contributed to the small differences in estimated SSDs. Forbes and Calow (2002) suggest that risk assessments based upon SSDs should be relevant to specific sites. However, there is a question over whether we should seek to protect what *is* currently present at a site, or whether we should protect what *could* be present at a site. The latter is likely to be more precautionary and will act as a driver to improve environmental quality, rather than just maintain the status quo. A more sophisticated treatment of site-specific assemblages is certainly achievable for UK lotic systems, by using RIVPACS (Wright *et al.*, 1994) to predict site-specific assemblages for SSD construction.

This work suggests that SSDs may be a robust approach for defining safe concentrations of chemicals. The valid theoretical criticisms of the approach do not appear to translate into major practical difficulties, at least for chlorpyriphos. We have also demonstrated some strategies for constructing SSDs that will help to overcome current deficiencies and make them more environmentally relevant and technically robust. Certainly, more

would need to be done to extend the range of chemicals and to explore more thoroughly the extent to which expert elicitation could be employed in informing species rankings in the way we have described for chlorpyriphos. The potential bias among agency biologists toward organic pollution effects is a difficulty. It is possible that such biases may be less for pesticides if it was possible to elicit input from biologists working on the development of new products. A direct comparison with EQSs derived in the conventional way is not possible for chlorpyriphos because such a standard has not yet been developed, although a predicted no-effect concentration (PNEC) is expected to emerge from a EU review of this substance being undertaken under the Plant Protection Products Directive.

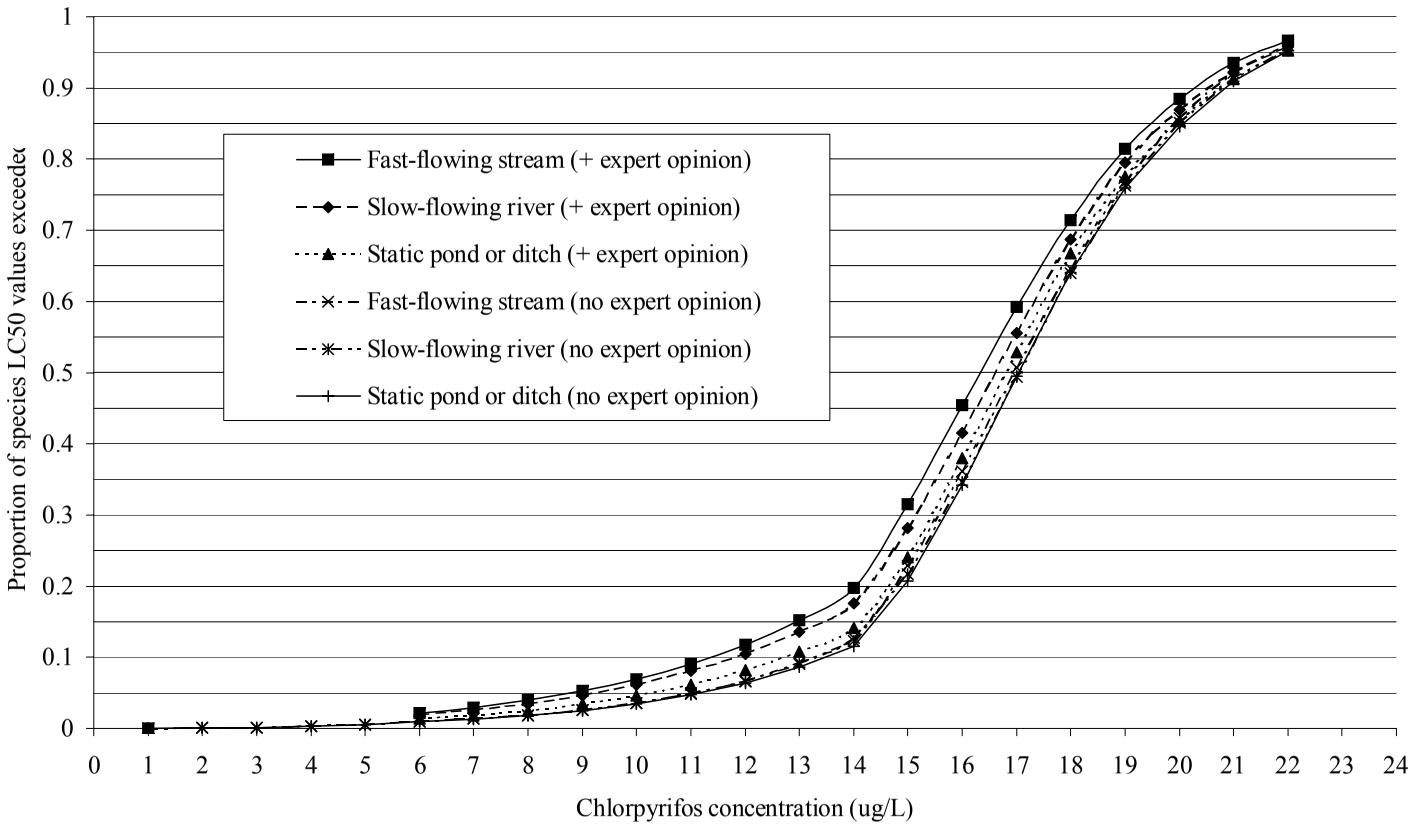


Figure 3.1: SSDs for chlorpyriphos for different habitats and with/without expert opinion

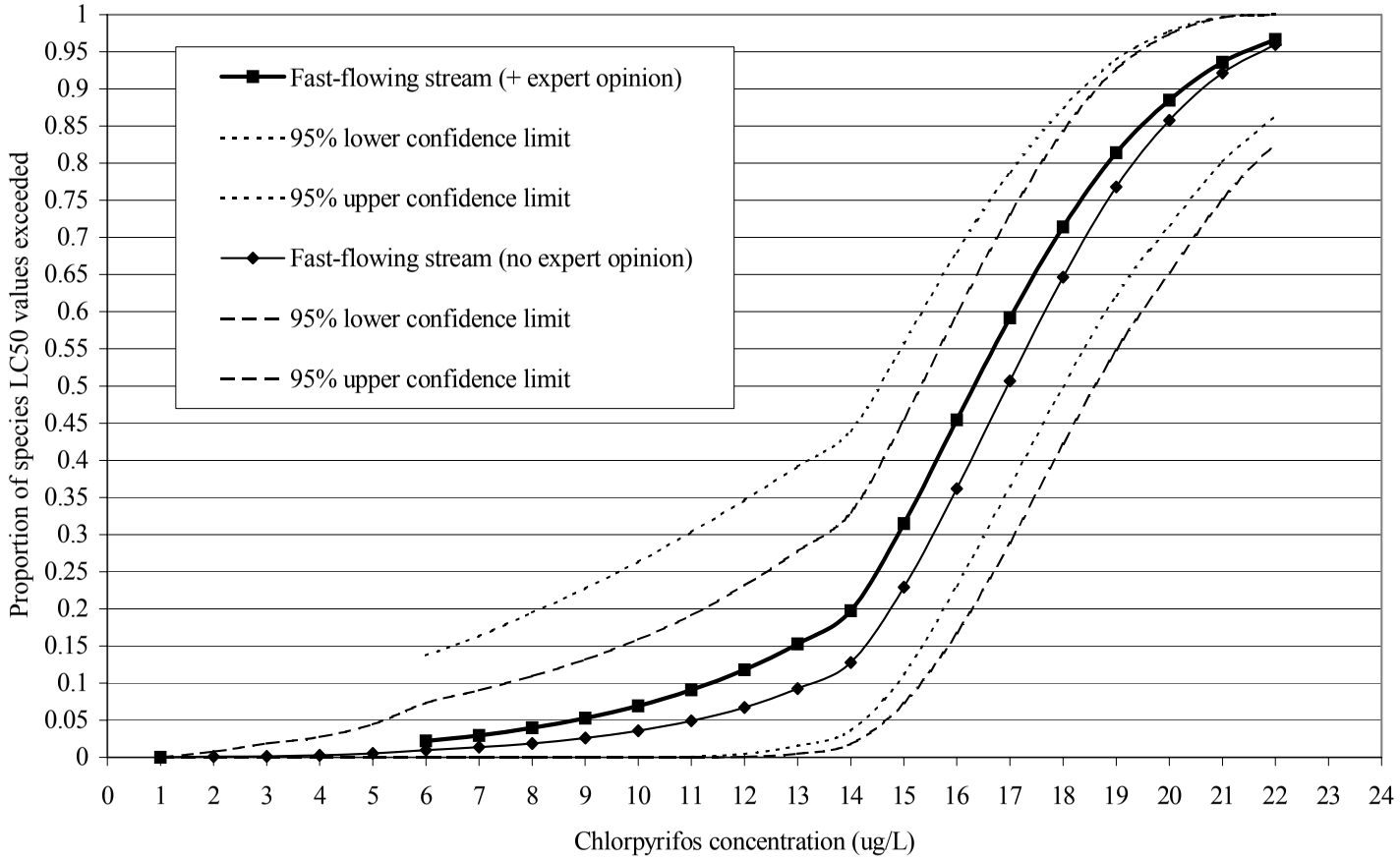


Figure 3.2: SSDs for chlorpyriphos – fast-flowing stream species

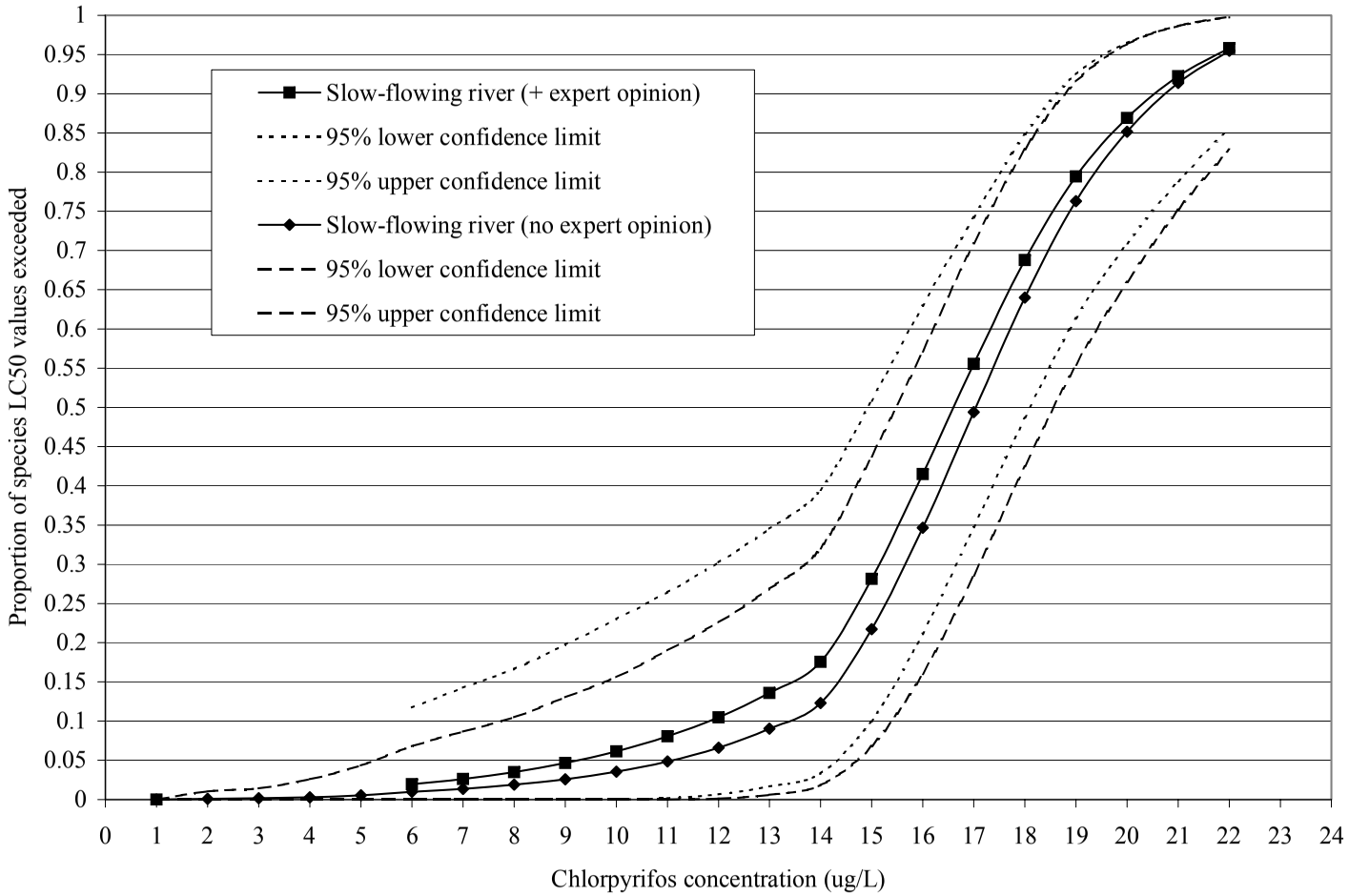


Figure 3.3: SSDs for chlorpyriphos - slow flowing river species

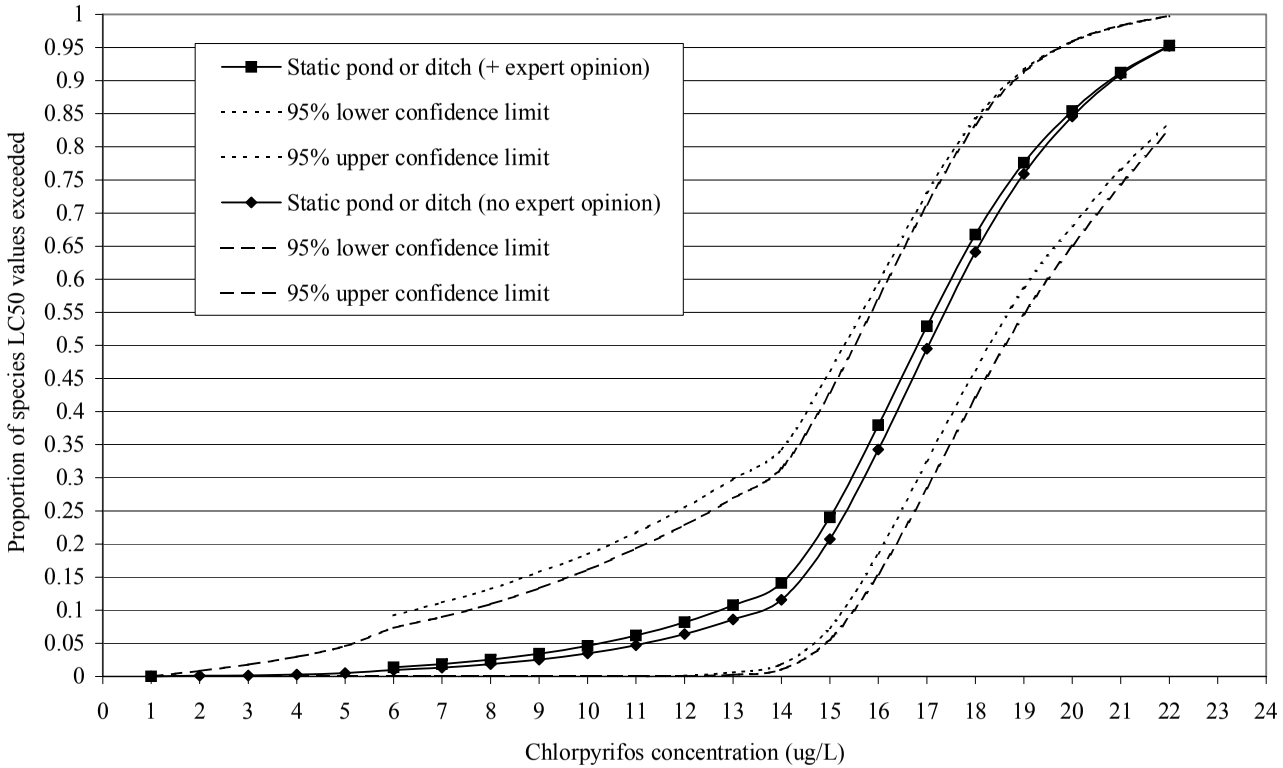


Figure 3.4: SSDs for chlorpyriphos - static water (ponds/ditches) species

4. EMPLOYING SPECIES SENSITIVITY DISTRIBUTIONS IN SITE-SPECIFIC RISK AND COST-BENEFIT ASSESSMENTS

4.1 Introduction

In assessing the risks posed by the release of a chemical to a watercourse, the risk assessor has a number of possible approaches that he/she can call upon. These range from a simple deterministic approach based on single point estimates of exposure and effects concentrations (typically both are conservative estimates) through to fully probabilistic approaches where all the inputs are described in terms of distributions (e.g. of toxicity data, emission rates, dilution rates, usage rates). In this way, it is argued they deal with uncertainty in these inputs that a deterministic approach fails to do and, as a result, lead to more realistic outcomes. An intermediate approach is to use a point estimate for one component and a distribution for the other. Whilst there is a degree of choice in the approach to be used, it is clear that probabilistic approaches are more demanding in terms of the quantity of data required. In this Chapter, we have adopted the intermediate option in which point estimates of exposure are compared with distributions of toxicity data, expressed as an SSD.

Notwithstanding legitimate concerns about the accuracy of predictions made with SSDs (Forbes and Calow, 2002), SSDs are potentially useful tools in the assessment of risk reduction options because, unlike pass/fail thresholds typical of current EQSs, they explicitly relate the effect of concentration to the proportion of species affected by a chemical. In Chapter 3, we showed how SSDs may be constructed using a variety of ecotoxicological endpoints. Below, examples of SSDS for two substances, ammonia and chlorpyriphos, are taken and applied to realistic scenarios for their release into surface waters. In both cases, risk reduction options are available (to reduce environmental exposure) and we illustrate how the environmental benefits of these measures in terms of the proportion of species protected may be predicted. In addition, the costs of achieving these predicted improvements in species diversity are also shown. In a third example we illustrate how SSDs may be used to describe the environmental consequences (in terms of the proportion of aquatic species affected) of different levels of chlorpyriphos usage on a regional scale.

4.2 Case 1 - Measures to Reduce Emissions of Ammonia from a Sewage Treatment Works

4.2.1 Introduction

Ammonia is a common and extensively regulated substance arising from sewage treatment works (STWs). Along with sanitary determinands such as BOD and suspended solids levels, it is one of the key determinands on which operators of STWs focus.

Concentration of ammonia in the final effluent from activated sludge plants (normally expressed as ammoniacal nitrogen) can vary considerably, between 25-30 mg/L for non-

nitrifying plants achieving only modest removal of ammonia, to around 2.5 mg/L for plants with normal nitrifying capability. STWs operating dedicated nitrifying facilities may achieve more efficient removal and final effluent concentrations in the region of 1 mg/L may then be achieved (Dee, WRc plc, *pers. comm.*). In practice, few activated sludge plants produce a partly denitrified effluent that is intermediate between a nitrifying and non-nitrifying plant but operating conditions can cause the process to swing between these extremes. Therefore, under good operating conditions (favourable sludge loading, sludge age, aeration and wastewater retention time) ammonia concentrations as low as 2.5 mg/L may be achieved but this could increase 10-fold if stable conditions cannot be held.

4.2.2 Upgrading of a sewage treatment works

The case study under consideration is a STW with little nitrification capability that, under certain operating conditions fails to nitrify at all. It produces a final effluent with an average unionised ammonia concentration of 20 mg/L, which discharges to a small watercourse providing approximately 20-fold dilution. This results in an in-stream ammonia concentration of *ca.* 1 mg/L after mixing and is thought to be responsible for the poor quality of a downstream coarse fishery. In addition, housing development in the area is likely to exacerbate this situation. Remedial action to reduce the impact of ammonia emissions, especially during the summer months when available dilution is reduced, is under consideration and investment in the STW plant is one of the options under review.

Figure 4.1 illustrates an SSD for unionised ammonia based on time-independent LC0 values for a range of freshwater fish species (as described in Section 2.5.4). The in-stream concentration of unionised ammonia (ca. 1.0 mg/L) is predicted to affect a high proportion of fish species – indeed the predictions are that effects on survival of most (ca. 70%) fish species would occur as a result of long term exposure under these conditions.

Two risk reduction options are available, based on increasing nitrification at the plant through increased aeration capacity and reduced sludge loading rates. These are illustrated in Figure 4.2. Option A is predicted to reduce final effluent concentrations to approximately 2.5 mg/L whilst the more expensive option would reduce final effluent concentrations to just 0.8 mg/L. Reference to Figure 4.1 suggests that, after 20-fold dilution in the receiving water, these final effluent concentrations would correspond to reduced risks to survival of fish species with ca. 20% of species at risk from option A and less than 5% at risk with option B. Clearly, these reduced risks carry a cost (Figure 4.2). The decision facing the regulator and Sewerage Undertaker is whether or not the additional cost of implementing option B over option A (a difference of £9,544 in annual operating costs and a difference of £308,000 in capital costs) is justified by the environmental benefits, in terms of the reduced risk to approximately 15% of species.

Figure 4.1 also shows the 95% confidence intervals on the fitted regression. A decision based on these leads to a more precautionary outcome (but one whose magnitude has some technical basis). Whilst suggesting the level of protection could be substantially less, the relative difference in improvement for options A and B is similar to that based on a 'face value' assessment of the SSD. Nevertheless the uncertainty in the predictions would lend weight to questions about the true benefits of implementing option A over

option B. In a comprehensive assessment of costs and benefits, estimates of uncertainty in the predictions can be used more formally to describe 'worst case' and 'best case' scenarios.

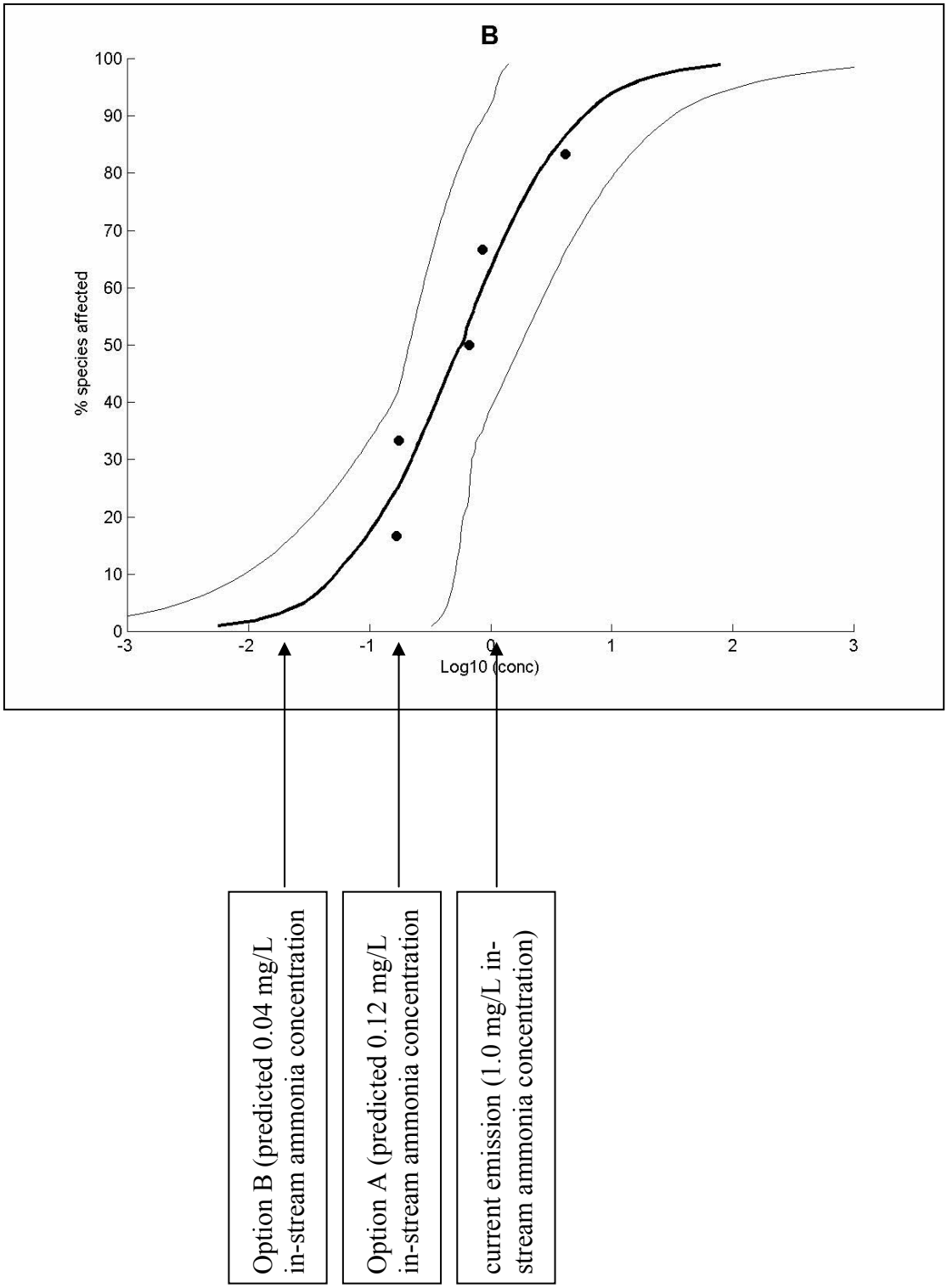


Figure 4.1: Species sensitivity distribution for unionised ammonia (mg/L) based on time-independent LC0 values estimated for freshwater fish species (see text for details of superimposed STW emission reduction scenarios)

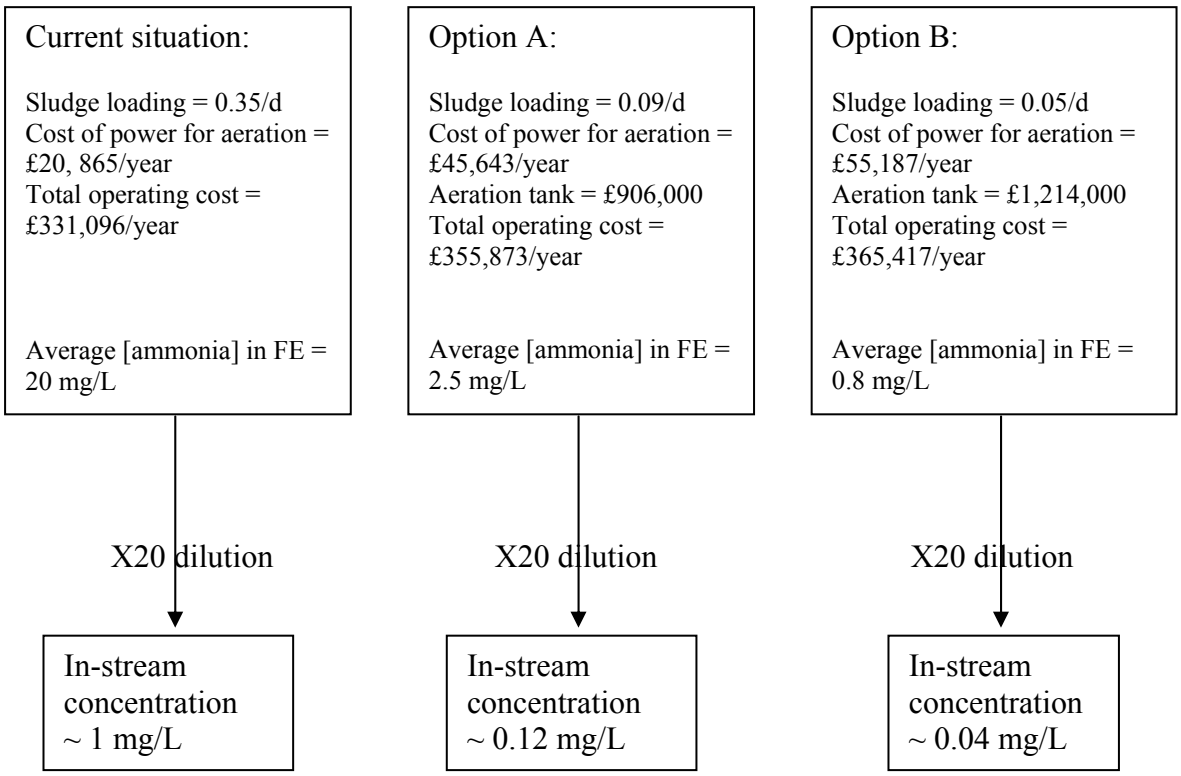


Figure 4.2: Options for reducing ammonia emissions from a weakly nitrifying STW

4.3 Case 2 - Buffer Zones to Reduce Spray Drift of Chlorpyriphos Applied to Top Fruit

4.3.1 Introduction

Chlorpyriphos is approved for use in a variety of crops in the UK to control insect pests. Crane *et al* (2002) describe a study of the risks to surface waters from early season applications of chlorpyriphos to top fruit in the UK using a fully probabilistic approach. They showed that contributions from surface run-off and drain flow in a sandy loam soil could effectively be discounted because of the high sorption potential for this compound. For the purposes of this example, we consider a neighbouring static surface water (e.g. large ditch complex), an example of an important habitat because ditches can support large numbers of invertebrate and plant species (Biggs *et al*, 1994). The lack of significant recharge of such water bodies also makes them vulnerable to pesticide impacts.

4.3.2 Buffer zones in top fruit

Crane *et al* (2002) predicted spray drift from a top fruit crop and then related the resulting concentrations in the neighbouring waterbody when different widths of a buffer zone between the crop and waterbody were imposed (Table 4.1). Below, these data have been supplemented with an estimate of the economic loss to a grower (in terms of the value of the lost fruit yield) of different buffer zone widths so that an economic dimension can be introduced.

The economic data are based on DEFRA (2001) and for the UK crop of Cox apples in 1999. A value (farm gate prices) of £6,770/ha is estimated. A typical planting density in an intensive system is 3.5m between rows and 1.75m between trees within a row. For the purposes of this worked example, we have taken a 5ha (500m x 100m) orchard with a 100m ditch running along the entire length of the shorter side. Assuming the rows are perpendicular to the ditch, each 10m increase in buffer zone would effectively result in a loss of yield of £677. Thus a buffer zone of 20m would mean a financial penalty to the grower of £1354 and a buffer zone of 50m would mean a penalty of £3385. This may overestimate the financial implications to the grower because it makes no allowance for possible alternative uses of the buffer zone and the income associated with those. Nor are the savings in planting costs discounted by not having to plant the buffer zone. Nevertheless, it does serve to illustrate the way in which a financial element can be introduced into decision-making about the benefits to the local aquatic biota of different widths of buffer zone.

Table 4.1: Predicted spray drift of chlorpyriphos from early season top fruit application

Buffer width (m)	Economic cost, in terms of lost yield* (£)	Estimated mean concentration in ditch (μg/L)	Estimated 95th percentile concentration in ditch (μg/L)
5	334	23.3	46.7
10	667	9.97	21.1
20	1354	3.04	6.89
40	2708	0.20	0.64
50	3385	0.16	0.56
75	5073	0.07	0.23

For the purposes of this example, the distribution of possible concentrations in the ditch at a particular buffer width is simplified into a single point, corresponding to the estimated mean concentration from Table 4.1. It is now possible to compare these predicted mean concentrations at different buffer widths with the SSD based on time-independent LC0 values (Figure 2.2). This is redrawn in Figure 4.3 with the financial

loss to the grower associated with different buffer widths superimposed on the horizontal scale. In this way it is possible to relate changes (increases) in invertebrate diversity to changes in buffer width and the financial penalties faced by the grower in bringing these about. Clearly, there is a high risk to biota from use of this compound in top fruit (this is confirmed in a separate study by Crane *et al* (2002). It should be noted, however, that the time-independent LC0 provides a conservative estimate of effects, especially as chlorpyriphos is known to be lost from waterbodies as a result of sorption and hydrolysis, and so is unlikely to persist in the environment (Giddings *et al*, 1997). This particular expression of the toxicity of chlorpyriphos may therefore overstate the actual risks to freshwater biota.

For clarity, just two scenarios are illustrated. These show that a buffer zone of 50 m is predicted to result in impacts in over half of the species present (at the LC0 level of protection). This equates to lost income to the grower of approximately £3385. Extending the buffer zone to 75m would result in a significant reduction in risk to biota (32% of species being affected at the LC0 level) and a reduction in income of £5073 to the grower. This corresponds to an improved protection of species diversity of approximately 35% for an additional financial cost to the grower of £1688. Further scrutiny of the species that are likely to be affected (i.e. those occurring in the lower tail of the SSD) along with an assessment of local freshwater biota would be useful so that any particularly important or desirable (e.g. rare) species may be accounted for.

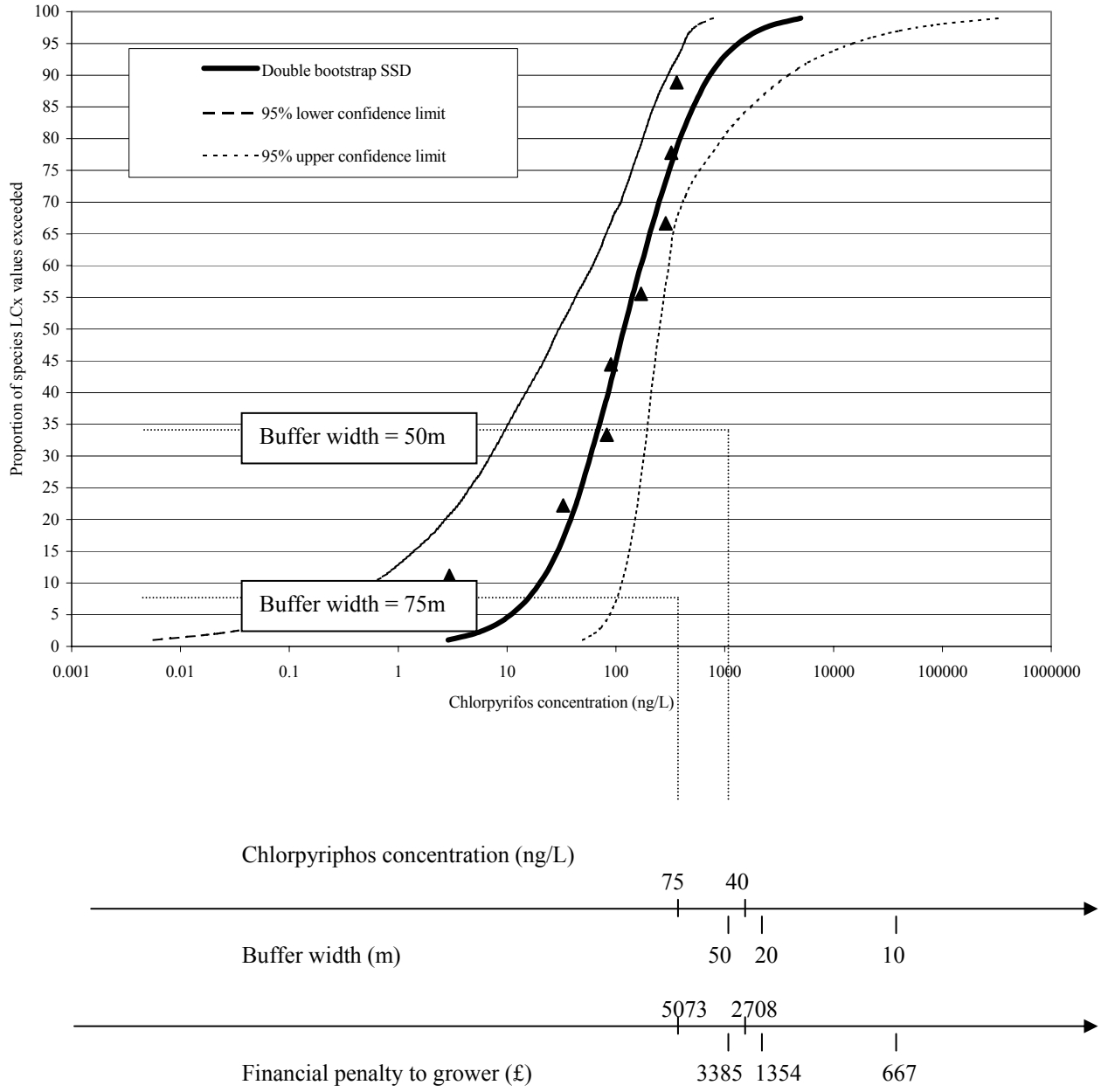


Figure 4.3: Costs to the grower of achieving different levels of reduction in impact of chlorpyriphos on aquatic biota (LC0 over infinite time) in a neighbouring watercourse as a result of different buffer zone widths

4.4 Case 3 –Illustrating Impacts on Aquatic Biota Based on Chlorpyriphos Usage

4.4.1 Introduction

The POPPIE (Prediction of Pesticide Pollution in the Environment) system has been developed by the Environment Agency to predict pesticide pollution in controlled waters in the UK. It involves a combined database of pesticide usage, models to predict pesticide transfer to surface waters and a GIS system covering England and Wales. Below, we give an example based on chlorpyriphos to illustrate how POPPIE outputs may be used in conjunction with SSDs to estimate impacts on aquatic biota. It is important to stress that the resulting estimations are subject to some inaccuracies because chlorpyriphos exposure is probably systematically under-estimated. This is because POPPIE model inputs are based on surface and drain flow, but it is known that the main source of aqueous contamination by chlorpyriphos is spray drift (Crane *et al*, 2002). Furthermore, predictions of concentrations in watercourses are remote from the point of use, where they are expected to be higher. For these reasons, the resulting risks to aquatic biota will be underestimated. Nevertheless, the example is shown to illustrate how SSDs may be used within a geographical context.

4.4.2 Describing impacts on a geographical scale

POPPIE outputs based on chlorpyriphos usage in 1999 and inputs from surface and drain flow only show low levels of surface water contamination across most of the UK but heavier levels of contamination in parts of Yorkshire, north Cheshire and particularly in the Cambridgeshire/Lincolnshire fens (Fig. 4.4). Predicted concentrations in watercourses are highest during July, presumably coinciding with heaviest use. In the fenland area, predicted concentrations are as high as 20.5 – 42.9 ng/L. We can illustrate how such exposure data may be used to determine the level of risk to aquatic biota by comparing the lower bound[10] of the concentration bands shown in Figure 4.4 with the SSD based on time-independent LC0 data for freshwater species exposed to chlorpyriphos (Fig. 4.3). This can be expressed in terms of the proportion of species where the time-independent LC0 lies below the lower bound in different catchments and therefore whose survival may be compromised.

The proportion of species judged to be ‘at risk’ are superimposed onto Figure 4.4. In those catchments where heaviest contamination of surface waters is predicted, up to 15% of freshwater species are judged to be ‘at risk’. The absolute values are not the main concern here because, as explained above, the predicted concentrations are likely to be an underestimate, particularly in water courses close to the point of application. Rather, this example is intended to show how SSDs may be used to ‘map’ risks on a geographical scale, in contrast with the local, site-specific assessment of risks shown in Case 2.

[10] This represents a conservative assessment of the possible impacts in terms of number of aquatic species affected

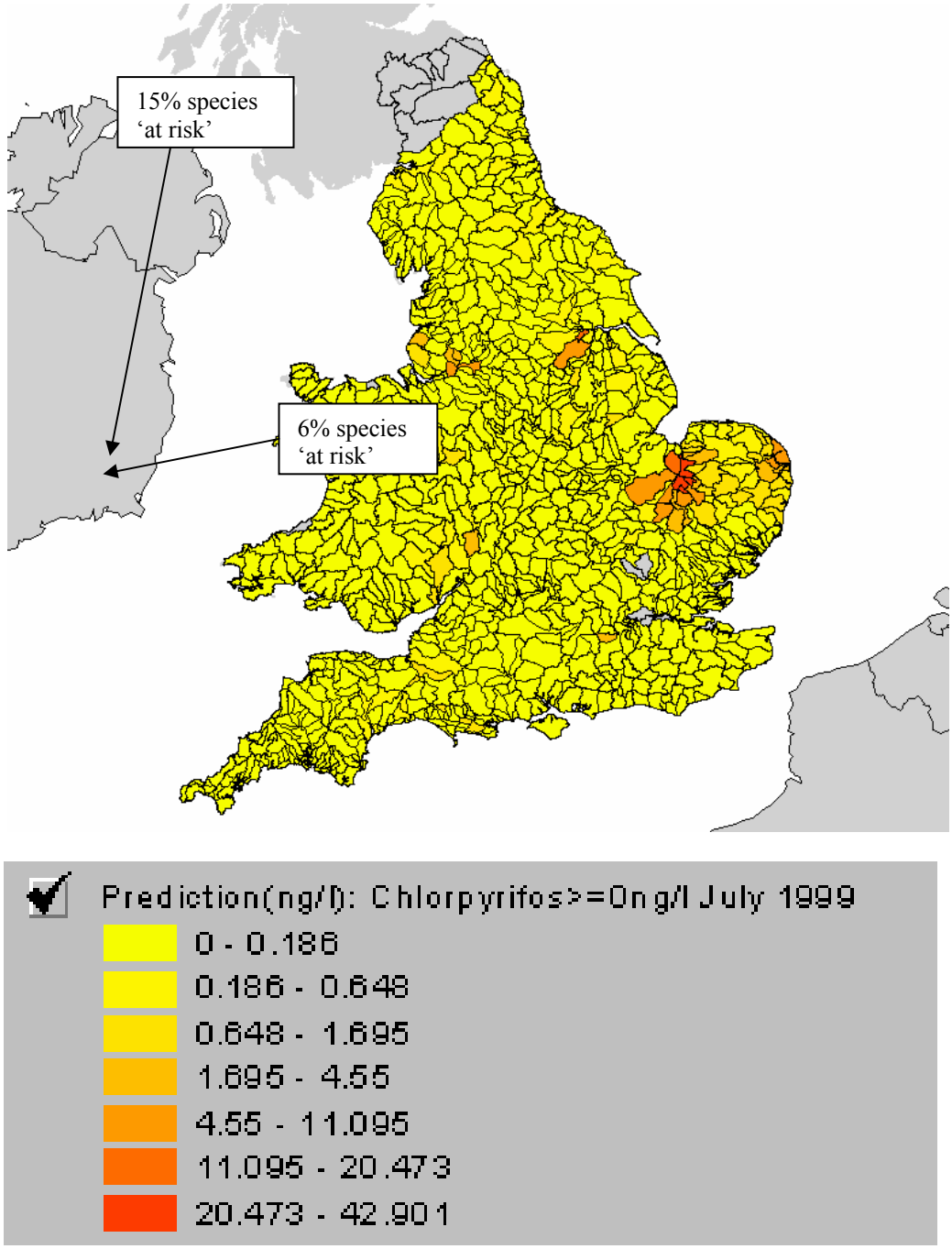


Figure 4.4: Illustration of the estimated risks to freshwater biota from chlorpyriphos (POPPIE predictions of concentrations in watercourses, July 1999)

4.5 Summary

The examples shown above are effectively risk assessments, some containing an economic element. Elements of probabilistic approaches have been used but they could be further refined by taking a probabilistic approach to both the description of effects (i.e. the SSD) and the environmental concentrations of the chemical such as distributions of pesticide exposure (Case 2), or distributions of effluent and receiving water flow rates (Case 1). Although rather simple, the examples show how a risk-based expression of an environmental threshold can be used to inform decisions about risk reduction options and help take account of their economic implications. Case 3 also shows that constructing SSDs from toxicity data can also be used to illustrate the effects of environmental contamination in a way that simple pass/fail thresholds cannot, in this case illustrating a landscape-level assessment of risk.

Where controls on emissions are sought, the probability of some acceptable level of impact may be defined that does not necessarily correspond to zero impact on all species. Indeed, such approaches could help in devising a programmed reduction in emissions over time in which the aim might be to achieve certain levels of species protection at agreed stages (see example below). This could include impacts on a very small proportion of species at some point. Such an approach is not possible when only pass/fail thresholds are available because they tell the interested parties nothing about the consequences of exceeding a particular level of a chemical in the environment. The 95% confidence intervals around the SSD (or concentration-response) can also be useful because uncertainties on the benefits side of the assessment can be quantified, allowing 'worst-case' and 'best-case' projections to be made.

A more speculative possibility relates to implementation of the Water Framework Directive. Under this Directive, Competent Authorities are required to characterise the status of waterbodies both in terms of their chemical and biological quality. The target is good ecological and chemical status by 2015 and excellent status for special sites (e.g. SACs). However, we are not aware of any current procedures for relating different biological qualities with the extent of chemical contamination. Ecological quality targets can be started in terms of the range of species or families predicted by habitat 'template' models e.g. RIVPACS[11], PSYM[12] or FABSIM[13]. If points on the vertical axis of the SSD (i.e. the proportion of species affected) could be 'calibrated' against biodiversity indices in catchments for different classes of ecological quality, then it may be possible to provide such a link. The SSD approach could be employed using a number of species from a range of phyla as suggested in Table 4.2 (the thresholds shown are for illustrative purposes only) and including representatives of different trophic levels (Forbes and Calow, 2002). In this way, chemical thresholds (for substances that are not already covered by EU Quality Standards) needed to achieve a particular ecological class could be estimated, and thereby a 'biological target' is translated into a more manageable 'chemical target'.

[11] River Invertebrate Prediction and Classification System

[12] relates to invertebrate fauna of ponds

[13] relates to fish fauna of rivers and streams

Table 4.2: Linking ecological quality (biodiversity) and chemical quality (SSD parameters)

Quality rating	Number of BMWP families	Equivalent SSD %
Excellent	50	100
Good	45	90
Fair	35	70
Poor	20	40

If such an approach could be developed, a phased approach to risk reduction could be employed with a series of incremental improvement targets for a particular chemical in rivers classified as poor or fair, and with the ultimate aim of achieving good ecological quality by 2015. Such an approach is not without its difficulties however. First, existing biodiversity and habitat template models were developed with organic pollution in mind rather than specific chemicals. Secondly, it would be necessary to demonstrate a link between the presence of a particular chemical and impaired ecological quality to avoid the risk of unnecessary clean-up and overlooking other chemicals of concern.

A possible weakness of the use of SSDs in risk and cost-benefit assessment approach is that the risk assessor must make decisions about what level of impact (e.g. the proportion of species affected) is acceptable in a given situation. Under current approaches with pass/fail thresholds, the decision is essentially quantal i.e. does the environmental concentration exceed or fall below a threshold value? The approach outlined here puts a greater burden on the risk assessor to justify his/her decisions. This takes us back to the issue raised in Section 1.5.1 about what constitutes an 'acceptable risk' and emphasises the importance of developing such criteria in the derivation of environmental standards and risk assessment.

The other main limitation is one of data availability and the taxonomic diversity amongst test data. In practice, the number of data points and the biological diversity represented will often be narrow with the result that we cannot always be confident that the resulting SSDs are good representations of the local ecology. For this reason, we suggest that, at least for the foreseeable future, SSDs should be considered only when circumstances warrant their use. This is developed further in Section 6.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Current Methodology for Deriving EQSs

EQSs for the protection of aquatic life have been derived in the UK using the same basic approach for approximately twenty years. Aquatic toxicity data lie at the heart of the standard-setting process and it is inconceivable that this situation will change in any significant way in the foreseeable future. However, there have been significant advances in the conduct and interpretation of aquatic testing in recent years and also a questioning of the relevance of much of the test data that is routinely generated. Two themes are noteworthy in relation to EQSs.

(a) The first concerns the ecological relevance of the endpoints that are measured, and particularly a recognition that estimates of effects on populations rather than individuals are more useful in regulatory chemical control schemes.

(b) The second concerns a growing reliance on risk assessment for regulatory decision-making and, more recently, the development of more sophisticated probabilistic approaches to ecological risk assessment.

In this project we have explored whether these developing themes may be usefully incorporated into the way EQSs are derived and expressed.

5.1.1 Strengths and limitations of the current approach

The current UK approach to deriving EQSs is essentially based on application of safety factors to a critical dataset, representing the most sensitive species/endpoints. In this respect it is not unlike other approaches used for standard setting throughout the developed world and, in common with these methods, the UK approach is subject to a number of shortcomings, summarised below:

Potential shortcoming	Implications
Lack of clarity about protection goals	A risk of inconsistency in stringency of resulting EQSs
	Risk that most meaningful endpoints are not always used
Reliance on NOECs as summaries of toxicity	Inefficient use of information
Use of default safety factors	Risk of over-conservatism - investment in remediation when it may not be needed
	Possible false impression of accuracy – does not convey uncertainty
Reliance on critical data	Large datasets are not always adequately exploited. Disincentives to generate new data
Expression of EQSs as simple thresholds	Limited utility in site-specific risk assessment and cost-benefit assessment

Nevertheless, the current UK approach contains some elements that set it apart from more mechanistic applications of the critical data/safety factor paradigm. These centre on:

- A high level of investment of effort in the quality assessment of critical data and a consideration of the biological significance of the toxicity estimates at our disposal
- Flexibility in the size of safety factors applied
- Technical peer review
- An understanding of the strengths and shortcomings of different expressions of toxicity

These go some way to ameliorating the shortcomings listed above, especially the risks of over-conservatism (because default safety factors are used as a 'last resort') and the errors that can result from placing undue emphasis on NOECs. It is reasonable to suggest that the UK approach represents amongst best practice, given the fundamental framework it works with. However, is there room for improvement?

5.2 More Relevant Endpoints?

5.2.1 Population-level endpoints

A basic premise to our exploration of endpoints used as the basis for EQS-setting has been that the aim of EQS should be to achieve sustainability of populations of aquatic organisms. With the emphasis of ecotoxicity testing on the responses of individuals to chemicals, this introduces a degree of uncertainty that could be reduced if more relevant endpoints were available. This seems a reasonable approach for most species. However, it may be argued that, for some species, particularly those long-lived, rarer species with low fecundity, typically occupying higher trophic levels (e.g. top predators), protection at the individual level is also appropriate.

A useful estimate of chemical effects on population size is the intrinsic rate of population increase, r, or its log equivalent, λ. Values of r of zero imply no increase in population over time and a negative value, implies a projected decline, ultimately leading to extinction. Thus, the endpoint can be used to convey the 'risk of extinction' of a particular species associated with a particular chemical concentration. Approaches based on Life Tables for estimating r from long-term studies with aquatic organisms have been refined. These allow the uncertainty associated with such estimates to be quantified using a 'double bootstrap' approach (Appendix A). It has been applied to two datasets from long-term aquatic toxicity studies, one in which *Daphnia* were exposed to zinc sulphate and another in which fathead minnows were exposed to the steroid oestrogen, 17α-ethinyloestradiol. The approach is workable in practice although with significant input of specialist modelling expertise. In the case of the 17α-ethinyloestradiol study, the approach was helpful in formulating proposals for thresholds for this substance.

Extending this approach to cover a range of species has proved difficult, largely because it has proved extremely difficult to obtain the data required for analysis for more than a few species. Where such data are available, they tend to be for a narrow range of species (e.g. cadmium). For this reason, estimation of r as a basis for EQS-setting is considered impractical under current circumstances except in specific cases where:

a) there is a question about the ecological relevance of existing endpoints

and

b) attention can be focused on a small number of species

These criteria were met in the example described for 17α-ethinyloestradiol. Some reassurance may be drawn from work by Calow *et al* (1997) and Forbes and Calow, (1999) suggesting that conventional endpoints describing effects on reproduction and survival will, in most cases, provide acceptable surrogates for effects on the more ecologically relevant summary, r. In other words, effects on survival or reproduction were often seen at concentrations similar to, or lower than, those seen on r.

While practicalities prevent estimates of r being incorporated into an SSD-based methodology for effects assessment, recalculating data to estimate r remains a viable option for deriving EQSs when they are based on a critical data/safety factor approach. In this application, such analyses can be useful in informing the derivation of EQSs.

5.2.2 Time-independent expressions of survival data

Although not as appealing as population viability as a protection aim, defining concentrations of chemicals that will have no adverse effects on the survival of aquatic organisms is nevertheless important. However, conventional fixed-time effects concentrations (e.g. a 96h LC50) are almost meaningless as risk assessment objectives and form the subject of much of the extrapolation that is carried out when deriving EQSs. Expressions that are closer to the protection aim are therefore to be welcomed.

We have demonstrated that more useful summaries of survival data can be generated through time to event analyses of survivorship data because they describe effects that are closer to the protection aims of EQSs. The summaries we have estimated are time-independent LC0 values using a refinement of the two-step linear regression method described by Mayer *et al* (2002). Although raw data are again required, our experience is that suitable datasets are more likely to be located than are suitable datasets for estimation of r (Section 5.2.1) but the general availability of such data remains unclear. The resulting analyses may be used to inform thresholds derived for particular species of interest (e.g. particularly sensitive ones) but sufficient datasets may be collected to understand the extent of interspecies differences in these endpoints and to construct SSDs. This was achieved for two of the substances investigated, chlorpyriphos and ammonia.

A separate and rather specific aspect of exposure is when organisms are subjected to repeated toxic challenges. This can be important where episodic discharges are anticipated or in tidal situations where sessile organisms are exposed in a predictable way to a body of contaminated water. Currently, standard test protocols do not

accounted for such a scenario although methods for predicting the effects of episodic exposure have been advanced recently (e.g. Karman, 2000). We have not been able to examine such scenarios within this study but raise them here as a different aspect of temporal exposure that may warrant consideration in certain situations.

5.3 Species Tolerance Distributions

SSDs offer potentially powerful features in risk assessment because they explicitly relate chemical concentration to a meaningful measure of biological impact (species diversity). For risk assessment purposes, SSDs conventionally employ long-term NOECs but can be based on acute data or even effects concentrations (e.g. LC50s). However, they are subject to some important criticisms, particularly about how representative they are of assemblages of organisms in nature (Forbes and Calow, 2002).

We have been able to construct SSDs using more useful summaries of survival data (time-independent LC0 values) for two substances. We have also sought to address the issue of species 'representativeness' used in constructing SSDs by the use of Bayesian methodology to incorporate both expert judgement of species tolerance distributions and empirical data into SSDs. Our experience shows that very similar results resulted when SSDs were based on empirical data alone. This may have been because the empirical data and expert judgement happened to match each other well. If so, this should provide reassurance that the empirical toxicity data were reasonably representative of assemblage tolerance distributions. However, there are some questions about a possible lack of chemical-specific experience of many field biologists and possible biases resulting from experience based largely on organic pollution.

Forbes and Calow (2002) suggest that risk assessments based upon SSDs should be relevant to specific sites. However, there is a question over whether we should seek to protect what *is* currently present at a site, or whether we should protect what *could* be present at a site. The latter is likely to be more precautionary and will act as a driver to improve environmental quality, rather than just maintain the status quo. A more sophisticated treatment of site-specific assemblages is certainly achievable for UK lotic systems, by using RIVPACS (Wright *et al.*, 1994) to predict site-specific assemblages for SSD construction.

This work suggests that SSDs may be a robust approach for defining safe concentrations of chemicals. The valid theoretical criticisms of the approach do not appear to translate into major practical difficulties, at least for chlorpyriphos and the use of more relevant summaries of toxicity helps make them more environmentally relevant and technically robust. Certainly, more would need to be done to extend the range of chemicals and to explore more thoroughly the extent to which expert elicitation could be employed in informing species rankings in the way we have described for chlorpyriphos.

5.4 Assessing Risk and Cost and Benefits

The use of SSDs in addressing both the costs and benefits of different risk reduction options has been illustrated through a series of worked examples. The key point is that expressions that relate chemical concentration to some measure of impact provide a

more powerful way of addressing cost-benefit issues than do simple pass/fail thresholds, which are weak in this regard. We have demonstrated the use of SSDs for this purpose. However, one could legitimately use the concentration-response for a single species if there was reason to believe it was particularly sensitive to the substance of concern or that it was of particular importance (i.e. the protection goal was expressed in terms of protecting this particular species).

Added advantages of such risk-based expressions of biological impact are that:

- The uncertainty associated with the predictions can be quantified (unlike pass/fail thresholds) and this can be a useful feature in more formal assessments of cost and benefit.

- Phased programmes of risk reduction can be developed over time, in which incremental steps are implemented, each delivering an expected improvement in water quality. This can be monitored so that revisions to the programme of risk reduction may be made where appropriate.

- The use of SSDs may provide a useful tool for linking the chemical quality of water courses to their biological quality because both can be described in terms of species diversity. This would need to be explored further but may be relevant to the requirements of the Water Framework Directive where both chemical and biological quality must be considered.

The use of risk-based approaches in decision-making poses some difficulties as well. When pass/fail thresholds are used as the basis of effects assessments, decisions about what is, or is not acceptable, have already been made. In a risk-based approach, the risk assessor must make decisions about what level of impact (e.g. the proportion of species affected) is acceptable in a given situation. Unless clear protection objectives have been established, decisions about acceptable levels of risk may be difficult to defend. In this regard, the tiered approach outlined below may have some merit.

5.5 Recommendations

5.5.1 A 'combined' approach to the derivation and use of EQSs

The 'ideal' EQS suggested in Section 1.7 would be one in which ecologically relevant summaries are expressed in a risk-based framework with minimal reliance on default safety factors. There are clearly limitations to achieving this goal. While we can go some way down one route or the other, our experience suggests that it will remain impractical for the foreseeable future to integrate population-level endpoints within a risk-based framework such as an SSD. Nevertheless, there are ways in which the derivation of EQSs and risks to the environment are assessed that might be improved through a tiered approach.

A two step hierarchy is proposed, as outlined in Figure 5.1. Essentially this entails the development of a generic EQS expressed as a simple pass/fail threshold. It is against this generic EQS that receiving water quality would be assessed and discharge consents derived, much as they are now. However, in the event of marginal or non-compliance

with this generic EQS, a risk-based approach is invoked in which the relationship between chemical concentration and biological impact is explicitly described. Essentially, what is proposed is a site-specific assessment of the risk to biota and the costs of achieving a particular level of improvement, but only in those cases where there is a demonstrable failure to meet the requirements of an initial 'screen'. This could, in turn, lead to a range of site-specific measures, aimed at delivering an acceptable level of environmental protection (that might ultimately correspond to compliance with the generic EQS).

In many cases, a useful component of the site-specific assessment would be to consider the bioavailability of chemicals apparently exceeding generic EQSs. Local conditions may reduce bioavailability of metals in particular to an extent that they pose no risk to biota yet are in exceedance of generic EQSs when these are expressed as 'total' concentrations (e.g. Mathiessen *et al*, 1999). A risk-based approach would only be adopted if a risk to biota is confirmed, probably by a combination of laboratory and field studies to determine whether impacts are occurring in practice.

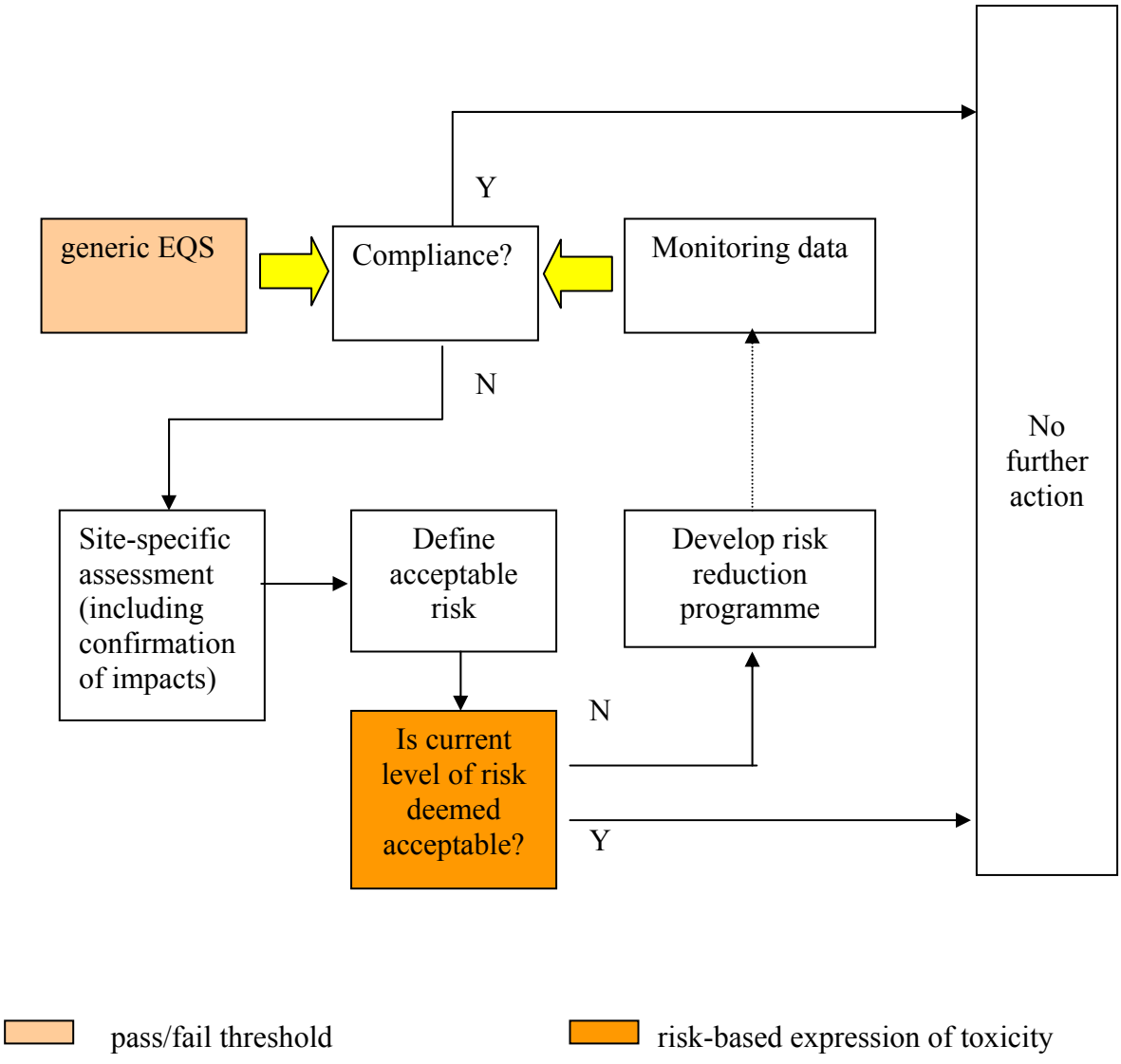


Figure 5.1: Proposed tiered approach to compliance assessment

5.5.2 Deriving a generic EQS

In a tiered system such as that proposed here, it is appropriate that the initial decision (based on a generic EQS) should err on the side of caution so that the risk of false negatives (i.e. failing to act when it is warranted) is minimised. Indeed, this is entirely consistent with tiered approaches to chemical risk assessment. For this reason, we propose that the generic EQS should be based on the critical data/safety factor paradigm in the expectation that this will tend to deliver more conservative outcomes and only requires the EQS to be expressed in terms of a simple pass/fail threshold.

Identification of critical data

To minimise the shortcomings of the critical data/safety factor approach, there should be careful scrutiny of ecotoxicity datasets. The following points of 'good practice' are suggested for those datasets that indicate greatest sensitivity. Overall, the underlying principle is to generate endpoints that are more relevant to the protection objectives of a generic EQS and, in so doing reduce uncertainty in the extrapolation process:

1. Eliminate the use of NOECs wherever possible, preferably by estimating the lower confidence limit on an EC10 or EC20 (Moore and Caux, 1997) or by using time to event analyses as an alternative. These alternatives will usually require access to raw data.

2. If this is not possible, substitute the LOEC and NOEC with the MATC (the geometric mean of the LOEC and NOEC) but bear in mind that the MATC will tend be less conservative than a NOEC. If a NOEC has been reported, there must be a LOEC, by definition.

3. The magnitude of biological effect at the MATC should be understood and a clear view formed about the power of the study generating these summaries (level of replication, within-treatment variance, level of significance chosen, disposition of test concentrations).

4. Summaries that are described as NOECs yet actually describe an absence of effect at the highest test concentration should be recognised as such (they are not NOECs).

5. Where r or λ have been estimated, they should be used alongside endpoints describing effects at the individual level.

6. Where unicellular algae or *Lemna* are amongst the most sensitive species, emphasis should be placed on toxicity summaries describing growth rates based on cell or frond number (e.g. 'ErC50') rather than final biomass ('EbC50').

7. Only endpoints of demographic significance should be considered amongst the critical dataset. By this is meant survival, reproduction and developmental effects, particularly where there is evidence of effects on the time taken to attain reproductive maturity. Effects on growth of individual organisms, behavioural endpoints and physiological/biochemical changes should not, generally, be prime determinands of generic EQSs. We would, however, expect

concentrations giving rise to such responses to fall within the safety factors applied to the critical data.

8. Where LC/EC50 data are reported, information on the slope of the concentration-response or LC/EC10 and LC/EC20 values should be sought. For critical datasets reanalysis of raw data to estimate these summaries is a simple task using widely available software.

9. It is important to remember that, in this approach, a small critical dataset is extremely influential to the standard-setting process. It is therefore appropriate to exploit these data for as much relevant information as possible. Where suitable data are available for sensitive species, we recommend that r or λ should be estimated and used to inform the setting of a generic EQS.

10. If suitable chronic data are not available, consideration should be given to estimating thresholds using time to event methods, to derive time-independent analysis of survival data, as described in Section 2.5. This approach is strongly recommended when survival data feature in the critical dataset.

Size of safety factors

The notion of a 'correct' size for safety factors in ecotoxicological effects assessment is almost futile because no one set will be universally applicable. It remains the case that selection of safety factors remains largely qualitative with little quantitative basis. There should be an attempt to strike a balance between under-protection (factors are too small) and over-protection (factors are too large) (Chapman *et al*, 1998).

However, certain principles and practical recommendations can be recognised:

1. The use of default safety factors should be minimised as far as possible. Rather, emphasis should be placed on the use of case-specific factors that use experimental data to inform the size of the safety factor to be used or, in some cases, to avoid the need for extrapolation i.e. data should supersede default values. Some practical examples of this include:

    a) the use of dose-response relationships to inform the size of factor used to extrapolate from an effects concentration to a no-effects concentration

    b) comparing SSDs for freshwater and saltwater organisms to decide on the suitability of freshwater thresholds for the protection of marine organisms (e.g. Wheeler *et al*, 2001)

    c) estimating no-effects concentrations (or concentrations specifying a low level of effect, e.g. EC10 or EC20) by re-analysing data rather than relying on a default factor to extrapolate from an EC/LC50

2. Extrapolation is uncertain and to rely on fixed safety factors gives a false sense of accuracy. The use of a range of factors is preferable, and this is consistent with the previous point. Current UK practice already employs a range of factors and this is a strength of the approach.

3. A ‘weight of evidence’ approach can be useful in which the outcomes from a variety of endpoints help formulate a final decision about a threshold (Chapman, 1996). The use of field data alongside laboratory toxicity data is a case in point, as would be the use of endpoints describing predicted declines in population size in conjunction with conventional toxicological endpoints.

5.5.3 Site-specific risk assessment

The tiered approach, advocated here, avoids treating EQSs as absolute values but seeks to recognise their limitations (with respect to their accuracy and utility) and treats the generic EQS more as a trigger for more detailed investigation, where there is a risk of environmental impact. Unlike the current situation, the generic EQS is not the final decision-making tool. By supplementing the decision-making process with additional risk assessment tools, more informed decisions may be reached that take better account of:

- costs and benefits
- the practicalities of achieving particular levels of environmental protection and
- the uncertainty that is inherent in the risk assessment process

What is needed at this stage is a way of quantifying the relationship between chemical concentration and biological impact, as illustrated in Chapter 4. Two approaches to site-specific risk assessments are envisaged, the use of (a) concentration-response relationships or (b) species sensitivity distributions. Ultimately, the level of risk reduction required is a policy decision but in practice this may be informed by the environmental objectives for the catchment in question. By considering the magnitude of effect and the associated confidence interval, informed decisions about risk reduction may be formulated, as described in Chapter 4.

The adoption of more probabilistic approaches to site-specific risk assessment may have the effect of providing incentives to generate new data. Where a high level of risk is indicated, but this is based on few data, the associated confidence intervals will be large. Additional data would help reduce that uncertainty. Finally, the use of either concentration-response relationships or SSDs in this way would seem to sit comfortably within the river basin plans concept under the Water Framework Directive. This is because they provide a means of selecting control measures that should reduce pressures on the biological quality of the catchment in a quantifiable way and that take account of costs and benefits in way that is not possible when effects assessment is based on pass/fail thresholds.

a) Concentration-response relationship

A concentration-response plot is developed for the most sensitive species (by definition, drawn from the critical dataset used to derive the generic EQS). Alternatively, a species of economic importance could be selected depending on the protection objectives for the reach/catchment in question. If the main issue is one of compliance with short-term EQSs, then this would usually be based on lethal toxicity data, but if compliance with

long-term standards is an issue, the concentration-response relation is better developed on the basis of long-term and probably sublethal toxicity (if data are available).

b) Species sensitivity distribution

An SSD is developed based on an endpoint that is relevant to the perceived risks (e.g. chronic toxicity or acute toxicity). Ideally, the input data would describe concentrations having no, or low, effects on the measured endpoints and they would be time-independent. A good representation of the local ecology will be important but the need for taxonomic diversity amongst these species also depends on the mode of action of the substance. For example, an emphasis on insects and crustaceans is entirely appropriate for an insecticide, although greater taxonomic diversity would be required for a substance that does not have a specific mode of action.

Ideally, SSDs would be based on the more demographically relevant endpoints discussed in Chapter 2. We have shown that this can be achieved in practice for time-independent survivorship endpoints with a modest investment of modelling expertise and where suitable datasets for sufficient species can be located. Even if only LC50 data are available, this can still tell the risk assessor something useful about the consequences of different risk reduction options although the conclusions are obviously less meaningful than those based on no-effect concentrations.

The level of effort to be invested in trying to secure data that provides greater taxonomic diversity (by experimentation or eliciting expert judgement as suggested in Section 3) or which might yield more relevant endpoints depends on the magnitude of the problem. Thus, a situation in which there are frequent exceedances of a generic EQS over a wide geographical scale, and especially where there is evidence to link the presence of that substance with deterioration in biological quality, would encourage a search for datasets from a wider range of taxa from which more relevant summaries of toxicity can be estimated. Reliance on existing toxicity summaries is more defensible when exceedances of EQSs are isolated, or there is little to implicate the presence of the substance in reduced biological quality.

We recognise that the construction of robust SSDs in such circumstances is not a trivial undertaking. There remains a question therefore about how feasible these proposals are in the current regulatory context and the benefits that would follow, compared to a regime that relied solely on generic EQSs. Nevertheless, we argue that the incorporation of a risk-based approach into an overall regulatory framework, coupled with greater clarity about protection objectives, is consistent with current regulatory thinking and the approaches described in this report provide practical pointers on how this might be achieved.

REFERENCES

Aldenberg, T and Slob, W (1993) *Confidence Limits for Hazardous Concentrations Based on Logistically Distributed NOEC Toxicity Data.* National Institute of Public Health and Environmental Protection (RIVM), Report No.719192992.

Baatrup, E., Bayley M., Sørensen F.F. and Toft G. Can animal behaviour predict population level effects? p 139 in: *Forecasting the Environmental Fate and Effects of Chemicals*, Rainbow, P.S., Hopkin, S.P. and Crane, M., Eds., Wiley, Chichester, UK, 2001.

Barnett V. and O'Hagan A. (1997) Setting Environmental Standards. The statistical approach to handling uncertainty and variation. Report to the Royal Commission on Environmental Pollution. Chapman and Hall.

Barnthouse, L.W., Suter, G.W., Rosen, A.E., and Beauchamp, J.J. (1987) Estimating responses of fish populations to toxic contaminants. *Environmental Toxicology and Chemistry* 6, 811-824.

Barnthouse L.W., Suter G.W. and Rosen A.E. (1990) Risks of toxic contaminants to exploited fish populations: influence of life history data uncertainty and exploitation intensity. *Environmental Toxicology and Chemistry,* 9, 297-311.

Bedaux J J M and Kooijman S A L M (1994) Statistical Analysis of Bioassays based on Hazard modelling. *Journal of Environmental Statistics* 1, 303-314

Berkson J (1994) Application of the logistic function to bioassays. *Journal of the American Statistical Society* 39**,** 357-365

Biggs, J., Corfield, A., Walker, D., Whitfield, M., and Williams, P. (1994) New approaches to the management of ponds. *British Wildlife* 5, 273-287.

Birch L.C. (1948) The intrinsic rate of natural increase in an insect population. *Journal of Animal Ecology* 17, 15-26.

Bliss C I (1935) The dosage-mortality curve. *Annals of Applied Biology* 22, 134-167

Brown A. R., Riddle A.M., Cunningham N.L., Kedwards T.J., Shillabeer N. and Hutchinson T.H. (in press) Predicting the effects of endocrine disrupting chemicals on fish populations. *Human and Ecological Risk Assessment*.

Bruce R D and Versteeg D J (1992) A statistical procedure for modelling continuous toxicity data. *Environmental Toxicology and Chemistry* 11, 1485-1494.

Calow P., Sibley R.M. and Forbes V. (1997) Risk assessment on the basis of simplified life-history scenarios. *Environmental Toxicology and Chemistry*, 16, 1983-1989

Cairns J. (1992) The threshold problem in ecotoxicology. *Ecotoxicology* 1, 3-16.

Carlson D. (1967) Fathead minnow *Pimephales promelas* (Rafinesque) in the Des Moines River, Boone County, Iowa, and the Skunk River drainage, Hamilton and Story Counties, Iowa. *Iowa State Journal of Science*, 43, 363-374.

Caswell H (1996). The analysis of life table response experiments, I. Decomposition of treatment effects on population growth rate. *Ecological Modelling* 46, 221-237.

Caswell H. (2001) Matrix Population Models, 2nd edition, Sinauer , Mass, USA.

Caswell, H. and John, A.M., From the individual to the population in demographic models, in: *Individual-based Models and Approaches in Ecology*, DeAngelis, D.L. and Gross, L.J., Eds., Chapman & Hall, New York, NY, 1992.

Chapman P.F., Crane M., Wiles J., Noppert F., McIndoe E. (1996) Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* 5, 169-186.

Chapman, P.M., Fairbrother, A., and Brown, D. (1998) A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environmental Toxicology and Chemistry* 17, 99-108.

Cox, D.R. and Oakes, D., *Analysis of Survival Data*, Chapman & Hall, London, 1984.

Cox, D.R. (1972) Regression models and life tables (with discussion), *Journal of the Royal Statistical Society,* B34, 187.

Crane M and Chapman P F (1996) Asking the right questions: ecotoxicology and statistics. *Ecotoxicology* 5**,** 137-138

Crane M, Chapman PF, Sparks T, Fenlon, Newman MC. (2002a) Can risk assessment be improved with time to event models? pp 153-166 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL.

Crane M and Grosso A. (2002) Time to event analysis of standard ecotoxicity data. pp 7-22 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL.

Crane M, Newman MC, Chapman PF and Fenlon J. (2002b) *Risk Assessment with Time to Event Models*. Lewis Publishers, Boca Raton, FL.

Crane M, Whitehouse P, Comber S, Watts C, Giddings J, Moore D, Grist E. Aquatic ecotoxicological risks from application of chlorpyrifos to top fruit in the UK. *Pesticide Science*, submitted.

Crane, M. (1997) Research needs for predictive multispecies tests in aquatic toxicology. *Hydrobiologia,* 346, 149-155.

Crane, M., Sorokin, N., Wheeler, J.R., Grosso, A., Whitehouse, P. and Morritt, D. (2001) *European Approaches to Coastal and Estuarine Risk Assessment.* pp 15-39 in: Coastal and Estuarine Risk Assessment, Newman, M.C., Roberts, Jr., M.H. and Hale, R.C. (eds.), Lewis Publishers,.

Daniels R.E. and Allan J.D. (1981) Life table evaluation of chronic exposure to a pesticide. *Canadian Journal of Fisheries and Aquatic Science* 38, 485-494

Day, K. and Kaushik, N.K. (1987) An assessment of the chronic toxicity of the synthetic pyrethroid, fenvalerate, to *Daphnia galeata mendotae*, using life tables, *Environmental Pollution,* 44, 13.

DEFRA (2001) Basic Horticultural Statistics for the UK. Calendar and Crop Years 1990/1 – 2000/1.

Dixon P M and Newman M C (1991) Analysing toxicity data using statistical models of time-to-death: an introduction. pp.207-242 in: *Metal Ecotoxicology: Concepts and Applications,* edited by M C Newman and A W McIntosh, Lewis Publishers, Chelsea, M1.

Dixon P M (2002) Just how much better is a time to event analysis? pp 69-87 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL.

ECETOC (1993) Aquatic Toxicity Data Evaluation, Technical Report No. 56

Emlen J.M. (1989) Terrestrial population models for ecological risk assessment: a state-of-the-art review. *Environmental Toxicology and Chemistry* 8, 831-842.

Fitter R, Manuel R. 1994. *Lakes, Rivers, Streams and Ponds of Britain and North-West Europe*. Collins Photo Guide, Harper Collins, Hong Kong.

Forbes T. L. (1993) The design and analysis of concentration-response experiments. pp 438-459 in: *Handbook of Ecotoxicology, Volume 1,* edited by Peter Calow, Blackwell Science Ltd, Oxford.

Forbes T. L. and Forbes V. E. (1993) A critique of the use of distribution-based extrapolation models in ecotoxicology. *Functional Ecology* 7**,** 249-254

Forbes V.E., Calow P. and Sibly R.M. (2001) Are current species extrapolation models a good basis for ecological risk assessment? *Environmental Toxicology and Chemistry* 20, 442-447

Forbes, V.E., and Calow, P. (1999) Is the *per capita* rate of increase a good measure of population-level effects in ecotoxicology? *Environmental Toxicology and Chemistry* 18, 1544-1556.

Forbes V.E. and Calow P. (2002) Species sensitivity distributions revisited: a critical appraisal. *Human and Ecological Risk Assessment* 8, 473-492.

Fraunhofer Institute (2001) Development of methods for the derivation of Quality Standards in the context of the Water Framework Directive. Draft Report EAF(2) – 06/03/FHI.

Giddings, J. M., Biever, R. C. & Racke, K. D. (1997) Fate of chlorpyrifos in outdoor pond microcosms and effects on growth and survival of bluegill sunfish. *Environmental Toxicology and Chemistry* 16, 2353-2362.

Giesy, J. P., Solomon, K. R., Coats, J. R., Dixon, K.R., Giddings, J. M. & Kenaga, E. E. (1999) Chlorpyrifos: Ecological risk assessment in North American aquatic environments. *Reviews in Environmental Contamination and Toxicol*ogy 160, 1-129.

Girling, A.E., Tattersfield, L., Mitchell, G.C., Crossland, N.O., Pascoe, D., Blockwell, S.J., Maund, S.J., Taylor, E.J., Wenzel, A., Janssen,C.R.and Juttner, I. (2000) Derivation of Predicted No-Effect Concentrations for Lindane, 3,4-Dichloroaniline, Atrazine, and Copper. *Ecotoxicology and Environmental Safety* 46, 148-162.

Grist E.P.M., Crane M, Jones C., Whitehouse P. (2002) Estimation of demographic toxicity through the double bootstrap. *Water Research* 37, 618-626.

Hartley H.O. and Sielken R.L. Jr. (1997) Estimation of 'safe doses' in carcinogenic experiments. *Biometrics* 33**,** 1-30

HMSO (2000) Government Response to the Royal Commission on Environmental Pollution's Twenty-First Report 'Setting Environmental Standards', Cm 4794, July 2000.

Huebert D.B. and Shay J.M. (1993) Considerations in the assessment of toxicity using duckweeds. *Environmental Toxicology and Chemistry* 12, 481-483.

Illing H.P.A. (1999) Are societal judgements being incorporated into the uncertainty factors used in toxicological risk assessment? *Regulatory Toxicology and Pharmacology* 29, 300-308.

Jagoe R H and Newman M C (1997) Bootstrap estimation of community NOEC values. *Ecotoxicology* 6**,** 293-306

Kammenga J. and Laskowski R. Demography in Ecotoxicology. John Wiley and Sons Ltd, Chichester UK, 2000.

Karman C.C. (2002) Using time to event modelling to assess the ecological risk of produced water discharges. In, *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL. pp 89-102.

Kooijman S.A.L.M. (1987) A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Research* 21, 269-276

Kuhn, A., Munns, W.R.Jr, Poucher, S., Champlin, D. and Lussier, S. (2000) Prediction of population-level response from mysid toxicity test data using population modelling techniques. *Environmental Toxicology and Chemistry* 19, 2364-2371.

Laenge, R., Hutchinson, T.H., Croudace, C.P., Siegmund, F., Schweinfurth, H., Hampe, P., Panter, G.H. and Sumpter, J.P. (2001) Effects of the synthetic estrogen 17α-ethinylestradiol on the life-cycle of the fathead minnow (*Pimephales promelas*). *Environmental Toxicology and Chemistry* 20(6), 1216-1227.

Laenge, R., Hutchinson, T.H., Scholz, N. and Solbe, J. (1998) Analysis of the Ecetoc Aquatic Toxicity (EAT0 Database.II- Comparison of Acute to Chronic Ratios for various Aquatic Organisms and Chemical Substances. *Chemosphere* 36, 115-127.

Leslie, P.H., Tener, J.S., Vizoso, M. and Chitty, H. (1955) The longevity and fertility of the Orkney vole, *Microtus orcadensis*, as observed in the laboratory, *Proceedings of the Zoological Society of London,* 125, 115.

Lin B., Miyamoto K., Yoshida K. and Nakanishi J. (2002) Risk assessment of 4-nonylphenol by a population-level approach using life cycle toxicity data for medaka (*Oryzias latipes*). SETAC Europe 12th Annual Conference, Vienna, May 2002.

Manly BFJ. 2002. Time to event analysis in ecology. pp 121-140 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL.

Martínez-Jerónimo, F., Villaseñor, R., Espinosa, F. and Rios, G. (1993) Use of life-tables and application factors for evaluating chronic toxicity of Kraft mill wastes on *Daphnia magna*, *Bulletin of Environmental Contamination and Toxicology,* 50, 377.

Matthiessen P.M., Reed J., and Johnson M. (1999) Sources and potential effects of copper and zinc concentrations in the estuarine waters of Essex and Suffolk, United Kingdom. *Marine Pollution Bulletin*, 38, 908-920.

Mayer FL, Ellersieck MR, Krause GF, Sun K, Lee G, Buckler DR. 2002. Time-concentration-effect models in predicting chronic toxicity from acute toxicity data. pp 39-67 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL.

Newman M C *Quantitative Methods in Aquatic Ecotoxicology.* Lewis Publishers, Boca Raton, FL, 1995.

Newman M C and Aplin M (1992) Enhancing toxicity data interpretation and prediction of ecological risk with survival time modelling; an illustration using sodium chloride toxicity to mosquitofish (*Gambusia holbrooki). Aquatic Toxicology* 23**,** 85-96

Newman M.C. Population Ecotoxicology, John Wiley and Sons Ltd, Chichester UK, 2001.

Newman M.C. and Crane M. (2002) Introduction to time to event methods. pp 1-6 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL..

Newman M.C. and McCloskey .J.T. (1996) Time to event analysis of ecotoxicity data. *Ecotoxicology* 5, 187-196.

Newman M.C. and McCloskey .J.T. (2002) Applying time to event methods to assess pollutant effects on populations. pp 23-38 in: *Risk Assessment with Time to Event Models* (Crane M, Newman MC, Chapman PF, Fenlon J, eds.), Lewis Publishers, Boca Raton, FL.

Newman MC, Ownby DR, Mézin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson B-A. (2000) Applying species sensitivity distributions in ecological risk assessment, assumptions of distribution type and sufficient numbers of species. *Environmental Toxicology and Chemistry* 19, 508-515.

Nyholm, N. (1990) Expression of Results from Growth Inhibition Toxicity Tests with Algae. *Archives of Environmental Contamination and Toxicology* 19, 518-522.

Odum E.P., Finn J.T. and Franz E.H. (1979) Perturbation theory and the subsidy-stress gradient. *BioScience* 29, 349-352.

OECD (1993) Guidelines for Testing of Chemicals. Volume 1.

OECD (1996) *Draft Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data, Braunschweig, Germany, 15-17 October 1996*. Organisation for Economic Co-operation and Development, Paris. 50 pp

Pesch, C.E., Munns, W.R. and Gutjahr-Gobell, R. (1991) Effects of a contaminated sediment on life history traits and population growth rate of *Neanthes arenaceodentata* (Polychaeta: Nereidae) in the laboratory, *Environmental Toxicology and Chemistry,* 10, 805

Pollard, S., Duarte-Davidson, R.,Yearsley, R., Twigger-Ross, C., Fisher, J., Willows, R. and Irwin, J.(2000) A Strategic Approach to the Consideration of Environmental Harm. Report Number 36 by The National Centre for Risk Analysis and Options Appraisal, Environment Agency.

Pyke, D.A. and Thompson, J.N. (1986) Statistical analysis of survival and removal experiments, *Ecology,* 67, 240

Royal Commission on Environmental Pollution (1998) Setting Environmental Standards. 21st Report of the Royal Commission on Environmental Pollution, The Stationery Office Ltd.

Sheehan D.M. (2000) *Proceedings of the Society for Experimental and Biological Medicine*, 224, 57-60.

Sibly R.M. (1999) Efficient experimental design for studying stress and population density in animal populations. *Ecological Applications* 9, 496-503.

Smith E.P. and Cairns J. (1993) Extrapolation methods for setting ecological standards for water quality: statistical and ecological concerns. *Ecotoxicology* 2, 203-219

Snell, T. and Serra, M. (2000) Using probability of extinction to evaluate the ecological significance of toxicant effects. *Environmental Toxicology and Chemistry* 19, 2357-2363.

Solomon K R, Baker D B, Richard R P, Dixon K R, Klaine S J, La Point T W, Kendall R J, Weisskopf C P, Giddings J M, Giesy J P, Hall L W Jr, and Williams W M (1996) Ecological risk assessment of atrazine in North American surface waters. *Environmental Toxicology and Chemistry* 15, 31-76

Solomon, K. R., Baker, D.B., Richards, R.P. Dixon, K.R., Klaine, S.J., La Point, Sprague, J.B. (1969) Measurement of pollutant toxicity to fish. I. Bioassay methods for acute toxicity. *Water Research* 3, 793.

Steen, R. J. C. A., Leonards, P. E. G., Brinkman, U.A.T., Barcelo, D., Tronczynski, J.,

Sun K, G F Krause, F L Mayer Jr, M R Ellersieck and A P Basu (1995) Predicting chronic lethality of chemicals to fishes from acute toxicity test data: theory of accelerated life testing. *Environmental Toxicology and Chemistry* 14, 1745-1752

Tanaka, Y. and Nakanishi, J. (2000) Mean extinction times of populations under toxicant stress and ecological risk assessment. *Environmental Toxicology and Chemistry* 19, 2856-2862.

US EPA (1991) Technical support document for water quality-based toxics control. EPA/505/2-90-001. PB91-127415.Washington DC.

US EPA (2000) Benchmark dose technical guidance document. Draft for external review. EPA/630/R-00/001. Risk Assessment Forum. United States Environmental Protection Agency, Washington DC.

USEPA (1984) Estimating Concern Levels for Concentrations of Chemical Substances in the Environment. Environmental Effects Branch, USEPA, February 1984

van Straalen, N. M. and van Rijn, J.P. (1998) Ecotoxicological risk assessment of soil fauna recovery from pesticide application. *Reviews of Environmental Contamination and Toxicology* 154, 85-141.

van Wijngaarden R, Leeuwangh P, Lucassen WGH, Romijn K, Ronday R, van der Velde R, Willigenburg W. 1993. Acute toxicity of chlorpyrifos to fish, a newt, and aquatic invertebrates. *Bulletin of Environmental Contamination and Toxicology* 51, 716-723.

Wagner C. and Lokke H. (1991) Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Research* 25, 127-1242

Walker, C.H., Hopkin, S.P.,Sibly, R.M. and Peakall, D.B. (1996) Principles of Ecotoxicology. Taylor and Francis, London, UK.

Walthall W.K. and Stark J.D. (1997) A comparison of acute mortality and population growth rate as endpoints of toxicological effects. *Ecotoxicology and Environmental Safety* 37, 45-52.

Wheeler J., Sorokin N., Leung K., Morritt D., Crane M., Cole S., Mitchell R., Holmes D., Pugh M., Rogers H., Whitehouse P. and Karman C. (2001) Marine risk assessment and ecosystem dynamics: comparison of marine and freshwater data and test methods. Report to CEFIC LRI ECO 1b, WRc Report No. CO 4972

Whitehouse P., Moran G. and Gowers A. (2000) Development of Environmental Impact Methodology for IPPC: Setting Environmental Assessment Levels. Interim Report to the Environment Agency, WRc-NSF Ltd.

Wright JF, Furse MT, Armitage PD. (1994) Use of macroinvertebrate communities to detect environmental stress in running waters. pp 15-34 in: Water quality and stress indicators in marine and freshwater systems (ed. Sutcliffe DW). Freshwater Biological Association, Ambleside, UK,

Young W.F., Whitehouse P., Johnson I., and Sorokin N. (2002) Proposed Predicted No Effect Concentrations (PNECs) for Natural and Synthetic Steroid Oestrogens in Surface Waters. Environment Agency Technical Report P2-T04/1, Environment Agency, Bristol, UK.

Zabel T.F. and Cole S. (1999) The derivation of Environmental Quality standards for the protection of aquatic life in the UK. *Journal of the Chartered Institute of water and Environmental Management* 13, 436-440.

APPENDIX A DOUBLE BOOTSTRAP MODEL FOR ESTIMATING R AND ASSOCIATED UNCERTAINTY

Introduction

This Appendix describes a 'double bootstrap' approach which can be used to generate a percentile confidence interval for quantifying the uncertainty associated with such an estimate (Davison and Hinkley 1999, Grist *et al*. 2001). The computer intensive method exploits empirical variation present in the sample data, on the basis that the sample distribution is representative of the (usually unknown) distribution from whence it came. It makes use of a 'bootstrap' followed by a bootstrap regression in which the individual data gathered in Life Table Response Experiments (LTREs) are used to generate a large set of curve replicates from which a pointwise percentile confidence interval is constructed.

The Leslie Matrix Model

Motivation for a matrix population model was driven historically by the desire to project natural population dynamics from life table data (Leslie 1945). In fact, the Leslie representation is an example of a state transition model (e.g. Zadeh and Desoer 1963) innovatively applied to biological populations by both Lewis (1942) and Leslie (1945). The 'Leslie' matrix **M** is specified by survival P_i and fecundity F_i parameters calculated across the m discrete classes of a life table defined by

$P_i \equiv$ the proportion surviving from age class i to age class i+1

$F_i \equiv$ the number of offspring produced by age class i

in a *fixed* discrete time interval (usually one day for a chronic test). Population dynamics are then described by the Leslie model

$$\boldsymbol{N}_{t+1} = \mathbf{M}\boldsymbol{N}_t$$

where $\mathbf{M}(P_i, F_i)$ is the Leslie Matrix and $\boldsymbol{N}_t$ is the population at time t.

In practice the vital rates P_i and F_i must be determined as average values taken over a cohort of n individuals held in a given life table experiment. Each of these parameters is calculated from the survival $l(t)$ and fecundity $m(t)$ values obtained for a life table cohort at successive observations over a protracted period. Clearly, discretization of time implies an absence of recorded information on the continuous functions $l(t)$ and $m(t)$ within the fixed interval spanned by each age class. Estimation of their average values is dependent on whether reproduction occurs continuously ('birth flow') or periodically ('birth pulse') within the interval (e.g. Caswell 2001). Hence it is necessary to know in advance the type of reproductive behaviour exhibited by the organism under scrutiny, in order to achieve an appropriate parameterization of the matrix **M**.

The iterative dynamics of the Leslie model can be projected and analysed to a large extent (Caswell 2001). Most importantly, in general the dominant eigenvalue λ of **M** gives a measure of the intrinsic rate of increase of the population according to whether

$\lambda > 1$ [$r > 0$] $\Rightarrow$ Population increases

$\lambda = 1$ [$r = 0$] $\Rightarrow$ Population is static

$\lambda < 1$ [$r < 0$] $\Rightarrow$ Population becomes extinct

This follows from the ergodic property that successive iterates $\boldsymbol{N_t}$ of any initial population $\boldsymbol{N_0}$ must eventually converge to the dominant eigenvector $\boldsymbol{V_\lambda}$ associated with λ. Once the stable ratio $\boldsymbol{V_\lambda}$ between age classes has been reached it will continue to be maintained thereafter with successive iterates of the population changing in absolute value by a factor of λ across all the age classes. It is in this sense that the value of $r = \ln(\lambda)$ is an estimate of the intrinsic rate of increase of the population.

Toxicity assessment from LTREs

Where survivorship and fecundity data are recorded for discrete age classes in the form of a life table, r may be determined from the dominant eigenvalue λ of the associated Leslie matrix **M** (for $r = \ln(\lambda)$). The effect of a toxic substance at population level can be estimated by comparing the control value r_0 with the value obtained for separate cohorts held in exposure to different concentrations of the substance (Caswell 1996). Under a range of *different* treatments in which R separate cohorts are held at toxic concentrations of $\{c_i\} = c_1, c_2, \ldots c_R$, giving rise to respective 'treatment population' growth rates $\{r_i\} = r_1, r_2, \ldots r_R$, the plot of $\{r_i\}$ *versus* $\{c_i\}$ for $i=0,1,..R$ depicts the toxicity of the substance on the natural population (Caswell 1996). If a model $r = f(c)$ is regressed to these data, the effect on r_0 of any designated concentration can be estimated from the curve of best fit. Conversely, the concentration required to reduce r_0 by a designated percentage x can be estimated. This leads logically to an 'ErCx' summary metric for the endpoint r (the Effect-on-r-Concentration percentage), which we define as:-

ErCx ≡ that concentration of toxic substance estimated to reduce r_0 by a specified percentage x

For example, the ErC100 is the concentration of toxic substance estimated to reduce r by 100% (therefore to $r = 0$), the ErC50 reduces r_0 by 50% and so on. But how much uncertainty is to be associated with such an estimate?

Standard statistical methodology can be used to obtain a $(1-2\alpha)$ % confidence interval at the α % significance level for a regression curve by assuming that corresponding residuals are drawn from a normal distribution (e.g. Sokal and Rohlf 1995). As well as being reliant on the underlying distributional assumption, this approach takes no account of the two independent sources of uncertainty that influence curve construction. First, the $\{r_i\}$ are point estimates and as such are subject to the combined uncertainty due to variabilities in individual responses (heterogeneity) and environmental conditions (environmental noise) as well as possible measurement and sampling error. There is therefore uncertainty in the vertical direction to be associated with each of the observational data points. This imputes a (second) level of uncertainty in curve configuration, because a 'best fit' curve is dependent on the placement of the points to which it is to be fitted. The purpose of this paper is to derive a confidence interval for the curve $r = f(c)$ which incorporates both sources of uncertainty in its construction.

The 'double bootstrap'

By definition, a life table is a summary schedule taken over a set of n individuals so that a large number of B resamples randomly drawn with replacement from this set give a bootstrap distribution of the uncertainty of the value of r (e.g. Caswell 2001). The first bootstrap in the 'double bootstrap' resamples the individual (or replicate) records in each of the R life tables in this manner to generate a first bootstrap distribution for each of the treatment population growth rates $\{r_i\}$. The set of R distributions are statistically independent because different cohorts are used in each treatment. It implies a second 'block bootstrap' can be used to generate resamples (each of size R) composed of one member randomly drawn from each of these R separate distributions (Efron and Tibshirani. 1993). A regression curve fitted to any particular block bootstrap resample is therefore a replicate of the best fit curve because it represents a possible variant of the curve configuration. A pointwise percentile confidence interval obtained by slicing vertically (or horizontally) through a large number of such curve replicates thus gives a measure of the uncertainty associated with any point on the original curve (e.g. Davison and Hinkley 1999). Lower and upper points at which a horizontal line of altitude $r(100-x)/100$ cuts the boundaries of this interval then represent a $(1-2\alpha)$ % percentile based bootstrap confidence interval for any ErCx estimate. Observe also that a bootstrap percentile confidence interval has a more intuitive interpretation than the confidence interval of standard 'frequentist' statistics in that it does imply a $(1-2\alpha)$ % probability of the estimate being contained within it (as opposed to a probability that $(1-2\alpha)$ % of the intervals will contain the estimate).

Discussion

Although the EC50 is traditional, there is a constant debate in the literature on what the x in the ECx should be (Chapman *et al*. 1996, Crane and Newman 2000). Moreover, the use of post-ANOVA testing to identify a concentration at which there is no effect (the NOEC) is a fundamentally invalid application of hypothesis testing. Failure to reject the null hypothesis of 'no difference' does not imply that there was no effect, but rather that there was a specified probability (the chosen significance level) of committing a Type I error. The term 'significant' implies a significant deviation from the null hypothesis but at a level that is arbitrarily specified. Hence an effect which is biologically but not statistically significant may pass undetected in the NOEC. Precisely this Type I error was encountered in the *Daphnia* example, where a 20% reduction (at the least) in the control population growth rate was not detected by the NOEC.

The 'ecological relevance' of r is clearly greater than an endpoint based upon survival or some other partial demographic endpoint (Newman 2001). However, an environmental regulator will still ask the question 'what level of effect is acceptable?' What is an acceptable reduction in the value of r? The ErCx is a sub-lethal metric that can be used to address this question. It is fully definable in the range $0< x<100$ whenever the relationship between population growth rate and concentration is not asymptotic. Subject to the usual caution to be exercised when applying any extrapolation method, the double bootstrap will provide both lower and upper bounds of the associated uncertainty of any ErCx. However, as with any 'frequentist' (non-Bayesian) statistical approach these confidence bounds cannot account for *inherent* model uncertainty, for example whether to choose a linear or quadratic regression model (e.g. O' Hagan 2001).

Resampling approaches such as the bootstrap have been applied across many diverse areas of ecology and ecotoxicology (Manly 1997a). These include population genetics (Rousset and Raymond 1997), time to event analysis (Newman 2001, Manly 2002), contaminated prey selection by predators (Mackay *et al*. 2002), construction of species sensitivity distributions (Jagoe and Newman 1997, Newman *et al*. 2000 & 2002) and conservation biology (Burgman *et al*. 1993). The advantages of resampling approaches are that they need not rely on distributional assumptions and can be easily implemented through software to make the necessary computations quickly (e.g., Ferson and Akçakaya 1990, Manly 1997b). The main disadvantage associated unfocused use of the bootstrap with small samples is that information on the true distribution of a variable, perhaps based on knowledge of underlying mechanisms, could be ignored. However, the same is true of inappropriate use of parametric distributions when these are unjustified. For example, Newman *et al*. (2000, 2002) found that a large number of species sensitivity distributions for different chemicals did not fit the standard lognormal distribution applied in North America, and concluded that use of a nonparametric bootstrap was preferable. In general, the associated summary statistics and confidence limits will be dependent upon the choice of statistical methodology and underlying model assumptions (e.g. Grist *et al* 2002, Verdonck *et al*. 2001). This importantly implies, as we have demonstrated in this paper, that there can be substantial differences in the 'safe' threshold levels that are ultimately estimated (Wheeler *et al*. 2002).

The E*r*C100 provides an upper bound on any perceived 'safe' threshold because it is the maximum concentration which a population can be estimated to endure before a projected decline towards extinction ensues. In reality, the E*r*C100 is unlikely to be sufficiently protective because of fluctuations in the population that will result from demographic and environmental stochasticity. Hence there is a requirement for the regulator to specify an x<100 which is perceived as sufficiently 'safe'. How can this be done?

The simplest approach is by appealing to whatever is considered to be an 'acceptable risk' of extinction (Ginzberg *et al*. 1982). Once this is specified, a brute force method through simulation studies can be used to determine a corresponding value of x. This would be based on a large number of realizations run for a long time from a large number of arbitrary initial populations. This empirical approach is used in the derivation of 'risk statistics' as employed for example, by the risk analysis package RAMAS (Spencer and Ferguson 1997). In the situation described here however, such an approach is complicated by the fact that specification of r does not uniquely specify an associated Leslie Matrix. A method is therefore required to extrapolate from the given Leslie Matrix obtained for each treatment (from which the $\{r_i\}$ are derived) to an appropriate Leslie Matrix corresponding to any specified value of r.

The generality of extrapolation methods in ecotoxicology has previously raised concerns (Smith and Cairns 1993). However, in estimating an E*r*Cx, there is scope in the laboratory protocol to ensure at least one treatment population growth rate is reduced to below zero. The relationship between r and c as described by the chosen function f is then firmly established as not being asymptotic with $r=0$ so that estimation of any E*r*Cx threshold for x<=100 is automatically enforced to be through interpolation rather than an extrapolation.

Conclusions

In summary, the methodology developed here demonstrates several important features to be addressed in the setting of environmental standards. A 'safe' threshold is dependent on the following.

1. The perceived relationship (as depicted by the function *f*) between population growth rate *r* and the substance concentration *c*. The choice of model function *f* is likely to be a subjective one based on available data.

2. The endpoint and its associated metric value considered 'safe'. For population growth rate *r,* the threshold concentration was specified in terms of an x% permissible reduction in the growth rate of the control experiment and denoted as the E*r*Cx. An upper bound for such an estimate is automatically the ErC100.

3. Inherent uncertainty due to the limitations and inaccuracies of available data. The double bootstrap provides an estimate of the uncertainty associated with the chosen E*r*Cx. The extent of such uncertainty is dependent also on the extent to which extrapolation or interpolation is required to 'reach' the perceived permissible reduction of x% at altitude *r*(100- x)/ 100. In general, but dependent on the configuration of data points, more uncertainty is associated with the former than the latter. This follows by virtue of lower and upper data constraints that will have to be met for any estimate obtained by interpolation.

References (Appendix A)

Burgman M.A., Ferson S., Akçakaya H.R. (1993). *Risk Assessment in Conservation Biology*. Chapman and Hall, London.

Caswell H. (1996). The analysis of life table response experiments, I. Decomposition of treatment effects on population growth rate. *Ecological Modelling*, **46**, 221-237.

Caswell H. (2001) *Matrix population models*, Sinauer Associates, 2nd edition, Sunderland Massachusets, USA.

Chapman P.F., Crane M., Wiles J., Noppert F., McIndoe E. (1996) Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicology* **5**, 169-186.

Crane M., Grosso A., Janssen C. (2000) Statistical techniques for the ecological risk assessment of chemicals in freshwaters. In, Sparks T, ed, *Statistics in Ecotoxicology*, John Wiley & Sons, Chichester, John Wiley & Sons, Chichester, UK. pp. 247-278.

Crane, M. and Newman, M.C. (2000) What level of effect is a No Observed Effect? *Environmental Toxicology and Chemistry,* **19**, 516-519.

Davison A.C, Hinkley D.V. (1999) *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, UK.

Efron B., and Tibshirani R.J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, New York, USA.

Ferson S., Akçakaya H.R. (1990). *RAMAS/Age User Manual*, Exeter Software, Setauket, NY.

Forbes V.E., Calow P. and Sibly R.M. (2001) Are current species extrapolation models a good basis for ecological assessment*? Environmental Toxicology and Chemistry,* **20**(2), 442-447.

Ginzburg L.R., Slobodkin L.B., Johnson K. and Bindman A.G. (1982) Quasiextinction probabilities as a measure of impact on population growth. *Risk Analysis* **21**,171-181.

Grist E.P.M., Leung K.M.Y., Wheeler J.R. and Crane M. (2002) Better bootstrap estimation of hazardous concentration thresholds for aquatic assemblages. *Environmental Toxicology and Chemistry,* **21***(7) (in press)*

Jagoe R.H., Newman M.C. (1997). Bootstrap estimation of community NOEC values. Ecotoxicology 6:293-306.

Kammenga J. and Laskowski R. (eds) (2000) *Demography in ecotoxicology*, John Wiley and sons Ltd, Chichester UK.

Leslie P.H. (1945) On the use of matrices in certain population mathematics. *Biometrika* **35**, 213-245.

Levin L., Caswell H., Bridges T., Dibacco C., Cabrera D., and Plaia G. (1996) Demographic responses of estuarine polychaetes to pollutants: life table response experiments. *Ecological Applications* **6**(4), 1295-1313.

Lewis E.G. (1942) On the generation and growth of a population. *Sankhya: the Indian Journal of Statistics* **6**, 93-96.

Mackay C.E., Colton J.A., Bigham G. (2002). Structuring population-based ecological risk assessments in a dynamic landscape. In, Newman MC, Roberts Jr MH, Hale RC (eds.) Coastal and Estuarine Risk Assessment, Lewis Publishers, Boca Raton, FL. pp 273-296.

Manly B.F.J. (1997a). *Randomisation, Bootstrap and Monte Carlo Methods in Biology*, 2nd Edition. Chapman and Hall, London.

Manly B.F.J. (1997b). *RT, A Program for Randomization Testing*, Version 2.1. Centre for Applications of Statistics and Mathematics, University of Otago, Dunedin, New Zealand.

Manly B.F.J. (2002). Time to event analysis in ecology. In, Crane M., Newman M.C., Chapman P.F., Fenlon J. (eds.) Risk Assessment with Time to Event Models. Lewis Publishers, Boca Raton, FL. pp 121-140.

MATLAB release 12, The Mathworks Inc (2000), MA 01760-2098, USA.

Mayer, F.L., Krause, G.F., Buckler, D.R., Ellersieck, M.R. and Lee, G. (1994) Predicting chronic lethality of chemicals to fishes from acute toxicity test data: concepts and linear regression. *Environmental Toxicology and Chemistry*, **13**, 671-678.

Newman M.C. (1995) *Quantitative Methods in Aquatic Ecotoxicology*. Lewis Publishers, Boca Raton, FL. USA.

Newman MC, Ownby DR, Mézin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson B-A. (2000). Applying species sensitivity distributions in ecological risk assessment, assumptions of distribution type and sufficient numbers of species. *Environmental Toxicology and Chemistry,* **19**, 508-515.

Newman M.C. (2001) *Population ecotoxicology*, John Wiley and sons, Chichester UK.

Newman M.C., Ownby D.R., Mézin L.C.A., Powell D.C., Christensen T.R.L., Lerberg S.B., Anderson B.A., Padma T.V. (2002). Species sensitivity distributions in ecological risk assessment: distributional assumptions, alternate bootstrap techniques, and estimation of adequate number of species. In, Posthuma L., Suter II G.W., Traas T.P. (eds.) *Species Sensitivity Distributions in Ecotoxicology*. Lewis Publishers, Boca Raton, FL. pp119-132.

OECD (1984) Guidelines for the Testing of Chemicals. 202 *Daphnia* sp. Acute Immobilisation Test and Reproduction Test. Organisation for Economic Co-operation and Development, Paris, France.

OECD (1996) *Report of the OECD Workshop on Statistical Analysis of Aquatic Toxicity Data*, Braunschweig, Germany, 15-17 October, 1996. Organisation for Economic Co-operation and Development, Paris, France.

O'Hagan A. 2001. Uncertainty in Toxicological Predictions: The Bayesian Approach to statistics. In Rainbow P.S., Hopkin S.P., Crane M., eds, *Forecasting the environmental fate and effects of chemicals*, Wiley, Chichester, UK.

Rousset F, Raymond M. (1997). Statistical analyses of population genetic data: new tools, old concepts. *TREE* 12:313-317.

Smith E.P and Cairns J. jr. (1993) Extrapolation methods for setting ecological standards for water quality: statistical and ecological concerns, *Ecotoxicology* **2**, 203-219.

Sokal R.R. and Rohlf F.J. (1995) *Biometry, the principles and practice of statistics in biological research 3rd edition,* W.H.Freeman and Co., New York, USA.

Spencer M. and Ferson S. (1997) RAMAS ® Ecotoxicology version 1.09, Applied biomathematics, Setauket, New York, USA.

Verdonck F.A.M., Jaworska J., Thas O., Vanrolleghem P.A. (2001). Determining environmental standards using bootstrapping, bayesian and maximum likelihood techniques: a comparative study. *Analytica Chimica Acta* 446:429-438.

Wheeler J.R., Grist E.P.M., Leung K.M.Y., Morritt D., Crane M. (2002). Species sensitivity distributions: data and model choice. *Marine Pollution Bulletin,* 45(1), part 12 (in press).

Zadeh L.A. and Desoer C.A. (1963) *Linear system theory*. McGraw-Hill, New York, USA.

APPENDIX B LIFE TABLE ANALYSIS OF DATA FROM CHRONIC STUDY WITH DAPHNIA MAGNA EXPOSED TO ZINC SULPHATE

Methodology

Experimental details

A 21 day LTRE (Appendix A) was performed with *Daphnia magna* juveniles under 24 hours old. Individuals were held in 5 different concentrations of zinc sulphate (Sigma, analytical grade). The experiments were performed in accordance with OECD guidelines (OECD 1984) with the following modifications:

1. 20 animals were exposed in control vessels and 10 animals in treated vessels.

2. The test medium was changed every Monday, Wednesday and Friday during the tests.

3. Animals were fed daily with 0.45 ml of *Chlorella vulgaris* suspension.

Each individual was held in a separate test vessel thus giving 20 replicates for the control and 10 replicates for each concentration. The zinc sulphate (Zn) concentrations were 0.022, 0.046, 0.1, 0.22 and 0.46 mg/L. All dilutions were in carbon filtered tap water, which was also used to culture the test organisms. The survival and reproductive output of each individual was noted on a daily basis throughout the 21 day study period.

Data analysis

Using survival and reproductive out put data, a life table for each treatment was constructed. Because reproductive timing occurred without periodicity within a daily time interval, the 'birthflow' average of Caswell (2001) was used to parameterize the corresponding Leslie matrix to the data of each life table. In the absence of further information, the choice of an appropriate function $r = f(c)$ describing the relationship between population growth rate r and exposure concentration c must hence be empirically based. We demonstrate how to proceed with the double bootstrap in two simple situations where f is perceived to be linear and nonlinear.

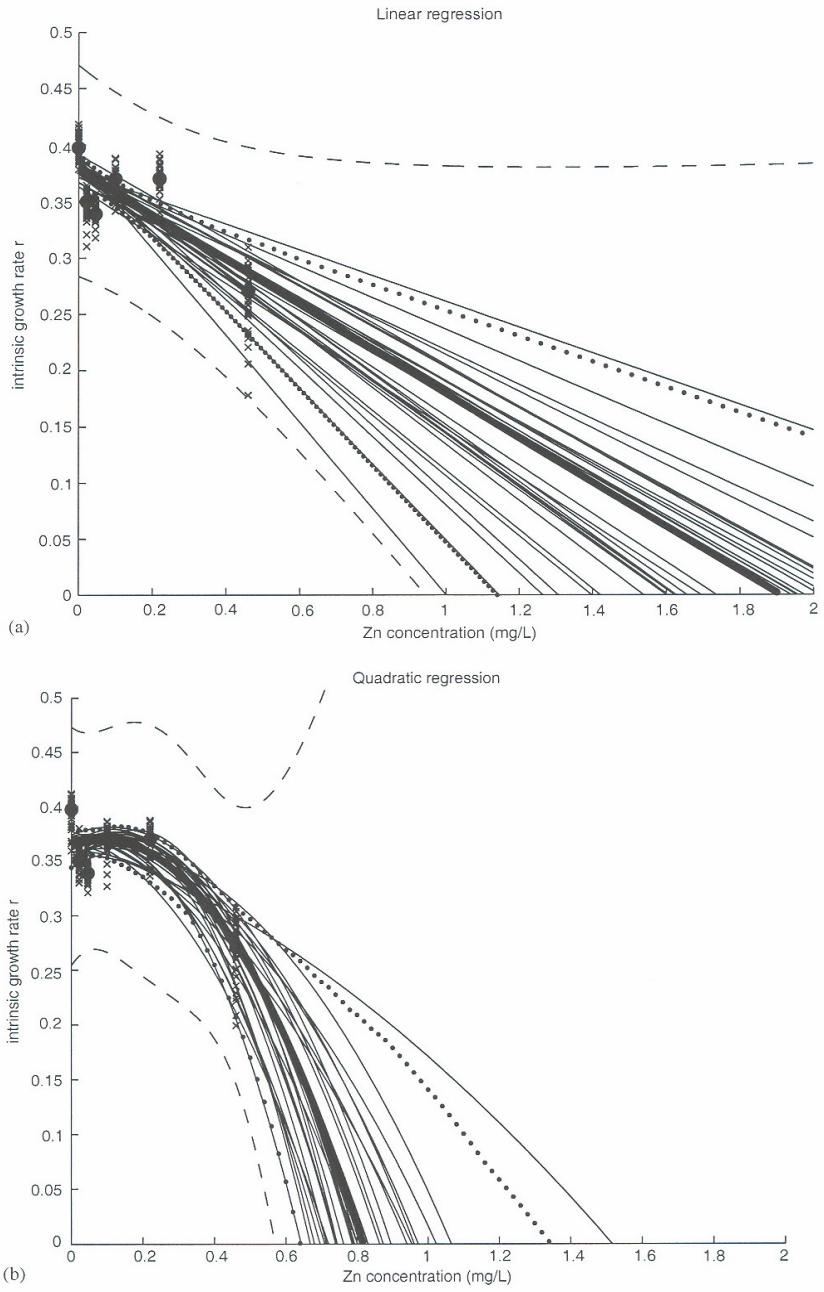


Figure 1: **Double bootstrap results obtained with (a) linear $f(x)=ax+b$, and (b) quadratic $f(x)=ax^2+bx+c$ regression models. In each frame, the curve of best fit (thick) is shown with 30 double bootstrap replicates (thin curves) and 95% pointwise percentile confidence intervals (dotted) obtained with B=2000. Normal distribution 95% confidence intervals (dashed) for the curve of best fit are also shown. • = data, x = resamples (30 shown). 'Best fit' curve parameters were (a) a= -0.198, b= 0.378, (r^2 = 0.630) and (b) a= -0.698 , b=0.125 , c= 0.364, (r^2 = 0.745) [NB: r^2 is the coefficient of determination in this context]**

Control and treatment population growth rates $\{r_i\}$ are plotted *versus* respective exposure concentrations $\{c_i\}$ for $i=0,1,..R$ (solid circles) together with respective 'best fit' regression curves (thick curve) in Figure 1(a) when f is the linear function $f(x)=ax+b$, and in Figure 1(b) when f is the quadratic function $f(x)=ax^2+bx+c$. Treatment population growth rates were relatively closely scattered to the control value so that any E*r*Cx beyond x=50 must be estimated by extrapolation whereas below x=50 it must be obtained by interpolation. However, neither description in view of these data turns out to be asymptotic with the horizontal axis (r =0). Therefore an implicit corollary in both cases is that toxic effect must ultimately *extrapolate* to a situation where population growth rate reaches zero. So, what is the concentration estimated to achieve this and the uncertainty associated with such an estimate?

If the safe threshold is perceived to be the E*r*C100, the concentration estimated to achieve this 100% reduction in population growth rate to $r = 0$ is estimated to be at the intercept of the best fit curve with the horizontal axis. Figure 1 shows the 95% pointwise percentile confidence boundaries (dotted lines) for each best fit curve obtained by the double bootstrap with B=2000 (30 replicates are superimposed) together with the standard 95% confidence limits based on normal distribution theory (dashed lines). B=2000 sufficiently reduced the Monte Carlo sampling error to permit stable bootstrap percentile confidence boundaries as in general (Efron and Tibshirani 1993). Thus in answer to the above question, the E*r*C100 is estimated to be 1.91 mg/L and within 1.14 to 3.28 mg/L with 95% certainty from description (a) or to be 0.82 mg/L and to fall within 0.64 to 1.33 mg/L from description (b). Regressions were performed using MATLAB (2000) in (a) by the method of least squares and in (b) by Gauss-Newton nonlinear regression with Levenberg-Marquardt modifications for global convergence.

It turns out that the 95% percentile confidence boundaries derived by the double bootstrap are narrower and wholly contained within the standard 95% confidence interval boundaries. Most strikingly, the upper boundary of the standard confidence interval fails to intersect with the horizontal axis in both model situations whereas that generated by the double bootstrap does so at a finite concentration value. Application of the double bootstrap would for example imply a maximum ' conceivably safe' concentration threshold for the E*r*C100 at either 3.28 mg/L (linear) or 1.33mg/L (quadratic) but an infinite concentration for such a safe estimate would follow from the standard methodology. Contrastingly, the lower confidence limits determined by the standard and double bootstrap approaches are comparable, and this time the standard approach is slightly more conservative (0.57 *versus* 0.64 mg/l for the quadratic model). A comparison between the linear and quadratic model estimates for the E*r*C100 reveals an order 2 reduction in going from linear to quadratic models in both the lower and upper limits. As is typically shown in Figure 2(a), the R independent first bootstrap distributions for each $\{r_i\}$ were generally skewed and therefore not normal (also see Caswell 2001). Second (block) bootstrap distributions, shown for the E*r*C100 in Figure 2(b) with linear and in Figure 2(c) with quadratic models were also typically skewed. The pointwise percentile confidence interval for the E*r*C100 in each case is by definition that interval spanned by the central 95% of the distribution .

A final comparison of standard metrics with the E*r*Cx is made in Table 1.

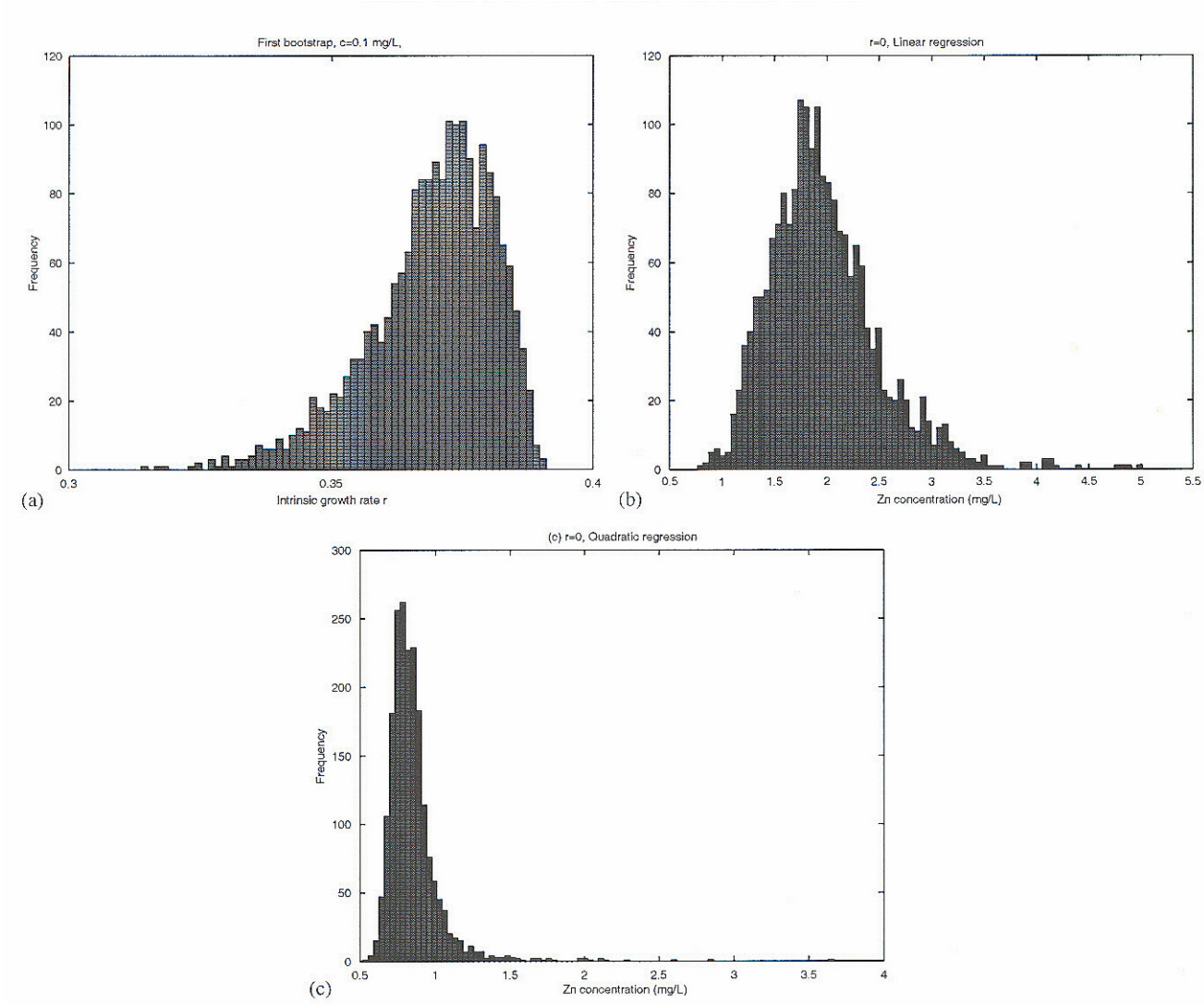


Figure 2: **Bootstrap distributions exhibiting typical skewness. (a) (First) Bootstrap distribution of *r* under 0.1 mg/L zinc sulphate (Zn) exposure. (b-c) (Second) Block bootstrap distributions of the E*r*C100 (i.e. the E*r*Cx at *r*=0) for linear and quadratic regression models respectively. *B*=2000 throughout**

Table 1: Comparison of NOEC, ECx and E*r*Cx estimates at x%=50, 20 and 10 with for *Daphnia magna* under chronic 21 day exposure to zinc sulphate solution (mg/L) with 95% confidence intervals. The E*r*C100 is also shown for linear and quadratic regression models. NOEC and ECx estimates were obtained after the data were arcsine transformed, by one-way ANOVA followed by Dunnett's test (e.g. Newman 1995)

	Reproduction inhibition	Population growth rate r	Population growth rate r
x%	ECx (95%CI)	ErCx (linear) (95%PP CI)	ErCx (quadratic) (95%PP CI)
10	0.07 (0.03 –0.57)	0.11 (0.04 –0.16)	0.21 (0.10 –0.31)
20	0.60 (0.06 –0.88)	0.29 (0.19 –0.44)	0.36 (0.27 –0.42)
50	1.10 (0.79 –1.22)	0.90 (0.55 –1.48)	0.58 (0.47 –0.81)
100	-	1.91 (1.14 – 3.28)	0.82 (0.64 – 1.33)
NOEC	> 0.46	-	-

Lethality is not a significant feature in this kind of chronic test where concentrations were selected in order to maximize reproductive information. Hence a meaningful ECx or NOEC for lethality is not calculable, since few individuals died over the course of the test. For reproduction inhibition however, the NOEC value turned out to be >= 0.46 mg/L because there was no significant difference between control and the highest tested concentration (0.46 mg/L) by one-way ANOVA followed by Dunnett's test (Newman 1995).

The value of the ECx and E*r*Cx each is given at a percentage x where x % is 50, 20 and 10 respectively. The E*r*C50 is that concentration where the control value of r_0 (= 0.3976) is halved to (0.3976/2) = 0.1988 which occurs at 0.90 mg/L for the linear and 0.58 mg/L for the quadratic regression models. Both of these estimates are therefore more conservative 'safe thresholds' than the EC50 of 1.10 mg/L. A comparison of the EC20 of 0.60 mg/L with respective E*r*C20 evaluations of 0.29 mg/L (linear) or 0.36 mg/L (quadratic) shows that with either regression model, the choice of *r* as an endpoint implies a substantial sub-lethal discrimination. In other words, toxic effects that produced a reduction of the order of 20% in the control population growth rate were undetected by either the EC20 or the NOEC. Although the EC10 estimate (0.07 mg/L)

was more conservative than the E*r*C10 of 0.11 mg/L (linear) or 0.21 mg/L (quadratic) estimate, conversely, this was offset by the uncertainty associated with their derivations. In particular, the confidence interval for either E*r*C10 estimate [(0.04-0.16) mg/L (linear) or (0.10-0.31) mg/L (quadratic)] was wholly contained within the much wider confidence interval for the EC10 (0.03-0.57).

APPENDIX C LIFE TABLE ANALYSIS OF DATA FROM LIFECYCLE STUDY WITH FATHEAD MINNOWS (PIMEPHALES PROMELAS) EXPOSED TO 17α-ETHINYLOESTRADIOL

Introduction

Data were acquired from Brixham Environmental Laboratory, arising from a study commissioned by Schering AG.

Experimental methods

Test system

The impact of 17α-ethinyloestradiol on the lifecycle of the fathead minnow (*Pimephales promelas*) was investigated using 5 experimental concentrations (0.2, 1.0, 4.0, 16.0, 64.0 ngl^{-1}) and a control treatment (dilution water only) delivered by a continuous flow-through system. A full description is given in Laenge *et al.* (2001) and only a brief summary is provided here:

> Four aquaria were set up per treatment. Each aquarium held two cups, each containing approximately 25 eggs. Eggs were examined daily and the number of dead eggs and hatched fry was recorded. On the fourth day, when sufficient eggs had hatched, the cups in each aquarium were merged, resulting in the release of 25 fry into the aquarium, randomly and equally selected from each cup. These fry were monitored daily for mortalities.
>
> On the sixtieth day the contents of two aquaria in each treatment were merged, resulting in release of 25 fry into an 'adult' aquarium, selected equally from each aquarium on the basis of health. Selection of the healthiest fish will have added bias to the results, the effects of which are considered in the discussion. Mortality was monitored until day 176. However, the 64 ngl^{-1} treatment was terminated on day 162 as the fish had not reached maturity and exhibited severe abnormalities.
>
> On day 176 eight fish were selected from each control, 0.2 and 1.0 ngl^{-1} aquarium, split into breeding pairs and placed in breeding chambers. Spawning tiles were added to the 4.0 and 16 ngl^{-1} treatments but fish were not paired. Mortality and egg production were monitored daily until the experiment was terminated on days 239 (16 ngl^{-1}), 289 (4 ngl^{-1}) and 305 (control, 0.2 and 1 ngl^{-1}). The 16 ngl^{-1} exposure was terminated prematurely as it was clear that the 4 ngl^{-1} was an effect concentration and so the 16 ngl^{-1} treatment was no longer required. On day 289 fish in the 4 ngl^{-1} exposure were placed in dilution water to depurate. This was to determine whether depuration would lead to normal growth and sexual differentiation.

Not all the data collected in the Laenge *et al* (2001) study were used in the analysis reported here. The data used in this analysis are summarised in Table 2.1.

Table 1: Data used in Life Table Analyses

Days post-hatch	Assessment
-4-1	Fo: embryo development, hatching
1-28	Fo: embryo-larval survival
56	Fo: sex determination
128	Fo: establish breeding pairs
140-300	Fo : spawning, fecundity
160-250	F_1: embryo development, hatching
160-260	F_1: survival up to 28d post-hatch and sex ratios

Data analysis

Data manipulations

Life tables of survivorship and egg production between daily observations up to the point when the experiment was terminated were constructed for each treatment. Because of zero fecundity and low survival for 16 and 64 ngl^{-1} treatments (hence the reason for their early termination) intrinsic population growth rate could not be meaningfully computed for these two highest exposures.

Data for the other treatments (control, 0.2, 1.0 and 4.0 ngl^{-1}) were smoothed by grouping into weekly age classes. The table entries were used to derive the probability of surviving up to age class x (the survivorship function, l_x) and the mean number of offspring produced by an individual in age class x (the maternity function, m_x). A corresponding Leslie matrix $\boldsymbol{A}^{(t)}$ for each of these treatments t =1 to 4 was parameterized. Because egg deposition occurred at the same time each day immediately after a photoperiodic stimulus, the matrix parameters were calculated by application of the postbreeding census birth-pulse formulae of Caswell (2001) for survival (P_i) and fertility (F_i) probabilities:

$$P_i = \frac{l(i)}{l(i-1)} \qquad (1)$$

$$F_i = P_i m_i \qquad (2)$$

for each age class i .

The dominant eigenvalue λ of each treatment matrix was computed as an estimate of the growth rate of a population exposed to each concentration of 17α-ethinyloestradiol.

Two types of analysis were then carried out:

- sensitivity analysis
- decomposition analysis

Sensitivity analysis

The sensitivity of λ to any component matrix parameter a_{ij} (e.g. survivorship and fecundity) was calculated as defined by Caswell (2001) as:

$$\frac{\partial \lambda}{\partial a_{ij}} = \frac{v_i w_j}{\langle \boldsymbol{w}.\boldsymbol{v} \rangle} \qquad (3)$$

where $\boldsymbol{w}$ and $\boldsymbol{v}$ are respectively the right and left eigenvectors of $\boldsymbol{A}$ associated with λ, and $\langle \, . \, \rangle$ denotes the scalar product of two vectors.

Decomposition analysis

The other analysis carried out was a decomposition analysis which shows the effect of each exposure treatment (0.2, 1.0 and 4.0 ngl^{-1} 17α-ethinyloestradiol) in comparison with the controls in this study. The reduction in λ in response to a given treatment, when compared with the control experiment, is 'decomposed' into a component effect (e.g. survivorship, fecundity) observed in each of the model parameters when compared with corresponding parameters in the controls. Each component effect (associated with each parameter change) is referred to as a 'contribution' towards the overall reduction in intrinsic growth rate (Caswell 2001). Contributions are, therefore, the differences between each of the parameters of a given treatment compared with corresponding parameters in the control experiment, but weighted by the sensitivity of λ to the treatment.

The decomposition analysis was performed by the first order approximation (Caswell 1996) for decomposing the effect of a given treatment t in comparison to the control treatment C,

$$\lambda^{(t)} - \lambda^{(C)} \approx \sum_{i,j} \left(a_{ij}^{(t)} - a_{ij}^{(C)} \right) \frac{\partial \lambda}{\partial a_{ij}} \qquad (5)$$

where the partial derivative in equation (5) may be evaluated at any matrix 'located' between the treatment matrix $\boldsymbol{A}^{(t)}$ and the control treatment matrix $\boldsymbol{A}^{(C)}$. Throughout this paper the partial derivative calculation was calculated at the mean matrix $(\boldsymbol{A}^{(t)} + \boldsymbol{A}^{(C)})/2$.

Results and Discussion

Life table analysis

We conventionally revert to the parameter $r = \ln(\lambda)$ for population growth rate (e.g. Kammenga and Laskowski, 2000; Newman 2001)) and proceed by assuming that $r = f(c)$ where c is concentration of the substance, and is a simple linear or nonlinear (quadratic) function. In the absence of further information, the choice of an appropriate

function to describe the relationship between population growth rate and exposure concentration must be empirically based.

If a model $r = f(c)$ is regressed to these data, the effect of any particular concentration can be estimated from the curve of best fit. In addition, the concentration required to reduce population growth rate by a particular percentage (compared with a 'base' concentration) can be estimated. Choosing the base concentration as the control would lead to a convenient E*r*C*x* summary metric. We define this as the concentration of test substance estimated to reduce intrinsic population growth (as represented by the value of *r* obtained in the control) by *x*% (Grist *et al.* 2002). For example, if x=100% then r=0, so extinction would be projected for a population exposed to the E*r*C100 concentration of toxic substance. Similarly, the E*r*C50 and E*r*C20 respectively correspond to concentrations of test substance estimated to reduce population growth to 50% and 20% of the intrinsic growth rate. The E*r*C20 is particularly useful because a 20% difference between treatment and control groups for such continuous toxicity data may reasonably be regarded as equivalent to a 'no effects' concentration (Bruce and Versteeg, 1992).

Table 2 summarises the effect of different concentrations on intrinsic population growth rate, r, and clearly shows the dose-dependant relationship between concentration and r.

Table 2: Relationship between 17α-ethinyloestradiol concentration and r

Population growth rate	Test Concentration (ng/L)			
	control	0.2	1.0	4.0
r/week	0.204	0.194	0.179	- 0.069
(95% bootstrap confidence interval)	(0.165 to 0.224)	(0.174 to 0.206)	(0.147 to 0.198)	(-0.070 to -0.043)
Reduction relative to control (r=0.204)	0 %	4.8 %	12.1 %	134.0 %

In Figure 3.1, the control and treatment population growth rates $\{r^{(i)}\}$ are plotted respective exposure concentrations (solid circles) together with respective 'best fit' regression curves (thick curve) for *f* as the linear function $f(x)=ax+b$. In Figure 3.2, the same plot is shown with *f* as the quadratic function $f(x)=ax^2+bx+c$. It is clear that both descriptions in view of these data must intersect with the horizontal axis (r =0). Therefore the E*r*C100 threshold can be estimated by interpolation to where population growth rate reaches zero. Estimates for this threshold concentration are shown in Table 3.2 for each model. It can be seen that after a 288 day exposure, a concentration of 4 ngl^{-1} 17α-ethinyloestradiol reduces intrinsic population growth rate to below zero. Concentrations for E*r*C50 and E*r*C20 are also shown in Table 3.

How much uncertainty is to be associated with these estimates?

Standard statistical methodology was used to obtain a 95% confidence interval for each model regression curve (dashed lines in Figures 3.1 and 3.2) by assuming that corresponding residuals are drawn from a normal distribution (e.g. Sokal and Rohlf 1995). As well as being reliant on the underlying distributional assumption, this approach takes no account of two independent sources of uncertainty that influence curve construction. First, the $\{r^{(t)}\}$ are point estimates and as such are subject to measurement error. There is hence an associated uncertainty in the vertical co-ordinate of each data point. This imputes a (second) level of uncertainty in curve configuration, because a 'best fit' curve is dependent on the placement of the points to which it is to be fitted.

To take these combined factors into account, an empirical confidence interval was also computed (dotted lines in Figures 3.1 and 3.2) using the 'double bootstrap regression' which incorporates both sources of uncertainty in its construction (Grist *et al.* 2002). The first bootstrap in the 'double bootstrap' is performed by resampling the individual (or replicate) records in each of the treatment life tables to generate a first bootstrap distribution for each of the treatment population growth rates $\{r^{(t)}\}$. These are statistically independent because different cohorts were used in each treatment. This implies that a second set of resamples, each composed of one member randomly drawn from each these distributions can be generated by a 'block' bootstrap (Efron and Tibshirani 1993). A regression curve fitted to any block bootstrap resample is therefore a replicate of the best fit curve because it represents a possible variant of the curve configuration. A pointwise bootstrap percentile confidence interval obtained by slicing vertically (or horizontally) through a large number of these curve replicates thus gives a measure of the uncertainty associated with any point on the original curve (e.g. Davison and Hinkley 1999). Lower and upper points at which a horizontal line of altitude *rx*% cuts the central 95% of these lines represents a 95% percentile based bootstrap confidence interval for the E*r*C*x* estimate.

Figures 3.1 and 3.2, and Table 3.2 clearly show that the double bootstrap method produces narrower confidence intervals for the line of best fit, especially with a linear rather than quadratic underlying model. In both cases the confidence interval is narrower than that derived by normal theory. However, caution should be exercised in asserting that a more accurate estimate is thereby obtained, because of the possible biases introduced by the way fish were selected for pairing in the experimental protocol i.e. only apparently healthy fish were chosen to continue in the study. This potentially imputes the assumed non-independence between the bootstrap distributions generated in the first part of the double bootstrap. Further, the declining number of surviving replicates became rather low for bootstrapping with increasing time and concentration (noticeable in the general decrease in vertical displacement of the x's at the 4.0 μgl^{-1} concentration as seen in Figures 3.1 and 3.2). For all these reasons, it is prudent to treat any threshold toxicity concentration estimated from these data as an upper bound of the true toxicity of this chemical.

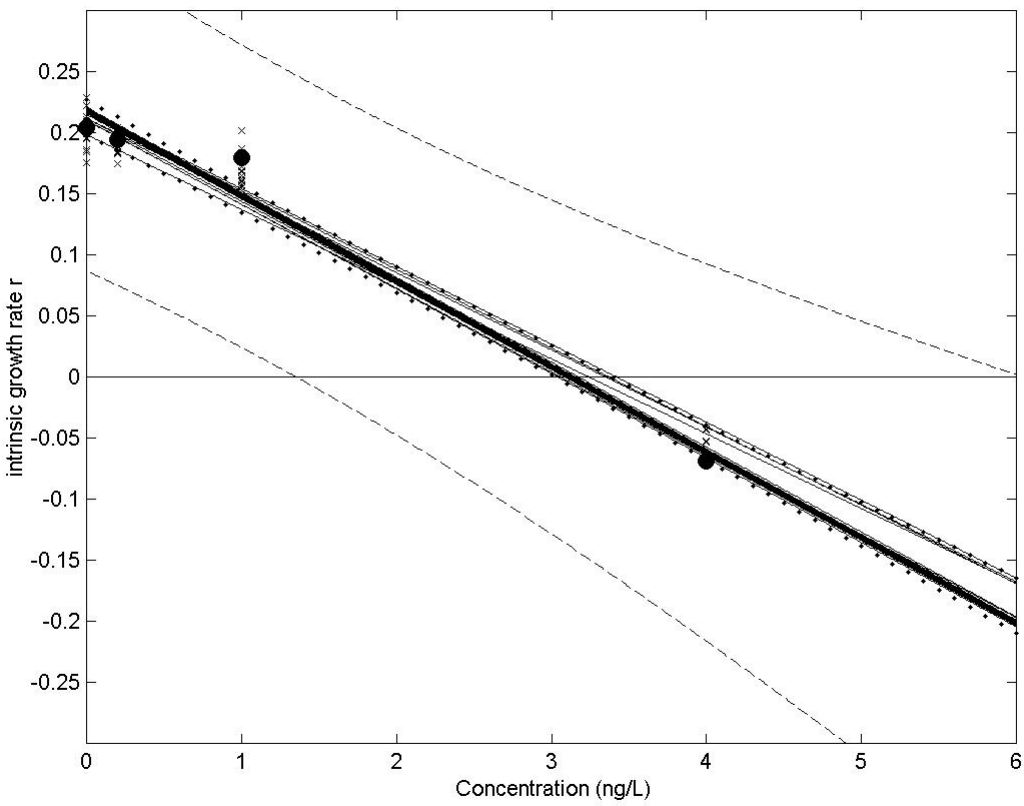


Figure 1: Effect of 17α-ethinyloestradiol on population growth rate (linear regression model). The line of best fit is shown along with 95% confidence limits calculated using normal probability theory (dashed) and pointwise double bootstrap method (dotted). Best fit parameters are a=-0.07, b=0.218; r^2=0.975

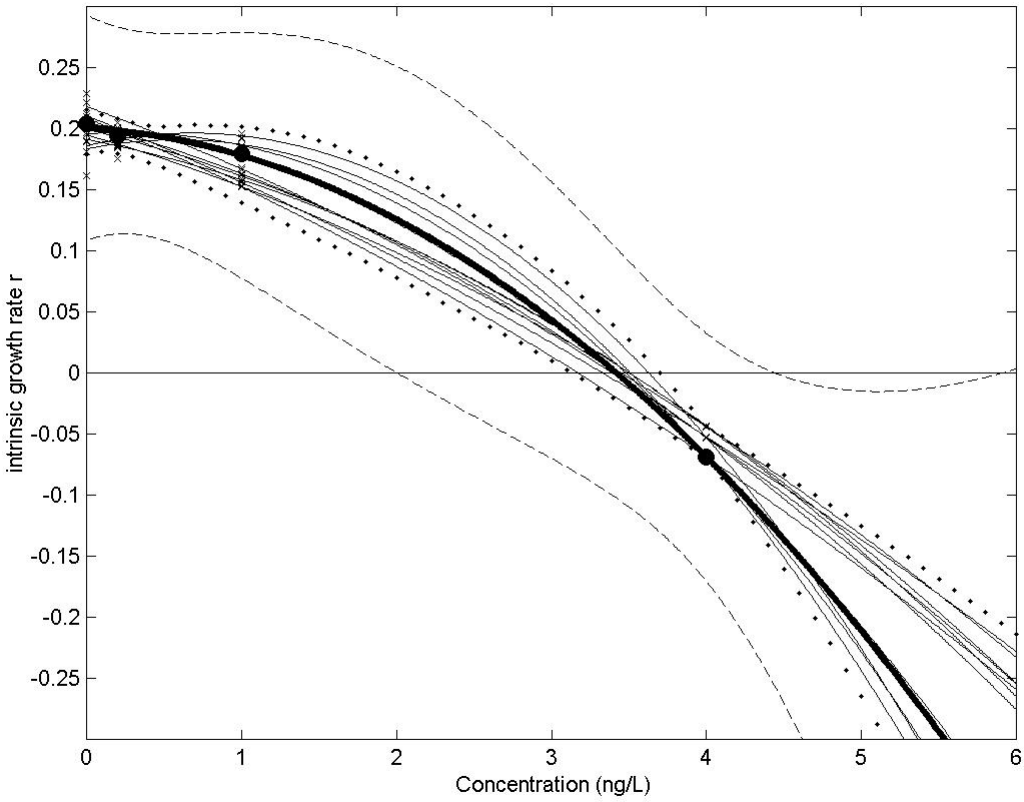


Figure 2: Effect of 17α-ethinyloestradiol on population growth rate (quadratic regression model). The line of best fit is shown along with 95% confidence limits calculated using normal probability theory (dashed) and pointwise double bootstrap method (dotted). Best fit parameters are a=-0.015, b=-0.0076, c=0.2006; r^2=0.999

Table 3: Estimates of E*r*C20, 50 and 100 and confidence intervals for linear and quadratic regressions fitted to population growth data

Reduction in *r* (x %)	E*r*Cx (ng/L)									
	Linear (R^2 =0.97)					Quadratic (R^2 =0.99)				
		Confidence interval					Confidence Interval			
		Bootstrap		Normal			Bootstrap		Normal	
		lower	upper	lower	upper		lower	upper	lower	upper
20	**0.78**	0.53	0.96	< 0	2.67	**1.35**	0.71	1.92	< 0	2.94
50	**1.66**	1.50	1.85	< 0	3.82	**2.33**	1.67	2.70	0.64	3.42
100	**3.11**	3.03	3.45	1.34	6.04	**3.41**	3.12	3.67	2.00	4.43

Sensitivity analysis

Fertility and survival sensitivities of fathead minnows in 0, 0.2, 1and 4.0 ngl^{-1} treatments are shown in Figures 3.3 (a-d). For all treatments the curve for survival is generally flat until first reproduction occurs. For the lowest three concentrations (a-c), sensitivity of λ to age classes beyond this generally decreases monotonically, which is a general property of such models when λ >1 and $P_{i+1} >= P_i$ (Caswell 2001). It is also a general property that sensitivity of λ to survival is greater than to fertility, at least up to the age of first reproduction, so the survival curve is at a higher altitude than the fertility curve. Fertility decreases exponentially, which is another general property of such models when λ >1.

The 4.0 ngl^{-1} treatment produced a markedly different response to the other treatments. Curves for fertility and survival of fathead minnows at 4.0 ngl^{-1} (for which r<0 [λ<1]) (d) show that no reproduction occurred during this treatment which is why it must follow that r<0 and that such a population must become extinct. Sensitivity of λ to survival at 4.0 ng/L is the same for all age classes (d), but sensitivity to fertility increases as age increases. The latter feature is a general property of a decreasing population (Caswell 2001). The large difference of the fertility curve in comparison with those at lower concentrations is thus largely attributable to the fact that the population is declining.

Up to the age of first reproduction, the sensitivity of λ to survival becomes decreasingly less with exposure to higher concentrations. For reproducing age classes the converse holds and sensitivity becomes increasingly greater with higher concentration. Under all treatments, sensitivity of λ to survival declines after the age of first reproduction. A possible interpretation for greater sensitivity at higher concentrations to fertility of reproducing age classes is that this is due to a corresponding decrease in fecundity. In other words, it indicates that population sensitivity to survival becomes more significant when there is less reproduction to offset mortality.

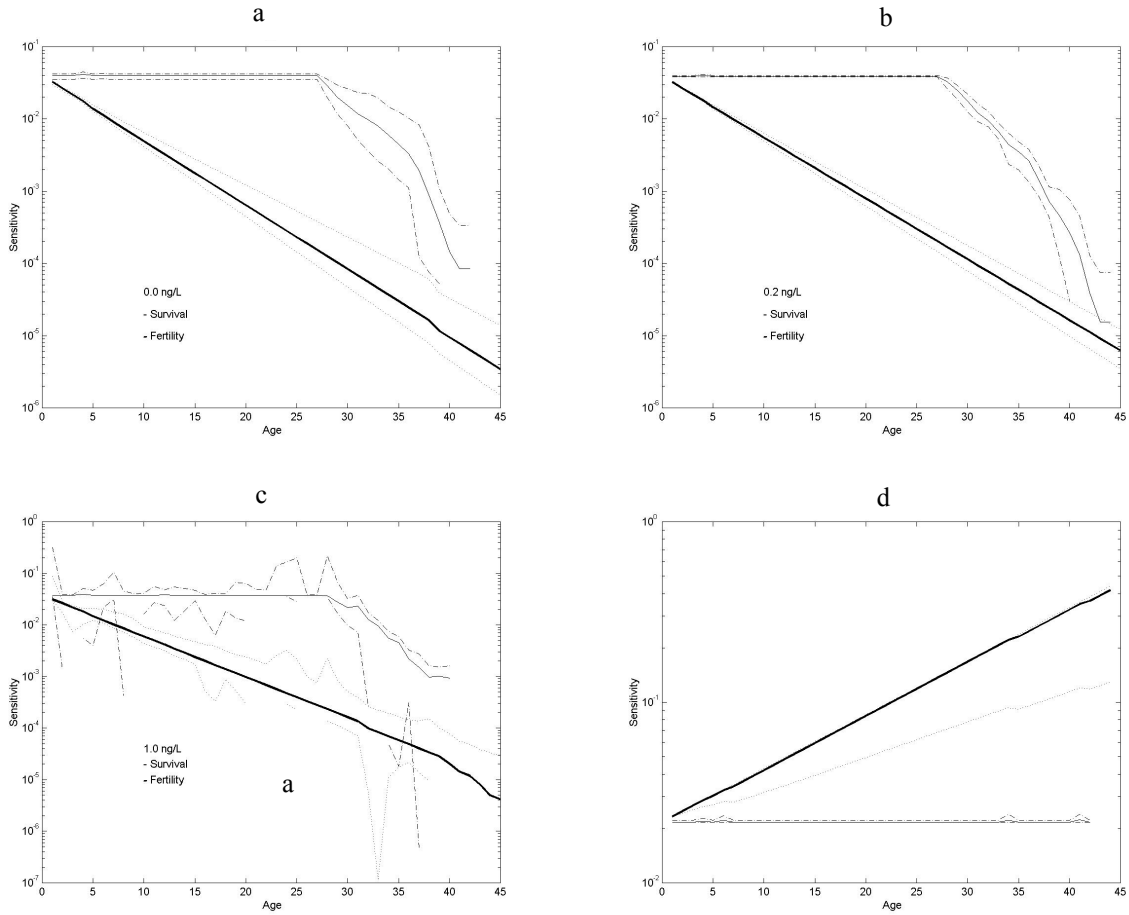


Figure 3: Sensitivities of λ to changes in the LTRE matrix survival (thin) and fertility (bold) parameters at different concentrations of 17α-ethinyloestradiol (a=control, b=0.2 ngl^{-1}, c=1.0 ngl^{-1} and d=4 ngl^{-1}). 95% bootstrap confidence limits are also shown for each parameter

Decomposition analysis

The contribution of changes in survival at each concentration of 17α-ethinyloestradiol is slight, as shown by the slight deviation of the survival curve away from the zero level across all age classes in Figure 3.4 (a-c). This implies that exposure to 17α-ethinyloestradiol at these concentrations had only a small effect on survival. There is a suggestion that, at around age 7 weeks, toxicity may initially produce a small increase in survival response at early ages, as shown by the small positive deviation.

The slight impact of changes in survival is in contrast with the negative contributions of fertility (Fig. 3.5, a-c), shown by the large decline at the start of reproductive age in each treatment (note that the curve pairs on each plot are directly comparable because the same vertical scale applies to each). This shows that the reduction in intrinsic growth rate is due mainly to a reduction in fertility and not survival rates, especially at

the age of first reproduction. Given the mode of action of 17α-ethinyloestradiol, this sensitivity of fertility compared with survival is to be expected. Thereafter, for older age classes the reduction in fertility becomes less acute as shown by the upward climb of the curve after the initial steep decline (Figs 3.5, a-c).

There is a greater negative survival effect on older age classes (after about 30 weeks) at a concentration of 1.0 ngl^{-1} compared with the control treatment. However, exposure to 4 ngl^{-1} compared with the control does not show this response because there were no mortalities observed at this concentration. It is necessary to recall that the protocol at 4ngl^{-1} was different, in that after day 176 approximately 50 individuals were held together in the same tank. This contrasts with the lower concentrations in which paired individuals were placed in much smaller breeding compartments, which may have influenced survival.

Conclusions

The analyses shown in Section 3 demonstrate that the observed reduction in intrinsic rate of growth with increasing 17α-ethinyloestradiol exposure is caused more by a general reduction in fertility than by a reduction in survival rates.

For the purposes of risk assessment or deriving environmental thresholds, the ability to derive confidence intervals is valuable because it allows the risk assessor to take explicit account of uncertainty in the estimates of toxicity. This cannot be done when toxicity is expressed in terms of a threshold such as an LOEC, NOEC or MATC (Crane and Newman 2000) but can when toxicity is expressed as a point estimate (e.g. an ECx). Indeed, this is a major argument in favour of point estimates as opposed to thresholds derived from hypothesis testing. In the following discussion, we have taken the lower pointwise confidence limit as a conservative estimate of the concentration on which decisions might reasonably be taken, effectively applying the precautionary principle in a transparent way and giving the ‘benefit of doubt’ to the environment.

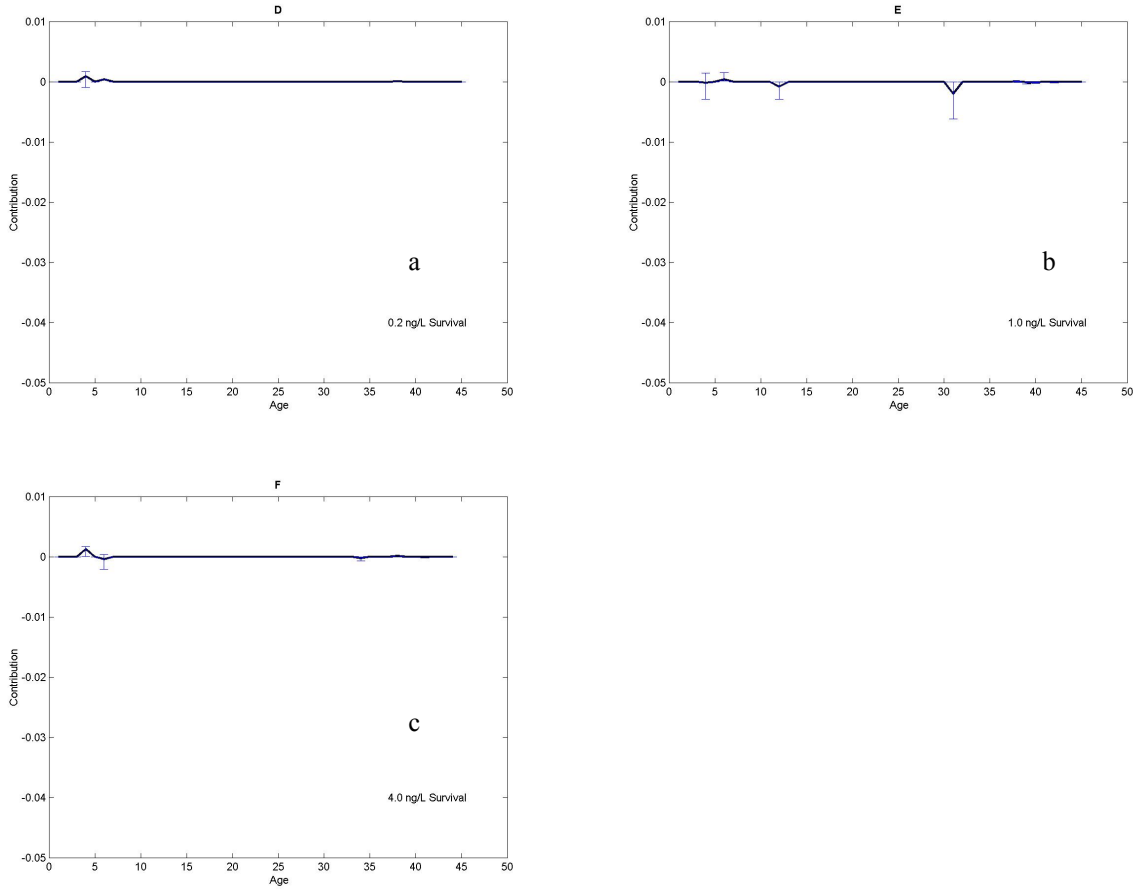


Figure 4: Contributions to changes in λ of the survival parameters for 0.2 ngl^{-1} (a), 1.0 ngl^{-1} (b) and 4.0 ngl^{-1} (c) 17α-ethinyloestradiol. Error bars are 95% pointwise bootstrap confidence intervals

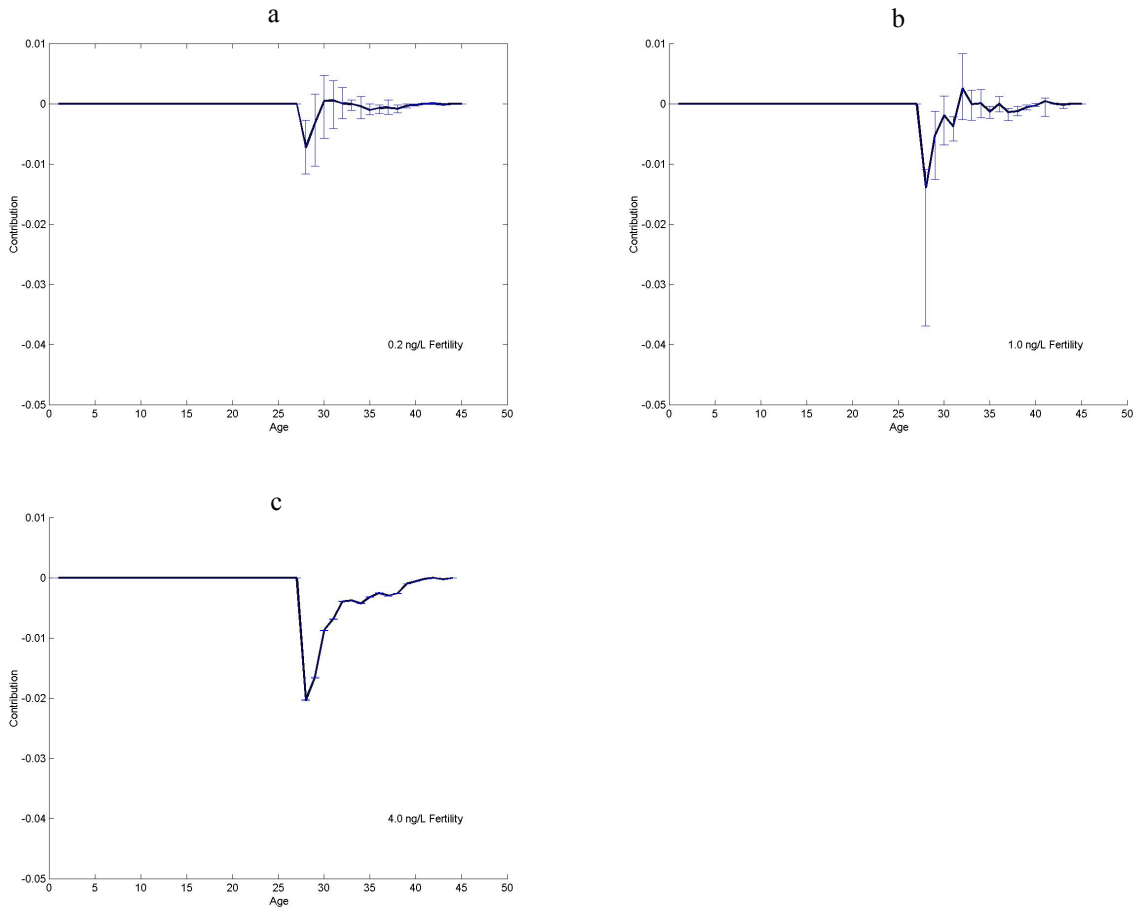


Figure 5: Contributions to changes in λ of the fertility parameters for 0.2 ngl^{-1} (a), 1.0 ngl^{-1} (b) and 4.0 ngl^{-1} (c) 17α-ethinyloestradiol. Error bars are 95% pointwise bootstrap confidence intervals

Even at 3 ngl^{-1} 17α-ethinyloestradiol (the lower 95% pointwise confidence limit on the estimated E_rC100), our analysis predicts a complete lack of recruitment through reproduction and, under these circumstances, extinction of populations of fathead minnows would be projected. This corresponds reasonably closely to the LOEC for feminisation of experimental fish from the same study (Laenge *et al*, 2001). The lower 95% pointwise confidence limit on the E_rC20 is 0.71 ngl^{-1} and may be taken as a conservative estimate of the 'no effects' concentration for λ, on the assumption that a 20% difference between treatment and control groups for such continuous toxicity data cannot be resolved (Bruce and Versteeg, 1992). This concentration approximates to the NOEC of 1 ngl^{-1} reported by Laenge *et al.* (2001) for sex reversal in fathead minnows, suggesting that the threshold concentration of 17α-ethinyloestradiol for effects on population growth is similar to that causing a high level of phenotypic sex reversal.

Results of this analysis indicate that toxicity summaries for 17α-ethinyloestradiol based on conventional reproductive endpoints in fathead minnow are in close agreement with those derived by demographic analysis, based on a long period of continuous exposure. The immediate conclusion is that environmental thresholds for 17α-ethinyloestradiol

based on conventional endpoints are likely to afford adequate protection at the population level, at least for fathead minnows. It is reasonable to suppose the same principle would apply to other fish species with similar life history strategies.

References (Appendix C)

Bruce R.D. and Versteeg D.J. (1992) A statistical procedure for modelling continuous toxicity data. *Environmental Toxicology and Chemistry* **11**, 1485-1494.

Caswell H (1996). The analysis of life table response experiments, I. Decomposition of treatment effects on population growth rate. *Ecological Modelling* **46**, 221-237.

Caswell H. (2001) Matrix Population Models, 2nd edition, Sinauer , Mass, USA.

Crane M, and Newman MC (2000) What level of effect is a no observed effect? *Environmental Toxicology and Chemistry* **19**, 516-519.

Davison A.C, Hinkley D.V. (1999). Bootstrap Methods and their Application. Cambridge University Press, Cambridge, UK.

Desbrow C., Routledge E.J., Brighty G.C., Sumpter J.C. and Waldock M.J.(1998) Identification of estrogenic chemicals in sewage treatment works effluents. 1. Chemical fractionation and *in vitro* biological screening. *Environmental Science and Technology* **32**, 1549-1558.

Efron B, Tibshirani R.J. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York, USA.

Forbes V.E., Calow P. and Sibly R.M. (2001) Are current species extrapolation models a good basis for ecological assessment? *Environmental Toxicology and Chemistry* **20**, 442-447.

Grist E.P.M., Crane, M., Jones, C. and Whitehouse, P. (2002) Estimation of demographic toxicity through the double bootstrap. *Water Research*, in press.

Kammenga J. and Laskowski R. (eds) 2000, Demography in ecotoxicology, John Wiley and sons LTD, Chichester UK.

Laenge R., Hutchinson T.H., Croudace C., Siegmund F., Schweinfurth H., Hampe P., Panter G.H. and Sumpter J.P. (2001) Effects of the synthetic estrogen 17α-ethinyloestradiol on the life cycle of the fathead minnow (Pimephales promelas). *Environmental Toxicology and Chemistry* **20**, 1216-1227.

Leslie P.H. 1945. On the use of matrices in certain population mathematics. *Biometrika* **35**, 213-245.

Levin L., Caswell H., Bridges T., Dibacco C., Cabrera D., and Plaia G. (1996) Demographic responses of estuarine polychaetes to pollutants: life table response experiments. *Ecological Applications* **6**, 1295-1313.

Newman M.C. (2001) Population Ecotoxicology, John Wiley and Sons Ltd, Chichester UK.

Purdom C.E., Hardiman P.A., Bye V.J., Eno N.C., Tyler C.R. and Sumpter J.P. (1994) Estrogenic effects of effluent from sewage treatment works. *Chemical Ecology* **8**, 275-285.

Routledge E.J., Sheahan D., Desbrow C., Brighty G.C., Waldock M. and Sumpter J.P. (1998) Identification of estrogenic chemicals in sewage treatment works effluents. 2. In vivo responses in trout and roach. *Environmental Science and Technology* **32**, 1559-1565.

Sokal R.R. and Rohlf F.J. (1995) Biometry, the principles and practice of statistics in biological research, 3rd edition. W.H. Freeman and Co., New York, USA.

Young W.F., Whitehouse P., Johnson I., and Sorokin N. (2002) Proposed Predicted No Effect Concentrations (PNECs) for Natural and Synthetic Steroid Oestrogens in Surface Waters. Environment Agency Technical Report P2-T04/1, Environment Agency, Bristol, UK.

APPENDIX D LETTER REQUESTING RAW DATA FOR DEMOGRAPHIC ANALYSIS

Dear <>

DEVELOPING MORE ECOLOGICALLY RELEVANT WATER QUALITY STANDARDS - ECOTOXICITY STUDY DATA

I am writing to you in connection with your publication in <> entitled <>. Specifically, we are keen to obtain more details of this study if possible, to help us in a research project we are undertaking. First, I will try and explain more about our project and then go on to indicate how you may be able to contribute.

The project is entitled 'Risk-based approaches to the derivation and expression of Environmental Quality Standards'. It is a collaborative study between Royal Holloway University of London, University of Sheffield, Virginia Institute of Marine Science and WRc-NSF Ltd and sets out to investigate ways of developing more ecologically relevant approaches to the derivation of water quality standards. A key aspect of the project is to describe chemical toxicity in terms of effects at the population level. This would represent a significant advance on standard practices that currently place heavy reliance on arbitrary safety factors. Thus, it is our main intention to derive more meaningful endpoints.

The information of greatest value to us concerns chemical effects on survival time, time-to-reproduction and fecundity in freshwater invertebrates and fish. We would be most grateful if any (including handwritten) raw data in connection with these responses could be made available to us. In particular, we seek records showing the responses of *individual* organisms to test concentrations with the progression of time. Our experience tells us that this information is likely to have been recorded as the primary raw data from many aquatic toxicity studies. Many conventional studies will have generated potentially useful data but will not always have reported results in this way. It is for this reason that I am writing to you now.

If you are able to assist in this request we will be very grateful. We will, of course, acknowledge your assistance in reports to our sponsors if you wish. If you would like to learn more about the project then we will be happy to oblige.

I hope you will be able to contribute to this initiative and I look forward to hearing from you shortly.

Yours sincerely

Dr Paul Whitehouse
Dr Eric Grist
Dr Mark Crane
Prof Tony O'Hagan

DDI Number: 01491 636648

APPENDIX E SUMMARY OF RAW DATA COLLECTED AND ITS SUITABILITY FOR ESTIMATING DEMOGRAPHIC ENDPOINTS

Chemical	Species	Common name	Mortality data provided?	Fecundity data provided?	'r' or λ calculated by author?	Reference
Ammonia	*Salmo trutta*	Brown trout	Yes	No	No	WRc (1990)
Ammonia	*Oncorhynchus mykiss*	Rainbow trout	Yes	No	No	WRc (no date given)
Ammonia	*Oncorhynchus mykiss*	Rainbow trout	Yes	No	No	WRc (1992)
Ammonia	*Rutilus rutilus*	Roach	Yes	No	No	WRc (1992)
Ammonia	*Cyprinus carpio*	Carp	Yes	No	No	WRc (1988)
Ammonia	*Rutilus rutilus*	Roach	Yes	No	No	WRc (1983)
Ammonia	*Cyprinus carpio (ELS)*	Carp	Yes	No	No	WRc (1986)
Ammonia	*Rutilus rutilus*	Roach	Yes	No	No	WRc (1979)
Ammonia	*Perca flubiatilis*	Perch	Yes	No	No	WRc (1979)
Phenol	*Brachionus rubens*	Rotifer	Yes	Yes	Yes	Halbach *et al.* (1983)
Phenol	*Brachionus calyciflorus*	Rotifer	No	No	Yes	Snell and Moffat (1992)
Phenol	*Pimephales promelas*	Fathead minnow	Yes	No	No	Broderius *et al* (1995)
Phenol	*Daphnia magna*	Water flea	Yes	Yes	No	Tisler *et al* (1999)
Phenol	*Zenopus*		Yes	No	No	Bernardini *et al* (1996)
Phenol	*Daphnia magna*	Water flea	No	Yes	No	WRc (1 No year given)

Chemical	Species	Common name	Mortality data provided?	Fecundity data provided?	'r' or λ calculated by author?	Reference
Phenol	*Daphnia magna*	Water flea	No	Yes	No	WRc (2 No year given)
Chlorpyriphos	*Brachionus calyciflorus*	Rotifer	No	No	Yes	Snell and Moffat (1992)
Chlorpyriphos	*Daphnia magna*	Water flea	Yes	Yes	No	Naddy *et al.* (2000)
Chlorpyriphos	*Brachionus calyciforus*	Rotifer	No	Yes	Yes	Preston *et al* (2000)
Chlorpyriphos	*Daphnia pulex*	Water flea	Yes	Yes	No	Van der Hoeven (1997)
Chlorpyriphos	*Wyeomyia smithii*	Mosquito larvae	No	Number of larvae per instar	No	Strickman *et al* (1983)
Cadmium	*Moina macrocopa*	Water flea	Y (Graphs)	Y (Graphs)	Yes	Wong and Wong (1990)
Cadmium	*Moina macrocopa*	Water flea	Y (Graphs)	Y (Graphs)	Yes	Hatakeyama and Yasuno (1981)
Cadmium	*Daphnia pulex*	Water flea	Y (Totals)	Y (Totals)	Yes	Bertram and Hart (1979)
Cadmium	*Daphnia carinata*	Water flea	Yes	No	Yes	Chandini (1989)
Cadmium	*Daphnia magna*	Water flea	No	Yes	No	Enserink *et al.* (1990)
Cadmium	*Daphnia magna*	Water flea	Y (Graphs)	Y (Graphs)	Y (Graphs)	Enserink *et al.* (1993)
Cadmium	*Daphnia magna*	Water flea	Y (Graphs)	Y (Graphs)	Yes	Bodar *et al.* (1988)
Cadmium	*Brachionus calyciflorus*	Rotifer	No	No	Yes	Snell and Moffat (1992)

Chemical	Species	Common name	Mortality data provided?	Fecundity data provided?	'r' or λ calculated by author?	Reference
Cadmium	*Daphnia magna*	Water flea	No	No	Yes	Van Leeuwen *et al.* (1985)
Cadmium	*Daphnia magna*	Water flea	No	No	Yes (3 expts)	Van Leeuwen *et al.* (1987)
Cadmium	*Daphnia pulex*	Water flea	No	No	Yes	Meyer *et al.* (1987)
Cadmium	*Daphnia pulex*	Water flea	No	No	Yes	Kluttgen and Ratte (1994)
Cadmium	*Daphnia magna*	Water flea	Y (Graph)	Y (Graph)	Y (Graph)	Barata *et al.* (2000)
Cadmium	*Daphnia magna*	Water flea	No	No	Yes	Knops *et al.* (2001)
Cadmium	*Brachionus calyciforus*	Rotifer	No	Yes	Yes	Preston *et al* (2000a)
Cadmium	?	?	No	Yes	Yes	Preston *et al* (2000b)
Cadmium	*Hyalella azteca*		Yes	No	No	Borgmann *et al* (1991)
Cadmium	*Gammarus pulex*	Freshwater shrimp	Yes	No	No	Borgmann *et al* (1989)
Cadmium	*Bulinus globosus*	Snail	Yes	No	No	Tomasik *et al* (1994)
Cadmium	*Hyalella azteca*		Yes	No	No	Jackson *et al* (2000)
Cadmium	*Daphnia magna*	Water flea	Yes	No	No	Weltens *et al* (2000)
Cadmium	*Oncorhynchus mykiss*	Rainbow trout	Yes	No	No	WRc (No year given)
Cadmium	*Rutilus rutilus*	Roach	Yes	No	No	WRc (1983)

Chemical	Species	Common name	Mortality data provided?	Fecundity data provided?	'r' or λ calculated by author?	Reference
Cadmium	*Noemacheilus barbatulus*	Stone loach	Yes	No	No	WRc (1975)
Nonylphenol	*Brachionus calyciforus*	Rotifer	No	Yes	Yes	Preston *et al* (2000)
Nonylphenol	*Ceriodaphnia dubia*	Waterflea	Yes	Yes	No	CMA (1995)
Nonylphenol	*Ceriodaphnia dubia*	Waterflea	Yes	Yes	No	England (ABC Labs) (1995)
Nonylphenol	*Daphnia galeata*	Water flea	No	No	Y (Graphs)	Tanaka and Nakanishi (2001)
Nonylphenol	?	?	No	No	Yes	Preston *et al* (2000b)
Nonylphenol	*Pimephales promelas*	Fathead minnow	Yes	No	No	Naylor *et al* (1991)
Nonylphenol	*Ishnura elegans*	Damsel fly	Yes	No	No	Sims *et al* (1998)
Nonylphenol	*Pimephales promelas*	Fathead minnow	Yes	No	No	CMA (1991)
Nonylphenol	*Gammarus pulex*	Amphipod	Yes	No	No	Sims *et al* (1998)
Zinc	*Moina macrocopa*	Water flea	No	No	Yes	Wong (1993)
Zinc	*Dinophilus gyrociliatus*	Polychaete	No	No	Yes	Crema and Mauri (2001)
Zinc	*Hyalella azteca*		Yes	Yes	No	Borgmann *et al* (1993)
Zinc	*Noemacheilus barbatulus*	Stone loach	Yes	No	No	WRc (1975)
Zinc	*Bulinus globosus*	Snail	Yes	No	No	Tomasik *et al* (1994)

Chemical	Species	Common name	Mortality data provided?	Fecundity data provided?	'r' or λ calculated by author?	Reference
Zinc	*Asselus aquaticus*	Amphipod	Yes	No	No	Migliore *et al* (1990)
Zinc	*Daphnia magna*	Water flea	Yes	No	No	Barrata *et al* (1998)
Zinc	*Daphnia magna*	Water flea	Yes	No	No	Weltens *et al* (2000)
Zinc	*Oncorhynchus mykiss*	Rainbow trout	Yes	No	No	WRc (No year given)
Zinc	*Daphnia magna*	Water flea	Yes	No	No	WRc (1998)
Zinc	*Rutilus rutilus*	Roach	Yes	No	No	WRc (1983)

APPENDIX F EXPERT ELICITATION OF SPECIES SENSITIVITIES TO CHLORPYRIPHOS

Taxonomic groups and associated predictions of sensitivity weighted according to experts' self-assessment of their expertise

Taxon	Type of organism	Fast-flowing stream	Slow-flowing river	Pond or Ditch	Expert weighted mean sensitivity
Aeshnidae	Dragonflies		✓	✓	5.555555556
Ancylidae	Limpets	✓	✓	✓	4.294117647
Anguillidae	Eels		✓	✓	4.166666667
Aphelocheiridae	Saucer bugs	✓			5
Asellidae	Water hoglice		✓	✓	4.083333333
Astacidae	Crayfish		✓		6.25
Baetidae	Mayflies	✓	✓	✓	5.793103448
Beraeidae	Caddis flies	✓	✓	✓	6.2
Brachycentridae	Caddis flies	✓	✓		6.272727273
Bufonidae	Toads		✓	✓	5.2
Caenidae	Mayflies	✓	✓	✓	5.863636364
Calopterygidae	Demoiselles	✓	✓		5.333333333
Capniidae	Stoneflies	✓			5.739130435
Chironomidae	Midges	✓	✓	✓	3.56
Chloroperlidae	Stoneflies	✓	✓		6.130434783
Clupeidae	Shad		✓		3.909090909
Cobitidae	Loach	✓	✓		4
Coenagriidae	Damselflies		✓	✓	5.555555556
Cordulegasteridae	Dragonflies		✓		5.588235294
Corduliidae	Dragonflies			✓	5.5625

Taxon	Type of organism	Fast-flowing stream	Slow-flowing river	Pond or Ditch	Expert weighted mean sensitivity
Coregonidae	Whitefish	✓	✓		4.272727273
Corixidae	Lesser waterboatmen	✓	✓	✓	5.111111111
Corophiidae	Shrimps		✓		5
Cottidae	Bullhead		✓		4.4
Cyprinidae	Carp	✓	✓	✓	4.076923077
Dendrocoelidae	Flatworms		✓	✓	5
Dryopidae	Water beetles	✓	✓		4.571428571
Dytiscidae	Diving beetles		✓	✓	3.909090909
Elminthidae	Riffle beetles	✓	✓		5.2
Ephemerellidae	Mayflies	✓	✓		6.5
Ephemeridae	Mayflies	✓	✓		6.44
Erpobdellidae	Leeches	✓	✓	✓	4.8
Escocidae	Pike		✓		4
Gammaridae	Shrimps	✓	✓		5.566666667
Gasterosteidae	Sticklebacks		✓	✓	4.125
Gerridae	Pond skaters	✓	✓	✓	3.6875
Glossiphoniidae	Leeches	✓	✓	✓	4.894736842
Goeridae	Caddis flies	✓			5.954545455
Gomphidae	Dragonflies		✓		5.647058824
Gyrinidae	Whirligig beetles		✓	✓	4.888888889
Haliplidae	Water beetles		✓	✓	5
Heptageniidae	Mayflies		✓		6.75
Hirudinae	Leeches			✓	5.210526316

Taxon	Type of organism	Fast-flowing stream	Slow-flowing river	Pond or Ditch	Expert weighted mean sensitivity
Hydrobiidae	Snails	✓	✓	✓	4.888888889
Hydrometridae	Water measurers		✓	✓	4
Hydrophiliidae	Scavenger beetles			✓	4.5
Hydropsychidae	Caddis flies	✓			5.916666667
Hydroptilidae	Caddis flies	✓	✓	✓	6.181818182
Hygrobiidae	Screetch beetles			✓	4.6875
Lepidostomatidae	Caddis flies	✓	✓		6.142857143
Leptoceridae	Caddis flies	✓	✓	✓	6
Leptophlebiidae	Mayflies		✓	✓	6.565217391
Lestidae	Damselflies			✓	5.882352941
Leuctridae	Stoneflies	✓	✓		6.041666667
Libellulidae	Dragonflies			✓	5.411764706
Limnephilidae	Caddis flies	✓	✓	✓	5.173913043
Lymnaeidae	Snails	✓	✓	✓	4.473684211
Mesoveliidae	Water bugs			✓	4.5625
Molannidae	Caddis flies		✓	✓	5.888888889
Nemouridae	Stoneflies	✓	✓		6
Nepidae	Water scorpions		✓	✓	4.8125
Neritidae	Snails		✓		4.875
Notonectidae	Water boatmen			✓	4.722222222
Odontoceridae	Caddis flies	✓			6
Oligochaeta	Worms	✓	✓	✓	2.777777778
Percidae	Perch		✓	✓	3.923076923
Perlidae	Stoneflies	✓			6.590909091

Taxon	Type of organism	Fast-flowing stream	Slow-flowing river	Pond or Ditch	Expert weighted mean sensitivity
Perlodidae	Stoneflies	✓	✓		6.590909091
Petromyzonidae	River lamprey		✓		4.230769231
Philopotamidae	Caddis flies	✓			6.047619048
Phryganeidae	Caddis flies		✓	✓	5.904761905
Physidae	Bladder snails	✓	✓	✓	4.823529412
Piscicolidae	Fish leeches	✓	✓	✓	4.941176471
Planariidae	Flatworms	✓	✓	✓	5
Planorbidae	Ramshorn snails		✓	✓	4.684210526
Platycnemidae	Damselflies	✓			4.857142857
Pleidae	Lesser backswimmers		✓	✓	4.285714286
Polycentropidae	Caddis flies	✓	✓	✓	6.181818182
Potamanthidae	Mayflies		✓		5.75
Psychomyidae	Caddis flies		✓	✓	5.761904762
Ranidae	Frogs		✓	✓	5.1
Rhyacophilidae	Caddis flies	✓	✓		5.863636364
Salmonidae	Salmon	✓	✓		5.4
Scirtidae	Beetles	✓	✓	✓	4.866666667
Sericostomatidae	Caddis flies	✓	✓	✓	5.954545455
Sialidae	Alderflies	✓	✓	✓	5.333333333
Simuliidae	Black-flies	✓	✓		5.5
Siphlonuridae	Mayflies		✓		6.526315789
Sphaeriidae	Mussels	✓	✓	✓	5.444444444
Taeniopterygidae	Stoneflies	✓	✓		6.19047619

Taxon	Type of organism	Fast-flowing stream	Slow-flowing river	Pond or Ditch	Expert weighted mean sensitivity
Thymallidae	Grayling	✓			4
Tipulidae	Crane-flies	✓	✓	✓	4.842105263
Triturus	Newts		✓	✓	5.1
Unionidae	Mussels		✓	✓	5.4
Valvatidae	Snails		✓	✓	5.176470588
Viviparidae	Snails		✓		5.333333333